Subject: Re: Caste in NFHS-1
Posted by Bridgette-DHS on Tue, 24 Oct 2023 16:15:01 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff members, Tom Pullum and Fred Arnold:

We generally do not encourage data users to use the individual caste/tribe names because it would typically take several months to determine what caste/tribe names from the NFHS surveys correspond with the actual caste/tribe names. For example, there are many different spellings of specific caste/tribe names. Accurate measurement of castes in India is not as easy as it would seem. There are more than 1,000 castes and sub-castes in India. In NFHS, we do not provide a list of the castes/tribes to interviewers because they would usually not be able to find a caste on the list even if it was on the list somewhere, especially given multiple different spellings of each caste/tribe and the fact that the NFHs surveys are administered in 20 different languages. It is also difficult to given interviewers a rule for what constitutes a caste since there is no clear definition. For example, Nepali is a caste, but Bengali is not. The caste/tribe (SPECIFY) names in NFHS are not part of the standard datasets.

The 5 NFHSs have included two variables (except just 1 in the first NFHS) that give rough classifications of the castes/tribes. They have different variable names and different value labels in the different surveys. If you want to combine them, you will have to develop new variables as recodes of those nine variables. The Stata lines pasted below will give you the value labels for these variables, which will help you to construct the variables you want.

It may actually not be helpful to append the files. In all countries with DHS surveys, MANY variables have changed category definitions from one survey to the next. The value label in the last survey appended will over-write all the previous value labels. You may not even realize that the codes have changed or the labels have changed. You need to be very careful when pooling surveys.

```
* Specify a workspace
cd e:\DHS\DHS_data\scratch

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR23FL.DTA", clear
keep if hv101==1
keep hv001 hv002 shcaste
gen NFHS=1
save temp_1.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR42FL.DTA", clear
keep if hv101==1
keep hv001 hv002 sh40 sh41
gen NFHS=2
save temp_2.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR52FL.DTA", clear
keep if hv101==1
```

```
keep hv001 hv002 sh45 sh46
gen NFHS=3
save temp_3.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR74FL.DTA", clear
keep if hv101==1
keep hv001 hv002 sh35 sh36
gen NFHS=4
save temp_4.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR7EFL.DTA", clear
keep if hv101==1
keep hv001 hv002 sh48 sh49
gen NFHS=5
save temp_5.dta, replace

use temp_1.dta, clear
append using temp_2.dta
append using temp_3.dta
append using temp_4.dta
append using temp_5.dta
label list

*Then construct unifying recodes for the different variables related to caste in the different surveys
```