Sample Design and Weighting Procedures for the Kenya STEP Employer Survey

David J. Megill
Sampling Consultant, World Bank
May 2017

1. **Sample Design for Kenya STEP Employer Survey**

The sampling frame for the Kenya STEP Employer Survey was based on a register of all establishments with 5 or more employees, in three different regions of Kenya: Nairobi, Mombasa and Other. The sampling frame had information on the geographic location, economic activity and the number of employees. Within each region the frame of establishments was stratified by three employment size groups: 5-19 employees, 20-99 employees and 100+ employees. Given that the size information was missing for some establishments, it was necessary to establish another stratum for these establishments (Size Not Available) in each region, and the eligible establishments in this stratum that were selected and interviewed were post-stratified into the correct categories for the analysis. Table 1 presents the distribution of the establishments in the frame by region and employment size strata.

Table 1.   Distribution of establishments in the sampling frame for the Kenya STEP Employer Survey by region and employment size strata

Region	5-19 employees	20-99 employees	100+ employees	Size N/A	Total
Nairobi	8,461	1,875	591	722	11,649
Mombasa	2,379	640	164	217	3,400
Other	3,857	913	217	294	5,281
Total	14,697	3,428	972	1,233	20,330

First it was necessary to allocate the sample by region and employment size strata based on the distribution of the frame and the sample size needed for each domain. Given that each region and employment size stratum was a domain of analysis, the sample of 500 establishments was allocated equally to all the main strata, with a smaller sample allocated to the "Size N/A" stratum in each region, as shown in Table 2. Given the distribution of the frame shown in Table 1, this resulted in a higher sampling rate for the larger employment size strata. A supplemental sample of 100% of the target sample size was also selected in each stratum as a reserve for possible replacements.

Table 2.   Allocation of sample establishments for Kenya Employer Survey by region and employment size stratum

Region	5-19 employees	20-99 employees	100+ employees	Size N/A	Total
Nairobi	50	50	50	20	170
Mombasa	50	50	50	15	165
Other	50	50	50	15	165
Total	150	150	150	50	500

Within each region by employment size stratum the establishments were selected systematically with equal probability, after sorting the frame geographically and by economic activity.

In the case of any sample establishment that could not be interviewed, a replacement establishment was selected from the reserve sample for the same stratum. During the data collection for the survey it was found that some sample establishments were not eligible for the survey because they were closed, had less than 5 employees, or were otherwise out-of-scope. It was necessary to take these cases into account in the weighting procedures, as described in the next section.

Following the implementation of the Kenya STEP Employer Survey, a total of 504 sample establishments were interviewed. The distribution of the final sample of establishments that were successfully interviewed by region and employment size stratum is shown in Table 3.

Table 3.   Final distribution of sample establishments with completed interviews for Kenya Employer Survey by region and employment size stratum

Region	5-19 employees	20-99 employees	100+ employees	Size N/A	Total
Nairobi	53	51	56	8	168
Mombasa	47	54	48	16	165
Other	56	54	47	14	171
Total	156	159	151	38	504

2. Weighting Procedures for Kenya STEP Employer Survey

In order for the sample estimates from the Kenya STEP Employer Survey data to be representative of the population of establishments, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample establishment is equal to the inverse of its probability of selection.

As described above, a stratified one-stage sample design was used for the Kenya STEP Employer Survey. The sample establishments in each region by employment size stratum were selected

2

with equal probability. The probability of selection for the sample establishments in each stratum was calculated as follows:

$$p_h = \frac{n_h}{N_h},$$

where:

$p_h =$ probability of selection for the sample establishments in stratum h

$n_h =$ number of sample establishments with completed interviews in stratum h

$N_h =$ number of establishments in the frame for stratum h

The basic weight for the sample establishments is the inverse of this probability of selection, and can be expressed as follows:

$$W_h = \frac{N_h}{n_h},$$

where:

$W_h =$ basic weight for the sample establishments in stratum h

Given that some of the establishments in the frame were found to be ineligible because they were found to be closed, had less than 5 employees, or were otherwise out-of-scope, it was necessary to adjust the weights to take this into account. The information on the interview status of all of the sample establishments that were contacted during the survey, including both the original sample and reserves for replacement, was used to estimate the proportion of the establishments in the frame for each stratum that were eligible. The following types of non-interviews were considered out-of-scope:

4. Firm stopped working
5. Firm is in bankruptcy
6. Firm has been blocked for more than two months
9. It doesn't fit the target group – it has less than 5 employees
10. It doesn't fit the target group – it has been operating for less than a year

In order to adjust the basic weights, the estimated proportion of the establishments in the frame for each stratum that was not eligible was estimated as follows:

$$p_{Eh} = \frac{n_{Eh}}{n_{Eh} + n_{NEh}},$$

where:

$p_{Eh} =$ estimated proportion of establishments in the frame for stratum h that are eligible

$n_{Eh} =$ total number of contacted sample establishments in stratum h was were considered eligible, including the establishments with completed interviews

$n_{NEh} =$ total number of contacted sample establishments in stratum h was were considered not eligible

The final weight for the sample establishments in each region by employment size stratum was calculated by multiplying the basic weight for the stratum by the estimated proportion of eligible establishments in the frame for the stratum, as follows:

$$W'_h = W_h \times p_{Eh},$$

where:

$W'_h =$ final (adjusted) weight for the sample establishments in stratum h

Appendix:

TABLE 1. OUTCOME OF SELECTION AND ACTIVATED REPLACEMENTS

| | STRATUM | | | | | | | | | | | | |
	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
ORIGINAL SELECTION	50	50	50	20	50	50	50	15	50	50	50	15	500
ACTIVATED REPLACEMENTS	30	34	44	10	16	14	7	4	19	12	11	2	203
SUBTOTAL	80	84	94	30	66	64	57	19	69	62	61	17	703
NO ATTEMPT	21	17	6	5	34	36	43	13	31	38	39	14	297
TOTAL	101	101	100	35	100	100	100	32	100	100	100	31	1000

| | STRATUM | | | | | | | | | | | | |
	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
FRAME ERRORS	21	26	30	12	3	0	1	1	2	2	2	1	101
Out-of scope (15)	0	0	1	1	0	0	0	1	0	0	0	0	3
Stopped working (4)	4	4	1	1	3	0	1	0	2	2	0	1	19
Wrong address, moved, not possible to contact (7)	17	22	28	10	0	0	0	0	0	0	2	0	79
REFUSALS	6	6	8	9	16	10	8	2	11	6	12	2	96
To enumerator (2)	0	0	1	0	0	2	0	0	0	0	0	0	3
To coordinator (3)	6	6	7	9	16	8	8	2	11	6	12	2	93
TOTAL	27	32	38	21	19	10	9	3	13	8	14	3	197

| | STRATUM | | | | | | | | | | | | |
	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
EFFECTIVE SAMPLE COVERED (ORIGINAL STRATA)	53	52	56	9	47	54	48	16	56	54	47	14	506