
Subject: Appending Multi-phase Nigerian DHS Surveys

Posted by [Oby](#) on Fri, 30 Aug 2024 07:41:15 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello,

I'm working with multiple NDHS datasets from different survey years (2003, 2008, 2013, and 2018) and have encountered some challenges. I'd appreciate your guidance on the following issues:

1. Inconsistent Value Labels Across Years:

When attempting to append these datasets using `bind_rows()`, I receive warnings about conflicting value labels for certain variables.

Question: Should I convert these labelled variables to factors using a function like `as_factor()` before appending, or is there a better approach to standardize value labels across these different datasets? What is the best practice for ensuring that the labels are consistent before appending?

2. Creating a Survey Design with Different Sampling Designs:

Each of the four surveys has a different sampling design. After appending the datasets, I need to create a combined survey design object for analysis.

Questions: a.) How should I go about creating a survey design object that appropriately accounts for the different sampling designs across the four surveys?

b.) Are there other specific adjustments or considerations I need to make when combining these datasets for analysis?

Thank you for your assistance! I look forward to your advice on these issues.

Regards,

Oby

Subject: Re: Appending Multi-phase Nigerian DHS Surveys

Posted by [Bridgette-DHS](#) on Fri, 30 Aug 2024 19:35:59 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

There have been many forum postings on both of these topics. Please search with keywords.

There is no automated way to reconcile changes in coding. Sometimes there is just a change in the numbering, but often the differences are due to consolidation or division of categories and there is no simple way to recode. Sometimes there is a change in classification--for example, sources of water that are classified as "improved" in one survey may be "unimproved" in another survey. This is a complication when analyzing multiple surveys that cannot be avoided.

You can more easily handle the different sample designs. It is simplest to construct a categorical variables for "survey" that takes the values 1, 2, 3, 4, for example. The clusters are v001 (or

v021). Then you construct a combined variable with "egen cluster_id=group(survey v001)". You need to do the same thing with the strata. Then use cluster_id and stratum_id in the svyset command.

Subject: Re: Appending Multi-phase Nigerian DHS Surveys

Posted by [Oby](#) on Tue, 01 Oct 2024 20:11:42 GMT

[View Forum Message](#) <> [Reply to Message](#)

Thank you for your response bridgette. I have a follow-up question. My analytic sample is currently married women(15-49 years) and according to DHS guidelines, I am to denormalise the weight using population data from UNPD World Population Prospects. However, there is no population data for currently married women in Nigeria. Since I don't have direct population data for currently married women in Nigeria, I plan to use the proportion of currently married women from the DHS data to estimate it, and then apply that proportion to the total population of women aged 15-49 for each survey year.

This is my code below (I am using R)

```
# Denormalising the weight using population data from UNPD World Population Prospects-----
```

```
# divide v005 by 1 million
```

```
ir_2003 <- ir_2003 %>% mutate(weight_norm = v005 / 1000000)
```

```
ir_2008 <- ir_2008 %>% mutate(weight_norm = v005 / 1000000)
```

```
ir_2013 <- ir_2013 %>% mutate(weight_norm = v005 / 1000000)
```

```
ir_2018 <- ir_2018 %>% mutate(weight_norm = v005 / 1000000)
```

```
# total population of 15-49 women for each survey year
```

```
pop_2003 <- 31544644
```

```
pop_2008 <- 35882027
```

```
pop_2013 <- 41018918
```

```
pop_2018 <- 47146390
```

```
# Estimate the population of currently married women for each survey year
```

```
pop_married_2003 <- pop_2003 * prop_married_2003 # (prop_married means proportion of currently married)
```

```
pop_married_2008 <- pop_2008 * prop_married_2008
```

```
pop_married_2013 <- pop_2013 * prop_married_2013
```

```
pop_married_2018 <- pop_2018 * prop_married_2018
```

```
# Denormalize the weights using the estimated population of currently married women
```

```
ir_2003 <- ir_2003 %>% mutate(weight_denom = weight_norm * pop_married_2003 / sum(weight_norm))
```

```
ir_2008 <- ir_2008 %>% mutate(weight_denom = weight_norm * pop_married_2008 / sum(weight_norm))
```

```
ir_2013 <- ir_2013 %>% mutate(weight_denom = weight_norm * pop_married_2013 / sum(weight_norm))
```

```
ir_2018 <- ir_2018 %>% mutate(weight_denom = weight_norm * pop_married_2018 /
```

sum(weight_norm))

Would this be accurate, or should I just use the data for the total population of women without applying the proportion?

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Bridgette-DHS](#) on Wed, 02 Oct 2024 10:44:53 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

The IR files for the Nigeria surveys include all women (de facto residents and age 15-49), regardless of marital status. It will be easier if you just change the weights (v005) in the IR files, using the estimated numbers of women age 15-49 in the population, regardless of marital status. THEN select the currently married women in the data files.

I think your approach is fine but is equivalent to what I describe, which is simpler. Whenever there are questions about whether to use weights A or weights B, I usually check whether the results are different. Usually there is no difference, or the difference is negligible and that's what I would expect here.

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Oby](#) on Wed, 02 Oct 2024 13:15:12 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Bridgette,

Thank you so much for getting back to me so quickly. I will try the two approaches and see if they return similar results. Thank you once again!!

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Oby](#) on Fri, 04 Oct 2024 16:52:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Bridgette,

I have a follow-up question about the survey design (please note that I am working with the Nigerian DHS for 2003, 2008, 2013 and 2018 ir file). I would like clarification on the correct approach for creating the strata_id for my pooled dataset. Based on my understanding from the survey reports, the stratification for all four surveys seems to be based on urban and rural areas within each state. However, v023 for 2003 is based on regions and urban/rural residence, for

2008 it's based on only states. Do I use the v023 that way?

I am considering constructing the strata_id in my pooled dataset by combining state (sstate) and urban/rural residence (v025) for all four surveys and then grouping that with the survey variable to create a unique strata_id for each survey phase.

Could you please confirm if this is the correct approach to creating the strata_id for the pooled dataset, or if there is another method?

Thank you for your assistance.

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Bridgette-DHS](#) on Mon, 07 Oct 2024 12:33:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

In general, the strata are constructed as combinations of v024 and v025. In most of the recent surveys, the strata are given by v022 or v023 (and they are the same) as combinations of v024 and v025.

The Nigeria 2003 survey is an exception. The sampling clusters are given by v023. This survey and other exceptions appear in the following file (attached below).

If you combine surveys, such as all the NG surveys, you do need to construct unique identifiers for combinations of cluster x survey and strata x survey. You can do this with egen group, as has been described in other forum posts. Note that even though the strata may be defined in the same way in the surveys that are being pooled, you do need unique identifiers because the populations are different.

File Attachments

1) [Survey_strata.do](#), downloaded 67 times

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Oby](#) on Mon, 07 Oct 2024 16:10:06 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your response Bridgette. The do file you provided says I should use v023 for 2003 and v022 for 2008, 2013 and 2018. However, for 2013 and 2018, v022 is not the same as combining v024 and v025 (i.e region and rural/urban residence), but rather it is the combination of v024, sstate and v025 (i.e region, state and rural/urban residence).

Also, for 2013, I noticed that there are 73 strata instead of 74 since we have 37 states (including the capital territory) in Nigeria, so merging it with v025 should ordinarily result in 74 strata as in

2018.

So I'm a bit confused as to whether to use the v023(for 2003) and v022(for 2008, 2013 and 2018) that way.

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Bridgette-DHS](#) on Tue, 08 Oct 2024 14:09:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

My previous response said "In general, the strata are constructed as combinations of v024 and v025. In most of the recent surveys, the strata are given by v022 or v023 (and they are the same) as combinations of v024 and v025. The Nigeria 2003 survey is an exception. The sampling clusters are given by v023. This survey and other exceptions appear in the following file (attached below)."

What I said is correct. I only took the time to look up the 2003 survey, because you had singled it out. You looked up the other three surveys and found that they, too are exceptions. Glad you did that.

The strata specifications in "Survey_strata.do" should be considered to be definitive.

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Oby](#) on Mon, 21 Oct 2024 08:08:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Bridgette. I have a follow-up question on the denormalisation of weights, you said that "it will be easier if you just change the weights (v005) in the IR files, using the estimated numbers of women age 15-49 in the population, regardless of marital status. THEN select the currently married women in the data files."

In the case where I am using complete case analysis and have removed 1803 out of 85274 records with missing/na values. Won't this affect the weight?

Subject: Re: Appending Multi-phase Nigerian DHS Surveys
Posted by [Bridgette-DHS](#) on Tue, 22 Oct 2024 11:57:26 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Yes, you can re-calculate the weights with those cases removed. The difference between removing them or leaving them in will be very small, in terms of any results.
