Sample Design and Weighting Procedures for Dry Zone Household Survey in Four Townships of Myanmar

David J. Megill
Sampling Consultant, Michigan State University
July 2017

1. Background and Objectives

The main objective of the Dry Zone Household Survey was to study the agricultural and socioeconomic characteristics of the rural households in four townships in the dry zone of three different regions. This included the townships of Magway and Pwint Phyu in Magway Region, the township of Myittha in Mandalay region, and the township of Budalin in Sagaing Region. The four main crops of interest for the agricultural activities were paddy, greengram, groundnut and sesame. In addition, crops of secondary interest were blackgram, pigeon peas, chick peas and sunflower. The study was designed to compare the socioeconomic characteristics of the farm households growing the primary crops of interest with farm households growing other crops, and to non-farming or landless households in these townships. Therefore most of the rural households in the four townships were in scope for this survey.

The sampling frame for the Dry Zone Household Survey was based on the summary data and maps from the 2014 Population and Housing Census of Myanmar for the four townships. Auxiliary information was used to classify the village tracts in the four townships by the level of intensity of the primary crops of interest.

The purpose of this report is to describe the sample design and weighting procedures for the Dry Zone Household Survey. The sampling methodology was developed in collaboration with Ellen Payongayong, Michigan State University Consultant. The sampling frame was developed with the collaboration of Yin Yin Kyaing, Deputy Director, Department of Population, Ministry of Immigration and Population.

2. Sampling Frame and Stratification for the Dry Zone Household Survey

A stratified three-stage sample design was used for the Dry Zone Household Survey in four townships. The sampling frame was based on summary data and maps from the 2014 Population and Housing Census of Myanmar for the four townships. The primary sampling units (PSUs) selected at the first stage were the village tracks in each of the four townships. In order to improve the efficiency of the sample design for covering the four main crops of interest (paddy, greengram, groundnut and sesame), independent information was obtained on the approximate area planted in these crops in each village tract. Based on the total area planted in the four main crops of interest, the village tracts in each township were stratified into the following three categories with a similar number of village tracks in each stratum: 1-high, 2-medium and 3-low.

At the first stage a sample of village tracks was selected systematically with probability proportional to size (PPS) within each township, where the measure of size was based on the number of households in each village track from the 2014 Census data. At the second sampling stage one enumeration area (EA) was selected with PPS within each sample village track.

A listing of households was conducted in each sample EA. The households listed in each EA were classified into two groups: farm households with one or more of the main crops of interest, and farm households growing other crops, or landless households. The purpose of this third stage stratification of the listing in each sample EA was to select a separate sample from the group of farm households growing any of the four crops of interest, and from all remaining households.

2. Sample Size and Allocation for the Dry Zone Household Survey and Sample Selection

Based on the survey objectives and resource constraints, it was decided to select a sample of 25 EAs in each township, and at the last stage a sample of 16 households was selected in each sample EA, for a total of 400 sample households per township, and a total of 1,600 sample households overall.

The allocation of the sample by stratum was designed to improve the efficiency of the sample design for covering the main crops of interest. In order to increase the number of sample households growing the four main crops of interest in the sample, at the first stage the 25 sample village tracts in each township were allocated to the strata as follows: high stratum – 12 sample village tracts; medium stratum – 8 sample village tracts; and low stratum – 5 sample village tracts. The sample village tracts within each stratum were selected systematically with PPS based on the number of households in the 2014 Census frame. A total of 15 sample village tracks were selected with a probability of 1 because their measure of size was larger than the sampling interval for the corresponding stratum; these are identified in the sample with an SR (self-representing) code of 1. Since one sample EA was selected within each sample village tract at the second stage, the allocation of the sample EAs by stratum was the same as that for the sample village tracts.

At the third sampling stage the allocation of sample households with and without the four main crops of interest to be selected from the listing in each sample EA was also determined strategically. A sample of 10 households was selected in each EA from the group of households with one or more of the four main crops of interest, and 6 households were selected from the remaining households (other farm households and landless households). Within each group the sample households were selected using random systematic sampling.

Table 1 shows the number of village tracts in the frame for each township and stratum, and the number of sample EAs and sample households by group within each stratum.

Table 1. Allocation of sample EAs and sample households for Dry Zone Household Survey by township and stratum

Township/ stratum	Number of village tracks in frame	Number of village tracks and EAs selected	Number of sample households with crops of interest	Number of sample households without crops of interest or landless	Total sample households
Budalin	54	25	250	150	400
High	18	12	120	72	192
Medium	18	8	80	48	128
Low	18	5	50	30	80
Magway	61	25	250	150	400
High	20	12	120	72	192
Medium	20	8	80	48	128
Low	21	5	50	30	80
Pwintbyu	52	25	250	150	400
High	17	12	120	72	192
Medium	17	8	80	48	128
Low	18	5	50	30	80
Myittah	82	25	250	150	400
High	27	12	120	72	192
Medium	27	8	80	48	128
Low	28	5	50	30	80
Total	249	100	1,000	600	1,600

5. Weighting Procedures for Dry Zone Household Survey

In order for the sample estimates from the Dry Zone Household Survey in four townships to be representative of the population, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household would be equal to the inverse of its probability of selection, calculated by multiplying the probabilities at each sampling stage. The sampling probabilities at each stage of selection were maintained in an Excel spreadsheet with the information from the frame for the sample village tracts and EAs within each stratum.

A stratified three-stage sample design was used for the Dry Zone Household Survey. Given that the self-representing village tracts in each stratum were selected at the first sampling stage with a probability of 1, the probabilities and weights are specified separately for the self-representing (SR) and non-self-representing (NSR) sample village tracts. At the last sampling stage a separate sample of households was selected for the groups of households with and without the crops of interest, so the probabilities and weights also vary by group within each sample EA.

The probability of selection for the sample households in the NSR sample village tracts can be expressed as follows:

$$p_{NSRhij(g)} = \frac{n_{NSRh} \times M_{hi}}{M_{NSRh}} \times \frac{M_{hij}}{M_{hi}} \times \frac{m_{hij(g)}}{M_{hij(g)}} = \frac{n_{NSRh} \times M_{hij}}{M_{NSRh}} \times \frac{m_{hij(g)}}{M_{hij(g)}},$$

where:

$p_{NSRhij(g)}$ = probability of selection for the sample households in group g (with or without crops of interest) of the j-th sample EA in the i-th NSR sample village tract of stratum h

n_{NSRh} = number of NSR sample village tracts selected in stratum h

M_{hi} = total number of households in census frame for the i-th sample village tract in stratum h

M_{NSRh} = total number of households in the census frame for all NSR village tracts in stratum h

M_{hij} = number of households in the census frame for the j-th sample EA of the i-th sample village tract in stratum h

$m_{hij(g)}$ = number of sample households selected in group g of the j-th sample EA of the i-th sample village tract in stratum h

$M_{hij(g)}$ = total number of sample households listed for group g in the j-th sample EA of the i-th sample village tract in stratum h

The three components of this probability of selection correspond to the individual sampling stages. It can be seen from this formula that the probability of selection for sample households in the NSR sample village tracts simplifies to a form that is equivalent to the probability based on selecting the sample EAs directly with PPS within each stratum at the first sampling stage. However, the stratification is determined at the level of the village tract, and all the sample EAs are assigned to the same stratum as the corresponding village tract.

For the self-representing village tracts, the first stage component of the probability is equal to 1, and one sample EA was selected within each village tract with PPS. Therefore the overall probability of selection for the households in the SR village tracts can be expressed as follows:

$$p_{SRhij(g)} = \frac{M_{hij}}{M_{hi}} \times \frac{m_{hij(g)}}{M_{hij(g)}},$$

where:

$p_{SRhij(g)}$ = probability of selection for the sample households in group g of the j-th sample EA of the i-th SR sample village tract in stratum h

The basic sampling weight for the sample households is calculated as the inverse of the corresponding probability of selection. Based on the previous expressions for the probabilities of selection for the households in the NSR and SR sample village tracts, the corresponding weights for the sample households can be calculated as follows:

Basic weight for sample households in NSR sample village tracts:

$$W_{NSRhij(g)} = \frac{M_{NSRh} \times M_{hij(g)}}{n_{NSRh} \times M_{hij} \times m_{hij(g)}},$$

where:

$W_{NSRhij(g)}$ = basic weight for the sample households in group g of the j-th sample EA in the i-th NSR sample village tract of stratum h

Basic weights for sample households in SR sample village tracts:

$$W_{SRhij(g)} = \frac{M_{hi} \times M_{hij(g)}}{M_{hij} \times m_{hij(g)}},$$

where:

$W_{SRhij(g)}$ = basic weight for the sample households in group g of the j-th sample EA of the i-th SR sample village tract in stratum h

Following the data collection for the Dry Zone Household Survey, it was necessary to adjust these basic weights to take into account any eligible sample households that could not be interviewed. The weights for the sample households in each group (households with and without the crops of interest) within a sample EA were adjusted as follows:

$$W'_{NSRhij(g)} = W_{NSRhij(g)} \times \frac{m_{hij(g)}}{m'_{hij(g)}} \text{ and } W'_{SRhij(g)} = W_{SRhij(g)} \times \frac{m_{hij(g)}}{m'_{hij(g)}},$$

where:

$m'_{hij(g)}$ = number of sample households with completed interviews in group g of the j-th sample EA of the i-th sample village tract in stratum h

In a few cases during the interview it was found that a sample household was misclassified into the wrong group (with or without the main crops of interest) for the sample selection. In such cases the probabilities and corresponding weights are still based on the original classification for selection, even if the household was misclassified.