Subject: Missing Region for 2015 SPSS India dataset
Posted by olympiaca on Tue, 03 Jan 2023 13:31:05 GMT
View Forum Message <> Reply to Message

Hi,

I am loading the 2015 India DHS .SAV Individual file into Rstudio. When it loads I see 32757 respondents. When I see what regions these individuals are from it says they are only from Andaman and Nicobar Islands, Andhra Pradesh, Arunachal Pradesh, and Assam but there are no individuals from any of the other regions. Is this because R Studio is not loading the full dataset or is there an issue with the data file?

Thanks

Subject: Re: Missing Region for 2015 SPSS India dataset
Posted by Trevor-DHS on Tue, 03 Jan 2023 14:45:56 GMT
View Forum Message <> Reply to Message

Please provide the filename to let us know exactly which file you are referring to. We are not seeing a problem with the files, so we suspect that it has been truncated when reading the file into R (possibly R is only reading the first part of the file). Letting us know the filename will permit us to check the file properly.

Also provide the specific commands in R that you are using to read the data into R, and we can check whether the problem is with those commands.

Subject: Re: Missing Region for 2015 SPSS India dataset
Posted by olympiaca on Sat, 07 Jan 2023 06:33:32 GMT
View Forum Message <> Reply to Message

Hi

i'm using the function : read.spss from the haven package

I2015 <- read.spss("IAIR74FL.SAV", use.value.label = TRUE, to.data.frame = TRUE)

and the file name is IAIR74FL

When i read it i'm getting a sample size of 32757

Thanks

Subject: Re: Missing Region for 2015 SPSS India dataset
Posted by Trevor-DHS on Mon, 09 Jan 2023 21:10:46 GMT
View Forum Message <> Reply to Message

I downloaded the SPSS file and tested this, first in SPSS itself, and then in R. In SPSS I get
699686 cases, and I am seeing data for all states.
I then tested the code you gave in R. First, read.spss is not in the haven package, but in the
foreign package. The haven package uses read_spss or read_sav, not read.spss. Also the other
parameters are different between the packages.

I was unable to load the data with read.spss from the foreign package. I got the message "Error:
cannot allocate vector of size 5.3 Mb".

I then tried with the read_spss from the haven package:
I2015 <- read_spss("C:/Users/21180/OneDrive - ICF/Data/DHS_SPSS/IAIR74FL.SAV",
user_na=TRUE)
and it successfully loaded the data showing "699686 obs. of 4796 variables".

I would first try reading the data with read_spss. If you are still only getting 32757 cases, then I
think the data file that you have is somehow truncated so I would suggest downloading the
dataset from our website again and trying to read it again using read_spss.