
Subject: Difference DTA and SAV

Posted by [victor](#) on Tue, 20 Dec 2022 18:52:23 GMT

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

I have noticed that data is different between SAV and DTA datasets. Is this not an error?

Specifically looking at Gambia 2013, in household member recode. The DTA dataset has an additional value in the hv140 variable compared to the SAV dataset.

As a result, the mean calculation for registered children is different between these two datasets. With the DTA dataset it is possible to reproduce the figures in the official report. With the SAV dataset this value is a percentage point higher. It seems to me that however that SAV is a better approximation as with the DTA dataset the respondents with value 9 are used in the calculation.

For the DTA file

hv140		n
<dbl+lbl>		<int>
1	0 [Neither certificate or registered]	3294
2	1 [Has certificate]	8494
3	2 [Registered]	1961
4	8 [Don't know]	208
5	9	398
6	NA	38336

For the SAV file

hv140		n
<dbl+lbl>		<int>
1	0 [Neither certificate or registered]	3294
2	1 [Has certificate]	8494
3	2 [Registered]	1961
4	8 [Don't know]	208
5	NA	38734