Subject: Sample Weight for Merged Dataset
Posted by Helen on Sat, 20 Nov 2021 21:39:46 GMT
View Forum Message <> Reply to Message

Hi, I will appreciate your expert advice about the sample weight, strata, and cluster variables that should be used to create an SPSS complex Samples analytical plan file. I have merged the male and female datasets from the eight sub-Saharan countries that I am including in my study. I merged the sample weight variables, strata, and PSU variables for the male and female datasets. I am worried that there is no difference between the population estimate and unweighted counts from Complex samples frequencies. Here are the syntaxes I used to create the Complex Samples plan file. I have also attached the SPSS output for the frequencies.

```
Frequencies Variables=V005 MV005 GENDER.
Compute totalsamplewegiht=0.
If (GENDER=0) totalsampleweight=V005.
If (GENDER=1) totalsampleweight=MV005.
Execute.
Frequencies Variables=totalsampleweight V005 MV005 GENDER.

COMPUTE WGT= totalsampleweight/1000000.

* Analysis Preparation Wizard.
CSPLAN ANALYSIS
  /PLAN FILE='D:\Angola\Angola_Complex_samples_File2.csaplan'
  /PLANVARS ANALYSISWEIGHT=WGT
  /SRSESTIMATOR TYPE=WR
  /PRINT PLAN
  /DESIGN STRATA=V023Merged CLUSTER=V021Merged
  /ESTIMATOR TYPE=WR.
```

File Attachments

1) Descriptives_Age_Gender.spv, downloaded 205 times

Subject: Re: Sample Weight for Merged Dataset
Posted by Bridgette-DHS on Tue, 23 Nov 2021 15:09:49 GMT
View Forum Message <> Reply to Message

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

We do not understand what you mean by "I am worried that there is no difference between the population estimate and unweighted counts from Complex samples frequencies." By "population estimate" do you mean the "weighted counts"? Your SPSS output does show a difference EXCEPT that the totals match. The total weighted number and the total unweighted number match--as they should, because DHS weights include a multiplier to force the mean weight to be 1 (with a multiplier of 1000000). The weighted counts are not population estimates. The weighted means, etc., are population estimates, but not the counts.

You have appended the files from the different surveys. There is a difference between merging and appending. What you have done is correct but should be described as appending, rather than merging.

The "v" variables for women almost always correspond exactly with "mv" variables for men. If you just drop the "m" for men, you can keep the "v" variables and not have to invent new names.

The cluster and stratum variables (v021 and v023) will repeat some values in the different surveys but have to be revised so the identifiers are distinct. This is done with "egen group" in Stata. In SPSS you have to do something else to renumber those variables. Let us know if you are not sure how to do that.

There are options for what to do with the weights when surveys are appended. Have you considered those options? Let us know if this is not clear.

Subject: Re: Sample Weight for Merged Dataset
Posted by Helen on Tue, 23 Nov 2021 20:38:20 GMT
View Forum Message <> Reply to Message

Hi,
Thank you very immensely for the expedited response. Yes, I was referring to the weighted counts. The similar counts in the SPSS output now make more sense.
To clarify further, I plan to examine outcomes for all and also carry out a disaggregated analysis for males and females. So, I renamed the variables of interest in both data files, with similar data characteristics, created a gender variable in each dataset, and then merged (appended ?). Sure, I will appreciate more information about renumbering the cluster and stratum variables in SPSS. Also, I am curious to learn the available options for handling the sample weights.

Take care.

Subject: Re: Sample Weight for Merged Dataset
Posted by Bridgette-DHS on Wed, 24 Nov 2021 18:11:43 GMT
View Forum Message <> Reply to Message

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

First, regarding the re-numbering of v021 and v023. Here's a clumsy way to do it. Say that your surveys are numbered 1 through 8, and that the codes for v021 and v023 are always less than 1,000 (you have to check that, and modify if there are exceptions). Then you can construct new variables, cluster_id and stratum_id, defined with cluster_id=survey*1000+v021 and

stratum_id=survey*1000+v023. These new variables will have unique numbers for the clusters and strata in the pooled data file.

If you leave the weights as they are now, then the total weight for each survey will just be the total sample size for that survey. This is probably the least acceptable thing to do, because sample sizes are determined by many different arbitrary considerations, such as the budget for the survey. There are two alternatives. One is to multiply v005 by a scaling factor such that the total weight for a survey becomes proportional to the population of the country at the time of the survey. The UN Population Division website gives population estimates. But if you do this, you will find that the results are dominated by the largest country, sometimes overwhelmingly. The second alternative is to re-weight with a survey-specific factor that gives equal weight to each country. I personally prefer this but not everyone does. There is a larger problem with pooled estimates. The surveys were done at different times and we almost never cover all the countries in a region or even a sub-region. Within DHS, we only construct a pooled estimate with many countries when studying something that is relatively rare within the individual countries. (Pooling successive surveys from the same country is more defensible.)

A colleague, Mahmoud Elkasabi, points out that men are often subsampled, and they often have a different age range than women. You need to take that into account. If the men have been subsampled, you need to scale up mv005. For example, it's a 50% subsample, multiply mv005 by 2. If you explore the forum, you should be able to find old postings that discuss all these issues. Good luck.
