
Subject: Stratification

Posted by [hlyons](#) on Tue, 04 Feb 2014 01:09:56 GMT

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

Hello -- I was wondering if there are some known "strata" variable errors in some of the surveys? My example is a random one, Zambia DHS IV:

Using the children's recode and R for computing, here is a somewhat fake example at the bottom -- fake in the sense the outcome (urban/rural split for children) is not something I'm really looking at. Basically, it looks like there are too many strata (v022). A less theoretical example would be DPT3 coverage nationally for urban and rural areas -- I think the standard errors for urban in the final report are more compatible with using v022. It's hard to say for sure because I do get some differences trying to duplicate the results in the final report. Maybe I'll write another post about that later...

Thoughts, anybody?

Thanks!

Hil

```
# R example
```

```
library(survey)
```

```
tmp.data = read.dta("ZMKR42FL.DTA")
```

```
# PSU's, checks out: 320
```

```
length(unique(tmp.data$v021))
```

```
# province, check out: 9
```

```
length(unique(tmp.data$v024))
```

```
# province and urban/rural combinations: 18
```

```
nrow(unique(tmp.data[,c("v023","v025")]))
```

```
# strata: 153 instead of 18
```

```
length(unique(tmp.data$v022))
```

```
# example of standard errors under two designs
```

```
# first, stratify on v022; second, on province and u/r combination
```

```
DHSdesign.v022 = svydesign(id = tmp.svy$v021, strata=~tmp.svy$v022, weights =  
tmp.svy$v005/1000000, data=tmp.data)
```

```
DHSdesign.prov.ur = svydesign(id = tmp.svy$v021, strata=~tmp.svy$v023+tmp.svy$v025,  
weights = tmp.svy$v005/1000000, data=tmp.data)
```

```
# proportion urban amongst children
```

```
svymean(~v025, design = DHSdesign.v022) #SE = 0.0099
```

```
svymean(~v025, design = DHSdesign.prov.ur) #SE = 0.0227
```
