Sample Design and Weighting Procedures for Shan Survey

David J. Megill
Sampling Consultant, Michigan State University
September 2018

1. Background and Survey Objectives

The purpose of this report is to describe the sample design and weighting procedures for the Shan Survey that was conducted by the Food Security Policy Project, which is implemented by Michigan State University (MSU) and the International Food Policy Research Institute (IFPRI) in collaboration with the Centre for Economic and Social Development (CESD) and the Myanmar Ministry of Agriculture, Livestock and Irrigation (MOALI). The main objective of the Shan Survey is to study the farms growing maize and pigeon peas as well as landless households in particular townships in three main geographic areas (zones) in Southern Shan State where these crops are concentrated: North, South and East zones. The frame for each zone includes the following townships:

1. North Shan: Taunggyi, Hopong, Pindaya and Yatsut
2. South Shan: Hsihseng, Pinlaung and Pekon
3. East Shan: Linkhae and Moe Nae

Initially the township of Namhsan (South) was also included in the East zone, but later it was found that most of the villages in this township were not accessible because of security issues, so it was dropped from the survey. For the township of Linkhae the village tracts of Dar Seik and Wan Sit were also excluded for the same reason.

David Megill, Sampling Consultant with Michigan State University (MSU), developed the sample design for the Shan Survey in collaboration with Daw Yin Yin Kyaing, Deputy Director of the Department of Population, Ministry of Labour, Immigration and Population, and Ellen Payongayong, Senior Surveys Advisor, MSU.

2. Sampling Frame

A stratified two-stage sample design was used for the Shan Survey in the three different zones. Information on the approximate area planted in maize and pigeon peas in each village tract within the three geographic zones of Shan State were collected from the Department of Agriculture within each township. This information was used to define the final list of village tracts to be included in the frame for the Shan Survey, as well as the stratification. The names of the village tracts were matched to the geographic names from the Census to identify the ward/village tract (WVT) codes.

The village tracts in the frame were stratified by the three zones, the predominant crop (maize or pigeon peas) and the level of concentration of the crop (high and low). The high level for each crop corresponds to the village tracts with the largest area of the crop that cover a cumulative 80% of the total area of the crop. The remaining village tracts for each crop were assigned to the low level. In this way a total of 11 strata were defined in the sampling frame. A 3-digit stratum code was assigned to each village tract, where the first digit was the zone (1-North, 2-South, 3-East), the second digit is the crop (1-maize, 2-pigeon peas), and the third digit is the level (1-high, 2-low).

The stratified frame of village tracts was matched to the Census frame in order to identify the EAs in each village tract, and the number of private households in each EA from the 2014 Myanmar Census frame. Following the stratification of the village tracts, the EAs within each village tract were assigned to the same stratum.

Table 1 shows the distribution of the village tracts, EAs and households by stratum based on the 2014 Myanmar Census frame. This table reflects the distribution of the final frame after the township of Namhsan (South) was excluded from the East Shan zone, as well as the village tracts of Dar Seik and Wan Sit in Linkhae Township, because of problems in accessing these areas due to security problems.

Table 1. Distribution of village tracts, EAs and households in sampling frame by stratum for the Shan Survey, based on 2014 Myanmar Census

Stratum code	Stratum	Number of village tracts	Number of EAs	No. private households (2014 Census)	Percent of households within frame
111	North Shan maize high	27	304	44,104	25.0%
112	North Shan maize low	30	199	28,402	16.1%
121	North Shan pigeon peas high	12	64	7,733	4.4%
122	North Shan pigeon peas low	16	97	12,348	7.0%
211	South Shan maize high	6	80	12,333	7.0%
212	South Shan maize low	34	402	57,083	32.4%
221	South Shan pigeon peas high	1	7	1,122	0.6%
222	South Shan pigeon peas low	2	24	3,610	2.0%
311	East Shan maize high	4	15	2,355	1.3%
312	East Shan maize low	7	12	1,557	0.9%
321	East Shan pigeon peas high	9	41	5,581	3.2%
Total frame		148	1,245	176,228	100.0%

This stratified sampling frame of EAs was used for the first stage selection of EAs within each stratum, using systematic sampling with probability proportion to size (PPS), based on the number of private households in each EA from the 2014 Census frame. The frame of EAs in

each stratum was ordered by village tract and EA codes in order to provide an implicit stratification and improve the geographic representativeness of the survey results for each township.

A listing of households was conducted in each sample EA to identify all the households within the EA boundaries for the second stage selection of households.

3. Sample Size and Allocation

At the first sampling stage a total of 100 EAs were selected for the Shan Survey. A higher sampling rate was used for the high level stratum for each crop. Within the crop by level strata the sample EAs were first allocated proportionally to the three zones, but this allocation was adjusted to slightly increase the sample for the smaller Shan East zone. At the second sampling stage 16 households were selected in each sample EA, for a total sample size of 1,600 households. The final allocation of the sample EAs and households by stratum is shown in Table 2.

Table 2. Allocation of sample EAs and households by stratum for Shan Survey

Stratum code	Stratum	Number of sample EAs	Number of sample households
111	North Shan maize high	35	560
112	North Shan maize low	6	96
121	North Shan pigeon peas high	8	128
122	North Shan pigeon peas low	4	64
211	South Shan maize high	11	176
212	South Shan maize low	12	192
221	South Shan pigeon peas high	2	32
222	South Shan pigeon peas low	2	32
311	East Shan maize high	8	128
312	East Shan maize low	4	64
321	East Shan pigeon peas high	8	128
Total frame		100	1,600

4. Sample Selection Procedures

At the first sampling stage the sample EAs were selected within each stratum systematically with PPS, based on the number of private households in each EA from the 2014 Census frame. The number of sample EAs selected in each stratum was based on the sample allocation specified in Table 2. This selection of sample EAs was conducted by the Department of Population of the Ministry of Labour, Immigration and Population, which is responsible for the

national sampling frame based on the 2014 Myanmar Census. They used the Complex Samples module of the SPSS software for the first stage selection of sample EAs using systematic PPS sampling within each stratum. The EAs in the sampling frame for each stratum were sorted geographically by the village tract and EA codes in order to provide implicit stratification and ensure that the sample is geographically representative within each township.

Following the selection of the initial sample of 100 EAs and the beginning of the fieldwork for the Shan Survey, it was found that it was not possible to access a total of 12 sample EAs in the East Shan zone due to security problems. These 12 sample EAs that could not be enumerated were located in the township of Namhsan (South) and in the village tracts of Dar Seik and Wan Sit in Linkhae Township. As described previously, the sampling frame was modified to exclude these areas. The 12 original sample EAs that could not be enumerated in the East Shan zone included 4 sample EAs each in strata 311, 312 and 321. Therefore it was decided to select 12 new sample EAs with this same distribution by stratum from the modified frame for these three strata after excluding the EAs in the specified problematic areas. The new EAs in each stratum were also selected from the modified frame systematically with PPS. One of the new sample EAs in stratum 321 could not be enumerated because of security problems, so the final number of enumerated sample EAs for the Shan Survey was 99. This missing sample EA was taken into account in the weighting procedures, as explained in the next section.

During the listing operation it was determined whether each household identified in the sample EA had a respondent available at home, or whether the household was temporarily absent during the period of the survey. The selection of households was limited to those with respondents available at home. However, for estimation purposes it is assumed that the characteristics of the households that were temporarily absent were similar to those of the households that were interviewed. At the second stage a sample of 16 households was selected in each sample EA using systematic random sampling with equal probability.

5. Weighting Procedures

In order for the sample estimates from the Shan Survey to be representative of all the households in the frame, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household is equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). A stratified two-stage sample design was used for the Shan Survey. At the first stage the EAs were selected within each stratum systematically with PPS based on the total number of households in each EA from the 2014 Census frame, and at the second stage the households in each EA were selected systematically with equal probability. Therefore the overall probability of selection for the sample households can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}}$$

where:

p_{hi} = probability of selection for the sample households in the i-th sample EA of stratum h

n_h = number of sample EAs selected in stratum h, specified in Table 2

M_{hi} = number of private households in the 2014 Census frame for the i-th sample EA of stratum h

M_h = total number of private households in the 2014 Census frame for stratum h

m_{hi} = 16 = number of sample households selected in the i-th sample EA of stratum h

M'_{hi} = total number of households listed in the i-th sample EA of stratum h, including those that were temporarily absent

The two components of this probability of selection correspond to the individual sampling stages. The basic sampling weight for the sample households is calculated as the inverse of this probability of selection. Based on the previous expression for the probability, the basic weight for the sample households can be simplified as follows:

$$W_{hi} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}}$$

where:

W_{hi} = basic weight for the sample households in the i-th sample EA of stratum h

Following the data collection for the Shan Survey, these basic weights were adjusted for nonresponse. In the case of stratum 321, in which one sample EA could not be enumerated, it was necessary to adjust the weights at the stratum level. In the case of sample households that could not be interviewed, the weights were adjusted at the sample EA level. The final weight for the eligible sample households in each sample EA were adjusted as follows:

$$W'_{hi} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}} \times \frac{n_h}{n'_h} \times \frac{m_{hi}}{m'_{hi}} = \frac{M_h}{n'_h \times M_{hi}} \times \frac{M_{hi}}{m'_{hi}}$$

where:

W'_{hi} = final adjusted weight for the sample households in the i-th sample EA of stratum h

n'_h = number of sample EAs successfully enumerated in stratum h

m'_{hi} = number of sample households with completed interviews in the i-th sample EA of stratum h

The sampling probabilities at each stage of selection and the final weights for the sample households were calculated in an Excel spreadsheet with the information from the sampling frame and the final survey results for all sample EAs.