

Application of Multiple Frame Designs in Agricultural Survey Programmes

Mirane M. Carrilho

Instituto Brasileiro de Geografia e Estatística (IBGE)

Rio de Janeiro, Brazil

mirane@ibge.gov.br

1. Introduction

This paper refers to the area sample design used for the *Special Agricultural Survey on the use of Pesticides and Environmental Control*. The survey design utilizes the area sample used for the Brazilian Current Agricultural Survey Programme, which consists of annual agricultural surveys based on multiple frame sampling methods conducted, since 1987, by the Brazilian Institute of Geography and Statistics (IBGE). For this reason, this paper provides also an outline of the Brazilian Current Agricultural Survey Programme. This large-scale, multiple-purpose, periodic survey programme allows the IBGE to obtain baseline statistics necessary for the evaluation and planning of the agricultural sector, for the States of Parana, Santa Catarina, Sao Paulo and the Distrito Federal, an important agricultural area of around 500,000 km² and 800,000 farms.

2. The Current Agricultural Survey Programme based on multiple frame survey methods

2.1 General characteristics of the Current Agricultural Survey Programme

- The area sample component of the survey design involves, in each State, a one-stage, systematic and self-weighted area sample of segments, stratified by land-use.
- The area sampling frame considers the territory divided into a number of strata defined by proportion of cultivated land, the predominance of certain crops or other land use characteristics. The strata and administrative divisions of each state were completely subdivided into geographically ordered areas called Primary Sampling Units (PSUs). The strata, PSUs and sample segments have identifiable physical boundaries that can be located both in the field and on the cartographic materials used for their identification (satellite images from the TM sensor aboard the Landsat-V, mosaics of aerial photographs and topographic maps).
- The survey design involved the construction and annual updating of complementary list frames of *special farms* for each of the main crops and livestock herds surveyed. Each list is formed by farms that correspond to a large percentage of the total area and production of the crop or a large percentage of the total livestock herd.
- The segments and the *special farms* are used annually to collect ground data. During the field work, a questionnaire is applied for each farm partially or totally included in the sampled segments and applied in each of the special farms. Also, all fields and the farms are located, delineated and measured on aerial photographs covering the sampled segments.
- The survey provides annual estimates and forecasts of planted and harvested crop areas, potential and actual yields and production for the major summer crops, planting and harvesting dates, quantity of seeds used, livestock estimates, fertilizers and pesticide use, as well as characteristics of the farms and farming systems.

- Multiple frame estimators, which combine for each survey variable an area sample estimate with farm sample estimates, are used in order to obtain the estimate for each State.

2.2 Area sample parameters of the Current Agricultural Survey Programme

The total area sample consist of 1,691 *segments*, distributed as follows: State of Parana, 525 segments (0.5% of the total area of the State); State of Santa Catarina, 430 segments (1.2% of the total area of the State); State of São Paulo, 546 segments (0.4% of the total area of the State); and Distrito Federal, 190 segments (4.7% of the total area of the State). Table 1 shows, for each State, the total area, the stratified area, the number of strata and PSUs, the number of segments in the universe and in the sample, and the inverse of the sampling fraction.

Table 1

STATE	Area (km ²)		Number of				1/f
	Total	Frame	Strata	PSUs	Segments		
					Universe	Sample	
Parana	199 323	194 797	7	6 926	99 540	525	189,6
Santa Catarina	95 318	94 016	9	3 797	36 531	430	84,9
Sao Paulo	248 255	237 8887	15	11 639	127 170	546	193,2
Federal District	5 794	5 794	10	652	3 163	190	16,4

2.3 Area sample estimators

The type of estimator used for each variable depends on the corresponding reporting unit (tracts or farms). Three types of estimators are used to associate the reporting units with the sampling units, namely, the closed, weighted and open estimators. The formulae for the direct expansion estimators for totals are based on substrata, considering the method of successive differences for the variance (Kish, 1965, 4.1.2, p.119). The estimator for the total of variable x is:

$$\hat{X} = \sum_{h \in L} \sum_{i \in L_h} \sum_{j \in n_{hi}} e_{hij} \sum_{k \in m_{hij}} t_{hijk} \quad , \quad \text{where}$$

\hat{X} : estimator for the total of variable x .

L : number of land use strata.

L_h : number of substrata of stratum h .

n_{hi} : number of sample segments in substratum i , in stratum h .

m_{hij} : number of farms included in segment j , substratum i , stratum h .

e_{hij} : expansion factor for farm included in segment j , substratum i , stratum h .

t_{hik} : value of the variable in farm k , segment j , substratum i , stratum h .

$t_{hijk} = x_{hijk}$, for closed segment estimators,

$t_{hijk} = x_{hijk} \cdot a_{hijk}$, for open segment estimators.

$t_{hijk} = x_{hijk} \cdot w_{hijk}$, for weighted estimators.

The farm headquarters is defined as the farmers' residence, and should be inside the segment in a rural area, or if 50% or more of the total area of the farm is inside the segment and the farmer residence is in a urban area.

The weighting factor w_{hijk} , is used for farm k , segment j , substratum i , stratum h and, $a_{hijk}=1$, if the headquarters of farm k belongs to segment j , and $a_{hijk}=0$ otherwise.

The variance formulae based on substrata and using the method of successive differences are analogous for the three types of estimators:

$$V(X) = \sum_{h \in L} \sum_{i \in L_h} (N_{hi})^2 \cdot [(1 - f_{hi}) / (2n_{hi}(n_{hi} - 1))] [\sum_{j \in n_{hi} - 1} (t_{hij} - t_{hi(j+1)})]^2, \text{ where}$$

N_{hi} : total number of segments in substratum i , stratum h .

n_{hi} : number of sample segments in substratum i , stratum h .

$f_{hi} = n_{hi} / N_{hi}$, is the sampling fraction for substratum i , stratum h .

t_{hijk} : as defined above.

2.4 Multiple frame estimators

The multiple frame estimator has the following form:

$$\hat{X} = X_j + X_a \quad \text{where:}$$

X_a , indicates the area sample estimator, and X_j , indicates the estimate for the special farms.

The calculation for the area sample estimator do not include the estimates obtained from the special farms. In the cases in which the list of special farms is large, a simple, stratified random sample is used to estimate X_j .

In some cases, significant improvements in the precision of the estimates were obtained by using multiple frame estimators instead of area frame estimators. For example, in the State of Parana, the CV for the 1995 estimate for area of cotton using a multiple frame estimator (13.8%) was much lower than the CV for the 2000 estimate based only on an closed segment area sample estimator (18.0%). The list of special cotton farms included 150 farms with an area of cotton ≥ 100 ha, corresponding to around 29,750 ha. A stratified simple random sample of 32 farms was used to obtain the list sample estimate. For the 2000 Survey only the area sample estimator could be used since it was not possible to update the list of special cotton farms.

3. The Special Agricultural Survey on the use of Pesticides and Environmental Control

The *Special Survey on the use of Pesticides and Environmental Control* was conducted in 1999 in the State of Parana (with an area of around 200,000 km²). The agricultural commodities studied were: cassava, cotton, maize and soybeans. The questionnaire were developed to obtain a set of questions to elicit information as to whether fertilizer had been applied and, if so, what kind and in what quantity. The survey included specific variables on the use of chemical products used, and variables on the characteristics of the farms using pesticides.

A weighted segment estimator (the weight being the ratio between the area of the farm included in the segment and the total area of the farm) was used to obtain the estimates for the specific variables on the use of chemical products. Table 2 shows some of those estimates for soybeans.

Table 2

Type of Product	Área (1000 ha) (1)				Quantity (t) (2)				Average Number of	
	Total	CV (%)	With Application	CV (%)	Active Ingrid.	CV (%)	Elabor. product	CV (%)	Applic.	CV (%)
Herbicide	3 183	5,2	3 054	5,3	4 368	7,3	10 330	7,3	2,88	2,6
Inseticide	2 614	5,8	2 538	5,7	1 143	8,3	2 675	8,8	2,15	3,2
Fungicide	786	13,4	719	13,0	217	17,0	452	15,6	1,31	4,9

(1) Areas with application of more than one product; (2) Kg ou liter.

Table 3 shows the results for pesticide frequency of use and the main destination of containers. The open segment estimator and a ratio estimator (defined as the ratio between the number of farms with information on the characteristic and the total number of farms) were used to obtain the estimates of the characteristics of the farms.

Variable	% of Farms	CV%
Frequency of Use		
Always	45,70	5,7
Sometimes	13,07	8,7
Never	41,23	6,3
Used in 1999	84,44	2,4
Destination of containers		
Stores	61,09	5,1
Burns	30,99	9,6
Triple washing	23,10	13,2

The results obtained from the Special Survey on the use of Pesticides and Environmental Control not only complements other studies conducted by IBGE but showed the advantage of using area sampling methods for environmental studies.

REFERENCES

González-Villalobos A., Moreira, M., Silva, G. y Biffi, A. J. (1988) *Desenvolvimento do Sistema de Informações Agropecuárias*. Annals of the Brazilian Symposium on Remote Sensing-National Space Research Institute, Publ.4940: Natal, Brazil.

Kish, L. (1965). *Survey sampling*. John Wiley and Sons.

FAO (1988). *Multiple Frame Agricultural Surveys. Volume II*. Statistical Development Series N° 10, Rome.

IBGE (1995, 2001). *Pesquisa de Previsão de Safras - Resultados 1995 e 2000*. Instituto Brasileiro de Geografia e Estatística.

IBGE (2000). *Uso de Agrotóxicos no Estado do Paraná 1999 - Resultados*. Instituto Brasileiro de Geografia e Estatística.

RÉSUMÉ

Cet article décrit l'échantillonnage de l'Enquête agricole sur l'usage des pesticides et sur le contrôle de l'environnement. L'enquête utilise la base d'échantillonnage aréolaire du Programme Annuelle d'Enquêtes Agricoles du Brésil, enquêtes agricoles sur base d'échantillonnage multiple, réalisés depuis 1987 par l'Institut Brésilien de Géographie et Statistique (IBGE). Pour cette raison, on décrit sommairement le Programme d'Enquêtes Agricoles du Brésil. Ce programme d'enquêtes annuelles, à grande échelle et à objectifs multiples, permet le IBGE d'obtenir les statistiques de base pour l'évaluation et planification du secteur agricole des états de Parana, Santa Catarina, Sao Paulo et le Distrito Federal, une region agricole importante d'environ 500,000 km² et 800,000 exploitations agricoles.