

A New Approach to Improving Sample Design for Crop Forecast and Post-Harvest Estimates in Zambia

**John Kalumbi
Deputy Director
Agriculture and Environment Division**

1. Introduction

The Agriculture and Environment Statistics Division of the Zambia Central Statistical Office (CSO) is responsible for the production of official statistics on food and agriculture. The Division has two branches namely: the Agricultural and the Environment Statistics Branches. The Agricultural Statistics Branch consists of two sections namely; large-scale farm section and small and medium scale farm section. The Environment Branch also has sections dealing with land degradation, air pollution, water sanitation, forestry and the wildlife and fisheries.

The CSO has been conducting annual agricultural surveys since the 1970/71 agricultural season and mostly these have been the Crop Forecast Survey (CFS) and the Post Harvest Survey (PHS) for small-medium and large-scale agricultural holdings. Both the CFS and PHS use the same methodology, i.e., survey sampling and are conducted in conjunction with Ministry of Agriculture and Cooperatives using questionnaires during the periods March/April and November/December each year, respectively.

2. Survey Methodology for the new Sample Design

During the census field operation, a cartographic operation is conducted to define Census Supervisory Areas (CSAs) which are further divided into Standard Enumeration Areas (SEAs) within a district. An SEA is the unit of data collection point at district level with well-defined boundaries identified on census sketch maps.

The previous CFS/PHS sample was based on the 1990 census data and only the rural districts in the country were covered on sample basis. A stratified three-stage sample design was used in which the CSAs were the Primary Sampling Units (PSU) selected at first stage with Probability Proportional to Size (PPS). The measure of size was the number of households (as listed in the 1990 Census) in each selected CSA, one sample SEA was selected at second stage again using PPS within each sample CSA. The total sample of 405 SEAs was drawn out of total 9,000 rural SEAs. Within each

sample SEA, the sample households were selected from the listing stratified by farm size category using Linear Systematic Sampling (LSS) procedure at the third stage.

The new CFS/PHS sample is based on the 2000 Zambia Census results and uses a stratified two-stage sample design. The SEAs are the Primary Sampling Units selected at first stage with PPS where the number of agricultural households in the SEA is the measure of size. The sample households are selected at the second stage from the listing stratified by farm size category. The new sampling frame also include a small fraction of urban SEAs (586) in which 70 percent or more households are agricultural according to the 2000 Zambia Census data. The new CFS/PHS sample has 410 SEAs (with a 5 percent urban sample) out of the 12,788 agricultural SEAs defined for the 2000 Zambia Census. The new sample design also allows for the replacement of any sample SEAs, which cannot be enumerated for one reason or another in order to maintain the effective sample size.

3. Stratification For new CFS/PHS Sample Design

The new sample SEAs were first stratified by district and a minimum of two sample SEAs was allocated to each district proportionally to the number of agricultural households. Eight crops were identified to receive special treatment in the new sample design in order to improve the precision of the survey estimates of crop area and production. These are crops with limited geographical concentration whose CVs previously were relatively high due to the smaller number of observations and these were: sorghum, rice, cotton, burley tobacco, virginia tobacco, sunflower, soyabeans and paprika. A crop stratum code was assigned to each SEA based on which of these crops was predominant.

4. Stratification of Households at the Second Sampling Stage

In the previous sample design, stratification of households in sample SEAs was by two farm size categories A: 0-4.99 hectares and B: 5-19.99 hectares. In the new sample design, stratification of households within each sample SEA is by farm size, number of livestock/poultry and the growing of special crops at the second sampling stage. The new sample design has three farm size categories stratification of households of A: 0-1.99 hectares; B: 2.00-4.99 hectares and C: 5-19.99 hectares. Then agricultural households listed in each sample SEA are assigned to one of the three categories A, B or C depending on the farm size, number of livestock/poultry and growing of special crops. Households with farm size category of 20 or more hectares are included in the special frame for large-scale farmers and are enumerated separately.

Further, households will be added to category C (if not qualified based on land area) if found raising the following minimum number of livestock/poultry: cattle-50; pigs – 20; goats –30 and poultry 50. Households are also added to categories B and C based on the special crops criteria. If the sample SEA only has 1 or 2 households with any of these special crops, these will be assigned to category C. If the sample SEA has

only 3 to 5 households with any of these special crops, such households are assigned to category B (if previously assigned to category A).

5. Sample Households Selection in Sample SEAs

A total of twenty households are selected from each sample SEA. The sample households distribution to each category is C-10, B –5 and A – 5. Where the number of households in category C does not permit selection of the ten households required, the shortfall from twenty will be allocated equally to categories B and A or category B will have one more sample households than category A if the difference cannot be allocated equally. Where there is no household in category C and less than 10 in category B, the difference from twenty will be met from category A. If all listed households fall in category A, then all 20 sample households will come from category A. For each category in the Sample SEA, selection of households is done using Linear Systematic Sampling procedure with a random start.

Appendix 1: Estimation Procedure

Estimating for the total involves the weighing of variables before aggregates are made at sub-national and national levels. The basic sampling weight is equal to the inverse of the probabilities of selecting the SEAs and households by category size.

The formula for estimating the survey estimate of a total is as follows:

$$\hat{Y} = \sum_h \sum_i \sum_s \sum_j W'_{shi} Y_{shij}$$

Where:

Y_{shij} = value of variable y for the j-th sample household in category s within the i-th sample SEA in district h

W'_{shi} = basic weight for the sample households in category s within the i-th sample SEA in district h (or inverse probability of selecting sample households in category s within the i-th sample SEA in district h.

$$= \frac{N_h}{M_h \times N_{hi}} \times \frac{N_{shi}}{n_{shi}}$$

Where:

M_h = number of sample SEAs selected in district h

N_{hi} = total number of agricultural households in the frame for the i-th SEA in district h

N_h = total number of agricultural households in the frame for district h

n_{shi} = number of sample agricultural households selected in category s from the listing for the i-th sample SEA in district h and

N_{shi} = total number of households in category s from the listing for the i-th sample SEA in district h

The survey estimate of a ratio can be expressed as:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}}, \text{ Where } \hat{Y} \text{ and } \hat{X} \text{ are estimates of totals for variables y and x, respectively}$$

Appendix 2: Distribution of New Agricultural Sample SEAs for CFS/PHS by Province, rural and Urban

Province	Rural SEAs	Urban SEAs	Total Sample SEAs	% Urban SEAs
Central	39	3	42	7.1
Copperbelt	25	7	32	21.9
Eastern	72	-	72	0
Luapula	42	2	44	4.5
Lusaka	23	3	26	11.5
Northern	70	2	72	2.8
Northwestern	30	2	32	6.2
Southern	45	1	46	2.2
Western	42	2	44	4.5
Total Zambia	388	22	410	5.4