

Stratified sampling by ALIS

Shoji KIMURA

ASEAN Food Security Information System, email: wood_v@yahoo.co.jp

This study report is written for a working level statistician who responses on conducting an area sample survey by ALIS in order to respond a request he or she should implement the stratified sampling on area sample survey by using ALIS.

1. Preface

The stratified sampling is an excellent method which helps to increase an accuracy of estimated data. I have already explained that the composing elements of a sampling error are a standard deviation and a number of samples^{note1}. The appearance rate of target crops in a sample dramatically increase by stratified sampling. In fact, a variation of sample data is more likely to become less, it is inversely to an increase of data appearance rate. It means the standard deviation becomes lower.

As a result, an error rate of estimated data becomes lower. This theoretical logic is very clear. So an academic statistician tends to recommend this method. However, we have to consider that a conducting of the stratified sampling means the needed statistical operation increases. It means that a risk of non-sampling error for estimated data increases surely. We are requested a comprehensive consideration for an adoption of the stratified sampling before an actual sampling survey.

2. What is the stratification?

First, we need to fully understand that “what is the stratification?” before we can make a decision on a stratification factor.

The stratification is;

- The process which it divides a mother population (framework) into the groups that having same nature.
- It conducts before the sampling.
- An element of the mother population belongs to only any one of the stratification.

I think you can easily understand these properties on the stratification. Well, what do you think next?

- It produces a weighted average of sample but an arithmetic average.

This property helps your understanding on the stratification. I would like to add some explanations for this property.

I have already mentioned “*Sample method is a method which looks the average of sample*^{note2}”. In fact, the statistical data is estimated by a sample average multiply number of framework. In case, a total number of frameworks are 1000 and a number of samples are 100, we can consider that this sample average becomes the arithmetic average. On the other hand, if we make a stratification framework which consisting of 500 from an original framework, the weight of sample average becomes 2 times with the original sample for data estimation. So we can consider that the stratification produces the weighted average of sample. Let clarify a merit and a demerit of the stratification before studying of a specific stratified sampling. We can consider the following as the merit;

- It can increase the estimation accuracy with less variation because we can get the standard deviation by the weighted average sample.

And, as the promising story;

- It may be efficient than a single mother population.

As the demerit;

- It needs a correct information on the mother population. Needless to say, you cannot introduce the stratification into the mother population without the correct information on the mother population. In generally speaking, it is very hard job to make a clear content of the mother population.

Other one is;

- It has a possible of high cost comparing with a single population. This cost means a cost and labor in order to clarify the mother population information. In addition, we have to accept the risk of increasing the non-sampling error.

^{note1} Study report “Accuracy calculation on area sample survey”

^{note2} Study report “Accuracy calculation on area sample survey”

3. Stratification of area mesh which including agricultural land or not including agricultural land

From here, I indicate some stratification standards on the area survey in case of using ALIS. The first is the stratification of an area meshes which including of agricultural land or not including of agricultural land. Although you may feel oddly, ALIS has already adopted the stratified sampling. In fact, the area framework of ALIS is not the original framework. The framework of ALIS is a result of the stratification as the area mesh including of agricultural land. Therefore, the sample average of ALIS becomes the weighted average with an average of basic method.

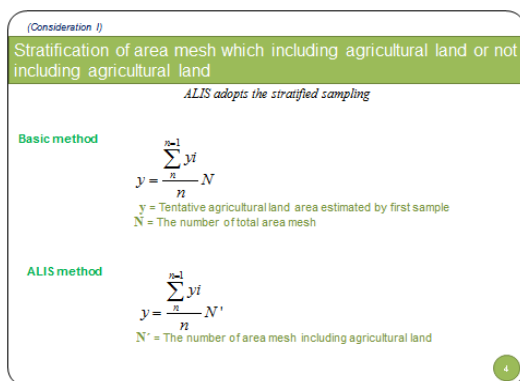


Figure 1

It may be such an excessive talk of ALIS, ALIS is imposed a lot of risks on non-sampling error such as a misunderstanding picture which has been adopting this stratified sampling. And it can say that this risk is the most notable risk on the ALIS operation. We have to mind anytime on a relationship with the sample survey operation and the non-sampling error.

4. Stratification of agricultural land to paddy field and upland

I think this stratification (in topic number 3.) will not satisfy on your request. So we have to consider more on the stratifications of ALIS framework. The most thinkable stratification is the stratification by a paddy field and upland. However, we regret to note that a present program of ALIS does not have a function which recognizing more stratifications as the different elements. In addition, if ALIS has the stratification function by program modification, there is a problem that remains to be solved.

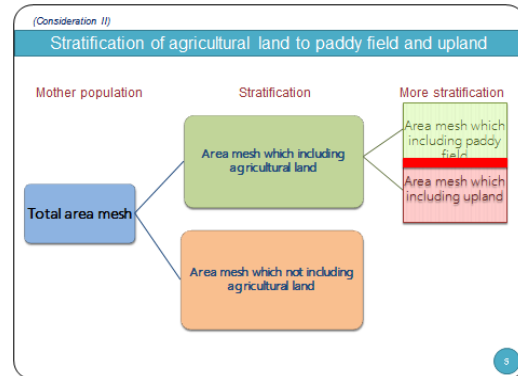


Figure 2

The unit of ALIS framework is the area mesh. It means that this area mesh as the unit of framework has a possibility to be stratified to different stratification. In fact, the area mesh has a possibility to be mixed of the paddy field and the upland. Please remember “what the stratification is?” *An element of the mother population belongs to only any one of the stratification.* In fact, strictly speaking, the stratification of the paddy field and the upland is impossible as long as using the area mesh as the unit of framework.



Figure 3

In fact, this area mesh is the area mesh as the unit of framework. And this stratification is the area mesh

including agricultural land. In case, we consider more stratifications of the paddy field, certainly this area mesh includes of the paddy field.

However, this area mesh includes of the upland too. In fact, this area mesh has the both stratification elements. This phenomenon goes against a concept of the stratification as *an element of the mother population belongs to only any one of the stratification*.

However, we can easily image that the stratified upland framework will increase the data accuracy on crops planted area like maize or sugarcane. So I recommend a method which makes the two kinds of framework not stratification from framework.

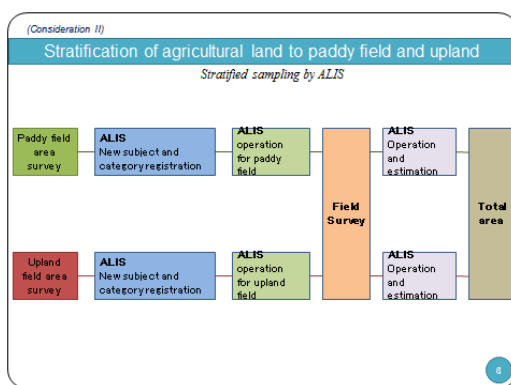


Figure 4

The first one is for the paddy field and the second is for the upland by registering different survey name to ALIS. In fact, it divides an original area survey into a paddy field survey and an upland survey and makes the framework respectively. Each framework element becomes the area mesh including the paddy field or the upland. After that an operator conducts ALIS operation by each survey. And you can conduct the field survey as the integrated survey. Finally, you can get the area data of the paddy field and the upland and then crops planted area with high accuracy. In addition, you can get total agricultural land area too by aggregating each survey data.

Strictly speaking, this operation is not the correct stratified sampling from original mother population, however, the important point is to make the different stratification of the paddy field and the upland, in order to get high accuracy data as the result. On this mean, ALIS function can follow the stratified sampling of paddy field and upland.

Needless to say, this stratification leads to increase sample appearance rate of upland crops. It means the data accuracy will increase dramatically. On the other

hand, we will have a new task on this stratification. We have to conduct a new operation which divides an agricultural land into the paddy field and the upland. It means the risk of the non-sampling error will be increased.

5. Stratification by crops area

Next consideration on the stratification is crops area. I think most of statisticians image this stratification as the element of stratified sampling. However, I do not recommend this stratification as the method which the unit of mother population is the area mesh like ALIS. Because this stratification is requested many information and labors to establish the borderline of each crop area. In addition, the determination of number of framework often does not stabilize by staff's arbitrariness and different information source. This situation should avoid on sample survey. If you have enough information to make the correct borderline of crop area, it would be no need area sample survey at all, the collecting information is enough to get the area data.

Anyway, I consider two cases for stratification by crops area.

Case 1 is; direct method which subdivides the province border-line by crops area. It requested more budgets to modify the ALIS program.

Case 2 is; the ingenious of ALIS function. The method which indicates of crops area by using a wide mesh is not using border-line.

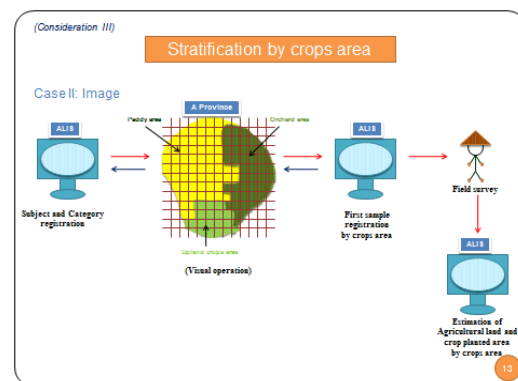


Figure 5

In case 2, the method of thinking is the same with the stratification of the paddy field and the upland. The operator register a new survey name and target crops and operate only the wide meshes which regarded as the target crop area. Needless to say, the determination of border wide mesh becomes a visual operation. The theory is very simple and ALIS

operation may not be difficult with operation skill up. It may be worth doing although I do not recommend the sub-region statistics by sample survey.

6. Stratification of main product province and non-main product province

The final consideration is about the stratification of main product province and non-main product province. Although this method is out of stratified sampling, the method concept is the same with stratified sampling and this method is used often to get the data of minor crop.

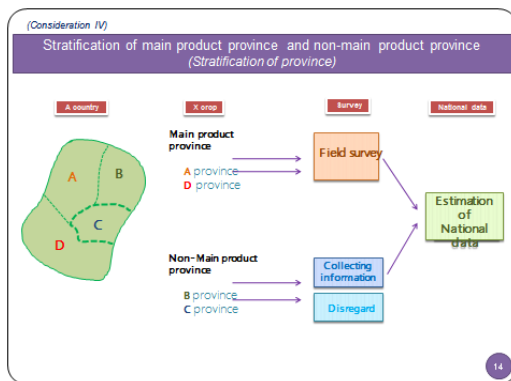


Figure 6

The concept of the method is very easy. The method is which you conduct survey for only main product province in case the productive condition is biased. In this case, the data of non-main product province is estimated by collecting information or disregarded data itself. This method is useful for the estimation of wide-area data like nation.