LIBERIA INSTITUTE OF STATISTICS AND GEO-INFORMATION SYSTEM (LISGIS)

Liberia 2007 CWIQ Survey

**INSTRUCTIONS FOR SELECTION OF HOUSEHOLDS WITHIN CLUSTERS/
ENUMERATION AREAS (EAS)**

SAMPLE DESIGN

The CWIQ survey will be carried out on a sample of 3,600 randomly selected households located in 300 randomly selected clusters. The clusters have been selected from each locality (County). The clusters are Census Enumeration Areas (EAs) from the 2007 Demographic and Health Survey (DHS). In each County the primary sampling units are Enumeration Areas (EAs) and the secondary sampling units are households. Twelve (12) households will be selected in each of the 300 EA using systematic sampling. As at the time of the 2007 LDHS, the number of EAs in each County, total number of households and number of EAs sampled in each County are given in the table below.

Distribution of Total Number of EAs and Sampled EAs by County

County	Total number of EAs	Sampled EAs
Bomi	189	8
Grand Cape Mount	171	10
Gbarpolu	139	28
Montserrado	171	12
Margibi	279	19
Grand Bassa	409	15
River Cess	94	8
Sinoe	134	16
Grand Gedeh	180	22
River Gee	109	12
Grand Kru	100	7
Maryland	147	27
Bong	540	15
Nimba	803	22
Lofa	429	13
Greater Monrovia	708	66
TOTAL	4,602	300

The sampling method for selecting households to be interviewed in the above survey is **Systematic Sampling.**

WHAT IS SYSTEMATIC SAMPLING?

It is a statistical method involving the selection of every k^{th} element from a sampling frame, where k, the sampling interval, is calculated as follows:

k = Population size (N) / Sample size (n).

Example 1

If N = 250 listed households in an EA and the number of households to be selected in an EA is n = 15

The Sampling interval k = (N/n) = (250/15) = 16.67 [Between 16 and 17].

As a rule you should choose a **Random Start** between 01 and 17 (rounded up number)

If 12 is chosen as a random start, then the numbers chosen in the samples will be as follows:

12, 29, 46, 63, 80, 97, 114, 131, 148, 165, 182, 199, 216, 233, and 250.

Example 2

Suppose you want to sample n = 8 households from an EA of N = 120 households.

k = (N/n) = 120/8 = 15, so every 15th households is chosen after a random starting point between 01 and 15. If the random starting point is 11, then the households selected are 11, 26, 41, 56, 71, 86, 101, and 116.

HOW DO WE CHOOSE A RANDOM START?

A random start is chosen from a table of random numbers available in most textbooks on statistics. This is to avoid bias in choosing random starting point.

AVOID SELECTION OF LDHS SAMPLED HOUSEHOLD

In the selection process if an LDHS is selected simply select its neighbour and continue the selection process. Do not change the sampling interval.

CHARACTERISTICS OF SYSTEMATIC SAMPLING

Each element in the population has an equal probability of selection. It is similar to random sampling. It is much more efficient (under certain conditions e.g. variance within sample>variance of population) and much less expensive. The distribution of the sample is spread throughout the sampling frame at equal intervals.

SELECTING 12 HOUSEHOLDS FOR A SMALL CLUSTER
e.g. LESS THAN 30 HOUSEHOLDS PER EA

The procedure is the same as that outlined above except that we have to maintain the decimal points preferably to two decimal points for the Sampling Interval. In selecting EAs for the CWIQ survey, this factor was taken into account to avoid having a very small number of households for the CWIQ sample.

SELECTING FIVE SUBSTITUTES

For this selection k = (N/5)
Where N=Number of households listed and 5 is the number to be selected.

Take a random start between 1 and k
Reject if the household is one of the selected 12 households or one of the LDHS selected household. Simply replace by the next household on the list (the neighbour). Maintain the interval and continue the selection process.