

---

Subject: Pakistan: Standard DHS, 2006 Dataset  
Posted by [yiningw](#) on Mon, 15 Feb 2021 14:52:37 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear DHS Representative,

I am working with the Pakistan 2006 dataset, specifically PKPR53FL.DTA and had a question about the education indicators. For individuals recorded between ages 5 - 18, only ~16000 out of ~230,000 students are recorded as being currently in school (var hv121). It seemed odd because the numbers for the 1990, 2012 and 2017 are much closer in estimate (eg., ~21000 out of ~34000 in 2012). The difference in 2006 is a factor of 10.

In addition, there is discrepancy in the 2006 data where observations for highest education attainment indicator can be primary school or higher, but other schooling related indicators (hv12x) are 0.

I could not find anything in the documentation that explains this discrepancy, and wanted to reach out to seek some clarification on these variables.

Thank you!

---

---

Subject: Re: Pakistan: Standard DHS, 2006 Dataset  
Posted by [Bridgette-DHS](#) on Thu, 18 Feb 2021 19:56:43 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The Pakistan 2006 survey had a relatively complex design, described in the paragraphs from the main report that I will paste below. Most households had a "short" questionnaire, and about 10% (10 out of 105) had a "long" questionnaire. Much more information was collected in the households with the long questionnaire, and they were the only households that included interviews with women.

I cannot locate a variable that conclusively specifies which households had the long questionnaire. Normally, I would expect that children in the households with the short questionnaire would have a dot in Stata (.), for NA or Not Applicable, for variables such as v121. However, it appears that the NA children were given a 0 rather than a dot.

If a woman in a given household was eligible for interview on all other criteria (hv104=2, hv105>=12, hv105<=49, hv103==1, and hv115>0; this was an ever-married survey and the age range started at 12 rather than 15) BUT she has hv117=0, then the household was in the "short" category. However, this is an indirect approach to determining whether the household was short or long, and it is not foolproof. I know there are some surveys in which a short vs long identifier was omitted and I hope this was not one of them. If I can find the identifier I will update this post by including it.

On page 5:

Quote:The second stage of sampling involved selecting households. In each sample point, 105 households were selected by applying a systematic random sampling technique. This way, a total of 102,060 households were selected. Out of 105 sampled households, ten households in each sample point were selected using a systematic random sampling procedure to conduct interviews for the Long Household and the Women's Questionnaires. Any ever-married woman aged 12-49 years who was a usual resident of the household or a visitor in the household who stayed there the night before the survey was eligible for interview.

On page 6:

Quote:The Short Household Questionnaire was administered in 92,340 households to list all the usual members and visitors. Likewise, the Long Household Questionnaire was used in the 9,720 households where the Women's Questionnaire was also administered. In addition to some basic information collected on characteristics like age, sex, marital status, education, and relationship to the head of the household of each person listed, another purpose of the two household questionnaires was to record births and deaths that occurred since January 2003 and, for verbal autopsies, to identify any death of child under age 5 since January 2005 and any death to a woman age 12-49 since January 2003a.

In addition, the Long Household Questionnaire collected more details, e.g., current school attendance, survivorship status of parents of children under age 18, and the registration status of each person. It also identified eligible ever-married women age 12-49 for interview with the Women's Questionnaire. The Long Household Questionnaire also collected information regarding various characteristics of the dwelling unit, such as the source of water; type of toilet facilities; type of cooking fuel; materials used for the floor, roof, and walls of the house; ownership status of various durable goods; ownership of agricultural land; ownership of livestock/farm animals/poultry; and ownership and use of mosquito nets.

---

Subject: Re: Pakistan: Standard DHS, 2006 Dataset  
Posted by [yiningw](#) on Fri, 19 Feb 2021 22:32:33 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you, Tom and Bridgette! This is helpful.

---