**Merging the BaSE household-level and child-level data**

The Bharat Survey for EdTech (BaSE) contains two datasets—at the level of the household and and at the level of the child. Both the datasets include sampling weights, and names of states, districts, villages, and settlements in addition to survey responses. A guide to merging the two datasets has been provided below.

**Types of files**

While merging data files, it is important to know the relationship between the two files to be merged. The household dataset contains unique surveyed households, while the child dataset may contain more than one observation for each household (instances where there was more than one child, eligible to be surveyed, in the household).

**Steps for merging the two datasets.**

All statistical packages (STATA, SPSS, SAS) have commands that allow the merging of the files. Regardless of the statistical package in use, it’s important to follow these steps:

1. Determine the common identifier between both the datasets. In the case of the BaSE survey, the unique ID of the household is ***base\_hh\_id.***
2. Determine the base (primary file). The base file establishes the unit of analysis. Usually, when the relationship is that of many to one [M:1], the base file is one with many entities. So, the base file here should be the child dataset.
3. Use the *BASE-merging\_codes.do* file to merge the two datasets.

**Understanding the data**

A total of 6030 households were surveyed as a part of BaSE from 6 states of India (Gujarat, Mizoram, Uttar Pradesh, Madhya Pradesh, Odisha, and Telangana). A minimum of 1 and a maximum of 3 school-going children were included from each household. On average, 1.6 children were surveyed in each household. A total of 9867 children from 6030 households were surveyed. Table 1 summarizes the data.

***Table 1:*** *Summary of the dataset*

|  | **Household** | **Children** | **Merged dataset/Output** |
| --- | --- | --- | --- |
| Sample size (N)/Number of observations | 6030 | 9867 | 9867 |
| Number of variables | 86 variables | 174 variables | 254 variables |