**DHS data**

**Dependent variable-Dropouts**

In the household dataset I found the information of children, youngsters and adults, the age range varies from 0 to 98 years old.

I have two approaches to define the dependent variable:

For the first approach, I transformed the dataset to long format and subsetted it to respondents between 5 and 24 years old, as the DHS does to analyze dropouts in their report, the sample size is 59,529 respondents. I used a variable that specifically asks children about their school status and the categories are:

0 Never attended. Children with no education.

1 Entered school. Children who did not attend school the previous year but are currently enrolled. 2 Advanced. Children at a current level that is higher than the previous year

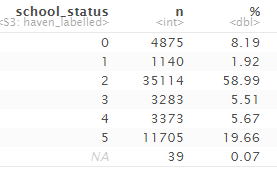
3 Repeating. Children who are at the same level than the previous year or at a level less than the previous year.

4 Dropout. Children who were at school the previous year but not currently attending school.

5 Left school 2+ years ago. Children who are not currently attending school and did not go to school the previous year.

8 Don’t know

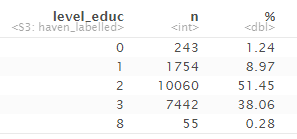
Here is the variation of the variable:



59% of the children are in category 2 which is “Advanced: Children at a current level that is higher than the previous year” and category 5 “Left school 2+ years ago”. Dropouts has 6% of the sample which represents 3,283 respondents.

For the second approach I used the level of education of the household member, I subsetted the dataset to respondents between age 18 and 24, assuming that people in this range of age should have achieved at least secondary education, if not they can be considered as dropouts, the sample size is 19,554. The categories of the variable are 0. No education, 1.Primary, 2.Secondary, 3.Higher, 4. Don´t know, Missing.

The variation of the variable is:



I could use category 0 and 1 as dropouts, which would represent 10% of the sample, 1,997 respondents.

These are some of the variables in the household dataset that I identified and can be useful for my analysis.

**Household dataset**

|  |  |
| --- | --- |
| **Original variable** | **Description** |
| Hhid  Case identification |  |
| Hv002  Household number |  |
| Hv025  Type of place of residence |  |
| HVIDX Line number of the household member. |  |
| HV101 Relationship to the head of the household |  |
| HV102 Whether the member is a de jure household member, i.e., whether the member is a usual resident of the household. |  |
| HV104 Sex of the household member |  |
| HV105 Age of the household member |  |
| HV106 Highest level of education the household member attended | This is a standardized variable providing level of education in the following categories:   1. No education 2. Primary 3. Secondary 4. Higher.   8. Don´t know  9. Missing  Any member below the lower age limit for the education questions is classified  in the "No education" category |
| HV107 Highest year of education gives the years of education completed at the level given in HV106 |  |
| HV108 Education in single years. |  |
| HV109 Educational attainment | Recodes the education of the household member into the following categories: None, incomplete primary, complete primary, incomplete secondary, complete secondary, higher education. |
| HV110 Whether the household member is still in school. | All members aged equal to or older than the upper limit (usually 25 years) for this question or who have not attended school are coded 0 (Not in school). |
| HV111 Whether the mother of the household member is still alive. | BASE: All children in the household aged less than 18. |
| HV112 Line number in the household of the mother of the member. | This variable is code 00 if the mother is not a member of the household. BASE: All children in the household aged less than 18. |
| HV113 Whether the father of the household member is still alive. | BASE: All children in the household aged less than 18. |
| HV114 Line number in the household of the father of the member. | This variable is code 00 if the father is not a member of the household. BASE: All children in the household aged less than 18. |
| HV115 Marital status of the household member |  |
| HV121 Household member attended school during current school year. |  |
| HV122 Educational level attended during current school year | With the same standardized levels as explained for HV106. |
| HV123 Grade of education at the level of education attended during current school year |  |
| HV124 Education in single years during current school year |  |
| HV125 Household member attended school during previous school year. |  |
| HV126 Educational level attended during previous school year. |  |
| HV127 Grade of education at the educational level during previous school year |  |
| HV128 Education in single years during- previous school year. |  |
| HV129 School attendance status | 0 Never attended. Children with no education. 1 Entered school. Children who did not attend school the previous year but are currently enrolled. 2 Advanced. Children at a current level that is higher than the previous year 3 Repeating. Children who are at the same level than the previous year or at a level less than the previous year. 4 Dropout. Children who were at school the previous year but not currently attending school. 5 Left school 2+ years ago. Children who are not currently attending school and did not go to school the previous year. 8 Don’t know |

**Women´s dataset**

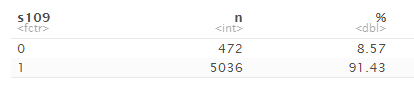
|  |  |
| --- | --- |
| **Original variable** | **Description** |
| caseid |  |
| V002  Household number |  |
| V012  Respondent´s  Current age |  |
| V106  Highest level of education | 0 No education  1 Primary  2 Secondary  3 Higher  9 Missing |
| V107  Highest year of education | 0 No years completed.  1:15  99 Missing |
| V133  Education in single years | 0:20  97 Inconsistent  99 Missing |
| V149  Educational attainment | 0 No education  1 Incomplete primary  2 Complete primary  3 Incomplete secondary  4 Complete secondary  5 Higher  9 Missing |
| s106a | Level of school completed |
| S106b | Grade of school completed |
| S109 | Currently attending school |
| S110 | Number of years stopped school |
| S111 | Reason to stop school |
| S112 | Mother current marital status |
| S113 | Father current marital status |
| S115 | Mother marital status when respondent was 12-14 years |
| S116 | Father marital status when respondent was 12-14 years |

**Female sample**

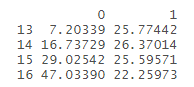
Size of sample respondents age 12 to 16🡪 5,508

Currently studying in school (0= no, 1= yes)🡪 472 are not studying which corresponds to 9% of the total sample.

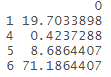
**Variation of the variable**



**Cross table age and studying status**

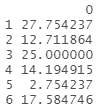


**Cross table ethnicity and share of women that are not studying**



I didn´t find in the recode map the codification of the variable but reading the DHS report in the ethnic section they mention 6 categories: indigenous communities, gypsies, raizal, palenquero, Afro-Colombians, and none. I assume that the variable corresponds from 1 to 6 to these groups. So, 71% of the share of women that are not studying do not belong to a particular ethnicity, 19% are indigenous, and 8.6% Afro-Colombian.

**Cross table region and share of women that are not studying**

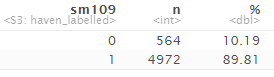


Based on the DHS report the 6 regions are: 1. Atlantic, 2. Oriental, 3. Bogota, 4. Central, 5. Pacific, and 6. Orinoquia and Amazonas. 27% of the share of women that are not studying come from the Atlantic coast, follow by 25% from Bogota, and 17% from the Amazon region.

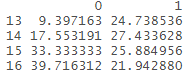
**Male sample**

Size of sample respondents age 12 to 16🡪 5,536

Currently studying in school (0= no, 1= yes)🡪 564 are not studying which corresponds to 10.2% of the total sample.



**Cross table age and studying status**



**Cross table region and studying status**

