TECHNIQUES OF SAMPLING AND STATISTICAL EVALUATION FOR CENSUS WORK

Eli S. Marks, Case Institute of Technology

The past 20 years have witnessed a virtual revolution in techniques of census-taking and tabulation. On the mechanical side, we have the extensive use of electronic data-processing machines by a variety of countries and the use of mark-sensing devices in the United States and Canada to eliminate the expense and error associated with punching and verification of cards. While these developments have very important implications for census work, the equipment involved is rather expensive and not readily available in industrially undeveloped countries. While some of the smaller, non-industrialized countries may look to electronic equipment for the solution of their census problems, most such countries must continue to rely for many years to come upon hand and punch-card tabulation techniques.

The restriction of available mechanical devices to those of 20 or 30 years ago does not, unfortunately, mean a restriction of census problems to those of 20 or 30 years ago. For one thing, almost all of the smaller countries have experienced a considerable population growth - in the case of the newer countries, one could appropriately consider the rate of growth infinite! More important, however, is the increased demands placed upon a census. Public and private agencies in almost all countries have become aware of the importance of census data as a basis for action and the demands for more, for more timely, and for more complex, statistics have increased exponentially while the available personnel and facilities have increased (at best) arithmetically.

Along with the increased demand for statistics, there has been of recent years an increasing sophistication with respect to the pitfalls of inaccurate statistics. In general, this sophistication has not kept pace with the increasing demand and it is not unusual for the hard-pressed census statistician to be confronted with requests for "figures - any figures, deliver them yesterday; and never mind these fool questions about how accurate they should be; we must, of course, have perfect accuracy!" In spite of this type of behavior, most "producers" of census statistics, and an increasing number of users of census statistics, are conscious of both the unattainability and the superfluity of "perfect accuracy." We may yearn for the "good old days" when a census was accurate by definition (and official decree) but most of us are willing to face up to the realities and are unwilling to deliver statistics of unknown but dubious accuracy, regardless of the pressures involved.

In the situation of increasing demand for both quality and quantity of statistics and of sharply limited resources, the producer of census data must turn more and more to modern techniques of sampling and evaluation -- sampling to meet those demands which can be met and evaluation (if for no other reason) to counter the "unmeetable" demands! Even the large and industrially developed countries are relying more and more on these techniques and for the smaller undeveloped countries they are indispensable. Furthermore, even those countries which can afford the latest model of digital computer that high-pressure salesmanship can supply, are discovering (or will discover) that electronic gadgets are a supplement to, rather than a substitute for, human ingenuity and sound statistical methods.

In the search for solutions to the problems of demand, other countries have looked, of course, to the United States. Fortunately, in the area of statistical developments, the United States has retained its pioneering vigor and can offer more than mechanical devices and public relations "gimmicks." In particular, the work of the United States Census Bureau in the area of sampling and statistical evaluation, has provided a model which can profitably be studied even by countries which may be more advanced in other statistical fields.

The application of one country's methods to another country's problems is, however, neither simple nor painless. Mistakes have been made in a too-literal following of United States instruction manuals and statistical texts. In the light of these mistakes, the experience with applying United States developed sampling and evaluation techniques to the Chilean censuses of 1960 should prove illuminating - both for its indication of what to do and its indication of what not to do. My own experience with the Chilean censuses related primarily to two problems and my remarks are, therefore, directed primarily at these two problems.

The two problems with which I was concerned in Chile were (a) the development of a measure (or of measures) of the completeness of the census coverage and (b) the selection of a sample to provide rapid and (reasonably) accurate advance tabulations of the census data. The latter problem is a particularly difficult one for all but the handful of countries which can afford the newest and fastest electronic data-processing equipment. And some of these latter are discovering in 1961 and 1962 what the United States discovered in 1950 and 1951 -- that having high-speed equipment and having the ability to use equipment at high-speeds are not the same thing.

Since a paper on the evaluation of the Chilean censuses of 1960 is to appear in the near future, I shall confine my remarks to the problem of sampling for advance tabulations. The general lessons to be learned from this problem are applicable also to the problem of evaluation although there are, of course, techniques specific

to each problem.

The problem of selecting a sample for advance census tabulations is -- in its design aspects -- an extremely simple one. A complete sampling frame exists and the sampling theory is straightforward and well-developed, since concern is entirely with estimating the results of the census and not with determining some "true value" which may be undefined (or even undefinable). But even during the stage of sample design, both the Chileans and I were (fortunately) aware of the problems of carrying-out a sample design. I say, "fortunately," because other countries have tried to use samples designed purely from a sampling cost and variance standpoint without considering the costs and biases of carrying-through on the sample design.

The Chileans had had an experience with the problems of carrying through a sample design in connection with the sample for advance tabulation of their 1952 censuses of Population and Housing. There a well-designed sample came to grief in execution and ended with a bad bias, whose correction required about two years and the services of a top United States Census Bureau sampling expert.

The problem in 1952 was that the clerks selecting the sample (a 2% household sample selected systematically for the whole country) tended to avoid the larger households, substituting the schedule immediately preceding or following. This type of bias is somewhat hard to detect without special control measures and was, in fact, not detected until the sample for the entire country was selected and punched and the preliminary totals compared with hand counts made previously.

The experience of the 1952 census was recent enough to provide a vivid object lesson. My own indoctrination prior to going to Chile had included a first hand briefing on the incident and many of the 1960 personnel of the Chilean statistical office had been there also in 1952. The head of the processing section remembered quite vividly the disorganization of his own work attendant upon the belated discovery of the bias and was determined to avoid a similar fiasco in the 1960 censuses. Quite independently of me, he hit upon one leg of what I think of as the tripod on which a sound sample must rest -- simplicity -- the other two being good design and good control. Unfortunately, in stressing this leg of the tripod he sawed off the "good design leg" although his basic analysis of the situation was sound.

The processing section head's proposal was that we sample entire enumeration districts (or, in the local terminology, enumeration zones). Here the problem of keeping track of what was selected and of avoiding substitutions and other biases would be greatly reduced in magnitude. However, although the Chilean enumeration districts average about one-tenth the size of United States enumeration districts (i.e., average about 20 households), the intraclass correlation is quite large and, on some characteristics, might mean that an enumeration district sample would have to be 5 to 10 times as large (in terms of number of persons) as a sample of households in order to attain the same accuracy. For that difference in sample size we could buy a very adequate third leg -- good control -- to stabilize our tripod of good design and simplicity.

The preceding statement about sample sizes is deceptively easy to make now but the difficulties of arriving at this conclusion illustrate the problems of applying for the first time in a country, techniques which are so "obvious" to "experts" that they never bother to explain them in detail! On my first visit to Chile, in March-May of 1960, the question of sampling enumeration zones or households was raised (as noted above). Before trying to answer this question, I naturally wanted to examine the variances and sample sizes involved. At this point I discovered that there were no enumeration zone totals available from the previous census and there were no distributions of households or families by size or other characteristics. We laid out, therefore, a program of preparing distributions of enumeration zones and households by size and numbers of males and females for four provinces (the major administrative subdivisions of the country).

Since these data had not previously been entered on punch cards and since the punching and tabulation facilities of the statistical office of Chile were overburdened already with other work (current surveys and census pre-test tabulations), the tabulation of the desired distributions was set up as a hand tabulation, with two of the office's limited number of clerks assigned to the job.

It developed that some of the materials for parts of two of the four selected provinces could not be located and these areas had to be omitted, curtailing further the already limited information which would be available for making a critical decision. In addition, I failed to point out the "obvious" fact that "open-end" intervals could introduce serious errors into a variance computation if they contained any appreciable number of cases. I also forgot the obvious fact that clerks must have tabulation intervals specified in advance and cannot be expected to notify the statistician in charge that there is a large number of cases in an open interval. By the time I discovered the problem, it was too late to do anything except try to make allowances for the range of possible error.

Eventually the decision was made to sacrifice some simplicity to good sample design and to use a sample of schedules. The requirements for sampling indicated different sampling rates would be desirable for different provinces since separate estimates by province were to be prepared and the provinces vary considerably in population. To balance these complexities, it was decided to use a systematic sample with separate starts in each comuna (equivalent to United States county). The system was to number consecutively all schedules of the comuna and then sample those schedules

with (randomly) prespecified ending numbers -- the number of "endings" varying from 2 to 18. The plural random starts, although adopted primarily in the interests of simplicity and ease of sample control, do have the advantage of giving a design which permits unbiased estimates of sampling variance.

All of the above is extremely simple. The formulae for anticipated sampling error and required sample sizes appear in any sampling text and selection of suitable "endings" requires only a table of random numbers -- although we did have to resort to the Inter-American Statistical institute's center at the University of Chile to obtain the latter! The important job was that of "sample control." This is a dull, routine task. It involves no new and exiting discoveries, no fascinating mathematics or other theory. It involves, for example, spelling-out the meaning of "number consecutively all schedules of the comuna" in full (and dull) detail, so that not even the most unskilled and lazy clerk can find an excuse for biasing the numbering -- "continue the numbering from one zone to the next within the comuna; start the first zone with 000-001; do not omit or duplicate any number, etc."

Sample control involves such trivia as checking that the highest number assigned agrees with an independent hand-count of the schedules and instructions on the procedure to be followed if it does not -- "look for omitted and duplicated numbers, assign duplicates the lowest unassigned number, assign omitted numbers to the last schedule(s) cancelling the number(s) previously assigned to it (them)," etc. These are "idiot" instructions but are unfortunately, necessary and cannot be replaced by the assumption that the use of discretion by the clerks will be equally satisfactory even though this assumption may sometimes be true.

In addition to the checks on numbering and the simplification of the sampling process, the sample control included a comuna-by-comuna check to assure that number of schedules, number of households (different from number of schedules since some households had two schedules), and number of persons were within sampling error bounds of the expected values. For the number of schedules and number of households, binomial sampling error formulae were applicable but the variance of numbers of persons had to be estimated. This estimate was obtained from the 1952 size of household distribution mentioned above. The controls described were implemented by a form on which population and sample figures were entered and which provided formulae for computing the sampling error "tolerance limits" and space for entering the results of each step in the computation.

All of this may seem like "idiot" procedure. Unfortunately, experience demonstrates conclusively its necessity. I understand that a very similar sampling for the Argentinian census ran into difficulties due to omission of some of these controls. Furthermore, the comuna-by-comuna checks in Chile pointed-up a defect in the sample design. This defect had to do with the sampling of institutions. The first schedule of an institution was the regular Population and Housing schedule with space for twelve persons. Continuation sheets, however, had space for listing 66 persons. In the design of the sample we considered drawing a sample of persons within institutions instead of a sample of schedules but rejected this alternative in order to keep the sampling procedure simple and uniform.

The sample checks for the first province completed, indicated several comunas where failure of the sample to satisfy the estimated "tolerance limits" could be traced to the large variance introduced by sampling whole schedules in institutions. Steps should have been taken immediately to reduce this variance by modifying the sample design. Unfortunately, there were delays in communication so that remedial action was delayed until it was too late to remedy more than a part of the problem. However, the mechanism for detecting trouble was present and this is the first half of avoiding trouble!

Another problem in the use of United States sampling techniques in the Chilean context centered in the estimation of sampling errors. Before leaving Chile I spelled-out the procedure for computing these estimated sampling errors. The defects in my assumptions did not become painfully apparent until seven or eight months later when people tried to follow my instructions. Then I received a letter which stated (among other things): "On calculating the sampling errors for housing, I have no problem. For population errors, difficulty arises because the process is a very long one, people-and-time-consuming, and I am very short in both respects...We have no tabulator with summary punch."

Again I had applied assumptions based on my United States experience in a situation where those assumptions were inaccurate. I had assumed that a job which was easy in the United States would present no difficulties elsewhere, forgetting that a "trivial" job can become gigantic if one lacks equipment, personnel, and money.

The conclusion to be drawn from the above experiences is not that United States or Canadian census experience is inapplicable to other countries. On the contrary, I found a great many unanticipated similarities to United States experience in the Chilean setting -- including a hot controversy raging on my arrival about the utility of the inquiry about "condition of dwelling unit" which made me feel right at home! Furthermore, almost all the principles of sample and survey design that I learned in the United States were directly applicable to the Chilean situation. What are not directly applicable are the specific procedures developed in the context of a highly mechanized economy. A brief consideration of the realities of the Chilean situation indicated that there was a solution to the problem of calculating sampling errors (from United States experience) which was applicable to Chile. What is needed is then the use of advanced techniques but with the necessary adaptations to local conditions. Above all, we need good design and simplicity and good control in the United States, in Chile, or in Ghana.