

SOME OBSERVATIONS ON THE WORLD CENSUS OF AGRICULTURE, 1960 80

by

P. C. MAHALANOBIS

Indian Statistical Institute, Calcutta, India

I shall avail myself of the present opportunity to examine some aspects of the FAO proposals for the World Census of Agriculture 1960 from the point of view of the needs of an under-developed country like India. The aim of the Census is to collect basic data and improve international comparability and accuracy of agricultural statistics all over the world.

BASIC DATA

We have to recognize that basic data are broadly of two types relating respectively to:—

- (A) Utilization of land (area under forests, food crops, cash crops, fallow, horticulture, live-stock etc.); the out-turn of various crops etc.
- (B) Number of persons living on the land (classified by sex, age, employment status etc.); and information on holdings (operational, ownership, land tenure, rent etc.).

In India we are in need of both types of information. Because of periodical shortages of food, information of type (A) has a higher priority in India. Such information is also likely to have a high priority in other underdeveloped countries which are of a predominantly agricultural character. We should therefore suggest including Type (A) information in the basic list (and not merely in the Extended List).

The "holding" is a convenient unit of enquiry and analysis for information of Type (B) and can also be used to collect information of Type (A). It is, however, possible to use "fields or plots of land" as the unit of enquiry and analysis for information of Type (A). The FAO proposals seem to look upon the holding as the only unit of enquiry and analysis. This emphasis on holdings may have arisen from the fact that in the advanced countries the holding has been the traditional unit of enquiry and analysis. It is also possible that in the advanced countries information of Type (A) can be collected with acceptable accuracy through interviews on the

basis of holdings. This is not however a sufficient reason to recommend its universal use in the underdeveloped countries. Experience in India has shown that information on crop acreages and crop outturn can be collected ~~far more~~ accurately by physical observations on fields or plots of land than by the interview method on the basis of holdings.¹ It would be desirable, therefore, to broaden the concept of the unit of enquiry and analysis to include the field or plot of land (in addition to holdings).

It is also necessary to recognize that physical observations on crops may give more accurate information on the acreage and outturn of crops in under-developed countries. The interview method must, however, be used to collect information of Type (B). Ideally, in an under-developed country like India it would be desirable to use both the methods: (a) physical observations on crops, and (b) interviews on the basis of holdings respectively for the two types of information. To use both methods at the same time would be expensive. In this situation it is necessary to explore possibilities of economy by either reducing the items of information and/or by using a more efficient statistical technique for such sample surveys.

I am glad to note that FAO has broadened the concept of census to include a sample census. There is a good deal of literature on the relative advantages and disadvantages of a sample survey in comparison with a complete census. I shall, therefore, restrict myself to some observations on sample surveys of agricultural holdings by the interview method on the basis of our experience in India which may be of interest to other under-developed countries.

SAMPLE SURVEYS OF AGRICULTURAL HOLDINGS

A sample survey is usually less costly than a complete census because the expense of covering all units would be greater than that of covering only a sample fraction (unless savings made on account of the smaller number of units to be investigated are offset by the increase in cost due to more complicated processing of the sample data). But economy is not the only consideration. The most important point is whether the accuracy of the results would be adequate for the purpose in view. Indian experience has shown that a properly designed and properly conducted sample survey can often give more accurate results than a complete enumeration. This point deserves some further consideration.

It is now fully recognised that a large-scale complete census is subject to many sources of errors. There are "coverage errors" due to either some of the holdings being overlooked and some others being included more than once. There are "identification errors" due to mistakes in identifying the units or items under investigation. There are "ascertainment errors" due to mistakes in ascertaining the information

¹ "Area under production" is included in the Short List. It is not clear whether the implication is that such data can be collected fairly accurately by the interview method. This is not corroborated by the Indian experience. To obtain reliable information, for crop-acreage it is necessary to make direct observations on fields; and for "yield per acre" it is necessary to harvest sample-cuts of crops.

due to misunderstanding of bias on the part of either the interviewer or the interviewee or gross mistakes due to negligence or dishonesty. (For example, a good deal of probing is necessary to collect information on the size of the holding in India, and which involves a listing of all the plots in the holding, ascertaining the area of each plot and adding them up, checking with the land records where available, and then a reconciliation of discrepancies by a second visit. Attempts to collect the same information by two investigators would sometimes show a difference in the results obtained by them.) There are "processing errors" at the stage of compilation, tabulation, and analysis. Some of these errors tend to cancel out but there are other components (for example, those due to bias) which are of a non-random character and tend to remain more or less fixed and constitute a "systematic error" which would affect both complete censuses and sample surveys. If the quality of the operational work in collecting and processing the data is high then the magnitude of the systematic error would be small and vice versa.

It is possible at some additional cost to improve the quality of work and thus reduce the systematic error by employing more qualified workers, giving them more training, improving the methods of inspection and supervision, and by establishing more effective controls etc. The larger the scale of operations the more difficult and more expensive it would be to attain a high quality of work. A staff of the best available 100,000 workers would be of lower average ability than a more highly selected staff of 10,000; and so on. With given resources, the systematic error would increase with an increase in the volume of work. This usually puts a limit to the usefulness of a complete census; with a given cost the systematic error of a complete enumeration may be so large that it would not be worth while carrying it out.

In a sample survey both the volume of work and the cost would be greatly reduced, and it would be possible to improve the quality of work so that the systematic error can be reduced. It is also possible to reduce the errors of sampling by increasing the size of the sample until the sampling error becomes less important than the systematic error. When this stage is reached there would be no further gain by increasing the size of the sample or in having a complete census. The essential point is simple. A complete enumeration would, it is true, eliminate the sampling error but, because of its large scale of operation, would usually have a large systematic error. A sample of a suitable size, because of its much smaller scale of operation, would have a comparatively small systematic error and also a sampling error which is even less important. The total error of a sample survey would be, therefore, often much smaller than the systematic error alone of a complete enumeration. We thus get an apparently paradoxical result, namely, that we can introduce some sampling error (that is, use a sample survey) in order to reduce the total error and do this at a small fraction of the cost of a complete enumeration.

A sample survey also has two other great advantages. The results can usually be made available far more quickly than the results of a complete enumeration. In a properly designed and properly conducted sample survey it is also possible to make a valid estimate of the margin of error and hence decide whether the results are

sufficiently accurate for the purpose in view.¹ In a complete census (unless superposed checks are used in the form of a sample) it is not possible to make any estimate of the error.

It is, of course, true that a sample survey cannot give reliable estimates separately for very small geographical or administrative areas. To attain the required precision the sample may have to cover most or all the units of enquiry in the geographical or administrative area under survey, in which case it may be simpler to carry out a complete enumeration. The above consideration sometimes leads to the claim that a complete census has to be carried out because it is necessary to have data for "small" administrative areas. If it is really essential to have such detailed information then a complete census is, of course, indispensable.²

The demand for detailed information for small administrative units, however, requires careful examination. First let us consider estimates at the national or large regional levels. It has been already stressed that national totals obtained by a complete census are often not likely to be sufficiently accurate for practical purposes, while a good sample survey can be relied upon to give more accurate results at a fraction of the cost. Again, if annual or periodical inter-censal estimates are needed then it would be clearly desirable to use sample surveys.

Secondly, a large number of tables required at the national or regional levels are usually not needed or tabulated for the smallest administrative units although the basic data to prepare such tables are all available.³ Sometimes sample tabulation is used, for purposes of economy, to process the census data which would indicate that the detailed information is not really needed. Also, if the administrative unit is not very small it would be often possible to use the sampling method.

Although emphasis is often laid on the need of securing detailed information for small administrative units it is not at all clear that such detailed information based on a complete census is actually much used in practice. It would be most instructive if the FAO would make a special study of the manner in which the detailed informa-

¹ There is an upper limit to what one may meaningfully desire as regards accuracy. Concepts and definitions are not perfect; and it is impossible to define with mathematical exactitude what we want to measure. The world, whether it be in respect of holdings or any other thing, is in a constant flux while survey operations must cover a period of time. The world must have changed by the time the survey results become available. Also, in actual practice we are not concerned so much with what the position was exactly at the time of the enquiry but what it would be when action is to be implemented. The desire for accuracy must be taken into consideration and be consistent with the above-mentioned uncertainties (which may be called the "physical uncertainties").

² It may be noted, however, that because of smallness of the unit not only the systematic component of ascertainment error, but also the random component may contribute substantially to the error in estimates derived from complete enumeration.

³ I may note that in a complete census certain figures, for example, sub-totals by villages, have been traditionally calculated as an essential intermediate step for obtaining national totals. I have a suspicion that in some cases these intermediate sub-totals are wanted simply because of the census tradition.

tion obtained for small administrative units during the World Agricultural Census 1950 was actually used in different countries of the world. Such a study would show to what extent it is really necessary to have a complete census.

COMBINED SAMPLE SURVEY OF POPULATION AND AGRICULTURE

The importance of the population and the agricultural censuses is well recognized. In the recent draft proposals of FAO it has been suggested that information on agricultural (or farm) population and employment in agriculture should be collected through a population census with provision to link it up with the size of agricultural holdings. It is not clear why the combined enquiry should be restricted only to the information on agricultural population and employment. Why not have a fully integrated survey of both population and agriculture? There is no technical difficulty in carrying out a combined sample survey of population and agriculture. Such a combined survey would not only reduce costs but would also supply much information which cannot be obtained from two separate surveys. The only obstacle possibly is institutional. In many countries the two censuses are carried out by two different Government departments and both departments would probably like to continue their separate censuses. In underdeveloped countries in which there is no tradition of having two separate censuses it would seem most desirable to think of an integrated survey.

In India, although decennial population censuses are being taken since 1872, and seasonal crop censuses are also being carried out over a large part of the country every year, it was considered desirable to set up the National Sample Survey (NSS) in 1950 to collect information on a wide range of subjects. In one of its "rounds" the NSS collected data of a much wider scope than what is usually collected in either a census of population or a census of agricultural holdings. In such an integrated survey it is possible, for example, to have tabulations not only on operational holdings but also on ownership holdings, non-agricultural holdings, total land holdings etc. The integrated survey introduces great flexibility in the definitions of different sectors and types; and helps to reduce errors of coverage and of classification. In India (and in many other underdeveloped countries) with a very large number of small holdings there is no obvious line of demarcation between agricultural and non-agricultural holdings and there are a large number of marginal cases. A decision as to whether a holding is or is not agricultural must rest on a multiplicity of information on the holding. Attempts to cover only the agricultural holdings would, under usual survey conditions, lead to large errors of "identification" because decisions would have to be made on the basis of insufficient evidence. In the NSS, information is collected on all types of holding which makes it possible, at the stage of tabulation, to decide on the correct classification.

INTERNATIONAL COMPARABILITY

The FAO has been continually stressing the importance of international comparability of the results of the agricultural census. While I am in entire sympathy with this aim in a general way I think it is necessary to define and explain the precise meaning and purpose of international comparisons. There is much need and scope for building up a conceptual framework for this purpose. FAO should initiate research studies based on the analysis of available data, especially the results of the World Census of 1950, to formulate methods of international comparisons at a concrete level for future guidance.

I may give one example. The FAO has laid down, on grounds of international comparability, the classification of holdings by size as most important. It has been pointed out, however, that certain countries would prefer to have other classifications, for example, those based on subsistence and commercial farming or on value of production or on value of sales. I venture to think that higher priority should be given to fulfilling national needs than to international comparability. Some compromise solution may also be found. For example, if the number of basic size-classes is reduced to three or four (which every country would try to provide) it may be possible to have cross-classifications by three or four size-classes jointly with other types of classification like subsistence and commercial farming etc. (which each country would be free to decide). Every country would be, of course, free to provide sub-divisions of the basic size-classes (for which also suitable standards may be prescribed).

There are also fundamental technical questions relating to the real significance of cross-classifications by size. For example, the average acreage per holding under a particular crop may change progressively from one size-class to another with increasing size of the class but would such changes have any real significance? If in each size-class, say, about 30 per cent of the total land is under one specified crop, it would be proper to think that there is really no change in the pattern of cultivation so far as this particular crop is concerned although the average acreage per holding would increase progressively with increasing size-class. In such cases it would seem appropriate to give the data on the basis of "per-unit-area" of each size-class. This method of comparison is quite important and it would be desirable to specify "per-unit-area" as basic statistics. Incidentally, comparisons on the "per-unit-area" basis would also indicate which size-classes need to be retained. It would be extremely useful if FAO would study this question on the basis of the 1950 census.

ACCURACY OF AGRICULTURAL STATISTICS

In an underdeveloped country like India the highest priority and importance must be given to improving the accuracy of agricultural statistics. Usually, there is a varying degree of appreciation of the existence and magnitude of inaccuracies; and views are often expressed purely on the basis of subjective evaluation. It is not

enough to draw attention to the need of making an assessment of the degree of inaccuracy of the data. It is necessary to prepare a minimum programme of objective evaluation of the accuracy of the data and request each country to fulfil this programme and publish the results of the assessment, that is, include such a programme of evaluation in the Short List.

Various methods can be used for this purpose. For example, duplicate sets of data may be obtained by independent visits by different investigators to some (or even a very few) areas or villages selected, if possible, at random. Also, supervisors or inspectors may carry out a post-enumeration check in some (or even a few) villages. When the sampling method is used, a most convenient and powerful control can be introduced by the use of inter-penetrating subsamples each of which would be drawn with replacement and would therefore supply an independent estimate. A comparison of the different sub-samples would immediately reveal to what extent the results are in agreement and would also supply a valid estimate of the margin of uncertainty. Indian experience has shown that this is an extremely effective method for assessing (and hence controlling) the internal consistency of the data. I have also already mentioned that direct physical observations on crops by the investigators (independently of the cultivators) have been found in India to be a reliable method of estimating the crop-acreage and crop-outturn. This method also may be used as a check in some cases.

RÉSUMÉ

Quelques observations sur le recensement agricole mondial de 1960

Ce mémoire discute les propositions de FAO quant à le recensement mondial de l'agriculture de l'année 1960 à l'égard spécial des nécessités et les conditions dans pays sous-développés tel que l'Inde par exemple. La statistique agricole se compose de deux divisions, une (A) à l'égard d'utilisation de terre et de récolte diverse, et l'autre (B) à l'égard de la dépendance de la population sur la terre et, de renseignements sur les parcelles agricoles (holdings). Quand « la liste extensive » des propositions de FAO comprend les informations du type (A) (données des produits) j'ici suggère que la liste fondamentale duise comprendre les informations du type (A) en considération de fait qu'il y a des pays exposés périodiquement à le déficit de nourriture.

A l'épreuve dans l'Inde les informations du type (A) peuvent être ramassés exactement par le moyen d'observation directe

sur le terrain ou sur les entrevues à l'égard de parcelles agricoles (holdings). Il est nécessaire qu'on fasse des préparatifs pour la collection des informations du type (A) à l'observation d'après la nature directe des champs et des petites pièces de terre, et du type (B) par la méthode d'entrevue sur la base des pièces de terre affermée. Tant s'en faut qu'on parle de deux recensements qu'au contraire il ne sera pas faisable pour la plupart des pays économiquement sous-développés de s'engager à même un recensement. La solution se trouve évidemment dans l'application des méthodes d'échantillonnage.

D'ailleurs comme il est dans les autres cadres, il est probable que les enquêtes échantillonnées agricoles soient plus effectives que le recensement complet, à cause de meilleure maîtrise des erreurs d'opération. Les demandes des informations en détail du sujet des petites unités

administratives (que sont obtenues seulement par recensement global) sont examinées d'une manière critique. Il semble nécessaire de faire une investigation jusqu'auquel point les données à les petits unités administratives recueillies dans le cours du recensement agricole mondial de 1950 ont été véritablement utilisées. Il est plus désirable qu'on prenne l'enquête intégrale de la population et de l'agriculture en ensemble, mieux que la faire séparément. De cette manière il est possible de réduire les coûts et de garantir un rapport vrai entre les idées et les définitions.

Pendant que les comparaisons internationales sont très importantes il est nécessaire qu'on se rend compte des besoins particuliers et des conditions singulières des pays divers. On doit ajuster aussi le programme. A propos, il y a beaucoup de besoin et assez d'espace de construire une charpente des concepts pour déterminer et expliquer les sens et le but des comparaisons internationales.

Un programme minimum d'évaluation objective de la précision de la statistique agricole doit être inclus dans la liste courte (Short List).