
Subject: Difference between data and report
Posted by [Aamna](#) on Sun, 02 Oct 2022 01:39:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

I am doing an analysis by using Pakistan's children data set for 2017-2018 (PKKR71FL). I conducted crosstabs to ensure some numbers. I compared my raw data set's low birth weight numbers with the country's survey report. I have not made any changes in the dataset. I conducted crosstabs on weighted and unweighted data, but the results were the same.

I found discrepancies between the two. In total, I have 11 observations less than the survey report's total. I also saw that there was a discrepancy at the category levels. I do not understand why there is a difference between the two. Am I using the wrong dataset, or is there any other issue?

Here is an example of one cross tab. (tab m19 v101, mi). I did it without missing too.

Raw Dataset Survey Report (table 10.1 last column)

Punjab 468 961
Sindh 495 514
Baluchistan 177 117
KPK 136 59
ICT 385 40
FATA 20 1
Total 1681 1692

I am so confused and stressed concerning this difference. My work has reached a standstill due to this issue. I really appreciate your help regarding this issue.

Aamna

Subject: Re: Difference between data and report
Posted by [Janet-DHS](#) on Tue, 04 Oct 2022 15:09:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member Tom Pullum:

The Stata code pasted below matches table 10.1 for the categories of v101 (which is the same as v024):

* construct table 10.1 in the report on the Pakistan 20117-18 survey
use "...PKKR71FL.DTA", clear

```
label list M18  
gen size=.  
replace size=1 if m18==5  
replace size=2 if m18==4
```

```
replace size=3 if m18<=3
replace size=4 if m18==8 | m18==.
label define size 1 "very small" 2 "< average" 3 "average + " 4 "DK/missing"
label values size size
tab v101 size [iweight=v005/1000000], row
```

```
label list M19
gen weight_measured=0
replace weight_measured=1 if m19<9996
label define noyes 0 "No" 2 "Yes"
label values weight_measured noyes
tab v101 weight_measured [iweight=v005/1000000], row
```

```
gen low_birth_wt=0 if weight_measured==1
replace low_birth_wt=1 if weight_measured==1 & m19<2500
label values low_birth_wt noyes
tab v101 low_birth_wt [iweight=v005/1000000], row
```

Subject: Re: Difference between data and report
Posted by [Aamna](#) on Wed, 05 Oct 2022 02:19:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

Thank you for the reply. I ran the codes, and it created the variables. Unfortunately, it has not resolved my issue. I have attached the table from the survey report as a file. I have put an arrow beside the column I am trying to replicate with my data. I am still not able to get the same numbers. The number of births is still not similar to the table in the survey report. What am I doing wrong? Why am I still not getting the number of birth weights recorded for each region? Is there a way I can find out how many birth weights were recorded for each region? I appreciate your help again.

Aamna

File Attachments

1) [Low Birth Weight.docx](#), downloaded 161 times

Subject: Re: Difference between data and report
Posted by [Janet-DHS](#) on Wed, 05 Oct 2022 13:32:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member Tom Pullum:

The Stata code I gave yesterday has three parts. The third part produces the percentages and the n's for the last two columns of table 10.1. If you look at the Stata results, you will see weighted n's that round to 961 for Punjab, 514 for Sindh, etc. I gave the steps for region (v024=v101) but the modifications for wealth quintile, etc., should also match the table.
