Subject: Survey Design Issue
Posted by hamid on Fri, 19 Mar 2021 21:44:45 GMT
View Forum Message <> Reply to Message

Recently I am experiencing some issues in setting the DHS survey design in R using the Afghanistan 2015 Standard DHS7.

After loading the Stata format individual recode dataset (AFIR71FL) into R:

```
library(foreign)
library(survey)

data <- read.dta("AFIR71FL.DTA")
data$wt <- data$v005 /1000000
```

I simply run the following command (as indicated in the DHS documentation):

```
DHSdesign<-svydesign(id = data$v021, strata = data$v023, weights = wt,  data=data)
```

and I get the following error:

```
Error in svydesign.default(id = data$v021, strata = data$v023, weights = wt, data = data) :
  Clusters not nested in strata at top level; you may want nest=TRUE.
```

which is basically telling me that PSU ids are not unique across Strata (Province +rural/urban).

Some additional notes:
- The same code worked just fine over the past month (working on it everyday);
- I tried a fresh install of R + packages without any success;
- R & packages are up to date;
- I get the same error using different machines;
- The AfDHS file in use is the most recent;
- Using the option "nest=T", as suggested by R, is not a working solution as it creates problems when running "svyglm" regressions.

I am wondering if this is a R:suvery package bug or something else.

Thanks,
Hamid

Subject: Re: Survey Design Issue
Posted by Bridgette-DHS on Mon, 22 Mar 2021 20:32:57 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Senior Sampling Specialist, Mahmoud Elkasabi:

Apparently this is due an error in the IR dataset (I assume it exists in the other ones as well; I haven't checked). One woman in cluster 476 is coded as rural in v025 although the cluster is urban. This causes the error message.

Here are two possible solutions:

1- Use the nest option after you re-construct v021 and v023:

```
IRdata$STRAT <- as.integer(factor(with(IRdata, paste(v024, v025))))
IRdata$CLUST <- as.integer(factor(with(IRdata, paste(v023))))

DHSdesign<-svydesign(id = IRdata$CLUST, strata = IRdata$STRAT, weight = IRdata$v005,
data=IRdata, nest = TRUE)
```

2- Recode a new v025 variable (say v025r) where you assign all cases with v021=476 to v025r=1 (otherwise v025r=v025) and then proceed with the code below:

```
IRdata$STRAT <- as.integer(factor(with(IRdata, paste(v024, v025r))))
IRdata$CLUST <- as.integer(factor(with(IRdata, paste(v023))))

DHSdesign<-svydesign(id = IRdata$CLUST, strata = IRdata$STRAT, weight = IRdata$v005,
data=IRdata)
```

Subject: Re: Survey Design Issue
Posted by hamid on Tue, 23 Mar 2021 09:12:18 GMT
View Forum Message <> Reply to Message

Thanks a lot. This actually works!

I prefer the second approach better. This issue does not apply to the couples recode (probably because this woman is not listed there).
As already written, the nest=T option creates problems when running some R function (e.g., lmtest::waldtest would produce an unequal sample between models error).

Kind regards,
Hamid
