
Subject: District-Level WaSH Indicators

Posted by [Mikaela22](#) on Fri, 01 Mar 2024 13:45:34 GMT

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

Hello!

Project: I am combining DHS WaSH indicators from the 2011 Mozambique DHS with data from a cluster-randomised trial in Mozambique assessing the performance of various treatment strategies on Schistosomiasis prevalence. I am attempting to model individual-level infection status after 5-years of mass-drug administration to see if there is any effect modification of the treatment strategy (villages were randomised to different treatment strategies) by different WaSH indicators at the district-level, specifically using an improved water / sanitation source.

I will be using multi-level logistic regression to capture the clustering of the data i.e., (1) individuals in (2) villages (the treatment-level) in (3) districts.

The cluster-randomised trial was conducted in one province in Mozambique, so I am only working with 8 districts and attempting to calculate a district-level indicator e.g., percentage of households in that district using an improved water source. I have used GPS data to locate the clusters in corresponding districts and have followed the suggested methodology (the complex sample design weighting) to generate estimates. However, as has been extensively discussed previously, the SEs are too large to be usable.

I propose the following methodology to resolve this and would appreciate some input:

- Use a bootstrap (I saw a link to a wild bootstrap mentioned in a previous post?) to calculate more precise standard errors - how would I go about using the sampling weights here?
- Use weights within the multi-level logistic regression model to account for the uncertainty around the district-level estimates.

I understand that using DHS data in this way to generate district-level indicators is not ideal, however, this project is more for hypothesis generation and identifying areas for future research.

Do you have any comments on what I have proposed, or is there anything else I should be thinking about in terms of using this data and conducting this analysis in the best way?

I appreciate any feedback!

Kind regards!

Subject: Re: District-Level WaSH Indicators

Posted by [Janet-DHS](#) on Wed, 06 Mar 2024 21:18:02 GMT

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

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

The setup for a bootstrap that matches the sample design would be complicated. It's easier to

get the estimates with a model that includes svyset--which you are using. I will paste below the lines to do this. Just for an illustration, I use the Mozambique 2011 data, with subpopulation hv024=1 (Niassa). The outcome y is 1 if the source of drinking water is an unprotected well (hv201=32), which is the largest category. The model has no covariates. The lines show how to extract the proportion of households with y=1 in Niassa, as well as the lower and upper bounds of a 95% CI for that proportion. I show how to do this with logit or logistic models. You also get the standard error on the logit or odds scale but I would not recommend the se on the scale of a proportion (also not on the odds scale). CI yes, se no. Hope this helps.

* Open HR file, cases are households

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\MZHR62FL.DTA" , clear

* Specify outcome and subpopulation

gen y=0

replace y=1 if hv201==32

gen Niassa=0

replace Niassa=1 if hv024==1

* Prepare svyset

svyset hv001 [pweight=hv005], strata(hv023) singleunit(centered)

* Logit model

svy, subpop(Niassa): logit y

matrix T=r(table)

matrix list T

* Extract P, L, and U as saved results

* P, L, and U are the point estimate and the lower and upper bounds

- * of a 95% confidence interval for the proportion of households in
- * Niassa whose main source of drinking water is an unprotected well.

scalar b=T[1,1]

scalar P=exp(b)/(1+exp(b))

scalar b=T[5,1]

scalar L=exp(b)/(1+exp(b))

scalar b=T[6,1]

scalar U=exp(b)/(1+exp(b))

scalar list P L U

- * Equivalent using logistic

svy, subpop(Niassa): logistic y

matrix T=r(table)

matrix list T

scalar odds=T[1,1]

scalar P=odds/(1+odds)

scalar odds=T[5,1]

scalar L=odds/(1+odds)

scalar odds=T[6,1]

scalar U=odds/(1+odds)

scalar list P L U