منظمة الأغذية والزراعة للأمم المتحدة

联合国粮食及农业组织

Food and Agriculture Organization of the United Nations

Organisation des Nations Unies pour l'alimentation et l'agriculture

Organización de las Naciones Unidas para la Agricultura y la Alimentación

APCAS/10/20
April 2010

Agenda Item 8

ASIA AND PACIFIC COMMISSION ON AGRICULTURAL STATISTICS

TWENTY-THIRD SESSION

Siem Reap, Cambodia, 26-30 April 2010

The Use of Area Frames for Agricultural Census

by

Robert Hale, Mathematical Statistician
International Programs Office
National Agricultural Statistics Service
United States Department of Agriculture

Current information on the agricultural situation in a country is critical. In some cases a method to estimate the current situation is to measure change from baseline information. The baseline information in agriculture is usually from the latest agricultural census. One of the best methods to conduct an agricultural census is to delineate census blocks or enumeration areas across the entire country. These blocks are designed to be easily located and each contains a fixed range of households. The process is well described in several Food and Agriculture Organization (FAO) publications. This is a reasonable method if time and funding are available. The problem is that the process to delineate all of the census blocks is time consuming and costly. The same is true for collecting data from all of the census blocks.

An alternative method is to construct an area sampling frame that will allow for efficient sampling to replace the complete enumeration of a census. The area frame can be designed to stratify the land in the country by factors important to the local situation. For example, a stratum could be designed for areas with a high percent of the land cultivated or with a high percentage of the land devoted to fruit trees or where much of the land is in pasture and livestock would be prevalent. Then the sample can be designed to focus more attention to these areas. However, all types of land should be represented in the sample. This report will use as an example the National Agricultural Statistics Service (NASS) assisted 2000 National Agriculture Census conducted in Ecuador.

There are several steps to construct and sample an area frame. First is to set the stratum definitions that fit the need. In the U.S the stratum definitions have to do with the percent of the land cultivated and the population density. The following stratum definitions were used in Ecuador.

Stratum	Definition
10	Predominantly grassland
20	Predominantly temporary crops
30	Predominantly permanent crops
40	Predominantly native vegetation
51	Very little agriculture
52	Water
53	Lowlands with no agriculture
54	Highlands with no agriculture (steep slope and forests)
60	Aquaculture areas (shrimp)
71	Urban areas
72	Rural populated areas
81*	Moderate agricultural areas with agriculture in < 2 km blocks
82*	Moderate agricultural areas with agriculture in > 2 km blocks
83*	Sparse agricultural areas with agriculture in < 2 km blocks
84*	Sparse agricultural areas with agriculture in > 2 km blocks

* Used only in three selected provinces

Associated with setting the stratum definitions is determining how to sample the frame. In some cases the sample may be a point that determines the farm to be surveyed. For an agricultural census the sampling unit is typically a fixed size block of land. The boundaries of the sampling unit can be determined several ways, but the best method for an agricultural census would be to have the boundaries on some type of physical features that the interviewers can see. The features are roads, railroads, foot paths, rivers, streams, etc. that are permanent.

Once the strata are set the size of the sampling units needs to be determined. The size can, and normally will, vary by stratum. It makes sense to sample a different size block of land in an agricultural area than in an urban area. The rules to determine the size vary depending on the number of interviews or the availability of physical features.

In Ecuador the sampling unit size was set to be 2 square kilometers in most agricultural strata. As more information became available, an adjustment was allowed in specific areas in certain provinces because the number of small farms was requiring too many interviews. Field supervisors were allowed to divide the sampling unit into 2 or 4 equally sized parts. Methods were provided to randomly choose one of the parts to be the new sampling unit. Care was taken to properly note the adjustment for summarization.

To construct the area frame, technicians will use satellite imagery or aerial photography to determine the land use. They can use maps, either printed or digital, to find boundaries. The task is to look for blocks of land that fit one of the stratum definitions. These blocks would typically be 6 to 8 times the size of the sampling unit. The process continues until all of the land in the province has been included in one and only one of the blocks. The work is usually completed at the lowest level of estimation needed. In Ecuador, the work was completed at the canton level, one administrative level below the province. These blocks are measured to determine the size and are assigned a size number representing the number of whole sampling units that would fit in the block. These blocks are called primary sampling units (PSU) because they will be used as the first stage of sampling from the frame. The size of each PSU is totaled to determine the total number of potential sampling units available in each stratum.

When the total possible number of sampling units (N) is known, the sample size can be determined. The size of the sample is determined by several factors. The primary factors are the desired precision of the results and the budget. There is nearly always a compromise between the desired precision and the available budget. The sample can be apportioned to the strata when the total sample size is determined. In Ecuador the total sample was set to 12,277 sampling units.

The selection of the actual sampling unit is a two stage process. The first step is to select the PSUs. The selection of the PSUs is based on the size of the PSU and is accomplished by either a simple random or systematic approach. In either case, a larger PSU has a greater chance of selection than a smaller one.

In an earlier step each PSU was assigned a size number based on the number of whole sampling units that fit in the PSU. Each chosen PSU is further subdivided into the number of sampling units based on the PSU size. Each of the sampling units should be approximately the same size, homogeneous, and based on good boundaries. Since only the selected PSUs have this additional work, a great deal of labor is saved over the complete enumeration. One sampling unit is randomly selected from the PSU and data is collected from that unit.

The interviewer is given a map to locate the sampling unit, an aerial photograph, if available, with the boundaries of the sampling unit, and a supply of questionnaires to record the data. The interviewer must account for all of the land within the boundaries of the sampling unit. A complete questionnaire is required for each agricultural producer. Also all of the non-agricultural land within the sampling unit is recorded. A quality check can be made when the data collection is complete. The total of all the agricultural and non-agricultural land should be nearly the same as the size of the sampling unit. In the example from Ecuador here were 145,219 interviews in the 12,277 area sampling units.

This is actually an example of an area frame survey. What is different to distinguish the example in Ecuador is the size of the sample. Typically a survey would sample a much smaller number of sampling units. This alone could be used and defended as a legitimate sample census. However, it can be improved by adding a list component to the process to make the multiple frame sample census.

Generally a complete list of all agricultural producers in a country does not exist. This is one reason for conducting an agricultural census. However, there usually is some register or list of the large producers. This could come from the legal responsibility requiring a license to operate a business. This list can be used to supplement the area sample to include the largest producers or producers of less commonly produced crops to improve the results of the sample census.

In Ecuador, the list added to the area frame were of known producers having more than 100 hectares, those producing rare commodities or commodities concentrated in small areas and those producers of non-traditional export products like flowers and shrimp. There were approximately 17,000 agricultural operations on the list and data were collected from all of them.

When the data collection is complete, both the area frame and list frame produce information to be included in the census. When two frames are used together to produce statistics it is called a multiple frame survey. Some additional decisions have to be made. One of the first is to decide which frame has precedence. For example, if an agricultural producer is in the list frame and is also contacted in the area frame, which information is used in the summary? Both cannot be used or duplication will exist. Typically the list information is given precedence. In the example from Ecuador, all agricultural operations on the list were included in the census. Therefore the data collected from the list frame can be used directly.

The data from the area sample must be handled differently. The first is to determine if any of the agricultural producers contacted in the area sampling unit were also on the list. When the data from the area sample is collected, the name and other identifying information of the agricultural producer should be recorded. This information will be used to compare to the list before summarization to determine if the producer matches a name on the list. This is not always an easy task because names can be similar. If the review determines that a producer in the area sample is the same as on the list frame then the area data is considered to be overlapped with the list and excluded from the area data. The remaining data is considered non-overlap.

Second, all of the data associated with the sampling unit must be expanded to represent the total. The factor used is the ratio of the total number of sampling units in the stratum in the first summary level to the number of sampling units in the sample. For example, if the total possible sampling units in a stratum were 50 (N) and the sample chosen were 10 (n) then the expansion factor would be 5 (N/n).

The total for any given administrative level is the sum of the data from the agricultural operations on the list and the expanded data from the non-overlap agricultural producers from the area sampling units. This combination takes advantage of the strengths of both frames. The list will provide data from the large and specialty producers. These will influence the results the

most. The area frame provides data representative of all producers not on the list. In statistical terms, the list provides data with little or no variation and the area provides complete coverage of the population (agricultural producers).

There are two types of error associated with any statistical survey; sampling and non-sampling. Sampling errors are those associated with collecting information from a sample rather than from every possible agricultural producer. Non-sampling errors are those associated with mistakes in collecting the data, such as transposing numbers or the producer forgetting to report about a separate parcel of land he owns. There will be non-sampling errors whether a true census or a sample census is conducted.

The sampling error should be zero if a true census is conducted. In a sample census the will be sampling error associated with the area frame sample that need to be calculated. There are standard mathematical formulas to use to calculate the sampling error. Generally a larger sample will produce a smaller sampling error. A measure commonly used to determine how well a survey estimated an item is called the coefficient of variation (CV). The CV is the ratio of the sampling error of an estimate to the estimate, expressed as a percent. The smaller the percent, the less sampling error is associated with the estimate.

There were approximately 17,000 agricultural producers on the list frame and 12,277 samples from the area frame in our example of Ecuador. The data gathered were used to effectively estimate 23 permanent crops, 34 seasonal crops, 19 types of flowers, and 13 livestock species at the province level. In addition, 10 types of farm equipment, 12 types of buildings, farm labor, producer demographics and economic indicators were estimated at the national level.

As a measure of how well the sample census performed, the census estimated 842,882 farms with a CV of 1.1% and 12,355,831 hectares of land in farms with a CV of 0.4% at the national level. Rice was grown on the most hectares in Ecuador in 2000. The estimate for hectares planted to rice had a CV of 1.5% and the number of rice farms had a CV of 1.7%. It is not too surprising to get this type of result on a crop as widely grown as rice. However, the pimiento estimate of just 956 hectares with a CV of 10.4% indicates the combination of list and area frame samples performed well. The livestock estimates also performed well. The cattle estimate had a CV of 0.8%, hogs 1.8% and sheep 2.2%.

The census conducted in Ecuador is a good example of how an area frame can be used to provide a sample to be used in combination with a partial list of agricultural producers to provide reliable national and provincial level information when budget constraints do not allow for a complete agricultural census. Complete results can be obtained from the following Internet site: http://www.sica.gov.ec/censo/index.htm