

# Kenya Household Master Sample Frame (K-HMSF)

seconds of 15 seco

## **Background to K-HMSF**

The Kenya National Bureau of Statistics (KNBS) is mandated, by the Statistics Act, 2006, to develop and maintain sampling frames for use by KNBS and other stakeholders. KNBS, therefore, has been developing household master sample frames after every population and housing census that conforms to spatial population distribution and current administrative units. A master sample frame (MSF) is a large representative sample from which other representative samples can be drawn for various surveys. Previously, the MSFs developed by the KNBS were named the National Sample Survey and Evaluation Programme (NASSEP) and the last one in the series was NASSEP V. Use of the name NASSEP was discontinued after the 2019 Census and the current frame is known as the **Kenya Household Master Sample Frame (K-HMSF)**.

## **What is K-HMSF?**

The Kenya Household Master Sample Frame (K-HMSF) is a household-based master sample developed and maintained by the Kenya National Bureau of Statistics. The current K-HMSF is based on the 2019 Kenya Population and Housing Census (KPHC) data.

## **Purpose of K-HMSF**

The master sample is a critical statistical infrastructure of the Bureau designed to facilitate the selection of representative samples for various household-based sample surveys.

## **K-HMSF Design**

K-HMSF is based on a two-stage stratified cluster sample design. The sampling units in the master sample comprise clusters as the primary sampling units and households as the secondary sampling units. Clusters are the smallest statistical units which can be an Enumeration Area (EA), part of an EA, or a combination of an EA. This is informed by the number of households in the EA. Clusters in the K-HMSF are standardized to

have one measure of size which is an average of 100 households, ranging between 50-149 households depending on the terrain and population density. The K-HMSF is composed of 10,000 clusters randomly selected with Probability Proportional to Size (PPS) from approximately 129,000 Enumeration Areas (EAs) which were developed during the 2019 Kenya Population and Housing Census (KPHC) Cartographic Mapping.

The master sample is explicitly stratified into 92 sampling strata, with counties forming the first level of stratification and the urban and rural areas within the counties being the second level of stratification. Nairobi and Mombasa are purely urban areas.

The sample frame is further randomized into four sub-samples (C1, C2, C3 and C4) each comprising 2,500 clusters and which can serve as independent sample frames. Any two or more sub-samples can be combined whenever a large sample size is required. The sub-frames are important for managing respondents' fatigue given the many household surveys undertaken in Kenya.

### **Process of K-HMSF Development**

To generate a list of households to serve as a sampling frame for the second stage selection for the various surveys, a household mapping and listing exercise is undertaken in the sampled Enumeration Areas to transform them into clusters. The listing entails identifying the EA boundaries using interactive maps, undertaking a household quick count, segmenting of large EAs, listing the households and numbering the dwelling structures at the doors.

Some of the information collected during the household listing includes names of household heads, household composition by age and sex, household telephone contact information and GPS coordinates of the structures where households reside.

Cluster maps are generated for each cluster showing the boundaries and structure numbers for ease of identification of selected households for the various household surveys. The listing is fully digitized, and a georeferenced database is maintained for the K-HMSF.

The dynamic nature of households underscores the need to keep the list of households up-to-date within the clusters. The K-HMSF clusters are

often updated by undertaking re-listing of households to reflect changes that have occurred in the clusters.

### **At what level of estimation can the frame support?**

The frame is designed to support surveys that provide estimates at national, rural-urban and county levels.

### **Frames for generating estimates lower levels domains**

Ad-hoc (stand-alone) sampling frames are usually developed, from the list of all EAs developed for the 2019 KPHC, to cater for surveys designed to provide estimates of population parameters for the domains below the county levels.

### **Accessing K-HMSF for Use in Surveys**

To access the sample frame, one needs to make a formal request to the Director General, KNBS with a concept note detailing the survey objectives and its design. The KNBS does not share the master sample frame in its entirety with external partners or users but only draws and provides a sample from the frame based on the client's sample design. Any data collection in the clusters within K-HMSF is done in collaboration with KNBS.