

Understanding Society User Support - Support #50

Comparison to LFS

05/10/2012 10:58 AM - Sung Park

Status:	Closed	Start date:	05/10/2012
Priority:	Urgent	% Done:	100%
Assignee:	Redmine Admin		
Category:	Data inconsistency		
Description			
<p>Hello: I am looking at the report "APPENDIX: UNDERSTANDING SOCIETY DESIGN OVERVIEW" which I believe is Chapter 15 of the Early Findings Report (http://research.understandingsociety.org.uk/findings/early-findings). I was attempting to replicate the Understanding Society column results from Table 8 ("Comparison of Great Britain Labour Force Survey sample and Understanding Society general population sample on key characteristics").</p> <p>I used the following code to replicate the marital status distribution using the dataset indresp.dta, using the design weight for individuals as described in the report:</p> <pre>svyset a_psu_dv [pweight=a_psnenus_xd], strata(a_strata_dv) svy, subpop(if(a_month >0 & a_month<13 & a_emboost==0 & a_xtra5min_dv==0 & a_gpcomp==0 & a_country==1 & a_pmarstat>0 & a_pmarstat<7)):tab a_pmarstat</pre> <p>However, I am getting slightly different percentages than what is shown in the table. Is there something wrong with my criteria/code?</p> <p>Also, in order to replicate the sex and age distributions, do I need to combine the indresp.dta file with the youth file and use the same weights and criteria above?</p> <p>Many thanks for your help.</p>			

History

#1 - 05/10/2012 11:26 AM - Redmine Admin

- Status changed from New to In Progress
- Target version set to M1
- % Done changed from 0 to 40

Sung,

There could be different reasons to why you get slightly different results. The most important is probably versioning. We correct known data issues for the latest (re-)releases. Please ensure you are using the latest version (for M1 the full release dated November 2011). Please also provide your results in detail.

Also, in order to replicate the sex and age distributions, do I need to combine the indresp.dta file with the youth file and use the same weights and criteria above?

The INDALL file holds data on all enumerated adults, youths and children.
Jakob

#2 - 05/10/2012 12:01 PM - Sung Park

Hi Jakob:

I downloaded the data in Jan 2012, so I think it is the most current version. Using the INDALL file gives me results that are much closer in terms of %. However, the final sample Ns seem off (for example, my results below don't match the N=34,502 for the sex distribution). I show my results below. Many thanks!

```
svyset a_psu_dv [pweight=a_psnenus_xd], strata(a_strata_dv)
svy, subpop(if(a_month >0 & a_month<13 & a_emboost==0 & a_country<4)):tab a_sex

(running tabulate on estimation sample)
```

Number of strata	=	1319	Number of obs	=	60597
Number of PSUs	=	2639	Population size	=	31028112
			Subpop. no. of obs	=	31166
			Subpop. size	=	16038346
			Design df	=	1320

sex	proportions
male	.4831
female	.5169
Total	1

Key: proportions = cell proportions

Note: 457 strata omitted because they contain no subpopulation members.

#3 - 05/10/2012 12:30 PM - Redmine Admin

- % Done changed from 40 to 60

#4 - 05/18/2012 02:48 PM - Peter Lynn

Sung,

The early findings volume was published in Feb 2011 and the analysis was done in Oct-Nov 2010 using the data which formed the first data release in Dec 2010. That release only included the year 1 sample and did not include the ethnic boost sample. If you are using the Jan 2012 release, you will get much larger sample sizes and different percentages unless you exclude the EMBS and the year 2 sample. Also, I suspect that the table 8 analysis was unweighted, as the purpose was to show the representativeness of the responding sample.

Peter

#5 - 05/21/2012 11:11 AM - Sung Park

Thanks, this makes sense.

#6 - 05/21/2012 11:44 AM - Redmine Admin

- Status changed from In Progress to Closed

- % Done changed from 60 to 100