Subject: Nepal DHS 2022 - children's Disability
Posted by UnseenDisability on Sat, 17 Feb 2024 15:54:45 GMT
View Forum Message <> Reply to Message

Hello all,

I am new to using DHS data and will do some data analysis for a project using STATA, which I've been learning for a few years but am no expert.

I will do a project using the DHS Nepal 2022 dataset. For my ethics application, I need to know how many children (>5, <18) in the dataset have some degree of hearing loss. I'd like to do a crosstab of the variables ha50 (under age 18) and hdis4 (difficulty hearing) and then separately ha50 and hdis3 (wear hearing aid). I have not otherwise merged any datasets or done any analysis, since I have not received the ethics approval yet. I just need to find out how many children in the dataset suffer from some degree of hearing loss, to determine whether or not I should include data from other DHS surveys (other countries). I am using the NPPR82DT dataset. All of the variables I need to use are in this dataset and none of them are empty, but when I try to crosstab them, I get the error "no observations".

Any ideas? The disability data are not commonly used, and I couldn't find anything in the other forums.

Thanks in advance

S

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by Janet-DHS on Wed, 28 Feb 2024 18:59:55 GMT
View Forum Message <> Reply to Message

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

Here is the distribution of hdis4, the question about difficulty hearing, in that data file. About 3% of children are reported by the household respondent to have some degree of difficulty. The disability variables are under-used, as you say, and I encourage you to try to work with them. The table below gives weighted frequencies.

. tab hdis4 hv104 if hv105>5 & hv105<18 [iweight=hv005/1000000], col

+-------------------+

| Key |

|-------------------|

| frequency |

| column percentage |

+-------------------+

| sex of household

have difficulty | member

hearing | male female | Total

---

no difficulty hearing | 3,281.377 3,280.552 | 6,561.929

| 97.23 97.43 | 97.33

---

some difficulty | 77.762576 71.904984 | 149.66756

| 2.30 2.14 | 2.22

---

a lot of difficulty | 14.765722 14.205368 | 28.97109

| 0.44 0.42 | 0.43

---

cannot hear at all | 1.030425 .42826 | 1.458685

| 0.03 0.01 | 0.02

Total | 3,374.936 3,367.091 | 6,742.027

| 100.00 100.00 | 100.00

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by UnseenDisability on Fri, 28 Jun 2024 08:44:17 GMT

View Forum Message <> Reply to Message

Thank you so much for this (much earlier) response. I am now working with the data and have found that fully half of the hearing disability variable data is missing. Is this correct? I am not sure if I'll be able to use the data with this level of missingness, and may have to use other datasets. Do you have any advice on the dataset?

Thanks in advance

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by Janet-DHS on Tue, 02 Jul 2024 14:32:45 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:
These are not "missing" cases, they are "not applicable" (NA) cases. If you look at the household questionnaire in appendix E, pages 616-617, you will see that the disability module was only applied in a subsample of households, and never to children under 5. Individuals who were not in the subsampled households, or were age <5, were given an NA code (a dot in Stata). The cases that are not NA are a representative sample of the population.

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by UnseenDisability on Tue, 02 Jul 2024 14:55:46 GMT
View Forum Message <> Reply to Message

Thank you for your response! I am still a bit baffled, though, because when I look at the same variable in the Pakistan PR file, I get very different output. The following output is only for children aged 5 to 17:

*observe missing values in Nepal PR file (2022) in hdis4

. svy: tabulate hdis4, missing
(running tabulate on estimation sample)

Number of strata = 14 Number of obs = 15,276
Number of PSUs = 476 Population size = 14,646.99
Design df = 462

---
have | difficult | y hearing | proportion
---|---
no diffi | .488
some dif | .0107
a lot of | .002
cannot h | 1.0e-04

```
        . |      .4992
          |
    Total |         1
----------------------
Key: proportion = Cell proportion

 *observe missing values in Pakistan PR file (2017) in hdis4

.
.
.
. svy: tabulate hdis4, missing
(running tabulate on estimation sample)

Number of strata =  12                        Number of obs   =     26,738
Number of PSUs   = 458                          Population size = 25,147.104
                                    Design df       =        446

----------------------
have      |
difficult |
y hearing | proportion
----------+-----------
 no diffi |      .9869
 some dif |      .0082
 a lot of |      .0024
 cannot h |      .0012
 don't kn |    6.5e-05
        . |      .0013
          |
    Total |         1
----------------------
Key: proportion = Cell proportion

Note: 4 strata omitted because they contain no population members.
```

Can you please help me understand why the output for "." is so different? The variables and values seem to otherwise be the same in these two surveys...

Here, I've done the analysis a bit differently (I have recoded hdis4 into hearing disabled or not, and age_cat==2 is children aged 5-17), but hopefully it can be understood as well:

```
. svy:  tabulate survey_id hearing_cat if age_cat==2, percent missing
(running tabulate on estimation sample)

Number of strata =    16                        Number of obs   =     47,764
Number of PSUs   = 1,037                          Population size = .570221056
                                    Design df       =      1,021
```

	hearing_cat				
survey_id	hears no	difficul	don't kn	.	Total
Nepal	21.92	.5755	0	22.42	44.91
Pakistan	54.36	.6506	.0036	.0699	55.09
Total	76.28	1.226	.0036	22.49	100

Key: Cell percentage

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by Janet-DHS on Wed, 03 Jul 2024 19:21:01 GMT
View Forum Message <> Reply to Message

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

I would find it much easier, when first looking at new data, to do just "tab hdis4" for the unweighted distribution and "tab hdis4 [iweight=hv005/1000000]" for the weighted distribution. You don't gain anything with the full svyset adjustments (for tables).
I believe your problem is with the stratification adjustment in svyset. Please try adding "singleunit(centered)" to the end of your svyset statement. I think "svy: tab" will then work in both of these surveys.

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by UnseenDisability on Thu, 11 Jul 2024 09:28:04 GMT
View Forum Message <> Reply to Message

Just to follow up, does it look as though I am doing something wrong here, or is there truly 49% missing data for the variable hdis4 in the Nepal 2022 PR dataset?

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by Janet-DHS on Tue, 16 Jul 2024 19:22:32 GMT
View Forum Message <> Reply to Message

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

If you look at the top of page 616 in the final report, you will see that there was subsampling in this survey, and the disability module was only applied in the households that were selected for the men's interview. This was half of all households, identified with hv027=1. It was also only given to household members age 5 and above.

In the Nepal 2022 survey, all of the disability variables are coded with a dot (in Stata), for Not

Applicable or NA, if hv027 is 0 or hv105 is <5.

Subject: Re: Nepal DHS 2022 - children's Disability
Posted by UnseenDisability on Tue, 16 Jul 2024 20:12:32 GMT
View Forum Message <> Reply to Message

Thank you so much for this explanation!