SAMPLE SURVEY STATISTICS TEACHING: AN ALMOST WORLDWIDE PROBLEM ON TEACHING AGRICULTURAL SURVEY METHODS

Álvaro González Villalobos
National Academy of Sciences of Buenos Aires, Argentina
alvarun@fibertel.com.ar

Most national or large-scale multiple-purpose Agricultural Surveys, because of their complex nature, should be based on multiple frame survey designs that use samples selected from an area frame, list frames of holdings and housing units, and point sampling methods. The problem is that, although such Agricultural Surveys imply important human and economic resources as well as time for their implementation and maintenance, the area and multiple frame survey methods already mentioned are simply not taught at most universities and statistical teaching institutions. In other words, it has been almost a worldwide tradition to teach in an inappropriate way agricultural survey methods. However those methods are widely used and involve specialized statistical knowledge for their implementation and analysis. In this paper we shall refer to the reasons of such lack of appropriate teaching as well as possible solutions.

INTRODUCTION

This paper refers to some of my observations and ideas on the teaching of Survey Sampling and particularly Agricultural Survey Sampling and its applications to large areas, regions and countries. It deals specially with the statistical survey methods needed for the national survey designs to be used in most countries in the world for their main national statistics (national censuses and probability sample surveys of agriculture, population and housing, fisheries, industry, health and environment), since to design and conduct those baseline national surveys and censuses require a large effort for the countries: a large number of persons, a long time, and important financial resources. Since baseline agricultural statistics are one of the important applications of survey sampling, it is needed to undertake decisive improvements in the teaching of survey sampling.

TEACHING OF GENERAL SURVEY SAMPLING

Professor Leslie Kish (who made during his life enormous contributions for the improvement of teaching of Survey Sampling) coined a title for one of his papers: “The Hundred Years’ Wars of Survey Sampling.” This seems to be a very appropriate phrase to start the considerations of this paper. In fact, the war continues since the introduction of large-scale sampling methods still has not been done in most countries of the world for many important statistics.

This is somehow a paradox in the era of informatics and communications (the era of John von Neumann) because nowadays there are more facilities than ever before for decisions to be taken based on accurate, reliable and timely statistics. All survey statisticians know that timely and reliable national statistics of a country can only be provided by the establishment of adequate national surveys based on probability sampling methods.

This lack of interest in most countries to promote teaching and training on survey methods is due in part to the wrong believe that the worldwide availability of computers, computerized instruments and statistical software (which has, of course, enormous advantages) can substitute knowledge of survey sampling. But nothing can substitute the knowledge of survey sampling or of sound statistical designs, and therefore for the purpose of designing and conducting probability surveys the new digital instruments and software by themselves are of no use.

Another reason of such lack of interest in promoting survey sampling is shown by the disappearance of important international statistical teaching centers that contributed for many decades for the formation of applied statisticians and in particular survey statisticians. This is the case of the disappearance of the CIE, the Inter-American Center for Statistical Training of the OAS (Organization of American States), who trained thousands of persons during decades from most countries of the American continent. In fact, even worst, statistics have disappeared from the OAS organization chart. Other example is the international support that used to provide in Paris

the INSEE (the French National Statistical Office) to all French speaking African countries. In this case, the training center was transferred to Abidjan, Ivory Coast, where the African students lost both the experience of the life in Paris and the knowledge of famous statisticians. Besides those examples, among others, it is worthwhile to recall that the financial support given to foreign students in the United States, the European Union and in international organizations have decreased considerably. There are still, of course, some excellent courses provided in national agencies and universities mainly in the United States.

TEACHING OF AGRICULTURAL SURVEY SAMPLING METHODS

The Use of Probability Sampling Methods

As already mentioned in general for most baseline national statistics, also for agricultural statistics the introduction of large-scale sampling methods has not been implemented in most countries of the world.

In general, in most developed countries great efforts have been done to introduce probability sampling methods. In the US (for example), a pioneer country in this area, already in the 1960's probability sampling techniques were massively introduced (substituting non-probability samples) for baseline national surveys in order to supply more reliable and timely information in its commodity estimating program (cf. e.g., USDA, 1983; Sivers, 2002).

It should be mentioned also, specially since this Conference is held in Salvador, that from all countries of Central and South America, Brazil is by far the country that has more qualified agricultural statisticians and personnel of other fields needed to implement agricultural surveys. In fact, Brazil, in this respect, is superior to most countries, even developed countries around the world, including countries in Europe, Asia and Africa. It is for this reason that in our view it will be highly beneficial to Brazil to expand and implement agricultural surveys based on multiple frame survey methods.

Most Important Agricultural Survey Methods Not Included in Classical Textbooks

Most important agricultural survey sampling methods are not included in the classic and most used books on Survey Sampling namely: Kish (1965), Cochran (1977), Hansen, Hurwitz, and Madow (1953), Deming (1960)**,** Murthy (1967), Raj (1972), Desabie (1966), Sukhatme and Sukhatme (1970) and Yates (1981). Even well known books on Agricultural Surveys did not include some of the most important survey methods already implemented in the US, for example the books of Zarkovich (1965) and Kish (1989). The same also applies for more recent books still not so widely used in most countries and teaching centers as the classical ones mentioned.

This is due to a number of reasons. A very important one, in my view, is an historical reason: the fact that in the United States, agricultural statistics were obtained by a different agency than the other baseline national statistics. And also the style of the statisticians involved some of them less linked to the universities and other teaching centers. As a result, the revolutionary research and applications undertaken at the USDA (the US Department of Agriculture) was not coordinated with the research and applications undertaken at the Bureau of the Census (BUCEN) and other national and private survey centers and universities. The statisticians in the later institutions were the ones that wrote the books from which most students learn survey sampling, and so they do not include fundamental agricultural sampling methods used worldwide.

The Bureau of the Census (BUCEN), for instance, designed and conducted all baseline Population and Housing Statistics, directed by the eminent statistician Morris Hansen until his resignation from the Bureau, and his foundation of Westat with other famous colleagues from the Bureau. Morris Hansen and his strong and large group of colleagues (like the great statistician Joseph Waksberg) continued working in the improvement of list frame sampling methods in a great variety of fields, and changed the way survey sampling is done today.

The USDA, on the other hand, designed and conducted all agricultural surveys except the Agricultural Census (conducted by the Bureau of the Census until recent years, when the responsibility of the Census was transferred to USDA). At the USDA also famous survey statisticians (from USDA or hired by other institutions) were in charge of the design of the

agricultural surveys, and they were the ones that changed the way of designing and conducting agricultural surveys, as it will be explained in the next point.

A Brief History of Agricultural Survey Methods Designed and at USDA, in Particular the Introduction of Area and Multiple Frame Surveys

- *Multiple list sampling methods:*

Among the survey statisticians that contributed significantly for the improvement of list frame designs, Fellegi and Sunter (1969) should be mentioned. Important improvements in multiple list frames were also accomplished by statisticians working at universities and national agencies for general surveys, not only those conducted by USDA (cf. e.g., Gurney and González, 1972).

- *Area sampling methods:*

Probably the main new, completely different survey methods introduced in the design of the Current National Agricultural Survey was *area sampling, not included in the above-mentioned classical textbooks,* although the research started in the 1940s.

These new methods were invented, and their application was initiated by a number of statisticians, for example: Houseman, Jessen, King, and Trelogan. It should be noticed that parallel to the work being developed at the US, Mahalanobis introduced the idea of area sampling in India.

- *Multiple frame sampling methods:*

Then another key contribution was made by Mr. H. O. Hartley 1960s, who introduced *multiple frame sampling*, combining an area frame with list frames.

A great development then took place at USDA and a multiple frame survey was implemented in the whole country. This development was due to a number of other distinguished statisticians who published many papers and training courses for that purpose: Caudill, Kibler, and Moore, in the 1960's, Vogel, Bosecker, Pratt, Rockwell, Huddleston, and Hill in the 1970s, Vogel, Nealon, Geuder, Fecso, Tortora, Ford, Cotter, and Kott in the 1980s, Kott and Vogel in the 1990s.

- *Multiple frame sampling methods combined with the use of satellite imagery:*

In the 1970s, another very important contribution to agricultural statistics was the effective use of satellite imagery to construct the area sampling frames for the design of the National Agricultural Survey and also for other purposes like estimating crop yields and special areas. This new contribution was done mainly by USDA statisticians that learned remote sensing techniques since NASA could not accomplished the proper methods to use for multiple-purpose agricultural surveys. The main statisticians that introduced remote sensing techniques were Wigton, Bormann, Hanuschak, Morrisey, Dodson, Allen, Clark, Geuder, Craig, Cotter, and Tomczak among others.

- *Multiple frame sampling methods combined with the use of GIS: CASS*

The last key development for the National Agricultural Survey in the US was the gradual introduction of *CASS* (Computer-aided stratification and sampling), an ad-hoc GIS developed for the construction of the area sampling frames and the selection of the samples. This implied a reduction of personnel, costs and time required in comparison with the previous methods. All sampling frames and the selection of the area samples is done using the *CASS*.

Area and Multiple Frame Agricultural Surveys in Other Countries and Regions

Over the years, starting in the 1970s, area or multiple frame sampling methods were applied in dozens of countries (or part of countries) and regions, adapting to the local conditions the methods applied in the United States. Moreover, during the last years multiple frame sampling methods were implemented using standard GIS software. A description of the survey design used in many of those countries up to 1998 can be found in González Villalobos and Wallace (1998).

Some of the survey statisticians that contributed in the design and implementation of area and multiple frame surveys in countries of Africa, America, Asia and Oceania, are the following: Bouzaffour, Carfagna, Chinnappa, Davidson, Dunkerley, El Sheikh Elbashir, Gálvez Calix, González-Villalobos, Hale, Huddleston, Julien, Khan, Kovar, Maranda, Martins Carrilho, Moore, Nazif, Otañez, Silva, Steiner, Stepancich, Théberge, Trépanier, Vogel, Wallace and Wigton.

In Europe, "square segments" were introduced in substitution to segments with identifiable physical boundaries, except in the case of Italy. Some of the survey statisticians responsible for those designs were: Ambrosio Flores, Carfagna, Delincé, Fournier, Fuentes, Gallego, and Hanuschak.

FINAL CONSIDERATIONS AND POSSIBLE SOLUTIONS

As a consequence of the strong arguments I have given so far, and taking into account the views of many of the colleagues I have mentioned, it seems appropriate that the lack of training on agricultural survey methods could be decisively improved by including in the "curricula" of survey sampling at least three methods that are of common use for the design of agricultural statistics: area sampling, multiple frame sampling and point sampling.

Concerning the arguments I have presented on the general lack of training on agricultural survey sampling methods, let me say that they are based on my work over the years in more than fifty countries, were I had the opportunity to discuss such a problem with almost of the well-known survey statisticians that I have mentioned. Let me just recall the opinion on this matter of two of the late great survey statisticians of the last century: I had the honour and pleasure to discuss many hours survey sampling problems with Morris Hansen. In particular, Hansen agreed with the existence of the problem. With Leslie Kish I had a strong relationship since 1975, and became also a close friend. He followed and inspired my work during his interesting and productive live. Shortly after I published, in collaboration with Wallace (and the contribution of many other colleagues) a book on multiple frame sampling (cf. González Villalobos and Wallace, 1996). Kish told me that he had not realized, until reading the book, the importance of multiple frame methods for agricultural surveys, and that was why he did not include those methods even in his book on "Sampling methods for agricultural surveys," already mentioned.

On the other hand, the teaching of agricultural survey sampling methods for large-scale surveys should now also take into account the teaching of informatics techniques, like the use of GIS, remote sensing for the analysis of satellite and aerial images and photos, and the use of instruments like GPSs, computerized data entry devices, etc. And not only because these techniques, software and computer instruments are useful in combination with the survey methods, but because they are *available* in most countries, even in very poor countries. But of course, for the purpose of having good statistics all those modern computers and instruments and their software and techniques are of no use if there is a lack of knowledge of survey sampling. And this is precisely the situation in most countries of the world. It should be noted that the improvement of the teaching of agricultural survey methods (among other important statistical methods) would not necessarily imply in a significant cost considering, for example, the resources involved in their application and the resulting economic advantages of having reliable and timely baseline data for a given country.

Another aspect of the application of agricultural survey methods is the possibility of having designs that combined the described multiple frame samples with list samples of housing units or households in the agricultural areas (e.g., Ernst, 1995, and Novaro, 2002). These methods could be very useful in many developing countries were there are scarce resources to conduct surveys.

It can be considered that this paper refers to a special aspect for improving the knowledge of the agricultural sector in many countries of the world where millions of people live in extreme poverty. It is necessary to know in order to improve.

In other words, it is necessary, in most countries, to improve agricultural, livestock and fishery statistics and the knowledge of the agricultural sector through the application of new methods and techniques on sampling, data collection, data analysis and dissemination. It is necessary through better statistics to improve the knowledge on the conditions of living of millions of persons that live in great misery.

Baseline agricultural statistics are specially important in order to be able to "foresee to provide" as said by Teixeira de Freitas, allowing to anticipate political changes in the benefit of humanity.

REFERENCES

Cochran, W. G. (1977). *Sampling Techniques* (3rd edition). New York: John Wiley and Sons.

Deming, W. E. (1960). *Sample Design in Business Research*. New York: John Wiley and Sons.

Desabie, J. (1966). *Théorie et Practique des Sondages*. Paris: Dunod.

Ernst, L. R. (1995). *Maximizing and Minimizing Overlap of Ultimate Sampling Units*. Washington: Bureau of Labor Statistics.

Fellegi, I. P. and Sunter, A. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64(238), 1183-1210.

González Villalobos, A. and Wallace, M. A. (1996). Multiple frame agricultural surveys - Current surveys based on area and list sampling methods, Vol. I; *FAO, Statistical Development Series 7*. Rome: FAO. Published in English, Arabic, Chinese, Farsi, French and Spanish.

González Villalobos, A. and Wallace, M. A. (1998). Multiple frame agricultural surveys - agricultural survey programmes based on area frame or dual frame (area and list) sample designs, Vol. II; *FAO, Statistical Development Series 10*. Rome: FAO. Published in English, Arabic, Chinese, French and Spanish.

Gurney, M. and González, M .E. (1972). Estimates from samples from frames where some units have multiple listings. *1972 Proceedings of the Joint Statistical Meetings*, Montreal. Alexandria, VA: American Statistical Association.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory. Vols. I-II*, New York: John Wiley and Sons.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. (1989). Sampling methods for agricultural surveys. *FAO Statistical Development Series No. 3*, Rome: FAO.

Kish, L. (1995). The hundred years' wars of survey sampling. *Statistics in Transition*, 2(5), 813-830.

Murthy, M. N. (1967). Sampling theory and methods. *Statistical Publishing Society*, Calcutta.

Novaro, S. F. (2002). Encuestas Agrícolas y Encuestas a Hogares: Estrategias Comunes. Paper published on the Final Report of the "*Reunión Regional Sobre Estadísticas Agropecuarias para América Latina y El Caribe,"* City of Panamá, Panama, December 2001.

Raj, D. (1972). *The Design of Sample Surveys*. New York: McGraw-Hill.

Sivers, L. A. (2002) International Programs Office, National Agricultural Statistics Service (NASS), United States Department of Agriculture (USDA), USA. Trabajo publicado en el *Informe Final de la "Reunión Regional Sobre Estadísticas Agropecuarias para América Latina y El Caribe,"* City of Panamá, Panamá, December 2001.

Sukhatme, P. V. and Sukhatme, B. V. (1970). *Sampling Theory of Surveys With Applications* (Second edition). London: Asia Publication House.

USDA. (1983). Scope and methods of the statistical reporting service. *Miscellaneous Publication Nº 1308*, Washington.

USDA. (1994). *CASS. Computer-Aided Stratification and Sampling*. Washington, DC: National Agricultural Statistics Service, United States Department of Agriculture.

Yates, F. (1981). *Sampling Methods for Censuses and Surveys* (4th edition). London: Griffin and Co.

Zarkovich, S. S. (1965). *Sampling Methods and Censuses*. Rome: FAO.