
Subject: Re: merge IR and KR: duplicate line number in IR
Posted by [bun_2019fall](#) on Sat, 28 Aug 2021 02:45:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Bruno,

Thank you for much for looking into the file, and I really appreciated it!

The "caseid" solution is almost perfect, except for one file "DRIR21FL":

use "V:\DHS raw survey\DRIR21DT\DRIR21FL.DTA", clear

*duplicates by caseid

duplicates tag caseid, generate(dup)

sort caseid

tab dup /*4 duplicates*/

*complete duplicates

duplicates tag, gen(dup1)

tab dup1 /*no duplicates*/

The "caseid" itself has 4 duplicates in this dataset, out of all DHS IR files that I have downloaded as of July 2021. I further checked, these observations are not complete duplicates. That said, to proceed with data analysis, I wonder if I should just drop the 4 observations? I wonder if DHS has any best practice regarding this issue? Thank you, again!
