Subject: Pakistan DHS from 1990 until 2017
Posted by dfigueiredo on Fri, 09 Aug 2024 16:51:03 GMT
View Forum Message <> Reply to Message

Hello,

I am using the Pakistan DHS from 1990 until 2017 and I have two questions regarding these:

1) Comparing the number of observations in the raw data with the ones reported in the DHS reports, I found several inconsistencies.

This is particularly true when looking at regions, separately. One example:

In the raw data there are 9177 children under 5 years old, 3705 in Punjab. If we restrict to children alive, then we have 8448 children and 3403 in Punjab. In the DHS report of 2006-07, table 10.5, there are in total 8367 children under 5 years old (presumably alive) and 4689 are in Punjab.

This discrepancy in the number of observations happens in all years, so I am wondering: Is there a reason for this difference? Or am I looking at the wrong numbers?

2) Is there any explanation for the fact that in 2012 and 2017 surveys, most of the observations (more than 2/3) have missing information for weight and height? The reason given in the survey is "no measurement found in household". What does this mean, exactly?

Thank you and kind regards,
Dalila

File Attachments
1) dhs_pk_2006_children0-5_alive.png, downloaded 284 times
2) dhs_pk_2006_children0-5_all.png, downloaded 271 times

Subject: Re: Pakistan DHS from 1990 until 2017
Posted by Janet-DHS on Tue, 13 Aug 2024 18:41:12 GMT
View Forum Message <> Reply to Message

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

Your "raw" numbers appear to be unweighted. The numbers in the reports are weighted. Differences between weighted and unweighted counts are greatest for region (v024) and strata (v023) because the sample design tended to over-sample the smaller regions/strata and under-sample the larger regions/strata. There have been many posts on the use of weights in DHS data.

It is more conventional to reserve the term "raw" for unedited data. DHS does not distribute data that are raw in that sense.

The "missing" responses for height and weight in the two most recent surveys are Not Appliable because of subsampling. In the HR and PR files, hv044 is 1 if the household was selected for the domestic violence module, and 0 otherwise. The selection extended to other topics, including anthropometry. If hv044 (v044 in the IR and KR files) is 0, the anthropometry variables have a dot (.) for Not Applicable or NA. You can confirm this, for example, with the following Stata line in the PR file: tab "hv044, summarize(hc2) means obs nolabel".