Sample Design and Weighting Procedures for Dry Zone Seed Survey in 6 Townships

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

1. Background and Survey Objectives

The purpose of this report is to describe the sample design and weighting procedures for the Dry Zone Seed Survey that was conducted by the Myanmar Ministry of Agriculture, Livestock and Irrigation (MOALI). The main objective of the Dry Zone Seed Survey is to study the variety adoption and quality seed demand for 8 different crops in 6 townships of the Dry Zone in Sagaing (Myinmu and Monywa), Magway (Magway and Pwint Phyu) and Mandalay (Madaya and Kyaukse). The following 8 crops are covered in this survey: paddy, greengram, groundnut, sesame, pigeon pea, chick pea, sunflower and blackgram.

David Megill, Sampling Consultant with the Michigan State University (MSU), developed the sample design for the Dry Zone Seed 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, Michigan State University (MSU).

2. Sampling Frame

A stratified two-stage sample design was used for the Dry Zone Seed Survey in the six townships. The primary sampling units (PSUs) were the enumeration areas (EAs) defined for the 2014 Myanmar Census of Population and Housing. In the rural part of each township the EAs are defined within the village tracts.

In order to stratify the frame of village tracts and EAs, MOALI provided a list of the specific crops of interest that are grown in each village tract within the six townships, as well as the approximate area planted in each crop. This information was used to stratify the village tracts based on the frequency of the different crops in each township. A special stratum was created for the village tracts with less frequent crops in each township, so that a higher sampling rate could be used for these strata. The more frequent crops in each township were covered in the "other" stratum, since these crops are found in most village tracts and did not require any special stratification. Two of the townships (Magway and Pwint Phyu) only have one stratum for all crops, since the different crops were fairly evenly spread across the village tracts within these townships. 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 Census data. The final list of strata is presented in Table 1, which shows the distribution of the number of village tracts, EAs and households in the 2014 Myanmar Census

frame by stratum. Following the stratification of the village tracts, the EAs within each village tract were assigned to the same stratum.

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

Stratum code	Stratum	Number of village tracts	Number of EAs	No. private households (2014 Census)	Percent of households within township
11	Myinmu Other	37	122	16,856	77.8%
12	Myinmu Sunflower	9	30	4,033	18.6%
13	Myinmu Chickpea	2	5	784	3.6%
21	Monywa Other	36	153	21,749	65.8%
22	Monywa Blackgram	19	80	11,284	34.2%
31	Magway All	61	369	48,171	100.0%
41	Pwint Phyu All	52	281	38,168	100.0%
51	Madaya Other	44	154	21,835	48.0%
52	Madaya Pigeon Pea	12	90	13,242	29.1%
53	Madaya Blackgram	14	74	10,406	22.9%
61	Kyaukse Other	58	225	32,669	71.2%
62	Kyaukse Groundnut	8	31	4,022	8.8%
63	Kyaukse Pigeon Pea	18	64	9,202	20.1%
Total frame		370	1,678	232,421	

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 cover all the households within the EA boundaries, and to identify the households that grow at least one of the crops of interest. Only the households growing these crops were eligible for selection at the second sampling stage.

3. Sample Size and Allocation

Since the survey results will be analyzed at the township level, a total of 24 sample EAs were selected within each township at the first sampling stage, for a total of 144 sample EAs. Following the listing in each sample EA to identify all the eligible households growing at least one crop of interest, 10 eligible households were selected in each sample EA. Therefore the

maximum sample size would be 1,440 eligible households. However, some EAs had less than 10 households growing one or more eligible crops, in which case all of these households were selected for the survey at the second stage.

The 24 sample EAs within each township were first allocated proportionally to the different strata, and then the sample allocation was adjusted to increase the number of sample EAs for the special strata with less frequent crops, with a corresponding decrease in the number of sample EAs allocated to the "other" stratum in the township. This procedure increased the number of sample observations for the less frequent crops.

The allocation of the sample EAs and sample households by stratum for the Dry Zone Seed Survey is presented in Table 1.

Table 1. Allocation of sample EAs and households by stratum for the Dry Zone Seed Survey in six townships

Stratum code	Stratum	No. of sample EAs	Maximum no. of sample households with crops of interest
11	Myinmu Other	14	140
12	Myinmu Sunflower	7	70
13	Myinmu Chickpea	3	30
21	Monywa Other	12	120
22	Monywa Blackgram	12	120
31	Magway All	24	240
41	Pwint Phyu All	24	240
51	Madaya Other	8	80
52	Madaya Pigeon Pea	9	90
53	Madaya Blackgram	7	70
61	Kyaukse Other	12	120
62	Kyaukse Groundnut	4	40
63	Kyaukse Pigeon Pea	8	80
Total sample		144	1,440

Following the survey implementation, it was found that some of the sample EAs had less than 10 households growing at least one of the crops of interest. There were also two sample EAs where there were no households growing any crops of interest. In one case the main occupation of the village was brick production, and in the other case the agricultural land was confiscated by the army and the university. The final number of eligible households with completed interviews was 1,388.

3. 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 is 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. The EAs in the sampling frame for each stratum were sorted geographically by the village track and EA codes in order to provide implicit stratification and ensure that the sample is geographically representative within each township.

Following the listing operation in each sample EA to identify all the households growing at least one crop of interest, at the second sampling stage a sample of 10 eligible households with the specified crops were selected in each EA using random systematic sampling. In the case of EAs with less than 10 eligible households listed, all of these households were selected with certainty at the second sampling stage.

4. Weighting Procedures

In order for the sample estimates from the Dry Zone Seed Survey to be representative of all the households growing the crops of interest in the six townships, 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). A stratified two-stage sample design was used for the Dry Zone Seed Survey. At the first stage the EAs were selected within each stratum with PPS based total number of households in each EA from the 2014 Census frame. However, at the second sampling stage only the households growing at least one of the six crops of interest were eligible to be selected. The overall probability of selection for eligible sample households can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{COIhi}}{M_{COIhi}}$$

where:

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

n_h = number of sample EAs selected in stratum h

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_{COIhi} = number of sample eligible households with crops of interest selected in the i-th sample EA of stratum h (generally equal to 10)

M_{COIhi} = total number of eligible households with crops of interest listed in the i-th sample EA of stratum h

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 eligible sample households can be simplified as follows:

$$W_{hi} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M_{COIhi}}{m_{COIhi}}$$

where:

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

Following the data collection for the Dry Zone Seed Survey, these basic weights were adjusted for nonresponse at the sample EA level. The final weight for the eligible sample households in each sample EA were adjusted as follows:

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

where:

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

m'_{COIhi} = number of eligible 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.