SAMPLING IN A GOVERNMENTAL STATISTICAL SYSTEM

By

W. EDWARDS DEMING

Graduate School of Business Administration
New York University

Delivered at the International Seminar
on Statistical Organization, conducted
by the United Nation, Ottawa
13-31 October 1952

Published by the STATISTISCHES BUNDESAMT (Abteilung VIII), Wiesbaden, July 1953

Sampling in a Government Statistical System

Berichterstatter: Prof. Dr. W. E. Deming (New York University)

The aim of this lecture: sampling as an integral part of a statistical system. The aim here of this lecture will be to explain the uses of sampling as an integral part of a government statistical system. It will not be an attempt to teach the techniques of sampling; rather, it will deal with sampling as an administrative device for scientific management of a statistical system. We may cover in one lecture a great many of the essential points about the uses of sampling from the standpoint of an administrator, but we could accomplish nothing we attempted to study the techniques in a short time.

In the first place, sampling should be used as an integral part of a statistical system. The surveys should be planned methodically, with the use of all the best techniques of questioning, instruction, interviewing, format, sampling, and tabulating. A complete census is no exception: as we shall see later, sampling is used to supplement, broaden, and speed the results of a complete census. Sampling, to be most effective in a complete census, must be built into the planning, long in advance, and not afterward as is done too often in a desperate attempt to salvage something from the complete census, after the field and office operations have broken down from sheer size.

Increased reliability, speed, and low cost are the chief characteristics of data obtained by sampling when it is carried out by rigid rules of selection and of estimation as dictated by statistical theory. Through the use of sampling, a government statistical system may now furnish current figures on changes in the population by area, by age groups, on employment and unemployment, on home manufacturing, on wages, prices, agricultural production, condition of crops, for the administration of government, manufacturing, distribution, and export trade.

Some of the most important decisions in government and in private business are made continually on the basis of information supplied by samples. This information may be economic or social, with respect to people or industrial or agricultural production, or it may be information on the quality or weight of an industrial product. Millions of dollars worth of product change hands every hour and are paid for on the basis of samples.

Where conditions change rapidly because of population shifts, or because of weather that causes failures of crops or surpluses thereof, or because of changes in patterns of trade, the administration of government or of commerce requires information that is current. Such information can be furnished only by samples, because a complete census is too slow or too rigid, and can not furnish all the types of information that will be required a few months or a few years following the census. A known and controllable margin of sampling error in the data delivered by a sample may be far preferable to the more exact and detailed data of a complete count, delivered too late to be of any use.

Because of sampling, emphasis in government statistics has shifted during the past 15 years to the satisfaction of current needs by means of monthly, quarterly, or annual surveys; while the historical value of complete counts of population, agricultural, and commerce has assumed less importance. Regular, complete, and reliable censuses are even more necessary than ever, but their utility is not for the current information that government and commerce require today. Rather, their utility lies in the fact that they make good samples possible for several years.

Samples may be used in the complete absence of census information, to obtain counts of the total number of inhabitants and of their frequencies by characteristics. This is a fact frequently not recognized. Of course, under such conditions the samples must be bigger than if there were good census information on hand to use in the planning[1]).

Even obsolete census information, or a poor census, or an old city directory, a map or a list of any kind, can be very helpful in the design of a sample. Errors, obsolescence, and simple omissions in such sources need not cause bias in the procedure of sampling, because in modern sample-designs these defects cancel out: the sample is self-correcting. (A further word on this point occurs later on.)

Sampling does not eliminate the need for a complete census. On the contrary, sampling enhances the usefulness of a complete census, and changes its character. The type of complete census that is most useful is one that furnishes certain basic figures for small areas, such as the number of people, the number of dwelling units, the percentage of the population that is foreign-born, the number of farms, or the number of acres that are cultivated. This type of information is very helpful in the design of samples of the population for various purposes. For sample-surveys of retail trades and services, lists and maps that indicate roughly the number of business establishments in small areas are likewise helpful.

What sampling is. From the standpoint of the administrator, sampling is a device for improving the usefulness of a government or private statistical system by speeding the statistical information, improving its reliability, and decreasing the cost of a unit of information. From the standpoint of technical statistics, sampling is a study for improving the procedures for collecting and interpreting statistical information with the aid of the theory of probability.

Sampling is a tool of precision. The design of samples is a professional job, requiring skill and ingenuity in theory, and maturity in practice. Even the simplest of sampling plans must be carried out rigidly, by a definite plan of selection and by a definitive plan for calculating the estimates and their margins of sampling error.

A brief outline of the steps in the design of a sample follows under:

a) A decision on what types of statistical information would be useful for rational decisions in a particular problem.

b) A decision on the lowest precision that would be satisfactory.
A survey that will deliver more precision than is necessary will increase unnecessarily the cost of the information, and will retard its date of delivery.

c) The formulation of procedures of selection and estimation, with the aid of the theory of probability, to obtain the required precision at low cost, and to make the best use of the available physical and human resources.

[1]) Outstanding examples of sample censuses of population and agriculture, accomplished in the absence of an earlier complete census, are furnished by the work of Mr. J. R. H. Shaul in Africa. His methods are contained in an article, "Sampling surveys in Central Africa", J. Amer. Stat. Assoc., vol. 47, 1952: pp. 239—254.

d) The interpretation of the information furnished by the survey in a form that will assist rational decisions and add to knowledge.

Sampling is thus a tailoring of the procedures to fit the requirements at low cost. At the same time, it gains speed, and brings improvement in quality.

Statistical data are for purposes of action; not merely to fill up books. Therefore, the first step in the planning of a sample is to enquire into the purpose of the survey- what information is needed, and why.? How will the information to be obtained assist the decisions to be made ? What precision is required ?

The estimates that are produced by a sample survey possess a controllable and measureable standard error, such as 10%, 5%, or 1%. This precision is not an accident: it is built into the design, in advance. It is then calculable more accurately afterward from the results of the sample itself, and not by comparison with census figures.

In any given instance it is therefore possible to design a sample that is tailored carefully to the requirements. For some problems great precision may be required, in which case the sample may be costly and slow. On the other hand, for other problems, no great precision may be needed; it may be sufficient to have the results in broad classes, in which case the sample may be smaller, less expensive, and more rapid.

In the data from any survey there will be uncertainties which arise from nonresponse, wrong information, refusals, failure to define the universe accurately, carelessness and blunders in performance, bad luck in planning and execution. The errors of sampling are only one kind of uncertainty.

The nonsampling errors may be much bigger than the errors of sampling unless proper care is taken to define the universe properly, to build a suitable questionnaire, to train the interviewers properly, and to reduce nonresponse, and to measure the bias therefrom. Through the use of correct sample-designs the nonsampling errors of this nature are usually decreased, as sampling affords the means by which to measure them, and to discover which types of procedures will reduce them.

It is important to note that the standard error is a measure of the error introduced by sampling. It is not a measure of the nonsampling errors, such as nonresponse, wrong information, bad planning, etc. This distinction will be clarified in more detail in the last part of this lecture.

Some advantages of sampling.

) Controlled precision, tailored to the requirements

b) Improved reliability

Through better controls and better interviewing, and through more rapid improvement of the techniques of questioning, interviewing, and processing.

c) Speed (and hence greater utility of the data)

d) Low cost (permitting expansion of the statistical program, and expanded usefulness)

e) Clarification of the universe to be studied, and clarification of the aims and purposes of the entire statistical program

This clarification could take place without the introduction of sampling, but it is hastened by sampling because the samples must be tailored to the requirements.

It is an astounding fact that wherever sampling has been used as an integral part of a statistical system, both in conjunction with complete censuses and for sample surveys, the result has been improved usefulness, and improved quality and reliability, with decreased lowered costs. This is so because samples are tailored to the requirements, and because the requirements must be clarified.

Speed is important because the social order is not static but dynamic. Census and sample are therefore both to some extent out of date by the time the data are processed and published, but the sample has a distinct advantage, as its time-interval is measured in days, weeks, or months, against many months or even years for a complete census. For this reason, in many types of enquiries, the complete census, by the time it is processed and tabulated, is often not as good a basis for action as the earlier returns of a small sample would have been. To the extent that sampling methods produce quicker results, they therefore often produce more accurate results for purposes of action and for prediction.

Sampling discloses shifts in population, employment, prices, business activity, agricultural production, etc. *While These Changes Are Taking Place* and not after it is too late to do anything about it.

What sampling is not. A sample is not the mere substitution of a partial coverage for a complete coverage. The following examples are *Not* samples:

1. The first 1000 returns from a census or survey of any kind.
2. 5000 interviews of people selected according to the judgment of the interviewer.
3. A convenient portion of a country or of a file of cards.
4. A sample drawn and attempted, but only partially completed.
5. 100 households selected according to the following plan:
   a) Count the streets in a city.
   b) Select 50 streets by random numbers.
   c) Show for each street the highest house-number.
   d) From each street select 2 house-numbers at random, and select 2 households at random from each house-number.
   e) Compute the simple mean of the sample and use it as an estimate of the mean per household in the city.

The above examples are *Not* samples because no person or household has in advance a known probability of getting into the sample; hence no calculation of the precision of the results can be carried out, either in advance or afterward. No proper size of sample can be specified for such procedures. Hence such procedures can not be admitted as possible where measurable and guaranteed reliability is requisite.

A further clarification is desirable. Sampling is not an indiscriminate drawing of some particular number of people or farms, at the whim of an expert in some realm of subject-matter, such as 1000 or 5000, nor is it some mystic percentage, such as 10% or 1%. The size of the sample required, and the manner of drawing it, and the formula for calculating the estimates and their margins of error, are decided by calculation with the aid of the theory of probability, in order to meet most economically the degree of precision that is required for some useful purpose.

Sampling is directed to a particular purpose, and is tailored to meet this purpose, no more, no less.

Many a time I have been called upon to design a sample of some designated size, only to find upon enquiry that the purpose of the study had not yet been stated clearly, and that there was serious doubt whether any useful purpose could be served by any sample. Under such circumstances it is my practice to discourage further consideration of a survey until the problem has received further study.

A sampling plan can not be judged suitable or unsuitable until the purpose and requirements of the survey have been stated. Thus, someone may tell me that he has designed a plan of selection that will use 200 groups of approximately 5 households each, contiguous or separated, to be drawn from lists of areas, statistics for which were given in the census of 1950; and that the plan of estimation will be by straight expansion, or by the use of ratios formed area by area. If he asks whether this is a good plan, I can not answer, because I know not the purpose of the survey and what precision he requires. I can only say that the estimates formed from such a sample will have calculable margins of error, by the use of statistical theory. Until the requirements are stated, there is no way to judge a plan either suitable or unsuitable.

Some of the uses of sampling in connexion with censuses of population. Our next step is to enumerate and to discuss some of the ways in which sampling is used in connexion with a complete census.

(a) In connexion with a complete census, to ask certain questions of only a sample of the population. (Aim: to broaden the scope of the census at low cost, without proportionate increase in the burden of response or in the time required for carrying out the work.)

(b) To tabulate a sample of a complete census. (Aim: to hasten the results for large areas; to save time and expense in the production of some of the tables that would ordinarily be produced from the complete census; to broaden the scope of the tabulations.)

(c) To collect information from only a sample of areas or of other units. (Aim: to replace a complete census by a sample census; to gain speed; to decrease the cost or to provide information under conditions that render a complete census either unnecessary or impossible.)

(d) To control the quality of the coding, punching, and tabulating, at various stages of the processing. (Aim: to investigate the effects of such errors on the published tables; to control and to improve the quality of the finished product; to save time and money in the completion of the work.)

(e) To investigate the quality and the meaning of the figures obtained in a census. (Aim: in samples of areas, selected with the aid of statistical theory, to carry out studies on completness of coverage, over and underenumeration, misunderstandings of the questionnaire, differences between interviewers; to test various methods of obtaining the information.)

Illustration of broadening the scope of a complete census and of tabulating a sample of the returns. Whenever sampling is used to broaden the scope of a complete census, the same sample of people, households, farms, or other units will also provide a sample for tabulations of the complete count. Thus, in the censuses of the population of the US in 1940 and 1950, the B-cards, which contained the information that was obtained for every person, plus certain supplementary information that was obtained only for the people in the sample served not only to provide tabulations of the supplementary information, but also — at great saving — for a large portion of the tabulations of the regular census information that was obtained for everyone, and which could have been tabulated 100 per cent if there had been any need of it. The substitution of the sample for the complete count enabled the programme of tabulation to be expanded greatly while costs were held within the inflexible budget.

Sampling is used now regularly for broadening the scope and the tabulation programme of a complete census in a number of countries, not only for censuses of population, but for censuses of agriculture, housing, manufacturing, and distribution.

Illustration of an advance tabulation of a sample. The samples mentioned in the preceding section served also to provide advance tabulations of information that was needed urgently.

A recent example of an advance tabulation of a sample of the returns of a complete census ist furnished by the census of population of Japan, October 1950. The chief aim of this particular sample tabulation was speed and simplicity, with a minimum of interference with the processing of the complete census, even at the sacrifice of efficiency. The first advance tabulation was based on a 1 per cent sample of the enumeration districts that contained private households, plus a 1 per cent sample of names from non-private households. Ten enumeration districts, drawn with random numbers from every consecutive districts of private households furnished ten 1 per cent samples of private households to which were added ten 1 per cent samples of non-private households. The first random number in each group furnished the 1 per cent sample, which was processed first, and followed by the other nine samples to make a 10 per cent sample which will provide the second advance tabulation. These tabulations were followed in turn by the processing of the complete count. It was my pleasure to suggest this advance programme of tabulation to the Japanese statisticians, and to work with them on the design of the procedures.

The first release of results from the 1 per cent sample for the whole country of Japan was published seven months following the census. Further releases followed within a few months. The tables include:

The number of inhabitans by sex and age in 5-year age-groups for urban and rural areas.

Characteristics of the labour force 14 years and over, by 11 broad industry-groups and by 8 broad occupations, by sex, urban and rural.

Characteristics of the labour force, age 10—13 and 14 years and over, by sex, urban and rural.

Marital status, 15 years and over, by sex in 5-year age-groups, urban and rural.

The 10 per cent sample will supply the same kind of information as the 1 per cent sample for prefectures and large cities as well as additional information that will not be prepared on a 100 per cent basis for the country as a whole.

Incidentally, the estimate of the total number of inhabitants of all Japan turned out by accident to differ from the total count by only 1 part in 800. However, sampling error of small classes are of course to be expected, as shown in a table in the releases.

In the Indian Population Census of 1951 a 10 per cent sample of households was separated out at the level of the lowest census administrative unit. Out of this 10 per cent sample various sub-samples were to be tabulated separately to provide advance information and variances for future sample design.

Sample censuses, independent of a complete count, have been taken on a number of occasions with excellent results and under diverse conditions. An early example is the censuses of 9 urban areas, described in a booklet *A Chapter in Population Sampling* by Deming, Hansen, Hurwitz, and Tepping (Bureau of the Census, Washington, 1947). The number of inhabitants was ascertained in each city with a standard error of sampling of less than 1 per cent in remarkable accordance with the precision aimed at in the planning. Besides the error of sampling there were, of course, the regular non-sampling errors of a complete census. These sample censuses were taken under favourable conditions, with a background of experience which provided information on variances and costs. Briefly, the method consisted of listing all the dwelling places in selected small areas within the city, then drawing a sample of dwelling places from this list and sending interviewers to the sample to obtain the information required for the census.

More recently, as mentioned earlier, in 1949—1951, Mr. J. R. H. Shaul has taken sample censuses of population and agriculture in Northern and Southern Rhodesia and in East Africa, without the benefit of a previous census, but with remarkable adaptation of theory to difficulties and limitations. The successes of these sample censuses open up the possibility of obtaining census information in many areas of the world in which no census has been taken recently.

In fact, at this date, Northern Rhodesia, Southern Rhodesia, East Africa, Gold Coast, Nigeria, Ruanda Urundi, and the Anglo-Egyptian Sudan are either using sampling methods or are preparing to do so.

Random samples vs. judgment samples. Modern sampling is tailored to the required precision in any particular case through the use of the theory of probability. It is here that modern sampling derives its efficiency. But the theory of probability applies only when the persons, households, areas, or other units were selected at random. The simplest and only satisfactory way in practice to make random selections is by the use of random numbers.

When we use random numbers we give every unit a serial number, and then, in effect, mix these serial numbers very thoroughly. The use of random numbers ensures a thorough mixing. This is how we achieve equal probabilities. It is as if we (a) wrote each serial number on a small disk like a poker chip, the disks being all uniform and physically similar; (b) placed these disks all in a smooth bowl; (c) stirred the disks for many hours; (d) drew a handful of disks from the bowl. The random numbers accomplish all this with simplicity.

Of course, in practice we often stratify the units. That is, we put some chips in one bowl, and some in another. We often introduce other complexities, such as subsampling. In other words, we make all possible use of judgment and of previous information in the form of census data for areas, or in the form of lists, directories, or maps. But the final selection of just what areas or of who is to be in the sample is not left to judgment: it is removed from human preference and judgment and is left to the random numbers instead.

The trouble with selections made by judgment is that the procedure is biased. What is worse, (a) a sample selected by judgment does not improve as the size of the sample increases, unless the sample is some very costly and unwieldly large portion of the whole, so that its accuracy is obtained by the force of sheer size, and not by design; (b) there is no way to assess the margin of error of the results. In contrast, when the selections are made with random numbers, the margin of error (standard error) is calculable from the returns of the sample.

One important point about the use of random numbers is that one person can use them as well as another. A new employee may learn to use a table of random numbers in a short time, and the samples that he draws will have the same validity as if an experienced statistician had done the drawing.

Brief sketch of procedures. In actual practice, a sample is a sample of units of some kind, which are drawn by random numbers, as mentioned earlier. The sample starts with a frame, which may be a list, a map, or directory, or other description of the universe concerning which information is desired.

It is really very easy to draw a simple but perfectly valid random sample of a region for information concerning the number and characteristics of the people, or for agricultural production, or to serve other needs. The procedures are especially simple when applied to an urban area, because maps are usually easier to find, and the work may be carried out with less moving about. The steps may be described in these terms.

a) Use the frame to divide up the area into several hundred or thousand pieces, called sampling units. Preferably these pieces should contain roughly equal numbers of people, or of acres, through the use of census or of any other information that can be obtained. However, the validity of the procedure is not destroyed by failure or inability to achieve rough uniformity.
b) Give a serial number to every area.
c) Draw a sample of these areas by the use of random numbers.
d) Interview the specified people in these sample areas.
e) Compute the average number of people or of anything else per area in the sample, or of the number of bushels of rice per acre in the sample, and then use this average as an estimate of the overall average value per sampling unit. As the number of sampling units in the entire frame is known, a simple multiplication gives an estimate of the total population that has any particular characteristic, or of the total number of acres in any particular crop.

By a valid sample I mean one that is drawn and estimated by a procedure that is unbiased or nearly so, and which has a computable and controllable standard error.

A complete count is defined as a complete coverage of all the pieces that constitute the frame.

In practice, for greater efficiency, we often use formulas of estimation that are more complicated than the simple average. Moreover, we may use more complex designs, with stratification and subsampling.

But the more complicated sample is better only in the sense of being more efficient: it gives a smaller margin of error for the same cost. The simple sample is still valid; it has a valid and calculable standard error.

A person may learn to draw simple random samples in a short time. But the design of highly efficient samples, by which to obtain a great deal of information from a very small sample, is a highly intricate professional job, and requires many years of study and experience.

Simple gaps in a listing may be filled in easily and automatically in the interviewing stage by use of the "half open interval", which specifies that the interviewing shall begin at a certain point on the list and extend up to but not include a second point. In this way, addresses not on the list still have the same probability as any other of being interviewed, provided they lie between two addresses that are listed.

For surveys of the population, if the frame is a map or a list of areas, the sampling units should preferably be roughly equal in size, or else their approximate sizes should be known. If the list is a list of business establishments for an economic or social survey, it is preferable that the list show the size of the establishments in total sales last year, or in some other meaningful measure. If these conditions can not be met, the sampling may still proceed, but the size and cost of the sample will be greater, because of the greater variability of the sampling units that will be formed within the frame.

Sampling sometimes proceeds in stages. The frame is then broken up into fairly large pieces (called primary or first stage units), each of which is to contain some certain number of smaller pieces (called secondary units), preferably of approximately uniform size. Serial numbers are then assigned to the smaller pieces. A primary unit is not actually broken up into the smaller pieces unless one of the secondary units is drawn into the sample. Thus, considerable time and expense may be saved.

It is difficult to lay down any general rules regarding the delineation and definition of the areas that are to be used in the various stages, because there is always a way of making use of whatever maps and lists exist. The primary areas may be large and preferably heterogeneous and they need not contain even roughly a uniform number of inhabitants, provided the number of inhabitants of each area is known approximately. In contrast, for convenience, the smaller areas delineated for selection in subsequent stages may preferably be very homogeneous and roughly uniform. Groups of households contiguous or separated, may be used as the ultimate sampling unit. These groups should also preferably be roughly uniform in size, although this is not very important where only proportions or averages are to be estimated from the sample, as when the total population is known from a concurrent or recent census.

The type and specification of sampling unit to be used at any stage will depend not only on what is preferable but on what maps and tabulations are available by area from previous censuses. The problem in sampling is to make the most effective use of whatever sampling units can be devised with the maps and information available, with consideration for the abilities of the workers who will take part.

Stratification is sometimes a very useful device for increasing the efficiency of a sample. On the other hand, stratification is sometimes ineffective and disappointing. As stratification is often expensive, it should not be used unless its cost will be wiped out by the savings that result from increased efficiency. The only safe plan is to make use of the proper statistical theory in advance, by which the gains to be expected from stratification can be calculated pretty accurately.

Stratification by geographic position is often good. Further stratification can not be used unless there is some rough information available from previous censuses or other source.

The use of ratios in place of straight arithmetic average in the estimation is sometimes very effective in increasing the precision of a sampling plan.

There are many possible devices to consider, and the statistician who designs samples will make use of any of them when theory indicates that a substantial saving will be the result.

Sampling must be carried out rigidly in accordance with appropriate plans. The procedure must be supervised very carefully. The estimates of the various totals, averages, and standard errors must be formed according to strict rules in conformity with the underlying theory.

Tabulation procedures must be considered. The procedures for preparing the estimates from the sample will depend on the tabulating equipment that is to be used, or on the abilities of the office force in case the work is to be done by hand. In turn, the sampling procedure itself will be geared to the procedures that are to be used in the tabulation. For example, if the tabulations are to be carried out by hand, the sampling may well be made uniform (such as 1 in 20) over all areas to avoid lengthy computations and errors.

Some considerations on the cost of samples. Many factors affect the cost of either a complete count or a sample. Consideration must be given to the size and complexity of the sample, which in turn are governed mainly by the aims of the survey and the procedures that will be most efficient. The aims of the survey might be to obtain a figure for the total population, or only to obtain proportions. The cost of obtaining an estimate of the total population within 2 per cent will be much higher than the cost of obtaining an estimate that is reliable within 5 per cent. If a complete count is taken concurrently, it will presumably give the total population accurately, in which case the sample is taken in order to provide the relative proportions of various classes, to broaden the scope of the census (by permitting additional questions on the questionnaire), or to provide measures of the completeness and reliability of the complete count, and to lessen the costs of tabulation.

Various procedures are used in sampling human populations, depending on the availability of lists, maps, prior census information on the characteristics of the population for small areas, trained personnel, instructions and training materials left over from previous similar surveys, tabulating equipment, and the like. Good detailed maps of small areas, preferably accompanied by rough measures of population, are particularly useful. When complete lists of households are already on hand by small areas, as appears to be so in some countries, the sampling plan may be relatively cheap and efficient. In general, the more information there is about the populations of small areas, the cheaper the sample required to produce the desired accuracy.

The population to be sampled may be largely urban or largely rural, or a mixture. Costs are particularly high if the area is mountainous, sparsely populated with meagre transportation facilities, or if there are many islands to cover. Such conditions raise the costs of both complete counts and samples.

Careful preparation is necessary if either a complete count or a sample is to yield useful results. Adequate time must be allowed for planning, drawing up the questionnaire, testing it, revising the definitions, training the workers, providing adequate controls, checks, and supervision of the work, both in field and office; otherwise no survey, complete count or sample, can succeed. Much of the planning for a sample is parallel with the planning for a complete count, and it would be wrong to suppose that sampling offers any short-cuts to the construction of the questionnaire, the definitions, training, and certain other phases of the preparation.

However, the field-work and office-tabulations of the samples are usually greatly simplified (although they must be carefully controlled) and it is in these operations where savings of time and money are often effected.

Further advantages of sampling. Increased reliability, speed, and low cost have already been mentioned. In addition, there is the reduced burden of response. A sample requires answers from only a fraction of the population, hence, on the average, a sample is much less of a burden upon the public than a complete coverage is. In countries where sample surveys are carried out continually, the selection of the households, farms, or other units can be so arranged that no family is included in more than one survey in several years.

The field-work for a sample does not require so large a staff as is required for a complete count. The result is a less severe demand on manpower. The same is true in the coding and tabulating, with the further result that less office space and equipment are required. Recruiting is therefore easier and quicker, and a higher grade of personnel may be selected for the work. Lessened cost and higher quality of work are direct advantages.

This statement is, of course, not true when sampling is used in conjunction with a complete count, as there must be enough workers to carry out and process the complete count as well as the sample, and the same field and office workers ordinarily do both of them simultaneously.

A further advantage, clearly obvious where sampling has become an integral part of a statistical system, is the fact that consideration of sampling errors forces attention upon other errors, such as those arising from faulty preparation of the questionnaire, the field and office procedures, and the analysis of the data. Care in the interpretation of sample data, and comparisons of the data obtained by complete counts, often disclose inadequacies in complete counts as well as in samples. The ultimate effect of the introduction of sampling has invariably been a refinement in both the sampling procedure and in the methods of complete enumeration, and better interpretation of the results from both.

Some disadvantages of sampling. There are instances where a sample may be more expensive than a complete coverage. Thus, for the so-called "block statistics" in the US, a sample would have to be so near 100 per cent that a complete coverage (100 per cent sample) would actually be cheaper. If an accurate figure (e. g. within 2 per cent) for the number of inhabitants is required in an area of 15 000 or fewer households, the cheapest solution may be a complete census.

The time and skill required for drawing up a sampling plan and for supervising its execution may lie beyond the capacity of the staff immediately available. It is then necessary to acquire the assistance of a statistician skilled in sampling, together with whatever number of assistants he requires. The sampling expert should be called early, as delay may seriously hamper his efforts and his results.

Sampling errors will be present in a sample, but the range of sampling variability is under control, as has been explained. A sample should be so designed that the errors of sampling do not impair the data for the uses intended. Moreover, an error of sampling, even as high as 10%, may be far preferable to the more exact results of a bigger sample or of a complete count finished too late to be of any use.

APPENDIX

Definition of precision and of the error of sampling. The administrative uses of data, whether in government or in private business, require the correct interpretation of the precision of a sample, or of the error of sampling, commonly measured by the standard error. To acquire an understanding of these things, it may be helpful to rehearse briefly the procedures of sampling, and the relation of a sample to a complete coverage. A sampling plan consists of three main parts:

1. A decision on the definition or delineation of the sampling units (single farm, group of farms, areas within areas). These rules, when applied, will divide the whole region into "sampling units." (In a 2-stage plan there will be further subdivision to form secondary sampling units within the primary units that fall into the sample, but we need not introduce any complexities for the present purpose.)

2. A decision on the procedure for drawing these units (e. g., the use of random numbers; equal probabilities within strata, whether to use strata or not and how to define them).

3. A decision on the particular formula to use for forming the estimate X for some particular characteristic of the farms of the sample. For example, X might be —

The total number of farms in the region, multiplied by the average number of acres under wheat per farm as found in the sample.

The total number of acres in all the farms in the region, multiplied by the average proportion of the acres under wheat in the farms that fell into the sample.

Now we are ready to define a sampling error. We imagine that a complete count has been taken over some particular region. We may suppose that this is an agricultural census. Interviewers were assigned to call on all the farms in the region. Perhaps they failed to find some farms: almost certainly there will be some nonresponse, perhaps some outright refusals, some information missing, some of the information wrong. Yet the census might be a very good one and extremely useful for many purposes.

The steps by which we shall define the error of sampling are these:

Step 1. Tabulate the results for the complete census. We shall use them to measure the errors of sampling.

Step 2. Decide on some particular procedure for a sample. This will consist of the three main parts as described above.

The question of economy and efficiency is of no moment at this point. Likewise, we care not at this moment whether the plan is single stage, multi-stage, stratified or not.

Step 3. Summarize the results of the complete census for every sampling unit, and for the entire region.

Step 4. Draw the sample by the procedure already specified in the prescribed sampling plan.

Step 5. Form the estimate X for some particular characteristic by the procedure already specified in the prescribed sampling plan.

Step 6. Let A be the result of the complete census for the entire region for the characteristic now under discussion. Then the error of sampling for this particular sample and for this particular characteristic is

$$\Delta X = X - A$$

The sampling error, defined in this way, is strictly the error that has arisen from the introduction of sampling to replace the complete census. The sample will contain nonresponse, incomplete information, careless entries, omissions, etc., and in fact all the other imperfections of the complete census, and in about the same proportion. In the tabulation of the sample there will arise the same questions as arise in the tabulation of the complete census with regard to editing, coding, filling in missing information, etc. The only difference is that the sample is afflicted with errors of sampling, *the sole reason for the disagreement* between X and A.

Definition of the expected value and the standard error of a procedure of sampling. In practice, we do not have the result A from the complete census for the total number of acres under wheat, or for the number of farms that sold milk off the farm last year—at least not until long after the sample-result X is calculated and put on the record. If we knew A in advance, there would be no need for the sample.

Without the complete census, we do not know whether X lies to the right or to the left of A.

We now recognize the fact that many other samples could have been drawn by the same rules of selection, from the same sampling units that were designated in Step 1. Hence, many sample results are possible, each with its own X for this particular characteristic.

Any one sample that we carry out in practice is one set of results drawn at random from a list of all the possible sample-results (values of X) that can be formed from the particular sampling units designated in Step 1, by following the rules for selecting the sample and for forming the estimate X. This list of all the possible sample-results would in practice be miles long. Thus, for the simple random selection of 100 units from 1000 units formed by Step 1, there would be $\binom{1000}{100}$ possible sample-results—a number that we can write easily in factorials, but too incredibly huge to write in any other way.

Suppose that we were to make up a distribution of the values of X derived or derivable from all the possible samples that we could form, according to the rules, from the sampling units designated in Step 1. In practice, this distribution will be practically normal. Anyhow, it possesses a mean EX and a standard deviation σ_X.

The mean EX is the *expected value of the prescribed procedure*; it is the mean of all the possible samples permitted by the procedure. (A single sample does not have an expected value; it is the *procedure* that has an expected value.)

The standard deviation σ_X is known as the *standard error of the prescribed procedure*—again, not of a single sample, but o all the possible samples permitted by the procedure.

If $EX = A$, our *procedure* is unbiased. In practice we often use procedures for which this equality does not quite hold. In other words, we sometimes use a formula that gives an estimate X' for which $EX' \neq A$ exactly. Such an estimate is *biased*. However, we never use such a formula unless we know in advance that EX' and A differ at most by some small amount less than trivial. It is important to note that the word "bias" here refers to the procedure of sampling, the selection of the sampling units and the formation of the estimate X. It does not refer to biases such as nonresponse, or misinformation; these kinds of biases are not attributable to the sampling.

It is an exciting fact that although we can not tell in advance of the complete census whether our one sample-estimate X lies to the right or to the left of EX, we may nevertheless, from our one sample, if it is big enough, estimate pretty accurately what σ_X is. In other words, a single large sample enables us to calculate the "spread" (more strictly the standard error) of the distribution of X. Knowing σ_X, or having a good estimate thereof, we may use the normal integral to determine approximately what proportion of all possible samples will fall beyond any specified magnitude of error to the right or to the left of the result A of the complete census. For example, we know that only about 2% of all possible samples will fall beyond $2\,\sigma_X$ in either direction from the result of a complete census carried out with the same procedures and care as were exercised on the sample, and that practically none will fall beyond $3\,\sigma_X$.

In practice, we of course very often draw the sample of areas first, cover only them, and omit to cover the areas not in the sample. This is usually what we mean by a *sample survey*. But the interpretation of the standard error is most straightforward if we adopt the above plan of definition, whereby a sample is a portion (up to 100%) of a complete census. We do not then commit the error of confusing a sampling error with other kinds of error.

It is not important to give the actual formulas here for estimating σ_X. Formulas for particular sample designs are known and used the world over. However, it may be desirable to make a few remarks concerning the use of any formula. First, to each procedure of sampling there corresponds a proper formula or set of formulas for estimating the standard error σ_X, and for calculating in advance the size of sample required to produce a certain desired standard error σ_X (such as 10% or 5%). The formula for estimating σ_X for one sampling procedure can not be expected to provide an estimate of σ_X for some other procedure.

Second, it is assumed that the farms in the sample are canvassed and interviewed as thoroughly as would be necessary in a respectable complete census, and that the estimates (X) are formed according to some particular theory of estimation that leads to a calculable standard error. A sample is not a convenient chunk of the whole, nor a sample merely attempted and not followed through.

Third, a small standard error does not signify that all the procedures were good. Thus, a small standard error does not tell us that the method of interviewing was good, or that the questionnaire was drawn up skillfully, or to the purpose.

Fourth, a small standard error attached to a sample does not imply that a second complete census would yield results in close agreement with the first complete census. It does imply that repeated samples *drawn from a census already taken* would be in good agreement.

Fifth, the above interpretations of the standard error hold whether the interviews of the questions are fine or coarse, objective or subjective, factual, opinionative, or dogmatic: and whether the work is carried out expertly or crudely. It is as if we had: *a*) tied a label to every unit, in the entire region, to show the result obtained in a complete census carried out by whatever procedures of interviewing or questioning are considered appropriate for the study; *b*) then had drawn a sample of these labels by the procedure of sampling prescribed; and *c*) formed the estimate X from the sample.

Thus, the principles and methods of sampling various kinds of material (human populations, farms, business and manufacturing establishments, manufactured product, records, physical materials, etc.) are all fundamentally the same.

A sixth remark is so important that we place it now in a separate section.

The standard error does not measure the reliability of a forecast

We shall learn how to design a sample so that it will estimate within (e. g.) 2 per cent, or closer if desired, what would have been the result of applying the same procedure of measurement or interview to every sampling unit in the frame. That is, through sampling, we can discover to within (e. g.) 2 per cent what would have been the result of asking every person in some region, by use of the same procedures and care as were exercised on the sample, how many families in an area own the homes that they are living in, how many homes are mortgaged, how much certain types of families spent for variuos items of food and clothing, how much wheat they raised last year, and how many acres were in wheat, and how many families intend to purchase certain items of household equipment next year. Yet the sampling error of such a survey is no measure of the reliability of someone's prediction of how many families will own their homes five years hence, what they will spend for various items of food, how many men will be employed or unemployed a year hence, etc. Often the sampling error was confused with a measure of the validity of a prediction of the condition of the frame a year hence. The distinction is important.

Good sampling is essential for good prediction but it is no guarantee of good prediction. Prediction of what people will do, or what a crop will be, even on the basis of a complete and perfect census, can fail for many reasons—unreliable methods of prediction, failure to understand the questionnaire and the information that was obtained, or because of unforeseen events, such as people changing their minds, and for many other reasons.