

---

Subject: how to reproduce malawi 2004 published sampling errors and deft values  
Posted by [ajdamico](#) on Tue, 29 Apr 2014 09:21:33 GMT

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

---

i am trying to figure out the stata (or any statistical language) setup required to match the statistics, standard errors, and design effects published in this official document:  
<http://dhsprogram.com/pubs/pdf/FR175/19AppendixB.pdf>

i have attached an easy-to-read PDF, a runnable do file, and the screen output when running that do file.

my comments throughout the script should explain all of my attempts at matching the standard errors and DEFT values for the..

Never married  
Children ever born  
Children ever born to women age 40-49

..rows of PDF page 5 of the appendix. (direct link - <http://dhsprogram.com/pubs/pdf/FR175/19AppendixB.pdf#page=5>). i am able to hit the standard errors, but never the DEFT values. and i am hitting the SEs without the jackknife technique, which is not what the paper says.

could i get some advice about the appropriate setup to hit these numbers on the nose? :)

possibly related- i noticed that the microdata i'm using (downloadable here [http://dhsprogram.com/data/dataset/Malawi\\_Standard-DHS\\_2004.cfm?flag=1](http://dhsprogram.com/data/dataset/Malawi_Standard-DHS_2004.cfm?flag=1)) have a date modified of 8/16/2011 but the report that this appendix b that i'm trying to replicate was published in december of 2005 (full report: <http://dhsprogram.com/pubs/pdf/FR175/FR-175-MW04.pdf#page=320>).. is it possible that any of the records were edited?

also possibly related- the paper says, "In the 2004 MDHS, there were 522 non-empty clusters. Hence, 521 replications were created." but the way i'm defining the clusters, i get 858 of them .. and stata indicates that means there are 858 replications calculated. perhaps i'm just missing some option?

thanks!!!!

---

### File Attachments

1) [20140429 malawi 2004 replication attempt.pdf](#), downloaded

1021 times

2) [20140429 malawi 2004 replication screen output.txt](#),  
downloaded 1353 times

3) [20140429 malawi 2004 replication attempt.do](#), downloaded 930  
times

---

---

Subject: Re: how to reproduce malawi 2004 published sampling errors and deft values

Posted by [Bridgette-DHS](#) on Tue, 06 May 2014 12:00:39 GMT

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

---

One of our Stata specialists is reviewing your post.

Thanks!

---

---

Subject: Re: how to reproduce malawi 2004 published sampling errors and deft values

Posted by [Trevor-DHS](#) on Mon, 12 May 2014 18:39:23 GMT

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

---

Thanks for your post and your detailed documentation on your questions. Firstly, for the design effect (Deft) estimation for proportions and means, you do not expect to get the exact numbers published in the DHS final report if you use a different software. Theoretically, the Deft is the standard error of an estimator with current survey design over the standard error of the same estimator if the sample was a simple random sample. The problem depends on how to estimate the standard error of the estimator if a simple random sample would be drawn. There is no standard method to produce this estimate, so the calculated Deft would have small differences.

Secondly, DHS surveys calculate the standard errors of simple estimators (proportion, mean, ratio) using Taylor linearization method as explained in the Appendix B of the DHS final report. Jackknife is only used for complex statistics that don't come down to a simple X/Y form. There are only a few statistics that we use Jackknife for - Age specific and total fertility rates, child mortality rates, and maternal mortality rates primarily. For the examples you gave we used Taylor linearization.

When I reviewed your approach to calculating the sampling errors, I came across a few issues:

1) We would now recommend that stratification to use is the stratification used in the sample design. The final report seems to indicate that the stratification was urban and rural areas within districts, with 17 of the districts grouped together into a "rest of Malawi" group. The dataset has a variable with the district codes used in the design (sdist) that can be used to create these strata, e.g.

```
egen strata = group(sdist v025)
```

Then you would use your svyset command:

```
svyset [pweight=weight], psu(v021) strata(strata)
```

We don't use the single unit parameter, and you don't need to with this specification as there are no strata with single PSUs in them..

This produces a slightly different (and more conservative) estimate of the standard errors than published.

2) If you wanted to produce the standard errors in the same manner as they were produced for the 2004 report, then you would need to use the implicit stratification approach that was previously used by DHS in earlier surveys, that is grouping neighboring PSUs into pairs or groups of three to form implicit strata (by contrast the stratification presented above is sometimes known as explicit stratification). V022 is supposed to contain that implicit stratification that was used at that time. Unfortunately I discovered an error in the creation of V022 in the recode dataset. I was though able to reconstruct the variable, and I have attached a .do file that contains the instructions for recoding v021 to produce v022 correctly. (See attached file: MV2004V022\_redo.do - make sure you drop the variable strata before using this file).

Once I use that file to recode v022 (creating the strata variable too), I can then set up my svyset command as before:

```
svyset [pweight=weight], psu(v021) strata(strata)
```

(note that we don't use v025 with v022 in the creation of the strata - it is just based on recoding v021).

Now when I use

```
svy: mean v201
```

I get a standard error and DEFT that matches the final report (the confidence intervals are slightly different as DHS uses +/-2SE, while Stata and other software often use +/-1.96SE).

Concerning your comment about the number of replicates when using Jackknife, there are two reasons for the difference: from combining v022 with v025 (which is not necessary), and v022 is incorrectly coded. With the correct coding of v022 given in the attached do file you should find a replicate for every cluster when using Jackknife.

I hope this helps.

## File Attachments

1) [MW2004V022\\_redo.do](#), downloaded 787 times

---

---

Subject: Re: how to reproduce malawi 2004 published sampling errors and deft values

Posted by [kate2](#) on Tue, 28 Oct 2014 20:46:20 GMT

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

---

This wasn't me asking the original question but this is a very useful reply. Thanks.

---