
Subject: Problems with Peru continuous survey files (contain same information)

Posted by [cportner](#) on Thu, 30 Apr 2020 15:40:30 GMT

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

Hello,

On the Survey Dataset Files page, Peru shows as having two separate continuous surveys for 2004/06 and 2007/08. The links go to differently named zip files, but the data files inside the zip files are identically named across surveys and contain the combined data across both surveys. That both zip files contain the same information is not what I would expect, given the naming convention and is different from other continuous surveys that use the same phase (such as Senegal). The inconsistency is a problem if you rely on the naming convention when doing large cross-country analyses.

Claus

Subject: Re: Problems with Peru continuous survey files (contain same information)
Posted by [Bridgette-DHS](#) on Fri, 01 May 2020 14:40:29 GMT

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

The dataset contains 2004-2008, but was duplicated to allow it to appear for 2007-08 as well as 2004-06. This was done because the Peru surveys were reported on the web page as 2004-06 and 2007-08 separately, following the pattern of the survey reports that were produced.

If you wanted to analyze 2004-06 you need to select on v007 <= 2006, and for 2007-08 select on v007 >= 2007.

Subject: Re: Problems with Peru continuous survey files (contain same information)
Posted by [cportner](#) on Fri, 01 May 2020 18:48:40 GMT

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

Bridgette,

Thank you for the response.

I am curious why Peru is different from any of the other surveys in this respect. As far as I can tell, all other surveys that are listed as separate on the website, and with separate reports, are also separated when it comes to the data files. The problem is not so much how to separate out the years but rather that when you download many surveys using a download manager it is easy to end up with both files and, therefore, duplicate data. Furthermore, it runs counter to the naming convention since the zip files are named differently from the data files contained in them.

All the best,

