SURVEY ON INTERNATIONAL MIGRATION IN SENEGAL (EMIS 2019)

Methodological note:

Preparation of the raw database for dissemination:

Poverty and Equity

The World Bank

Objective

The raw data collected by the interviewers in the field was further processed to obtain the public database, made available to external users. The chronology of these processing operations is briefly described in the following paragraphs to help users.

- **Preliminary Data Cleaning**

This step was conducted concurrently with data collection. It was initiated with a view to checking the day-to-day consistency of the information collected by the field teams. This stage focused on checks related to:

  o Amounts and other atypical numerical variables for confirmation or correction
  o Incomplete physical or phone addresses
  o Inconsistencies between related variables
  o Answers to Open questions /Other (specify), when too evasive
  o Elimination of duplicates and correct tracking of total non-response

The user should note that despite the corrections made, the data remains raw. The correct answers were collected from the previously surveyed household, by phone contact or a final visit to the household.

- **Restructuring of the database by section**

In order to be consistent with the logic of the questionnaire the database has been restructured into sections. The user will note that the sections combining individual and household levels have been split for easier manipulation. General household information has been retained in all sections, namely:

  o Household ID
  o Individual ID
  o Type of household
  o Weight
  o Region
  o Department
  o Commune
  o Area/Village

- **Households and Individual Weight**

A methodological note on the calculation of weights is available in the documentation.

- **Anonymization of data**

In accordance with the Act amending and supplementing Act No. 2004-21 of 21 July 2004 on the organization of statistical activities, in its Article 8-ter, data shall be made anonymous and shall be presented in such a way that the statistical unit cannot be identified, either directly or indirectly, taking into account all appropriate means that could reasonably be used by a third party(e.g.: household members names and telephone numbers, GPS coordinates and critical georeferenced variables).

- **Labelling of variables**

All the variables have been relabeled to provide the user with a readable and transparent framework.

- **Determining the type of household**
  Household type refers to the migration profile of the household. It is established based on the presence or absence of an international migrant or a returning migrant as well as the migration country of the latter. On this basis, the four profiles below have been identified. The user will also note that, in case of conflict between these profiles for the same household, the preferred profile is the one declared as the rarest during the preliminary enumeration of the sampled DRs. The profiles are listed according to the rarity scale, presented from the rarest to the least rare:
    - ✓ Returning migrant ECOWAS
    - ✓ International Migrant RDM
    - ✓ Migrant International
    - ✓ No international migrants

It should also be noted that the profile recorded in the database is the profile actually encountered during the data collection phase. However, this profile may differ from the profile recorded during the listing activity of the Enumeration Areas (EAs).

- **Validation of questionnaires**

The user should note that only households that fully responded to sections 1 to 4 were declared valid and retained in the final database.

- **Other Information Notes for Users**

This section provides the user with useful information to manipulate the data effectively.

  - o *Label value*

The user will note that *label value* has the same label as the variable whose modalities it labels, with the only difference that lowercase letters are used instead of uppercase characters.

Example: **Variable** p01_06 **Label value:** P01_06

- Section6

Unlike the variable labels in the entire database, the variables in Section 6 are preceded by the *s* character instead of the *p* character.

- Missing values

The user should note that quantitative variables (amounts or other open numeric variables) contain missing values that may be represented by a sequence of characters 9 whose length varies from 2 to the maximum number of characters contained in the variable. It is recommended to standardize the coding of missing values not to generate confusion with true values, or outliers.

- Merging with the community database

The key for merging the community and household databases is the ***com_men*** variable. The user will note that some households are not associated with any community. The reason being that the community interview could not take place in EAs due to various incidents.