
Subject: Are the caseid for IR file unique across rounds for DHS India?

Posted by [teeb](#) on Tue, 02 Aug 2022 14:13:22 GMT

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

I appended IR files from round 1 through 5 and found duplicates in terms of caseid when I checked for duplicates.

Is a caseid unique to the particular round? Or two women from different rounds can have same caseid? Were strictly different women surveyed in different rounds?

I also created a womanid by adding a round prefix to the existing caseid and the womanid did not have duplicates.

I'm not sure if this indicates an error on my part while cleaning data or the structure of the data is such.

Would appreciate any help.

Thanks

PS: in fact I went through caseids for the different rounds and I can see the caseids for IR file are not unique to a round. A woman in DHS 1992-93 has same caseid as that in DHS 1998-99. So the follow up question is are these different women or the same?

Subject: Re: Are the caseid for IR file unique across rounds for DHS India?

Posted by [Bridgette-DHS](#) on Tue, 02 Aug 2022 19:47:42 GMT

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

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

No women appear in more than one round. The samples in successive surveys are statistically independent. Even the sample clusters are different.

As shown below, there are no duplicates in the IR files for any of the India surveys. Duplicates would be indicated by any values greater than 1 when the number of cases with each ID are added up. Earlier posts have stated that v024 (region) must be included in the unique identifiers. I collapsed over "v024 v001 v002 v003". In another run I collapsed over "caseid" and got exactly the same result. There are no duplicates in the India IR files.

```
. local lpvs 23 42 52 74 7B
```

```
.  
. foreach lpv of local lpvs {  
  2.  
. use caseid v024 v001 v002 v003 using "C:\Users\26216\ICF\Analysis - Shared  
Resources\Data\DHSdata\IAIR\lpv'FL.DTA", clear  
  3.
```

```
. gen dups=1
  4.
. collapse (sum) dups,by(v024 v001 v002 v003)
  5.
. tab dups
  6.
. }
```

File Attachments

1) [collapse.jpg](#), downloaded 407 times

Subject: Re: Are the caseid for IR file unique across rounds for DHS India?

Posted by [teeb](#) on Wed, 03 Aug 2022 04:53:37 GMT

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

Thanks for the prompt response and for offering clarification on the sampling strategy.

If I understood the code correctly, this looks for duplicates in terms of caseid within each round.

However, my confusion arises from the fact that a caseid which appears in IR file of DHS 1992-93 also appears in IR file of DHS 1998-99. For eg, please see the attachments.

Would be grateful if you could offer some explanation.

Thanks again

File Attachments

1) [dhs_199293.PNG](#), downloaded 303 times

2) [dhs_199899.PNG](#), downloaded 398 times

Subject: Re: Are the caseid for IR file unique across rounds for DHS India?

Posted by [Bridgette-DHS](#) on Wed, 03 Aug 2022 19:38:23 GMT

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

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

As I said, each survey has different clusters, households, and individuals. However, v024 (region) has codes 1, 2, 3, etc. Within each value of v024, v001 (cluster) has codes 1, 2, 3, etc. Within each value of v001, v002 (household) is numbered 1, 2, 3, etc. Within each value of v002, v003 (line number) is 1, 2, 3, etc. (I am giving variable names in the IR file, they will be different in the PR and MR files.)

I think you are making this more complicated than necessary. The numbering system will lead to people in different surveys (in the same country or different countries) having the same ID code. The ID codes are not like passport numbers that are connected with specific individuals. They are simply a device for distinguishing different cases within the same survey.

Subject: Re: Are the caseid for IR file unique across rounds for DHS India?
Posted by [teeb](#) on Mon, 08 Aug 2022 09:55:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks a lot! This cleared up my confusion
