

Some remarks on the use of weights when combining different samples or using a part of sample

Erika Antal *

November 25, 2015

When only part of a sample is used or when different samples are merged, one should pay attention to how weights are used. This document is dedicated to clarify some essential issues about the use of weights in these cases. Of central importance is to identify the reference population of the subsample or of the merged sample.

- If using subsamples:
 1. When only a part of the sample is used, the reference population of this part is evidently not necessarily the same as that of the whole sample. Concerning the SHP samples, if one would like to take only the subsample of men for example, or only a group of persons in certain age categories or only the inhabitants of one (or more, but not all) region of Switzerland, etc. the reference population of these subsamples is not the same as that of the whole sample (the population older than 14 living in private households in Switzerland). For the subsample of men, the reference population is only the population of men older than 14, living in private households in Switzerland etc. For the group of persons in certain age categories, the reference population also only refers to these age categories, and so on. The basic weights delivered with the SHP data are calculated to adjust for the reference population of the whole sample and not for the one of the subsample. Consequently they are not completely suitable for these subsamples. Even if any additional information is available concerning the new reference population, the basic weights need to be slightly modified. For example in the case of the weights that keep the sample size ($\sum W_i = N$) the sum of these weights for the entire sample is equal to the sample size. So, when not all units are used, one should put some weights on the basic weights so that the

*Swiss Centre of Expertise in the Social Sciences C/O Université de Lausanne - Bâtiment Géopolis - CH-1015 Lausanne - Switzerland email: erika.antal@fors.unil.ch

sum of the new weights for the subsample is equal to the size of the subsample. That is:

$$W_k^{new} = SS^{new} \times \frac{W_k^{old}}{\sum_{l=1}^{SS^{old}} W_l^{old}}$$

where SS is the sample size, k =person k and l =the l^{th} person in the sample (ranging from 1 to SS). This means that for each person the new weight is calculated by dividing the old weight by the sum of all old weights (i.e. the sample size in this case) and multiplying it with the new sample size. If the subsample is, for example, 1/10 of the complete sample, all weights are recalculated to 1/10 of the original weight, together summing up to the subsample size.

2. Concerning the weights that inflate estimations to the size of the reference population (W.\$\$.P), the principle of the recalculation is the same as above. The sum of the weights for the new sample should be equal to the size of the new population of reference.
- If merging different samples one should keep the following considerations in mind:
 1. The reference population of the merged sample is always the **combination** of the reference populations of the different samples.
 2. In the case when the populations of reference are exactly the same (for the different samples and consequently also for the merged sample, i.e. both refer to the total population of 15 years and older living in private households in Switzerland) the delivered weights (W.\$\$.S in the SHP) can be used for analysis without modification.
 3. When the populations of reference are **mutually exclusive**, modifications are not needed either (for example if one combines a sample from Vaud with one from Geneva).
 4. In all other cases, one needs to modify the weights. This is done in several steps which we will illustrate with an example.

For example, if one wishes to combine the SHP sample with the survey of the canton of Vaud (another example would be the LIVES Cohort survey) the following procedure should be followed.

- The reference population of the sample of Vaud is the population of persons aged older than 14, living in private households in the canton of Vaud. Hence, it is different from the reference population of the SHP samples. The sample of Vaud can be combined with the whole SHP sample or with only a part of it (for example only the respondents from the SHP who lives in Vaud). In the case of an SHP subsample, first the new weights concerning the subsample of the SHP need to be calculated (see point 2). If the sample of

Vaud should be combined with the whole SHP sample, there is an additional step. First the SHP subsample with the same reference population as the Vaud sample is selected, and the weights are recalculated following the above mentioned procedure. This of course means that the weights of the SHP subsample and the weights in the rest of the SHP sample no longer add up to the total sample size (or population size, depending on the weights used). The rest of the SHP sample (so everyone who does not fall in the reference population of the Vaud sample) can also be considered as a subsample, so the computation of the new weights can be done as written above, in point 2, recalculating the weights to reflect the new sample size.

- When combining the SHP sample with the LIVES Cohort survey it is important to note that the reference population of the LIVES Cohort survey is the population of persons aged between 14 and 26 at the first wave of the sample (2013), living in private households in Switzerland. The combination of this sample with the SHP samples can be done in the same way as the combination of SHP samples and the Vaud sample as described above.