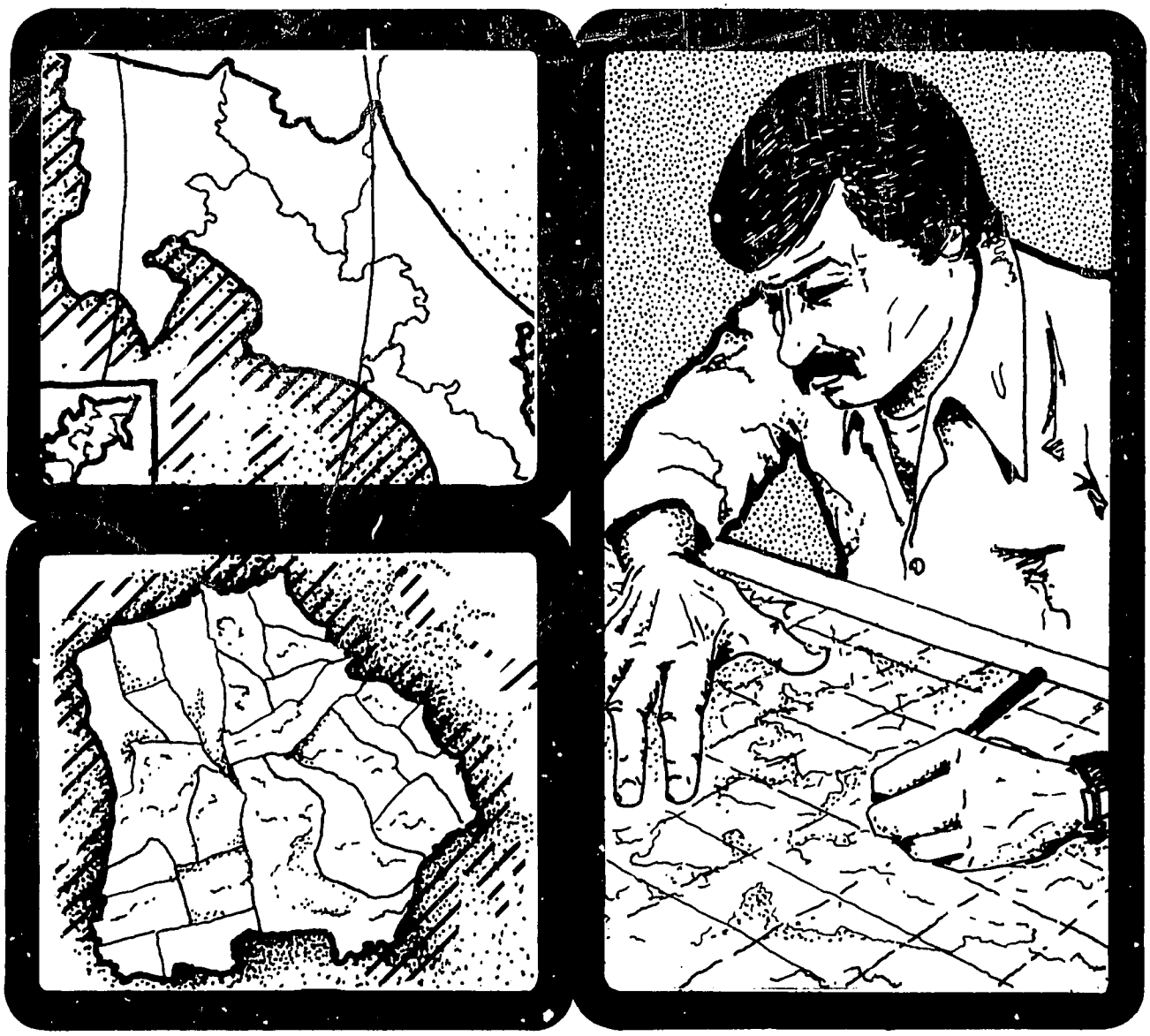
519.82
P819
PN-AAJ-840

PIADIC
Information Management Series

2

Area Frame Probability Samples



IICA
INSTITUTO INTERAMERICANO DE COOPERACION PARA LA AGRICULTURA

AREA FRAME PROBABILITY SAMPLES

Design, Construction and Use of Area Frames for Survey Sampling

PIADIC
INFORMATION MANAGEMENT SERIES

This series contains the experiences gained by
Programa Información Agropecuaria del Istmo Centroamericano

A joint project between:
Instituto Interamericano de Cooperación
para la Agricultura (IICA)
and
Regional Office for Central America and Panama (ROCAP)
(Contract AID 596 - 0048)

Title	Author/Editor	Series No.
Rural Area Profiling	Franklin E. Rosales	1
Area Frame Probability Sampling	PIADIC	2
Managing Information for Selective Dissemination	Finn Damtoft	3
PIADIC Bibliography	PIADIC	4

The series can be obtained from
IICA
Apartado Postal 55, 2200 Coronado, San José, Costa Rica
or
ROCAP
8a. Calle 7-86, Zona 9, Guatemala, Guatemala

Contents

Page	
3	ABOUT THIS MANUAL
3	Organization of Contents
3	Acknowledgments
4	INTRODUCTION
5	FRAME CONSTRUCTION
6	Stratification Concept
8	Uses of Aerial Photography
8	Geographic Institute Cooperation
8	Mosaic Construction
12	Regional Boundaries
13	Checking Frame Construction for Errors
13	Strata Definition
13	Block Size
14	Stratifying the Mosaic
15	Reviewing Mosaic
15	Count Units
16	Measurement and Records
19	Assigning Segments
21	THE SAMPLE
21	Calculation of Sample Size and its Distribution to Strata
24	Selection of a Replicated Sample
26	Subdivision of Count Units
27	Photo Enlargements
29	THE SURVEY
29	Questionnaire Design
30	Enumerators' Manual
30	Pilot Survey
30	Hiring and Training Enumerators and Supervisors
32	Enumeration Procedures
34	Planning and Organizing Field Work
35	Conduct of Field Work
36	Editing and Data Processing

(CONTENTS, cont.)

Page	
38	THE ESTIMATE AND VARIANCE CALCULATIONS BY REPLICAS
38	Summarizing Replicated Samples
40	Paper Strata
43	OTHER TECHNICAL ASPECTS AND SPECIAL PROBLEMS
43	Large-scale Producers
44	Problem Segments
44	Point Sampling
45	Preparation of Frames for Special Areas
46	Size-of-Farm Classification
46	Objective Procedures for Measuring Crop Yields
48	USEFUL REFERENCES

ABOUT THIS MANUAL

We've designed this manual as a working tool for training and field use in area frame probability sampling. It should prove especially useful for those countries or regions that would like to improve the precision of their information gathering —but are stymied by limited resources and a lack of trained technicians.

The manual's methodology and procedures have been tested over several years of extensive application in Central America. Our intended audience is the work team in other countries that will put these tools to immediate use in the field — thus we emphasize a "hands-on" practical approach. We assume that manual users will have had some previous training and that our manual will complement other instructional materials.

Although we have included a brief glossary of base concepts, readers interested in more thorough development of theory and concepts will find the bibliography a useful beginning. Materials and resources needed are listed in an instructor's guide.

Organization of Contents

Discussion is basically limited to "how-to" techniques and application of formulas. These are developed, in sequence of actual field use, over three basic sections:

Area Frame Construction: A step-by-step description of how maps and aerial photographs can be combined as a basis for (1) dividing the total "population" or "region", into (2) area sampling units that (3) have boundaries that can be easily and accurately identified on-site, by (4) interviewers who take the survey questionnaire to the field.

Sampling: Initial discussion deals with sample size and distribution of the samples to strata. Practical guides are given for calculating and selecting of random samples, segments, count units, etc. We describe the use of photographic enlargements in this process.

The Survey: Includes design of the questionnaire and enumerators' manual, use of a pilot survey, hiring and training of enumerators, enumeration procedures, plus planning and conduct of field work and editing and processing of the results.

In the final section we discuss estimate and variance calculations and other technical aspects and special problems. Thus, on completion of the manual we will have provided a precise means for locating our sample work area, developed our sampling procedure and questionnaires, selected and trained our enumerators, completed the field work, and processed the results.

Acknowledgments

This manual was prepared largely from materials developed by Mr. Montie Wallace, PIADIC technician, during several years of extensive training and application in the field in Central America and Panama. Mr. Roberto Alvarado also contributed to the development of this work.

INTRODUCTION

Although many developing countries are said to "lack" agricultural information, data usually are available for most variables; in fact several sets of figures may represent the same variable for different agencies and institutions. But how do we evaluate these estimates? Some are undoubtedly incorrect, while others may be valid.

In this manual, we suggest that scarce resources and talents be used to establish a unified survey system to provide more reliable data and to prevent duplication of effort. Since all decision-makers need selected data at specific times, an obvious need is for a system to gather, publish, and interchange primary agricultural data on a continuous basis.

Our key words here are "system" and "continuous". "System" suggests that the data is gathered and published in an organized manner, with ample time for planning and training for successful operations. The term system also implies studies focused on the most important data needs of the country. "Continuous" suggests that this service is performed on a standardized schedule that varies little from year-to-year — unaffected by changes in government, economic problems, etc. Such high-priority reports should be given permanency by means of an official decree or legislation so that data users are assured of a timely flow of primary agricultural data.

The term "primary agricultural data" includes data that come directly from the farmer and his farm, such as areas planted to specific crops, total production and use of production, livestock and poultry inventory, etc.

To provide this reliable flow of consistent data, the system requires an appropriate vehicle for the collection of primary data from all areas of the society, including farmers and rural homes. A census could be used to collect such data, but the cost of surveying every household is so expensive that developing countries have difficulty in obtaining the necessary budget to take censuses on a regular basis. There is also the problem of processing the data after a census has been taken — this often requires 2 years or more before data are available.

In this manual, we recommend the sample survey as a viable alternative or complement to a census. Although such a survey costs much less than a census, the wages, fuel, and vehicles involved in a national survey remain a significant investment. Thus, we must insure that the results are as reliable as possible and that they can be evaluated. To obtain this reliability, we use probability sampling. For practical purposes, such sampling requires a sampling frame.

An acceptable sampling frame can be defined as a list of *sample units* that: (1) includes all (100%) of the *population of interest,* (2) provides a clear-cut means of identifying each sample unit, and (3) arranges these characteristics so that probability sampling can be done efficiently.

A common sampling frame is a list of farms from the most recent census, often referred to as a "list frame". Such lists usually are deficient because: (1) censuses are never complete, thus any list taken from the census is incomplete — such lists rapidly become less exact as people sell their land, lose rented land, move, etc.; (2) a single unit may be represented by two or three names; and (3) few developing countries are able to handle the expensive and complicated process of preparing and maintaining list frames at a regional and national level. (This discussion relates to large populations; a list frame is usually the most efficient means of sampling small populations.)

The *area sampling frame* described in this manual has been developed to provide the desired characteristics for a sampling frame at the regional and national level. Our area frame sample units are not farms or farmers; rather they are pieces of land called "segments". These segments are identified by boundaries that are defined by physical terrain features such as roads, rivers, railroads, canals, etc. The frame is composed of all maps for a given country or specific region thus assuring total coverage of all units (Fig. 1).

The area sampling frame is simply a "vehicle" that can be used to select a probability sample of farms, farm families, and crop fields by relating them to these segments. Information is collected by interviewing operators of the farms, homes and associated land areas in each segment.

This identification of 100% of all sampling units (segments) in the area frame, without duplication or omission, insures an unbiased result. In the discussion that follows, we will find certain conditions under which the frame is less efficient. However, statistical analysis of results from a probability survey can indicate the areas where the frame is performing poorly, and can indicate the level of confidence that can be placed in the estimates obtained. Thus the area frame technique is self-correcting and can point out its own shortcomings so that they can be remedied.

Fig. 1 Area frame segments are identified by boundaries that are defined by physical terrain features such as roads, rivers, railroads, etc. Frame is composed of all maps for a given area to assure total coverage.

FRAME CONSTRUCTION

The basic objective of our area frame construction is to divide a "region" into uniquely defined segments so that we can select a sample of these segments. To improve the precision of our results, we will also try to group segments having similar characteristics, so that we can sample within each group. This grouping of units with similar qualities is called *stratification*.

Having stated our ideal objective, we quickly see that the work involved in dividing an entire country into, say, .5 km^2 segments, makes such a task impractical, if not impossible. Fortunately, stratification and unique identification of sample segments can be accomplished without physically placing boundaries around each segment.

Stratification Concept

First, let's consider stratification. If we could outline large areas of land with similar uses, or degree of use, we would expect that segments formed within that area or stratum would tend to have similar characteristics. *Intensity of land use* is the variable that we have used in our area frame to stratify successfully for multipurpose agricultural surveys. We consider "intensive use" as cultivation with annual or temporal crops, plus land in permanent crops such as coffee and citrus. However, to draw boundaries around each cultivated field would be more tedious and difficult than dividing the entire country into segments, so we must allow some tolerance in defining stratum "blocks" of a manageable size. A "block" is an area of one stratum completely surrounded by land in other strata. Commonly, a country will use three or four different classifications based on degree of intensity of use.

We can form a separate stratum with land that does not have any agricultural use, such as military reservations, national parks and lava flows.

Areas with high population density make up two strata: one stratum consists of the central part of larger cities, the other is made up of the outskirts of larger cities and the smaller rural towns, which have high population density along with small plots of agricultural production. The latter stratum is referred to as "ag-urban".

Areas covered by large bodies of water are grouped in another stratum that is not sampled. Agricultural land that will be covered by reservoirs for future hydroelectric dams should be placed in a special stratum, if the area to be flooded is known. Thus, we can sample the area while it is agriculturally productive, but can remove this stratum from the frame when the dam is built and the area filled with water.

Special strata can be established where countries have special needs. In El Salvador, for example, a special stratum was made for coffee production. Coffee is certainly of interest in other countries of Central America, but the establishment of a coffee stratum in the other countries was impractical because their plantations were widely scattered within a region. To have included all coffee production in these other countries, the "coffee" stratum would have covered the same area that would have been defined by using only strata of intensity of use.

Table I shows the strata definitions for the area frame projects in the Central American countries. The strata are similar, but with some differences to make the frames more effective under the conditions found in each country.

TABLE 1

DEFINITION OF STRATA BY COUNTRY

	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	Stratum 8	Stratum 9	Stratum 10
	Color % of cultivated terrain in use	Color % of cultivated terrain in use	Color % of cultivated terrain in use	Color % of cultivated terrain in use	Color use	Color use	Color use	Color % of cultivated terrain in use	Color use	Color use
1 Guatemala	Purple 60 100	Light blue 30 60	Orange 0 30	Brown Suitable and used land	Black Un-suitable land for farming	Green Populated areas	Dark blue Water			
2 Salvador	Purple 75 100	Purple 50-75	Light Blue 25 75% but at least 50% of coffee	Blue 15 50	Green Urban farming heavily populated	Green Urban zones	Green with shade Recreational and tourists zones	Orange 0 15	Un suitable land for farming	Water
3 Honduras	Purple Pink Blue 60-100	Light blue 30 60	Orange 0 30	Brown Suitable but not used land	Black Un suitable land for farming	Green Cities and towns	Dark blue Water	Blue Planned Water		
4 Nicaragua	Purple 60-100	Clear blue 25 60	Orange 0 25	Brown Land not suitable for farming	Green Cities and towns	Dark blue Water				
4 Costa Rica	Purple 60 100	Light blue 30 60	Orange 0 30	Brown Land not suitable for farming	Green Cities and towns	Green Urban Agriculture	Dark blue Water			

1 - Stratum VI is divided into VI_1 - Capital City; VI_2 other cities; VI_3 - suburban or agro-urban farming
2 - Originally there were II strata, but stratum II disappeared when a hydroelectric project flooded its area
3 - Stratum I is divided into I_1 - areas without coffee or irrigation; I_2 - areas with more than 30% cultivation of coffee; I_3 areas with irrigation
Stratum VI is divided into VI_1 - urban; VI_2 - Agro-urban
4 - The stratum of the city also is divided into several parts

Uses of Aerial Photography

We have found that the use of aerial photographs allows us to develop stratification without excessive field work — usually the more up-to-date the photographs, the better the stratification. In the heavily-populated older areas of a country, 10 to 20 year old photographs can be used effectively because the intensity of use of the land and field boundaries will have changed very little. However, in areas where forests are being turned into cultivated land, even last year's photographs may not show the full extent of the cultivation. In these colonization areas, field work is necessary to define the extent of cultivated land. If the direction of growth can be determined, it is advisable to make the cultivated stratum larger than actual size in the areas where expansion is expected. Some efficiency may be lost initially, but the stratification will serve for a longer period and efficiency will improve as the amount of cultivated land increases. Similarly, it is sometimes practical to place a "growth band" around expanding cities and towns.

If the aerial photographs of a country are new and have been made into *orthophotomaps* that are true to scale, we can build our area frames on such material. However, in most countries, aerial photographs are available on a country-wide basis only in "contact prints" of different ages and scales. A contact print is a photograph made directly from the aerial photographer's negative without enlargement or reduction (approximately 10" X 10" in size). The scale of a contact print is only approximate and varies from the center of the print outward, due to distortion in the image. Thus, we can see what the land looks like with a contact print, but we can't use it for precise measurement of land area. Also, an individual print, even in a scale of 1:60,000, does not cover sufficient area to allow efficient visualization of strata. Topographic maps in a scale of 1:50,000 fill the need for *cartographic material* that covers entire regions and countries in a standard scale. These prints and maps are usually available from the geographic institutes in each country (Fig. 2).

Geographic Institute Cooperation

Construction and use of the area frame depends upon the cooperation of the country's geographic institute. A working agreement should be drawn up when work begins on the frames, with one or two people assigned to work as liaison between the institute and the frame construction group. The liaison people must become intimately acquainted with the work of the institute to insure use of the most up-to-date photographs and maps. Their tasks will include the selection and ordering of contact prints and maps plus subsequent orders for enlargements of the selected sample segments.

Mosaic Construction

To "see" large areas of land, we can group a number of contact prints into "mosaics". When we unite a number of contacts (with a minimum of overlap), the combination forms a crude but useful "picture" of a large area. Since the topographic maps serve as the standardizing unit, a mosaic is best designed to cover the area of each topographic map, with a margin of at least the width of one contact print on all sides (Fig. 3).

To obtain the contacts for a mosaic we could simply order the prints to cover a specific map from the geographic institute. However, we would be buying approximately twice as many prints as needed for simple coverage, due to the overlap in aerial photography required for map-making purposes. Since the institute will usually have a complete set of contact prints, these can be efficiently reviewed to determine the number of photographs needed. The prints required for simple coverage of each map should be placed

Fig. 2 Prints and maps are usually available at geographic institutes.

together with a margin of one print around the border. Then an order placed for only the selected prints. The price charged for contact prints may vary from $1.50 to $3.00 each (U.S. dollars, 1981). A tremendous saving also can be realized if an arrangement can be made to provide photographic materials. (Most area frame construction is done as a government project, so the institute can be asked to furnish equipment and labor, if we furnish material. This relationship has worked well in El Salvador and Costa Rica.)

The contact prints will be handled and used extensively, so they should be made of double-weight paper, with a matte finish. (Kodak AZO E2 double-weight paper is used most often.) The strata boundaries are marked on the photographs with grease pencils. (Color used to identify each stratum is shown in Table I. Any other type of paper used should be inspected carefully to make sure that it is durable and that prints can be marked with grease pencils, or erased, without damage.

After we have obtained the contact prints necessary to cover a given topographic map, we are ready to construct a mosaic (Fig. 3). First, we'll need a cardboard backing upon which to construct the mosaic – preferable a good grade of posterboard between 1 and 2 mm thick. Ideally, we would buy the cardboard in 1 1/2 x 2 meter sheets, which is the size of our mosaic. However, this is not a common paper size, so we can join two smaller sheets with masking tape. This union becomes quite substantial when the photographs are glued on its surface. The edges of the cardboard also are reinforced with masking tape.

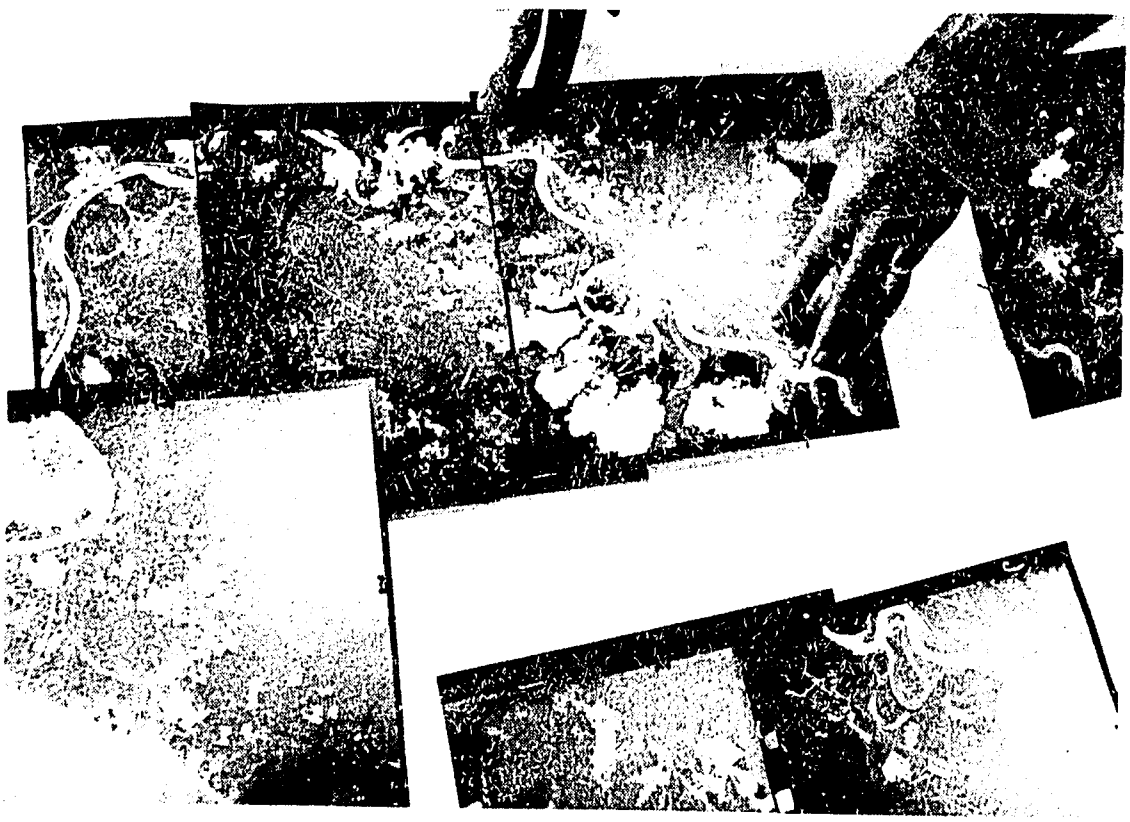


Fig. 3 Contact prints overlap to construct a mosaic.

The only practical way we can work with and store the photographs is to glue them in place. Before gluing, however, we will orient the photographs on the cardboard backing to correspond with the map, with North toward the top. (Fig. 4.) The photos are arranged by flight lines, with a minimum of overlap -- exactly like tiles on a roof. The second row extends slightly over the edge of the first row, etc. After the photographs are laid out, the flight number, photograph numbers, scale and date of photography are recorded on the margin of the mosaic. Individual data for each photo must be recorded on the back of each photo, then the photos are trimmed before gluing onto the mosaic.

After gluing the base print firmly in place, the next print can be put (without gluing) in the best approximate match. When the best match is made, the print can be marked along the edges and corners that overlap the cardboard. This will make it easier to position after the glue is applied.

When our initial layout appears to be satisfactory, we can begin gluing the first line of photos from right to left and from bottom to top, using a one-inch wide strip of rubber cement along only the top margin of each photograph. Rubber cement must be used, as it does not cause wrinkling and dries rapidly, but allows time for some adjustment. If the cement is applied correctly, prints can be taken up later without destroying them. When the first line of photos is in place, the second line can be glued from right to left. Prints are glued only at the top so that they can be lifted (the scene on the lower print could be better than that on the upper print). Immediately, we will find that it is impossible to match all points along the edges of the print due to distortion that occasionally is aggravated by a difference in scale caused by a variation in altitude of the plane from one flight line to the next. Therefore, an approximate match must be made that provides the best continuation of the scene from top-to-bottom and from side-to-side.

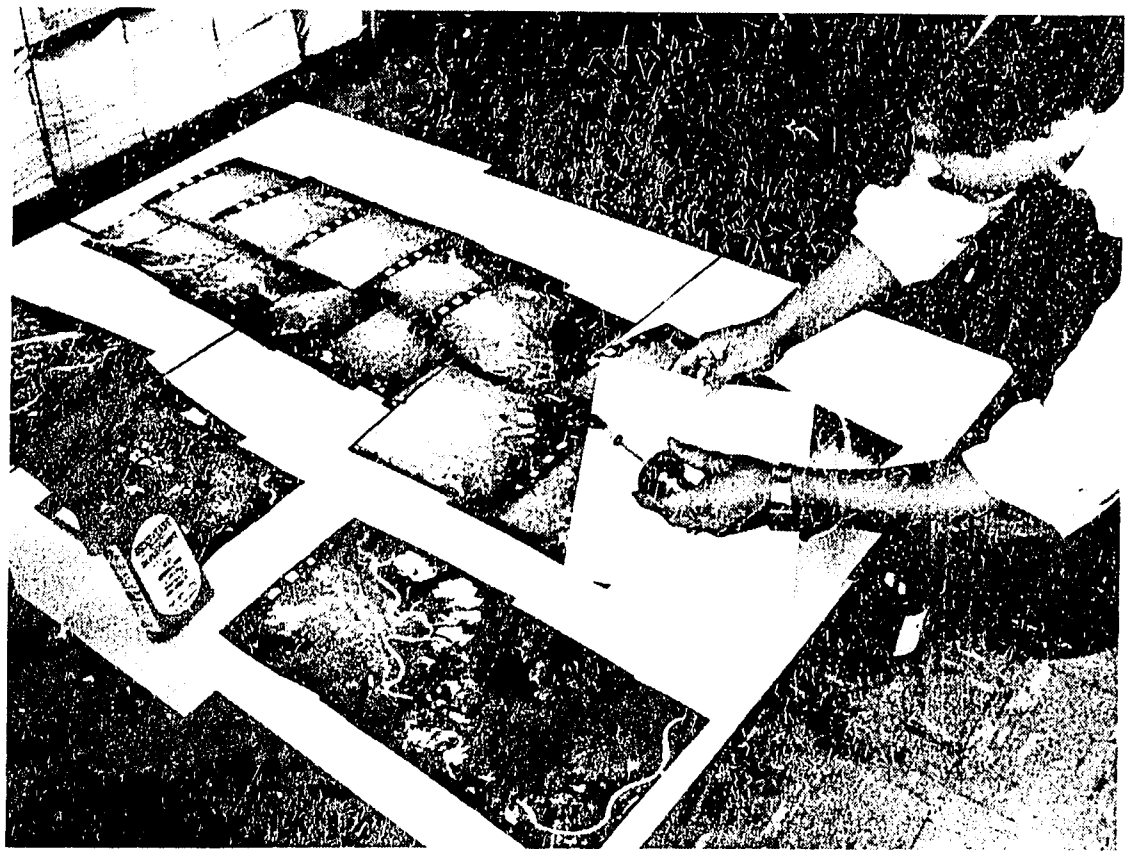


Fig. 4 Photos are arranged by flight lines, glued in place.

Checking Frame Construction for Errors

In our frame construction, all work will be reviewed by someone other than the person who did the original work. The supervisor will settle differences and spot-check as much work as possible in each phase. Teamwork is all-important. The purpose is to identify all mistakes; not to assess blame but to eliminate errors from the final product. This procedure should be established at the beginning of frame construction.

After we have drawn the regional and department or province boundaries on the mosaics, we must transfer these boundaries to the topographic maps. For this transfer, we will locate reference points along the boundary that are common to the photographs and to the map. Such reference points might be crossroads, schoolhouses, churches, cemeteries, points where roads or railroads cross rivers, and distinctive curves in roads or rivers. *Scaling rulers* are useful in converting distances. For example, using the adjusted scale, a reference point on the photograph can be used to locate a point that is the same distance from the same reference point on the map. Thus, we can precisely locate stratum boundaries on the map. If we find that points and boundaries differ in placement between the map and the photographs, we will consider that the photographs are correct.

We will make our first transfers on the map with an ordinary soft-lead pencil (HB). After the boundary transfer has been verified and approved, we will replace this initial penciled line with an appropriate color using a *color pencil*. Caution: grease pencils are *not* suited for use on maps. (Eagle verithin color pencils are excellent for use on maps. Any pencil used should be tested on a map to see if it makes an easily visible mark that can be erased without damaging the map. Blaisdell china markers have worked well on the photography.)

Strata Definition

After the national, regional, and departmental boundaries are drawn on the mosaics and maps, we can begin stratification. The team that does the stratification should first study and memorize the stratum definitions. Working with the photographs, all team members should reach consensus on the appearance of each stratum, then work can begin on individual mosaics.

Block Size

We should also determine the minimum *block size* to be used for each stratum. None of the strata will be "pure"; there are always bits and pieces of other strata within any given stratum. As a guide, our absolute minimum block size will be the equivalent of two segments — always keeping in mind the goal of including the largest possible area in the block.

Strata will be identified by color codes, both on the mosaics and on the maps, with specific colors assigned to the boundary and the stratum designations. Our color for a boundary will usually correspond to the color assigned to the higher-order stratum. Stratum designations are always in the color assigned to the stratum. To clarify the boundary between strata, we can lightly shade the inside of the boundary. All lines should be drawn to the side of, and almost touching, the physical feature that forms the boundary — thus work can be verified without erasures. A line drawn on contact prints will usually obscure any feature if drawn immediately upon that feature.

Stratifying the Mosaic

With the political boundaries drawn, we are ready to stratify a mosaic. First, we will outline the cities and high density population areas, then large bodies of water and large rivers. Next, we outline the most intensive-land-use stratum, followed by strata in diminishing order of intensity of use, so that the least-intensive-use stratum is last. After we outline the cities, town, and bodies of water, we will try to obtain a general idea of land use on the mosaic, before drawing any additional boundary lines. This task is easier if we identify some parts of a stratum and write in the stratum code or number in that area. The basic idea is to identify as many of these areas as possible for each stratum.

We should now have a fairly good approximation of the stratum blocks. Our next step is to refine the strata and identify the physical features that will serve as boundaries. Possible boundaries ranked in order of their acceptability are:

1. Paved highways
2. Secondary all-weather roads
3. Local farm-to-market roads
4. Permanent rivers and streams
5. Railroads
6. Permanent drainage and irrigation canals
7. Intermittent streams and rivers or prominent water-courses that carry water only during and immediately after heavy rains
8. Field boundaries
9. Trails and internal farm roads

As a last resort, we can use lines of trees and small gullies or watercourses, but only after a thorough search for more substantial features. The five top-ranked features are the most useful; every effort should be made to use these five for stratum boundaries.

None of our lines should subdivide a field. A *field* is a portion of a farm dedicated to a specific crop or land use, or portions with the same land use but physically separated by fences or other barriers. Different varieties of the same crop, or with different planting dates, constitute different fields. Often a well-defined path will angle across a field whose squared-off boundaries are equally visible. In this case, the trail is *not* acceptable as a boundary. Each line that we draw has the potential of being part of a segment boundary, thus we must be able to identify it and ask logical questions about the land use inside its limits. For example, if a farmer's field is divided by a line, it is difficult for him to answer questions that apply to the separate parts.

If recent photographs are available, they should be the basis for stratification. The definition of "recent" depends on the amount and rapidity of change in land use in the area. In some parts of a country that have been settled and cultivated for many years, 20-year old photography may still be relatively correct and therefore still useful to us. In other areas with recent deforestation, clearing, and colonizing, we could be misled by photographs no more than 3 years old. If we find that the photographs are relatively old and the area is changing rapidly, we will have to explore all available sources of information to guide our stratification.

Probably our wisest course would be to obtain new photographs for changing areas. Lacking this, satellite multi-spectral scanner MSS images and RBV data may show approximate areas of change in land use, if recent cloud-free images are available. We could also obtain information from experts familiar with the area of the country covered by the mosaic. After we obtain a general picture of the stratification, our next task will be to go to the field with maps and available photographs to locate the physical features needed for stratum boundaries. These boundaries are then established on the topographic maps and mosaics.

Reviewing Mosaic

Each stratified mosaic should be passed to another team member for review. Any questions or differences in interpretation should be settled between these two or in conference with the supervisor. The stratum lines must be shown completely on each print of the mosaic; for example, it is not sufficient to have lines on the upper print only, they also must be shown on the lower print, and vice versa.

The stratification cannot be finally approved until we have completed and checked the adjoining mosaics. These reviews insure that the boundaries continue using the same physical feature from one mosaic to the next. Thus, when we begin stratification on any mosaic, a first step will always be to locate strata boundaries from adjoining mosaics.

After stratification on a mosaic is approved by the supervisor, the strata boundaries are transferred to the topographic maps, using the same procedures described earlier for transferring regional and departmental boundaries. Where there appears to be a difference between the map and the photographs, the photographs are considered to be correct. When transfers have been reviewed and approved, we can replace the ordinary pencil line, with the appropriate color for the stratum, then lightly shade the *inside* of the boundary in the stratum color. First, we will carefully erase a small portion of the ordinary pencil line, then we will put in the stratum color, one small portion at a time.

Count Units

After all mosaics have been stratified and the strata transferred to maps, we can begin construction of count units or frame units. As with the other steps, we begin with the mosaic where each stratum is completely subdivided into count units. Count units are areas with readily apparent boundaries. An optimum size would be an area equivalent to 6 to 10 segments, but will usually vary in size from a minimum of two segments to the maximum equivalent of 15 segments. (Either extreme in count unit size should be avoided.) Count units are used to avoid subdividing an entire region or country into segments. We can determine the number of segments to be assigned to each count unit, thereby locating each segment and determining the total number in the universe.

The count unit boundaries are linked to physical terrain features as described earlier. Count units should be formed so that all are as nearly alike as possible in regard to agriculture and density of population. Ideally, each count unit should contain a representative mix of the stratum characteristics. (For example, a count unit made up entirely of pasture land within an intensely cultivated stratum could significantly increase the stratum variance, if it were selected to contain a sample segment.)

The ideal size of each count unit for each scale of photograph should be established by using a frame or grid to check each proposed count unit. However, obtaining good boundaries is more important than maintaining a precise size. Count unit boundaries are drawn in the stratum color. Count unit construction should begin in the urban strata, then proceed to the most intensive rural stratum and to other strata in order of diminishing intensity of land use.

All work must be reviewed by a second team member and approved by the supervisor. The main points to be considered in the review and approval are: (1) that the boundaries used are physical features that can be located in the field; (2) that the size of each count unit is within the established limits and (3) that the count units do not differ radically in agricultural or population characteristics.

Once approved, the count units are transferred to the topographic maps by identifying reference points that exist on both the maps and the photographs and by making careful use of a scaling ruler to establish relative distances. As with the stratification, the initial transfer is made to the map with an ordinary (HB) lead pencil. Once reviewed and approved, the count units are drawn on the map in the appropriate stratum color as described earlier.

In those areas where we had difficulty with stratification, we will probably have even more difficulty in forming count units. In these areas, we can treat stratum blocks as count units. The edge of the map becomes the outer boundary of the count unit in the absence of other limiting features. (A procedure for selecting and constructing segments under these conditions is described later, see section on Point Sampling.)

Measurement and Records

Next we number the count units in each stratum, within each level of sampling. *Level of sampling* refers to the geographic area for which independent samples are to be selected. In most countries this will be the region, but it also could be department or province.

We have found it convenient to use three numbers to identify each count unit; for example: 1-2-16. The first number indicates the stratum; the second, the order number of the count unit; and the third number represents the area in square kilometers. Thus, for this example, our count unit is number 2, the stratum is 1, with an area of 16 km^2. We arrange the count units in a serpentine pattern, beginning in the northeast corner of the region and ending in the southeast or southwest corner. We can better visualize the region for this numbering if we lay the maps for a region on the floor, joining them as closely as possible (Fig. 5).

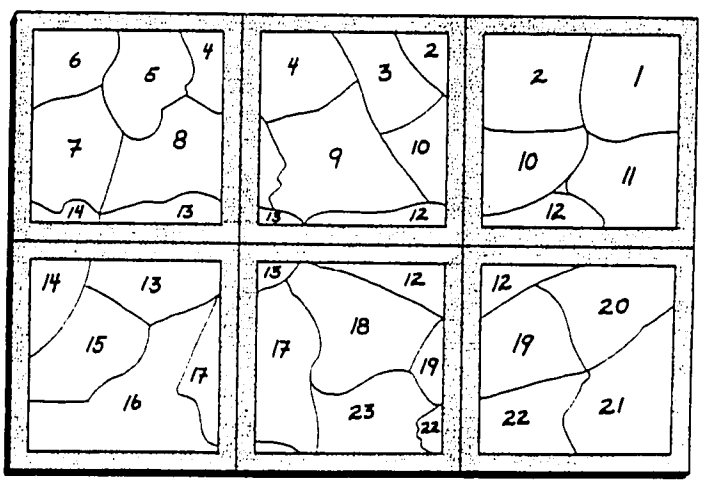


FIG. 5

Count units are numbered in a serpentine manner on map pages. Size of count units is exaggerated.

At the initial stage, we will have only the stratum and the count unit order number. A form must be developed for listing count units within each stratum, by region. This form will indicate the topographic map containing the count unit by name and number; the number of the count unit; space for a first planimeter reading (of area); a grid calculation of area; a second planimeter reading and a column for the final determination of the reading, and area of the count unit (Fig. 6).

AREA PROBABILITY SAMPLES
CONTROL SHEET NO. 2
COUNT UNIT SIZES

Stratum: _____ Region: _____

Province: _____

Count Unit	Name of Topographic Map	No. of Map	First Reading (Planimeter Reading)	First Reading Km²	(Grid Reading)	Second Reading Dif.	Second Reading (Planimeter Reading)	Second Reading Km²	Km²

FIG. 6

Form for listing count units within each stratum

As a control measure, only the supervisor has access to this form during the process of determining the final count unit areas.

We use a *planimeter* to measure the count unit areas. A first planimeter reading is obtained for each count unit on each map page. These readings are recorded on a separate piece of paper and delivered to the supervisor to be recorded on the control form. If a count unit should lay partially on one map and partially on another, the technician whose map has the largest percentage of the total area of the count unit will measure and record the portions and their total. Readings should be in km², accurate to within two decimal places.

Our measurement of an area by planimeter is an averaged reading established by a series of two or more measurements. Team members (who have no planimeter) measure count units with a grid, delivering the results to the supervisor as indicated above.

The supervisor collects the first planimeter readings and grid readings and lists them on his control form. This process insures that those doing the measuring are unaware of the results of their colleagues. If the difference between the first planimeter reading and the grid measurement of a count unit is less than 10%, the supervisor averages these readings to establish the area of the count unit. If the difference is more than 10%, the count unit requires another planimeter measurement by a different team

member. If the second planimeter reading is within 10% of the first, the final count unit area is the average of these two planimeter readings. Or, if the second planimeter reading is within 10% of the grid measurement, the final count unit area is the average of the second planimeter reading and the grid. If the second planimeter reading differs by 10%, or more, from that of both the first planimeter reading and the grid measurement, the process for measuring that count unit is repeated with completely new measurements until agreement is reached.

As count unit areas are determined and approved, they are entered on another listing sheet by stratum and region showing map identification, count unit number, count unit area in km², number of sample segments, accumulated number of segments — with up to 10 columns on the right-hand side of the list available for selection of replicas.

The count unit area decimal measurements are rounded to the nearest whole km² using standard procedures for rounding, as the transfer is made from the first control form. We also enter these rounded sizes on the maps (Fig. 7).

ACCUMULATED SAMPLE SEGMENTS

Stratum: _____ Region: _____

Province: _____

Name of Topographic Map No.	Count Unit	Area Km²	Sample Segments	Accumulated Sample Segments	REPLICAS									
					1	2	3	4	5	6	7	8	9	10

FIG. 7

Assigning Segments

Segments are assigned to the count unit according to segment area:

$$\text{Number of assigned segments} = \frac{\text{area of count unit}}{\text{area of 1 segment}}.$$

For example, if we are working in a stratum in which we want to use segments of .5 km^2, a count unit of 8 km^2 would yield 16 segments: $\frac{8}{.5} = 16$.

We usually determine the required segment size for a stratum by taking into account the following criteria:

(1) The availability of readily-identifiable boundaries that enumerators can find without loss of time. In the strata where land use is intensive, plentiful boundaries exist to form segments of 0.5 km^2 or smaller, if necessary. In the more extensive land-use strata, such boundaries are usually not available.

(2) A manageable number of agricultural tracts for which data must be gathered. A *tract* is defined as all land operated under one administration within a sample segment; it may consist of various non-connected parcels, but must always lay within the segment. Our definition of "a manageable number" depends on the amount and kind of data to be collected for each tract. We usually try to form segments that require an average of no more than one day for enumeration. In a standard multipurpose survey, this practice implies a total of no more than 20 tracts.

(3) A manageable number of resident farm operators. The operator living on his farm within the segment will, of course, be counted in the agricultural tracts, but usually will be asked for additional information that requires more interview time. The concept of "a manageable number" should include the additional enumeration time required for these resident operators. For a standard multipurpose survey, we have found that there should be no more than about 10 such operators per segment.

To assure the appropriate segment size for a particular survey, we have found it useful to test varying sizes along with the questionnaire and enumeration procedure to be used within each of the different strata. Table 2 shows some of the average segment sizes and acceptable variations in size used in Central America.

Our use of physical boundaries for defining segments precludes the formation of uniform segments of a specified size. Instead, a range or tolerance of sizes is established, with the assumption that the average size, measured over all segments, will be the specified size. Many sizes are acceptable, but we are shooting for about .5 km.

After determination of segment size, total area measurements at the regional and national level are checked against the official area measurements and measurements from other sources. Such regional differences should not exceed 10%. (Most of the frames we have constructed have been within 5% of the official measurements on the national level.) Any significant differences must be examined and corrected or justified. If changes are made, the segments must be reassigned correctly.

We can now accumulate assigned segments within each stratum by regions and our frame is almost ready for sample design and selection.

TABLE 2

SIZES OF SEGMENTS BY COUNTRY AND STRATUM (Km²)

Country	GUATEMALA			SALVADOR			HONDURAS			NICARAGUA			COSTA RICA		
Size Stratum	Minimum	Maximum	Optimum	Minimum	Maximum	Optimum	Minimum	Maximum	Optimum	Minimum	Maximum	Optimum	Minimum	Maximum	Optimum
1	0.75	1.25	1	0.75	0.75	0.50	0.25	0.75	0.5	0.5	1.0		0.25	0.75	0.5
2	1.0	3.0	2	0.25	0.75	0.50	0.75	1.25	1.0	0.5	1.0		0.75	1.25	1.0
3	2.0	6.0	4	0.25	0.75	0.50	1.50	2.50	2.0	1.0	2.0		0.5	2.50	1.5
4	2.0	6.0	4	0.75	1.25	1.0	3.0	5.0	4.0	Variable			Variable		
5	1.0	3.0	2			0.25	1.50	2.50	2.0			one block			0.01
6			1.0 block			0.10			One block						0.02
7				0.75	0.75	0.50									
8				1.0	3.0	2.0			0.5-2.0 according to stratum						
9				1.0	3.0	2.0									
10															
11				1.0	3.0	2.0									

THE SAMPLE

Calculation of Sample Size and Its Distribution to Strata*

Our calculation of sample size and distribution of samples to strata presents a difficult problem for the mathematical statistician. We often have poor estimates, or none at all, of the quantities needed in the formulas for the allocation procedure. The formulas themselves are straight-forward, and after a number is obtained for each variable in the formula, the procedures are easily affected. However, the actual numbers obtained for each variable may represent only a guess, thus the mathematical statistician must use these guesses as best he can.

We can get a better understanding of the problem if we use an example that assumes that the cost of collecting the data is the same per unit for each stratum. (The illustration still holds if the assumption is dropped, but the formula would have to include the cost per unit in each stratum.)

For our example, we can suppose that we will require estimates of population totals. Then the formula for the overall sample size is given by

$$\text{Formula 1:} \quad n = \frac{(\Sigma N_h s_h)^2}{s_h V + \Sigma N_h s_h^2}$$

where N_h is the total number of frame units in stratum h, s_h is an estimate of the population standard deviation in stratum h, and V is the desired variance of the population total. Our problem arises in obtaining the values of s_h in formula 1.

If we are fortunate, we may be able to use results of a previous survey of this population. Then we could use the values of s_h generated from that previous survey for estimates of our s_h's. We would assume that the population has not changed drastically since the previous survey, so that the values of s_h are still valid. We will assume that this is the situation for the data from Example 1 (next page).

Before proceeding with our calculations, we must specify the variance or coefficient of variation (C.V.) of the estimate of the population total. Suppose we desire a C.V. of 2%.

Then C.V. = .02 = $\frac{s}{\hat{y}}$ so that s = (.02) (26,681) = 533.62.

The variance would be $(533.62)^2$ = 284750.3.

* From dissertation by Dr. Robert Tortora, USDA/SAS Statistical Research Division, Washington, D.C., given at Statistical Seminar in San Jose, Costa Rica, April 1, 1981.

Example I

Average Number of Cattle per Farm and Total Number of Cattle

A district contains 2,072 farms which have been divided up into five strata on the basis of size of farm in acres. For the N_h farms in a stratum, a random sample of n_h farms is taken (see Table 1) and the number of cattle on each farm in the sample is determined. The sample mean $\bar{y}_h$ and the sample variances $s_{\bar{y}_h}^2$ are calculated from the sample in each stratum. The object is to estimate the average number of cattle per farm.

Table 1: Stratification of farms by area

Stratum	N_h	n_h	$\bar{y}_h$	S_h^2	s_h
Under 16	635	84	4.24	27.54	5.25
16-30	570	125	11.63	55.84	7.47
31-50	475	138	15.95	71.70	8.47
51-75	303	112	23.59	192.32	13.87
Over 75	89	41	29.61	334.93	18.30
Total	2,072	500			

We have

$$\Sigma N_h \bar{y}_h = 26{,}680.81 \text{ and } \Sigma N_h \bar{y}_h / N = 12.87$$

$$\Sigma N_h (N_h - n_h) \frac{S_h^2}{n_h} = 445{,}467$$

Thus the average number of cattle per farm is 12.87 and the variance of this estimate is 445,467/4,293,184 = 0.103761. The standard error is 0.322 and the coefficient of variation of the estimate is 2.5 percent. Furthermore, an estimate of the total number of cattle in the district is 26,681 and its standard error is 668.

If we substitute the appropriate values in formula (1), we obtain:

$$n = \frac{(17446.21)^2}{284750.3 + 171455.93} = 667$$

Now the total sample must be allocated to the strata. The formula is:

$$\text{Formula (2): } n_h = n \frac{N_h s_h}{\Sigma N_h s_h}$$

Substituting the appropriate values for each stratum in formula (2) gives the following allocation:

Allocation of 667 Sample Units to Strata

Stratum	Sample Size
Under 16	127
16-30	163
31-50	154
52-75	161
Over 75	62
	667

For the statistician who does not have estimates of variance from previous similar surveys, other alternatives are available.

Alternative 1 assumes that we have sufficient resources (time, money and manpower) to conduct a pilot survey to obtain estimates of variances. A pilot survey is a prerequisite to test survey concepts and procedures, and if adequate resources are available, we can make this survey large enough to obtain our variance estimates. Or we could divide the survey into two stages: an initial smaller stage to test concepts and procedures, and a larger second stage to obtain variance estimates.

Alternative 2 follows the approach suggested by Deming (1960) for estimating variances based on the sample unit distribution curve. If h is the range (largest value minus smallest value) of the distribution of sample units, the standard deviation is calculated as shown for the different distributions:

(Shape a) normal ($\overset{\frown}{\leftarrow h \rightarrow}$), then $\sigma = .167h$

(Shape b) equilateral triangle ($\overset{\wedge}{\leftarrow h \rightarrow}$) then $\sigma = .20h$

(Shape c) right triangle, skewed right or left

($\leftarrow h \longrightarrow$ or $\longrightarrow h \longrightarrow$), then $\sigma = .24h$

(Shape d) uniform ($\overline{\leftarrow h \rightarrow}$), then $\sigma = .29h$

We have found this procedure very useful with the area frame, closed-segment estimation procedure, because an upper bound, say B_i, is put on segment size in the it $_h$ stratum. Thus, in each stratum the range is automatically determined; i.e., $h = B_i$. This allows us to determine the shape of the distribution on the basis of the stratum definitions and/or observation of the aerial photography (if it is current) of some typical segments.

Alternative 3 makes use of a similar survey that might have been made in another country, using nearly the same concepts (e.g. the same stratification procedures). With this approach we might use that country's sample sizes, or approximate sizes that fit our budget. We caution that this alternative should be used with great care — and only after insuring that there are similarities in population characteristics that are to be estimated, with approximately equal stratum variance. This alternative calls for close communication and cooperation between countries and should *not* be attempted unless this exists.

Selection of a Replicated Sample

The type of sampling most often used with the area frame is called *systematic replicated sampling* or *interpenetrating sampling*. Replicated sampling involves the systematic selection of a number of independent samples or replicas from the same universe, rather than use of a single sample. For example, if we use replicated sampling and want a total sample size of n = 40, we could select 4 replicas, each having 10 units, or 5 replicas having 8 units, so that each replica is a valid sample of the universe.

As an example, let's take a population of 500 segments (i.e. N = 500). This could represent a stratum in a region of any country. Our optimum sample size is calculated at n=20. Replicated sampling will be used in which 4 replicas (independent samples) of n=5 will be selected. To systematically select samples of n=5, we first calculate the sampling interval: 500 : 5 = 100. This tells us that we are to select one n unit from every 100 segments. To select the sample, a *random number table* is used to choose a number between 1 and 100 (the width of the sampling interval); let's assume this number is 43. Now, assuming that the segments are arranged in an order that will be maintained, we know that segment number 43 is the first unit in the first replica. The second unit is determined by adding the sampling interval to the number of our first unit selected: 43+100=143. Segment number 143 is the second sample segment, in the first replica. (The third sample segment would be 143+100=243; the fourth, 243+100=343, and the fifth sample segment, 343+100=443.) Thus we will have identified all sample segments in the first replica. To select the other replicas, we would choose a random start between 1 and 100 for each replica and apply the sampling interval as above. No random start is repeated, as this would produce an identical sample. With this system, we have only 100 possible samples of n = 5.

For convenience, the sampling interval should be a whole number, and since we are working with area segments we can always arrange such an interval. For example, let's say that stratum I in Region II, of country X, has an accumulated total of 948 assigned segments. Our sampling team then decides that it is appropriate to use a total sample of N=50, distributed in 5 replicas of n=10. The sampling interval would be 948 : 10 = 94.8. In this case by adding two segments to the population (950) a whole number (95) sampling interval can be obtained. We could add these two segments by simply assigning an additional segment to two different count units in the stratum. These segments should be added to count units where it will least affect average segment size — usually these would be the larger count units. If our initial sampling interval had been calculated at 94.3, we could have obtained a whole number by removing 3 segments from the stratum. This could be done by removing one segment from three different count units, always selecting count units so that the average segment size would be least affected.

Figure 8 shows a work sheet in which 3 replicas of n=5 have been selected from a population of 200 segments. For this example, our upper number (19) in the column for replica I is the random start selected from a random number table. Our sampling interval is 40. Since segment number 19 is the first segment in the sample, the number 19 is listed next to count unit 2 that contains the 19th segment. (The column of accumulated sample units shows that count unit number I has segments I through 12; count unit number two contains segments 13 through 26; count unit three contains segments 27 through 44, etc.) Our next segment is selected by adding the sampling interval (200 : 5 = 40) to the random start: 19+40=59. This segment is in count unit 5 that contains segments 55 through 68, thus the number 59 is listed in the column for replica I as part of count unit 5. We continue this process of adding the sampling interval so that the 5 segments in replica I are selected; each of the selected segment numbers is listed next to the appropriate count unit. We follow the same procedure for succeeding replicas, using a different random start for each replica.

AREA PROBABILITY SAMPLES
CONTROL SHEET
ACCUMULATED SAMPLE UNITS

Stratum: _____ **Region:** _____

Province: _____

Name of Topographic Page	No.	Count Unit	Area km²	Sample Units	Sample Units Accumulated	REPLICAS 1	2	3	4	5	6	7	8	9	10
Concepcion 6330	III	1	6	12	12										
		2	7	14	26	19		21							
		3	9	18	44		38								
		4	5	10	54										
		5	7	14	68	59		61							
		6	8	16	84		78								
		7	4	8	92										
Ojancho 6329	IV	8	5	10	102	99		101							
		9	7	14	116										
		10	6	12	128		118								
		11	5	10	138										
Arroyo Grande 6329		12	4	8	146	139		141							
		13	8	16	162		158								
		14	6	12	174										
Peñas Blancas 6328	III	15	7	14	188	179		181							
		16	6	12	200		198								

FIG. 8

Subdivision of Count Units

We have now identified each count unit containing a sample segment. Our next task in sample selection is to subdivide only the selected count units into the assigned number of segments. As one team member subdivides these units, another member checks the results. After the team agrees that the subdivision is satisfactory, each segment is assigned a number. Then the supervisor is asked to provide a random number to select the segment. The supervisor notes the random number provided and the count unit; he handles the random numbers so that the technicians making the subdivision cannot unintentionally bias the selection. The subdivision of count units into segments is usually color-coded in yellow; first on the mosaic, then on the map. The segments selected for the replicas are given numbers color-coded in red. Selected segments should be outlined in red on two different sets of topographic maps. One set will be used by field supervisors to plan their work and to locate segments, the other map will be for office files as the permanent record of the area frame.

Our subdivision of count units is critical to the efficiency of our field work and our survey results. Ideally, subdivision would create each segment as a small representation of the count unit, so that each segment would contain its proper proportion of cultivated land, forest, pasture and occupied dwellings. Above all, we must not create segments that vary greatly — that is, so different that one or two segments might contain all the occupied dwellings, with other segments having all the crop land, and others the forest or pasture. Such segments could substantially increase the variance of survey results. And we must be extremely aware of the need for good segment boundaries. If the boundaries are bad, our enumerators will lose time in locating the segment and may make an incorrect identification that will contribute to the non-sampling error. Considerable administrative and supervisory time can be lost in correcting badly assigned boundaries.

Occasionally, we might not be able to subdivide a count unit into the assigned number of segments, if we use the boundaries that are visible on the photograph. In that case, we would make all possible divisions into individual segments, and assign the appropriate number of segments to the undivided area. Then we would make a random selection. If we are fortunate, one of the single outlined segments will be selected. If the random number should fall in one of the undivided blocks, our next step would be to go to the field and divide that block into the assigned number of segments — one of which would be selected at random. But, if this block should prove impossible to subdivide, the total block would be used as the segment. The enumeration result would then be reduced to a one segment equivalent; i.e. if the block is the size of 3 segments, the result of a total enumeration would be divided by 3, before being summarized.

For example, Figure 9 shows the subdivision of a count unit that has been assigned 9 segments. The portion numbered "3-4-5" could not be subdivided, so it was assigned segments proportionate to its area, and these segments were all given numbers as shown. If we now should randomly select one of the numbers 1, 2, 6, 7, 8, or 9, that segment number would become the sample segment, presumably with good boundaries. If the numbers 3, 4, or 5 were selected, we would go to the field and divide that block into its 3 assigned segments — then select one segment at random to be the sample. In the field, we might find that one segment could be identified, leaving a two-segment block that could not be subdivided. Again we could make a random choice: if the single segment is selected, we have a standard sample segment. If the two-segment block is selected, it becomes the segment and must be entirely enumerated. The enumeration results will be divided by two before summarizing.

FIG. 9

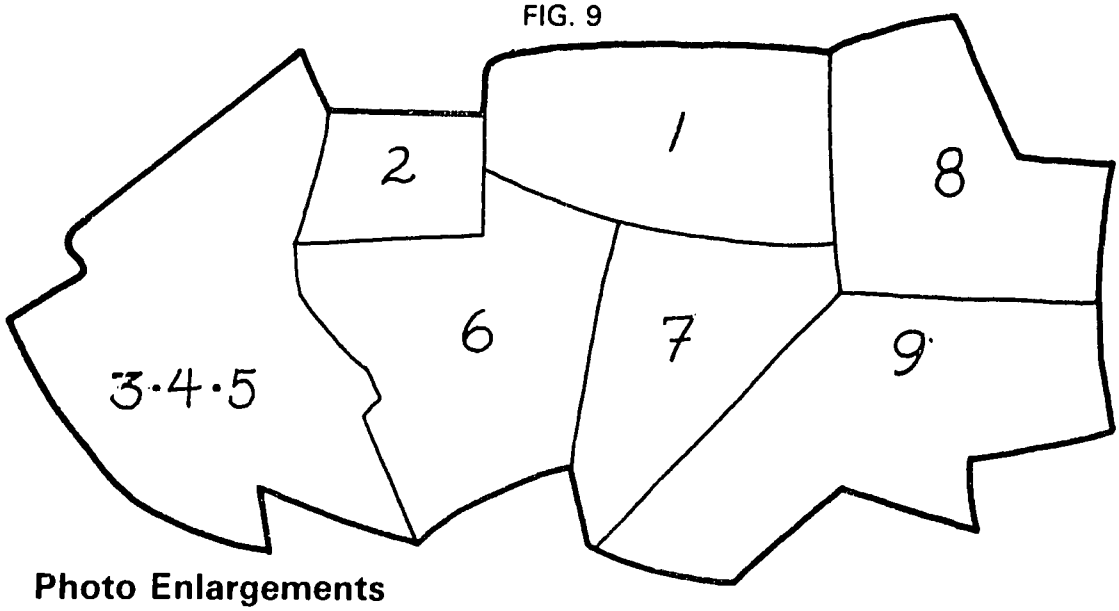


Photo Enlargements

After segments have been selected, we order photographic enlargements for these segments and their surrounding areas.

An approximate scale of 1:5000 has been found to be satisfactory for field work. This scale is convenient for the agency that makes the enlargements, since most of the photography scales are multiples of 5,000. (However, other scales may be used readily, when available.)

We also have found it convenient to use enlargements that are all of the same size (making it easier to construct protective portfolios and storage cabinets). A commonly available enlargement paper is available in sheets that measure 20" X 24", so this becomes a good size for the enlargements. Our experience suggests that we specify the required size — otherwise enlargements are likely to be delivered in varying sizes, with segments off-center. We obtain the best enlargements by selecting the contact print in which the segment is centered in the photograph and noting the identification data for that print. To achieve a standard size, we construct a frame from plastic or cardboard that serves as an outline for the print area that would be enlarged to 20" X 24" at a scale of 1:5000. If the contacts are in a scale of 1:20,000, our frame size would be 5" X 6". For a contact scale of 1:30,000 the frame would be 3.33" x 4". The segment should be centered in the frame to obtain as many points of reference as possible around the segment. Our results will be more precise if we indicate the area to be enlarged on the negative.

Our experience has been that, for purposes of control, it is better to order enlargements in small lots of 10 to 30, beginning as soon as a group of segments is selected. Production of enlargements and selecting of segments can be parallel processes and the geographical institute will appreciate your consideration of their work load.

When the enlargements are received, we outline each segment on the enlargement in red ink, with the ink line just outside, but touching, the physical feature that serves as a boundary. Before inking, the segment should be outlined on the enlargement in soft-lead pencil, with the validity of the transfer verified by the supervisor. We use ink instead of the grease pencil to prevent changes by field personnel and because it is more durable.

In addition to the segment boundaries, the enlargement should have an arrow indicating North in the upper right-hand corner, with a segment identification number in the upper left-hand corner. The segment identification number should show region, stratum, replica and segment order number within the replica. Thus, we use a 7 or 8 digit number, which provides a unique number in a system that will last for 99 years.

Although area segments are often measured on the maps, we have found it difficult to accurately locate some segment boundaries on the map, due to the lack of common reference points for the map and the photographs. Although the scale of the photographic enlargement is only approximate, since the scale of the original photography is only approximate, these enlargements are a true representation of the segment on the ground. Thus, we have found our measurements are more accurate on the enlargement. For this task, a relatively accurate scale must be established for each segment. We establish the scale of an enlargement by measuring between two points that can be positively identified on the ground and on the photograph. The line between these two points should be at points that are at average elevation of the segment (definitely not at either the lowest or highest point) and the distance between the two points should be at least 300 meters. The scale then is:

$$1: \frac{\text{Distance on the ground in meters}}{\text{Distance on the photo in millimeters}} \times 1000$$

If we measure 315 meters between the two points on the ground, and 55 millimeters between the points on the enlargement, the scale would be:

$$1: \frac{315}{55} \times 1000 = 1: 5727$$

With the scale established, an accurate measurement of the area can be made on the enlargement for use in control of the enumeration. The measurements for establishment of the scale can be done by the supervisors during the first survey, or can be done on a field visit to segments prior to any survey.

THE SURVEY

Questionnaire Design

Our basic questions as we begin to design a questionnaire are: What is the purpose of the survey? What kind of data tables will we want? What comparisons and analyses will we make with the survey results?

We prepare dummy tables with lists of all types of desired comparisons and analyses. And are ruthless in discarding data that has no specific use. Only after firm commitment to the data needed and its use, can we begin to formulate efficient questions. If our results are to be processed by computer, the programmer in charge of the processing should be included in the discussions of variables, tables, analyses, and questionnaire format.

Temporary enumerators are used in most surveys, thus we follow the general rule of writing each question so that the enumerator will get results, even if he only reads the question to the respondent. The questions are written simply and in phrases or terms that the respondent can understand. Some variables require the use of tables, however.

We use "skip" techniques and short enumerator instructions throughout the questionnaire. For example:

Do you have any hogs or pigs? _____ YES _____ NO - go to Section 2

continue

(Enumerator instruction) "All of the following questions refer to the entire farm as established in question 16".

For a general purpose national survey, we limit interview time to 30 minutes or less per individual farmer.

Regardless of the number of variables, the questionnaires are of a physical size that is easily manageable by the interviewer (i.e., fits on his clipboard).

Questions are arranged so that the interview flows naturally, with no skipping backward and forward between questions. If possible, the initial questions are the easier to answer, are of interest to the farmer, and are non-controversial. Enumerators obtain an answer for *every* question.

Where there are classifications of livestock by age or sex —or of land by land tenure— a space should be provided so that the enumerator can sum the classifications and verify the total with the respondent. If there is disagreement, the enumerator should re-check each classification until agreement is achieved. We reserve a space for enumerator observations on each page, so that they can provide explanations for unusual situations.

The questionnaire should be of good quality paper that can withstand erasures, rainstorms, and much handling.

Aside from the regular questionnaire on which the data are gathered, an area frame survey requires a tract control form that lists individual tracts within the segment, the operator's name, and the area of the tract. This form also may have a series of questions that determine whether the tract should receive a questionnaire. Listings of non-agricultural tracts will appear only on this control sheet.

Enumerators' Manual

We prepare an enumerators' manual along with the preliminary version of the questionnaire. The manual explains how to enumerate the segment and outlines exactly what is expected from each question. It defines all terms used in the survey, such as *cultivated land, annual crops, fallow land, interplanted crops, improved pasture, milk cow, beef cow, heifer, steer,* etc. In surveys that are done periodically to provide a data series, procedures must remain standard. Our enumerators' manual provides a means for making duplicate surveys despite changes in supervisors and administrators; it also is useful for those who interpret survey results.

Pilot Survey

We test all questionnaires, manuals, and survey procedures prior to their use in a major survey. A trial or pilot survey is made on a small number of segments that represent as many of the different agricultural situations as possible in the area to be surveyed. However, our previously established sample segments are not used, as they will be revisited during the survey proper. Enumerators for the pilot survey have background characteristics similar to those expected to serve in the survey proper. They receive the same training that is planned for the enumerators of the main survey. Staff who designed the questionnaire are best not used as the enumerators, as they might subconsciously bias the results.

During the pilot survey, all enumerators and supervisors review the questionnaire and survey procedures. They note interview time, fuel use, administrative needs, etc. Periodic meetings (or at least a final meeting) should be held in which everyone participating in the pilot survey has an opportunity to share opinions about the questionnaire and survey procedures.

We usually discover a number of questions that are awkward or inadequate, and may find that some field procedures require changes. Based on the results of the pilot survey, we prepare the final version of the questionnaire and enumerators' manual. The pilot survey also provides ideas on points that may require special emphasis during the training sessions.

Under certain conditions data gathered during the pilot survey can be used to check computer tabulation programs.

Hiring and Training Enumerators and Supervisors

As basic requirements, enumerators should be able to: (1) read and understand the enumerators' manual and (2) learn the use of the aerial photographs and maps. A simple quiz can be devised that tests reading comprehension, ability with arithmetic, and ability to use maps. We have included an example of such a quiz as Appendix A. Appropriate time limits and "passing" grades can be determined after some initial use of the quiz.

We have found that supervisors and enumerators should not be selected from the same source. Supervisors should receive additional training, and also should attend and assist with enumerator training.

Our enumerator training school covers the following subjects:

1. Purpose of the survey
2. Responsibilities as enumerators
3. Interviewing techniques
4. Definitions and technical aspects
5. Enumeration procedures and use of photography and maps
6. Questionnaire in detail (question by question)
7. Practice and/or simulated interviews
8. Field practice
9. Critique of field practice
10. Administrative procedures

In general, our aims with this approach are (by subject) to have the enumerator:

(1) fully understand the utility of the survey. If the enumerator is convinced that his work is worthwhile, he will carry this conviction to his respondents.

(2) conduct himself as a responsible person, by explaining the purpose of his visit, maintaining the confidentiality of individual data, and refraining from arguments.

(3) establish a good relationship with the respondent. He notices good points about the farmer or his family, without wasting his time. We suggest that there be no onlookers during the interview. (The interviewer should leave with a good word; he will probably be back.)

(4) understand enough about the sampling procedure to be able to explain random selection and the selection and the concept that all types of farms must be represented. One of the questions most often asked is "Why did you pick me?" Enough understanding of the technical aspects of the survey must be imparted so that the enumerator can explain his work. All terms used in the survey shall be defined, such as segment, tract, field, producer.

(5) understand and use the photographs and maps. In training, we use copies of a simple photograph and map so that all enumerators are visualizing the same area. Some training tips: relief on the photograph can best be seen, if the print is rotated until the shadows fall toward the viewer. Look for a number of physical features on the photo, such as rivers, large trees, lines of trees, cultivated land, paved roads, cemeteries, secondary roads, houses, cities and villages. Enumerators should locate reference points as indicated by the instructor and should be taught how to approximate distances on the photo using a ruler —and on the map, using its one-kilometer grid. Show procedures for orientating a photo or a map on the ground (without a compass) using reference points such as crossroads, lines of trees, and single trees. The sun also can be used to establish direction. First efforts of beginners in the field must be closely supervised as they learn to locate reference points.

(6) know the questionnaire perfectly before beginning enumeration. Each question should be individually discussed. Often we can send a questionnaire and enumerators' manual to each enumerator prior to the training school, so that they can complete an exercise in filling out the questionnaire prior to arriving at the training school. This encourages a thorough study of the manual. (The section that follows, Enumeration Procedures, provides more detailed instruction.)

(7) simulated interviews, with supervisors playing the role of enumerator and respondent. The enumerators individually note the answers in the proper spots on a blank questionnaire. Each of these questionnaires should be checked by the instructors — to evaluate the training and to determine areas that may require additional emphasis.

(8) obtain field practice near the training location. Enumerators are placed in small groups of 3 or 4, then go to the field with a supervisor to locate and enumerate a practice segment — one segment to each group. Each group uses exactly the same materials and same type of enlargements that they will use in the survey proper.

(9) return to the classroom where they receive a critique of their work, and have an opportunity for questions and to make comments.

(10) discuss administrative details, such as how and when they will be paid, what to do in case of accident, etc.

Enumeration Procedures

Before beginning enumeration we must be absolutely certain that we have identified all segment boundaries. The first time a segment is enumerated, we travel around it, either by vehicle or on foot, to be sure that the boundaries are identifiable. Following this initial check, the interviewers begin interviewing at the nearest house or with the nearest workers in a field.

When crops and land use are being studied, a first step is to determine the extension and location of the tract. To assure greater accuracy, we ask the respondent to accompany us. We outline his area of operation on the enlargement using a blue grease pencil. This tract is then identified with a letter (in sequence). Name and address of the tract operator are recorded on the segment control sheet, along with the identifying letter assigned to the tract. Then the control sheet questions are asked to determine if tract is to receive a questionnaire. If the tract is *not* to be given a questionnaire, the land use and area are entered on the control sheet.

Closed Segment: For tracts receiving the questionnaire, the enumerator then proceeds to gather data on activities *within* the tract. We call this *closed segment* enumeration. In this process, each field is outlined with borders in red. These fields are identified with a number and the required data is recorded. The red lines can be drawn at the same time that the tract is drawn. The data for each field will be recorded on the questionnaire.

Each enumerator estimates field areas with a transparent grid that can be used with the photographic enlargement. Use of the grid is not intended to replace the producer's estimates of the area in each of his fields; rather it is to assist the enumerator in detecting gross errors. It should *never* cause argument between the enumerator and the respondent. If the producer should estimate an area for a field that is obviously incorrect, he may simply be uncertain about the field in question, or perhaps he is including another field planted to the same crop. If the respondent insists on an obviously incorrect acreage, the enumerator should record the reply, but

supply his own opinion of the situation and the area — in the "observations" section of the questionnaire. (Remember: our grids are constructed for a specific scale, say 1:5000, and there will be few of the enlargements exactly to that scale. Thus, measurements made with the grid can be very imprecise and should serve only for detecting gross errors.)

We can use the closed-segment method to enumerate cattle, provided the animals are controlled in pastures and corrals. To use this method, we first take the data for all animals that are kept just in the tract within the segment. Next, we determine if any animals can freely enter or leave the segment. (For example, if the operator has opened a gate to allow the cattle to make use of a pasture outside the segment, or where the cattle are able to cross a stream or river that forms part of the segment boundary.) If the animals can move freely, we must determine the total area within the segment plus the area outside the segment over which these animals can move. Before the data are tabulated, a proportion of the total of these freely-moving animals will be assigned to the segment based on the proportion of the total area that is within the segment. This method is used *only* when the animals can move of their own free will into and out of areas contiguous to the segment. (If the operator moves the animals to a pasture across the road and shuts the gate, for example, the count on these animals depends on their location at the time of the interview.)

We do not use the closed-segment method for livestock or poultry that are herded on open range or allowed to run loose (nor for swine and poultry in Central America, since most are allowed to run free, returning to the farmer's house in the evening).

Open Segment: After we have taken all the required data for the tract, our enumerators can apply the *open segment* procedure used to collect data for the entire farm. Use of the open segment requires some unique reference point that will allow a farm to be identified with one (and only one) segment. In practice, the easiest reference point is usually the operator's house. However, this requires a sample in cities and urban areas, because some producers or operators live in cities and towns. Also, we must precisely determine who is the operator. Sometimes this is not a simple task. For example, let's say there are three brothers operating a farm. Each lives in a different house: two outside the segment, and one inside. When interviewed, they say that they share equally in the decisions. Who is the operator? One and only one must be identified, to avoid duplication. Although any one of the three can probably provide all required information about the farm, the location of the house of the one who is designated as the operator will determine whether an open segment questionnaire will be filled out for the farm. In making this decision, when all share equally in the management, the location of the residence of the eldest brother can be used. If he lives in the segment, the farm can be enumerated under the open segment approach.

If it is not feasible to sample in the cities, a series of reference points and a process of elimination can be used for selection of the farm. First, we would ask if the operator lives on the farm. If he lives on the farm, and the dwelling is within the segment, the farm is enumerated. If he lives on the farm, and the dwelling is outside the segment, we do *not* enumerate. If the operator does not live on the farm (let's say he lives in the city), then we ask if he has an administrator. If so, we ask if the administrator lives in or out of the segment.

If there is no administrator, our next question would be to find out if the farm has an occupied dwelling — and its location. If we find no occupied dwelling, we could determine the location of the largest structure on the farm. If the farm had none of the above, we would proceed to the location of the main entrance. If it is in the segment, we enumerate; otherwise we would not. Each step must be mutually exclusive. We have found this elimination procedure very difficult to apply, unless enumerators are especially well trained. Usually, we have used the residence of the operator and have sampled in the cities.

Weighted Segment: We can also use a third enumeration procedure known as the *weighted segment,* in which total data is gathered for any farm that has land in the segment. This collected data is "weighted" or allocated to the segment according to the proportion of the farm that is in the segment. If 10% of the farm is in the segment, 10% of the collected data will be assigned to the segment. Thus, for each tract in the segment, we must locate someone who can provide detailed information for the entire farm. This could substantially increase the enumeration time; but — in the example of the three brothers — we would not be concerned with which of the three is the operator. (The true weighted segment, as described here, has been used only experimentally in the United States. It should be tested before any large scale plans are made to use it.)

Planning and Organizing Field Work

We schedule our national survey so that all regions begin simultaneously, thus the data obtained represents comparable time periods. Our goal should be to complete the survey within two weeks. Using segments of the size shown in Table 2, we estimate 3 man-days per segment for the *first* survey. Our average of one-day of enumeration time per segment is reached after a number of surveys; however, much more time should be allowed for the initial survey.

A centralized organization for survey that should prove useful in some countries is shown in Figure 10. (By "centralized" we mean that all staff can operate from the capital city or other single headquarters to do the survey, and that there are no regional or departmental offices involved.)

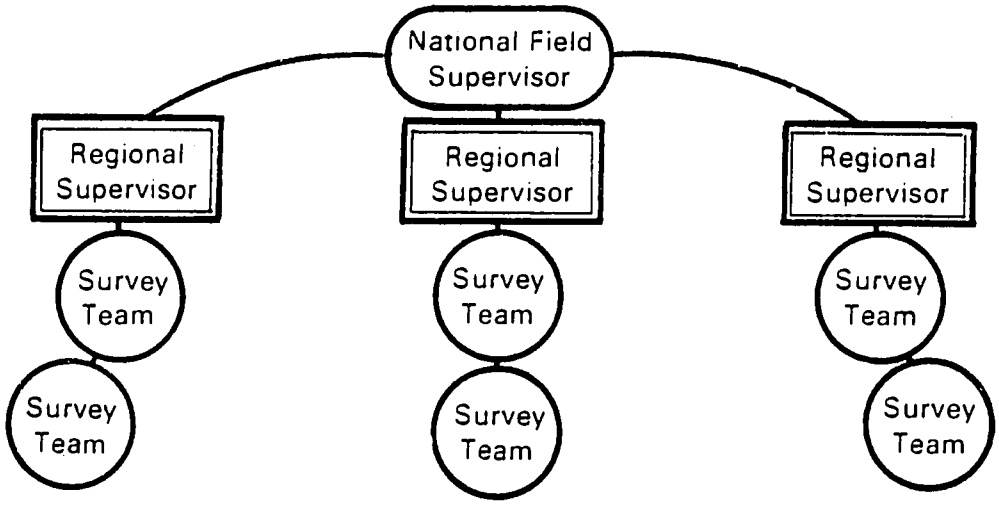


FIG. 10

A survey team consists of a supervisor, and 3 or 4 enumerators, (plus their vehicle). The number of survey teams assigned to each region will depend on the number of segments in the region and the time allowed to enumerate the region. Costs can be reduced by hiring supervisors who can drive the vehicle, thus no chauffeur will be needed.

Survey planning requires good sets of national and regional maps showing all roads, cities and towns, and the location of sample segments. Groups of segments can be assigned to teams, routes for arriving at the segments can be selected, and vehicle mileage estimated.

Segment assignments are prepared for each team, along with packages of blank questionnaires, control sheets, writing tablets, and the maps and enlargements of the assigned segments. Each enumerator is provided with a clipboard, writing tablets, pencils, small plastic ruler, red and blue grease pencils, an eraser that won't damage the enlargement, some type of carrying case to protect the enlargements, and a small grid for estimating acreage. Team supervisors will need the same equipment, plus green pencils for correcting questionnaires, and a 100 meter tape for establishing segment scale (if not done previously).

For the first survey in an area, we estimate a need for 20 questionnaires per segment, plus enough control sheets to handle 25 tracts in each segment. (In estimating quantities needed, we figure 10 questionnaires and control sheets for each person that participates in the enumerator training school. We add 15% for extras and keep the stencils, or masters, in case we need more.)

Segment control sheets are set up at the national and regional level to indicate team receiving the segments for enumeration, date received from the field, date returned for corrections, and date of final receipt. The regional control sheet shows date delivered to headquarters.

We arrange for vehicle fuel as needed. If coupons are used, we make sure they are accepted by service stations in all work areas. Also, provision is made for minor and major repairs, purchase of tires and batteries, oil changes, grease jobs and wash jobs.

A basic responsibility is to see that all staff receive salary and per diem as contracted. Prompt payment assures good team morale.

Conduct of Field Work

Team supervisors plan each day's work using the segment maps. They see that each enumerator is properly located and oriented in the appropriate segment. On the first visit to a segment, the supervisor may need to take measurements for establishing scale of the enlargement. Arrangements are confirmed for a time and place to meet the enumerators at the end of the day or when the segment is completed. The supervisor monitors progress in each segment, so that enumerators can be shifted to new segments to continue work.

The supervisor reviews the questionnaire for each completed interview —before leaving the tract— to insure that all questions are properly answered. As a segment is completed, he reviews all the questionnaires, making sure that all answers are reasonable and that all the tracts listed on the control sheet are shown on the enlargement; also verifying that each agricultural tract received a questionnaire. We have found it useful to put a check list for this purpose on the back of the control sheet.

This review of the completed segment is made while the supervisor and the enumerator are still at, or near, the segment. The supervisor looks at each field in all tracts, checking the enumerated size against the apparent size. Fields are summed to make sure that their area is equal to the tract area and tracts are summed to make sure that their area equals the segment total. The size of enumerated area of the segment is checked against the planimetered area of the segment (provided this planimetered area is based on the true scale of the enlargement as established by a field measurement). If the segment is relatively level, the size of enumerated and planimetered areas may vary plus or minus 10%. If the segment is mountainous with steep slopes, the size of enumerated area should not be less than the planimetered area, but could exceed it by 30%.

Each question should be reviewed for completeness and accuracy and all sums checked in *each* questionnaire. This review is particularly critical in countries where the roads and telephone service are bad, or non-existent. Only here can valid decisions be made, about data pertaining to the segment. All data filled in after leaving the segment (or corrections made later) will be averages, individual opinions or simply guesses. It is unlikely that time or money can be found to return to the segment, thus the supervisor is responsible for "getting it right the first time"

The regional supervisor keeps the regional work flowing, moving enumerator teams into appropriate areas. He spot-checks at least one completed segment for each team under his control as quickly as possible to detect any major problems. Spot-checks continue, as time permits, throughout the survey. Corrections by the team supervisor are pencilled in green; corrections by the regional supervisor are pencilled in blue.

Editing and Data Processing

Editing (critiquing) of questionnaires begins with the first completed questionnaires. We often find that some supervisors and enumerators have misunderstood instructions or otherwise not performed their assigned task. With early editing, corrections can be made for the balance of the survey. (Finished segments are systematically delivered to a central point or to the capital. In Guatemala, a team was sent to the field to perform a "pre-critica" that began as soon as some segments were completed.)

Editing is restricted to a minimum of changes and corrections and detailed editing instructions are provided for each question. Tolerance ranges are established for crop yields, pigs per litter, etc. Data reported outside these ranges are checked against enumerator observations; perhaps the crop was grown with a high level of technology, received irrigation, etc.

If a respondent refuses to provide information, the enumerator observes as much as possible, then notes the refusal and reports the "observed" data. The editor will complete the questionnaire, using the enumerator's observations and characteristics from adjoining farms.

Changes made during the final edit are pencilled in red —no data are erased from a questionnaire after it leaves the hands of the enumerator. Figures are corrected by marking through with a single line, then writing in the correction to either side or above the original. All original figures remain in legible condition.

The editing process requires checking of all items, including acreage of individual fields, tract totals and segment totals. One of the most frequent causes of disagreement between enumerated area and planimetered area is an underestimation of forest land, wasteland, or pasture land. If the cultivated land area is correct, review will usually indicate that other land was underestimated.

In the initial surveys, it is usually not practical to precode all items on the questionnaire. The necessary coding can be done along with the editing.

Our editing work flow is as follows: (1) completed segments are received and recorded as such on the master record, (2) clerks check all sums and indicate problems but make no changes, (3) one technician edits the segment and initials the segment envelope or editing record, (4) a second technician edits the same material and discusses any differences with the first technician; they agree on proper changes. The second technician also initials the segment, (5) the completed segment is delivered for machine tabulation.

Data processing should begin as soon as a number of segments have been edited, thus any problems within the programs are detected quickly. Also, the machine can accumulate "clean" data, so that the final results are obtained quickly when the final segments are delivered.

While computers are extremely efficient, even large surveys (6,000 questionnaires) can be processed by hand when computer time is not available or lack of programmers will cause extreme delays. Manual processing can provide an excellent learning experience.

THE ESTIMATE AND VARIANCE CALCULATIONS BY REPLICAS

Summarizing Replicated Samples

In summarizing a replicated sample, a first step is to obtain replica totals for each variable, by stratum. Critical review can begin with the examination of these replica totals. If any total is markedly different from the others, the segment data in that replica is carefully examined for possible errors. The next step is to review expanded data, by stratum, along with the coefficient of variation. Any stratum that appears to contribute inordinately to the overall variance is examined first for possible errors, with consideration given to increasing the sample to improve the next estimate.

The final regional and national expansions are checked against *all* available sources such as previous censuses, other surveys, export data, acreage and production registrations for some crops, and national balance sheets. Survey data often disagrees with many of these sources and comparisons help in review for errors.

Replicated sampling has a number of advantages:

(1) The results are easy to summarize. When we use the replica procedures, we begin to work with replica totals, instead of the values of individual sample units. Our total population becomes the number of replicas of a given size that can be selected from the population of individual sample units. Thus, standard statistical formulas are applicable, using replica totals instead of individual sample values.

(2) Replicas provide a useful tool for analysis. Each replica provides an estimate for the "universe" that usually consists of one stratum within a region. If the frame stratification, formation of count units, and formation and selection of segments has been done properly, we would expect the similar results for all replicas. If a replica should differ greatly from the others, it bears investigation for possible errors. In our example (Figure 11), replica 4 data should be reviewed, keeping in mind that there may be no errors and that the difference could be due to random variation.

(3) Sampling flexibility is increased. If we find after the first survey that a regional sample is too small, we do not have to reselect an entire larger sample; rather we maintain the replicas and add a replica or replicas until we have the number required. Conversely, we may find that the sample is larger than necessary in a region. In this case a replica can be dropped from the sample to save time and money. Budget considerations or other conditions may force us to reduce sample size, or perhaps we will find that the usual precision is not required for a certain estimate, hence we could use a reduced number of replicates.

This flexibility also makes it easier to rotate the sample. For example, we may find that continued interviewing over the years leads to an increased percentage of refusals, or some other problem that would make it desirable to shift the interviews to other farmers. Since a primary purpose of periodic surveys is to measure change over time, we would not want to change the entire sample, but could provide relief by allowing part of the sample to be replaced. The replicas can supply these replacements; we simply retire one replica or more depending on the percent rotation desired, and add new replicas. We identify and maintain all replicas retired from the regular survey system as they are valid samples, and may be used for special surveys as needed.

Some disadvantages of replicated sampling are:

(1) More office work is required. We select more than one sample, so a control system is required to identify and index all data by replicas.
(2) The variance estimate is unstable due to the reduced degrees of freedom used in replica calculations. In Fig. 11, the degrees of freedom are 4-1=3, whereas if the variance were calculated using individual segments, the degrees of freedom would be 20-1=19. Thus, more replicas provide better variance estimates. We calculate the number of replicas needed by first considering a rotation plan. For example, in the United States the rotation pattern requires changing 20% each year. In our example (Fig. 11), this would require five replicas. (Less than 4 replicas is hard to subdivide and more than 8 or 10 adds to the tabulation problems.)

SEGMENT NUMBER	REPLICA			
	1	2	3	4
1	21	17	73	38
2	39	15	19	47
3	11	86	37	23
4	57	31	16	14
5	32	28	25	17
Replica Totals	160	177	170	139

FIG. 11

(This example shows number of chickens per segment from survey)

An unbiased estimate of the replica mean is:

$$\frac{160 + 177 + 170 + 139}{4} = \frac{646}{4} = 161.5$$

An unbiased estimate of the segment mean is: $\frac{646}{20} = 32.3$

An unbiased estimate of the total number of chickens is provided by:

$N_r \bar{y}_r = 100 \times 161.5 = 16{,}150 = y'$

The variance is calculated in this manner:

$$\frac{S^2 = 160^2 + 177^2 + 170^2 + 139^2 - \frac{646^2}{4}}{4 - 1} = 273.67$$

Variance of the replica mean $= S^2_{\bar{y}} = \frac{S^2}{n} = \frac{273.67}{4} = 68.42$

Variance of the estimated total $= S^2_{y'} = N_r^2 S^2_{\bar{y}} = 100^2 \times 68.42 = 684{,}200$

Standard error of the estimated total $= S_{y'} = \sqrt{S^2_{y'}} = \sqrt{684{,}200} = 827$

Coefficient of variation of the estimated total $= \frac{S_{y'}}{y'}$

$$= \frac{827}{16{,}150} = .05 \text{ or } 5\%$$

Paper Strata

With replicated sampling, the use of paper strata is another tool available to the sampler to achieve a more efficient sample. In its simplest form, paper strata are created by the calculation of the sampling interval in systematic replicated sampling. If N=500 and we want replicas or independent samples of size 5, selected systematically, the sampling interval becomes 500 ÷ 5 = 100. In this case we have 5 paper strata, the first one includes segments number 1 to 100, the second paper stratum includes segments number 101 to 200, the third paper stratum includes segments number 201 to 300 and so on to form 5 paper strata.

By following the rules for numbering count units in a serpentine manner from the northeast corner toward the south within a region, a certain geographic ordering is established. The sampling interval will cut across this geographic ordering to form groups (paper strata) from north to south that could be a more efficient arrangement.

Paper strata can also be formed by grouping count units according to some known or detectable characteristic, such as production of a specific crop. For example, if it is possible to identify all or most of the count units that contain coffee, within a certain stratum, they could be placed in a specific "paper" stratum (or strata) that could be represented without regard to a geographic location. Such a grouping should improve the estimation for the classifying variable (coffee) and for other related variables, provided the initial determination for the production of the crop are valid.

The variance calculations using paper strata are as follows: (using data from Fig. 11)

	REPLICA 1	2	3	4	Stratum Segment Average	Est. birds per stratum ($N_h \bar{Y}_h$)	Stratum Variance (S_h^2)	Variance of estimated stratum mean ($S^2_{\bar{Y}_h} = \frac{S_h^2}{n}$)	Variance of estimated stratum total ($N_h^2 S^2_{\bar{Y}_h}$)
1	21	17	73	38	37.25	3,725	650.92	162.73	1,627,300
2	39	15	19	47	30.00	3,000	238.67	59.68	596,800
3	11	86	37	23	39.25	3925	1084.25	271.06	2,710,600
4	57	31	16	14	29.50	2950	393.67	98.42	983,200
5	32	28	25	17	25.50	2550	40.33	10.08	100,800
					161.50	16,150	2407.84		6,019,700

FIG. 12

In Fig. 12, the same data are being used but the calculations are made horizontally instead of vertically. Using paper stratum 1, the calculations are done as follows:

$$\text{Stratum Segment Average} = \bar{Y}_h = \frac{21 + 17 + 73 + 38}{4} = \frac{149}{4} = 37.25$$

$$\text{Estimated Birds per Stratum} = N_h \bar{Y}_h = 100 \times 37.25 = 3{,}725$$

The total number of segments in each paper stratum is $100 = N_h$

$$\text{Stratum Variance} = S_h^2 = \frac{21^2 + 17^2 + 73^2 + 38^2 - \frac{149^2}{4}}{4 - 1} = 650.92$$

$$\text{Variance of Estimated Stratum Mean} = S^2_{\bar{Y}_h} = \frac{S_h^2}{n} = \frac{650.92}{4} = 162.73$$

$$\text{Variance of Estimated Stratum Total} = S^2_{N_h \bar{Y}_h} = N_h^2 \, S^2_{\bar{Y}_h} =$$

$$100^2 \times 162.73 = 1{,}627{,}300$$

The estimated total number of birds and average number per segment remain the same as with the replica calculations, because they are from the same sample. Differences in the variance calculations indicate how the paper strata compare in efficiency with replicas.

We calculate the coefficient of variation of the estimated population mean by:

(1) summing the stratum variances and divide this sum by the number of replicas:

$$\frac{2407.84}{4} = 601.96$$

(2) extracting the square root = $\sqrt{601.96} = 24.53$

(3) calculating the C.V. = $\frac{24.53}{161.50}$ = .1519 or 15.2% is equal to the coefficient of variation of the estimated total.

The variance of the estimated population total, using the standard procedure for stratified sampling, is equal to the sum of the variances of the estimated stratum totals, or 6,019,700.

The coefficient of variation for the estimated population total $\bar{Y}$ is:

$$C.V._{\bar{Y}} = \frac{\sqrt{6{,}019{,}700}}{16150} = \frac{2453}{16150} = .1519 \text{ or } 15.2\%$$

In this case, the paper strata are considerably less efficient than are the replicas; the paper strata CV of 15% compares with a CV of 5% calculated by means of replicas.

OTHER TECHNICAL ASPECTS AND SPECIAL PROBLEMS

Large-scale Producers

If a few producers control a large portion of the production of any item, a random sample usually will not provide the proper representation of these large operations. For example, if a producer with 5,000 head of cattle were to fall in a sample, this number multiplied by the usual expansion factor of about 100 would completely distort the national estimate —both the estimated total and its variance. Livestock and poultry estimates are usually most affected by such sampling; however some crops such as rice can cause similar distortion, if the production area is represented by a few producers in a small section of the region.

To minimize the impact of large producers, we prepare a list of operators that will be completely enumerated or sampled to provide the "large-scale producer" portion of the regional and national estimates. Data collected on these large-scale producers in the process of the area frame survey must be eliminated from the area frame expansion. Thus, we must obtain a great deal of information about each of these producers. Where is the production located? Who are the administrators and farm or ranch managers? Where and from whom can you obtain information? What are the names and/or brands of the operations involved?

All questionnaires from the area sample must be checked against the large-scale producer list upon receipt in the office. If data are found relating to the large-scale producer or a portion of his operation, they are eliminated from the area expansion. The results of the estimation of the list is then added to the area sample expansions to provide regional and national estimates.

We often find it difficult to define the scope of the large-scale producer list. We usually start by looking at classifications by size of operation from the most recent census to determine the number of the largest categories for each variable we intend to estimate. If the sum of these categories appears to be a list of manageable size, we use it, if more are needed, we add the next largest category. As we make the list larger, we allow for a more efficient estimate of the total.

The list must be 100% complete and updated before the survey. As a very rough guideline, we might begin with the 500 largest operations. Experience with the survey will indicate whether or not we need to increase the size of the list.

While building the list, we obtain data such as size of herd, acreage of crop, and total land for each producer. The census list should be broken down by regions and circulated to extension agents, banks, and producer's associations for review. (They make additions, note changes in size of operations for producers on the original list, and eliminate names of producers no longer in operation. The list should be circulated in this manner once a year—or before each survey, if surveys are not done each year.)

A special questionnaire is made to enumerate the large-scale producers. This questionnaire is arranged to collect data for the total operation, but with data divided by regions; many of the largest operations will have production units in more than one region. Thus, the data can be correlated with the regional estimates from the area frame.

We also use a list of producers in conjunction with the area frame when we need to estimate rare crops or highly localized crops; for example: vegetables, cut flowers, specific breeds of livestock or poultry, etc. This list is created as described above and is used in the same way.

Problem Segments

Despite all of our precautions, we usually encounter problems with segment boundaries. For example, a fence or a line of trees that served as a boundary will be removed. However, we can often locate exactly where the boundary was. If this is the case, we use the original location as the boundary and enumerate the segment on that basis. The enumerator should indicate the closest existing boundary on the enlargement. If this new boundary does not increase the size of the segment beyond acceptable size, we can use it for future surveys. If the size limit is exceeded, we rework the count unit, forming segments of acceptable size.

If the enumerator is unable to locate a boundary or part of a boundary, he indicates the nearest real boundary and completes his enumeration on that basis, later informing the supervisor of this procedure. If the resulting segment exceeds appropriate size limits, the data must be reduced to a one segment equivalent before expansion.

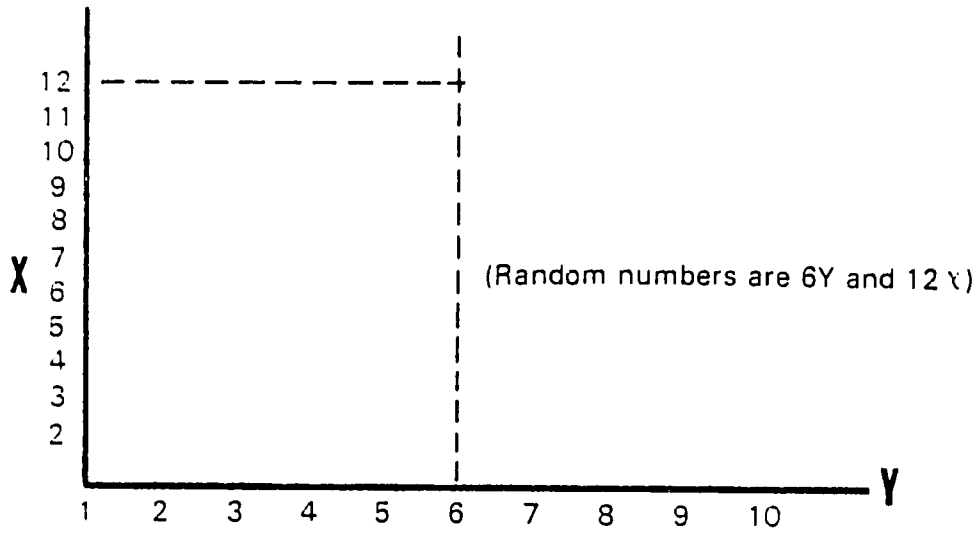
Segments with too many tracts also can cause problems. Sometimes these can be detected by inspecting the photographic enlargement—if so, the segment is subdivided prior to the survey. As a rule of thumb, a segment with more than 20 tracts is subdivided, with the objective of having 10-15 tracts in each division. Visual inspection of the enlargement allows us to subdivide the segment into approximately equal parts, depending on the number of tracts it contains. These subdivisions can form 2, 3, 4 or even more parts. For example, if the segment is divided into 3 parts, each part should contain approximately 1/3 of the total cultivated land, 1/3 of the dwellings, etc. One of these subdivisions is selected at random and becomes the sample segment and this selected segment is shown on the mosaic and on the maps. The other parts of the segment are removed from the enlargement, leaving only the selected portion for the use by the enumerator.

Each subdivided segment is given an adjustment factor for use in summarization. If the segment is divided into 2 parts, the enumeration result is multiplied by 2 before expansion; if divided into 3 parts, the enumeration result is multiplied by 3. In other words, the portion must be made the equivalent of a full-sized segment before entering into the normal expansion procedure. This adjustment factor must be recorded and used during the time the subdivided segment is used.

Point Sampling

We have found it difficult to construct count units and select segments in some regions when visible boundaries simply do not exist on the photographs. We have used a procedure known as point sampling in these situations, but only when we could find no other solution.

In point sampling, the stratification is done as usual, but no count units are formed. Instead, the stratum blocks as drawn on the map are measured and used as count units and these units are assigned the appropriate number of segments. The first step is selecting a count unit or, in this case, what really is a stratum block. Instead of attempting to divide the selected area, a grid with an X and Y axis is placed upon the map. A random number is selected along each axis and the intersection of perpendicular lines indicate a random point (as shown below).



This random point is then transferred to the photographs. The next objective is to form a segment around this point, with the point as near the center of the segment as possible. Using this system, a high percentage of the segments are formed in the field. Often boundaries become imaginary lines between two identifiable points. A segment may be formed that is 5 times the acceptable size. Thus we have the problem of subdivision, in reverse. The oversized segment is enumerated and the result divided by 5 to reduce it to segment size appropriate for that stratum.

Some of the selected points will be almost totally inaccessible mountain and/or forest regions. In such cases, we must determine if there is any agriculture or livestock within a radius of the point that would provide the segment area. Usually, these inaccessible areas have no agricultural production, which can be verified by inspection from a small airplane or helicopter. If there is production, we plan a way to enter the area for enumeration. If there is no evidence of production, these are considered zero segments for the purpose of the survey.

Preparation of Frames for Special Areas

After the national area frame is constructed, we sometimes find it necessary to survey in areas smaller than a region, province or department. After such an area is identified on 1:50,000 scale maps, we can quickly form a special area frame for the zone by simply identifying the existing count units that cover the area in question. The data for these count units is listed by stratum on another set of listing sheets, thus forming the sampling frame for the special area. Sample selection and preparation are as described previously.

Additional photography is required for the count units that are selected to contain a sample segment from the special frame. (The original mosaics are used only for the regional and national frame and sample.) The special frame count units are subdivided on the new photography, and all special frame materials are filed separately from the basic area frame.

A special frame need not be in one continuous area. It can be formed from count units or groups of count units scattered throughout the country; we need only identify the count units.

The existing count unit boundaries do not fall exactly on the boundaries of the special area; however, special areas usually are not precisely defined, so some imprecision in area is not a problem.

The use of an area frame becomes increasingly impractical as the size of the special area decreases. For national surveys that provide a data series it has been considered impractical to attempt to provide estimates for areas smaller than a region, because the sample size becomes so large that the cost of annual surveys is prohibitive. However, this limitation would not apply for special area surveys. These are usually "one-time" surveys in high-priority areas.

Size-of-Farm Classification

Sample design for regional estimates usually do not provide reliable classifications of data by size of farm. In some cases, it has been possible to make as many as three such sub-classifications with acceptable reliability. The sample is sufficiently large to make an acceptable regional estimate, but when sub-classified by farm size, the sample in each category is too small for reliability.

If size of farm data are required for as many as six categories, a rough guess is that our original sample would have to be at least doubled. However, the calculation of variance and computation of desired sample size would provide the only exact estimate.

Objective Procedures for Measuring Crop Yield

Many of the small farmers in developing countries cannot accurately estimate their crop yields or their total production. Or perhaps they can only estimate in a variety of units of measurement that must be converted to some standard unit. Thus, for example, the farmers often provide unsatisfactory estimates of yields of shelled corn per hectare in answer to our questions.

For more precise estimates, we can select a sample of fields in which small plots will be harvested by technicians to determine yield and total production. Our area sample provides a very efficient frame for the selection of sample fields.

Ideally, an area frame survey is conducted at the beginning of the main growing season, after crops are planted. Thus, for example, the enumeration process would provide the exact location of every field of corn in each segment in the sample. Thus, after the survey has been conducted we have available the elements for selecting a self-weighting sample of fields. Using corn as an example, we would first list total area planted to corn, by segment — along with the expansion factor of each segment. We multiply the corn area by the expansion factor and add another column to accumulate the expanded areas as shown in Figure 13.

Segment	Corn (ha.)	Expansion Factor	Expanded Area	Accumulated Area
1	21	61	1281	1281
2	17	61	1037	2318
3	19	61	1159	3477
4	32	78	2496	5973
5	26	78	2028	8001
6	9	109	981	8982
7	14	109	1526	10,508
8	18	109	1962	12,470
9	43	68	2924	15,394
10	27	68	1836	17,230

FIG. 13

To select a self-weighting systematic sample of 5 units, we first establish a sampling interval: 17230 ÷ 5 = 3446. Then we select a random number between 1 and 3,446, for example, 1799, indicating that the first sample is in segment number 2. We add the sampling interval (1799 + 3446 = 5245) and we see that the second unit is in segment number 4. (The third unit is in segment number 6, the fourth unit in segment 8, and the fifth unit in segment number 9.)

To continue this selection down to the individual field (using segment 9 as an example) we would make a listing of fields within the segment as shown below:

Field No.	Area in Ha.	Accum. Area
A-3	17	17
A-7	8	25
B-1	11	36
C-3	7	43

We simply select a random number between 1 and 43. If the random number happens to be 15, we have selected field number A 3, which can be precisely located on the segment enlargement. We then locate the sample plots in the field and harvest the corn when it is mature. A simple average of the yields provides the regional and national yield.

We can also use these sample fields for forecasting yield by correlating measurements of plant characteristics with final yield. (See suggested reference number 5 for detail.)

Useful References for Area Frames and Sampling

1. *"El Muestreo por Areas en la Agricultura"*
   Por Earl E. Houseman - 1975
   U.S. Department of Agriculture SRS No. 20-S

2. *"Cursillo de Capacitación sobre Conceptos de Muestreo para Encuestas Agropecuarias"*
   Por Harold Huddleston - 1977
   U.S. Department of Agriculture - SRS No. 21-S

3. *"Sample Survey Methods and Theory - Vol. I Methods and Applications"*
   Por Hansen, Hurwits & Madow - 1953
   John Wiley & Sons, New York

4. *"Sampling Techniques"*
   Por William G. Cochran - 1963
   John Wiley & Sons, New York

5. *"Técnicas de Muestreo para la Medición y Pronosticación de los Rendimientos de los Cultivos"*
   Por Harold F. Huddleston - 1978
   U.S. Department of Agriculture ESCS No. 9-S

6. *"El Valor Esperado de un Estimador Muestral"*
   Por Earl E. Houseman - 1974
   U.S. Department of Agriculture SFS No. 19-S

7. *"Principios de Fotogrametría"*
   Por Jaime Ignacio Roa M.
   SIREN
   Apartado Aéreo 53498
   Bogotá 2, Colombia