CONTRIBUTIONS TO STATISTICAL METHODOLOGY FROM THE FEDERAL GOVERNMENT

Barbara A. Bailar, American Statistical Association

1. INTRODUCTION

The Federal government has led the way in the development of statistical methodology in censuses and surveys. I will confine my remarks to examples from the Bureau of the Census and will discuss four main areas of work -- the development of sampling methods, non-sampling error, seasonal adjustment, and the development of methods to protect the confidentiality of respondents, usually called disclosure avoidance techniques. Finally, I will venture to hazard some guesses about future development.

2. SAMPLING

The story of sampling in the Federal government is primarily the story of a remarkable group of people at the Census Bureau, led by Morris Hansen and William Hurwitz. When one considers that the Census Bureau was committed to probability sampling in the early 1940's, one wonders: how could an innovation of this type have occurred so quickly in such a conservative institution? The adoption of innovative methods often takes a very long time and I suspect the Bureau is much slower in adopting and promoting new methodology today. Hansen has given three reasons why he thinks sampling was accepted relatively quickly by the subject-matter divisions of the Bureau [5]. They are: 1) support from the top, 2) conscious development of a team-work approach with the subject-matter divisions, and 3) the development of a corps of sampling experts (later, methods specialists) in the subject-matter divisions who were responsible to the Statistical Research Division (SRD) on technical matters. I think he left out one key ingredient and that is the force and the spirit of the dynamic duo and their cohorts.

In 1936, the Bureau began exploration of sampling and potential applications. Some sampling was already in use, but not probability sampling. There was judgment sampling and sampling of some large establishments. However, there was little or no theory to guide sampling approaches. In 1939, Congress authorized a national voluntary registration of the unemployed and partially employed. A questionnaire was to be delivered by the Post Office to every household. There was some concern that this voluntary registration could have some bias, so an enumerative check census was put in place in a sample of areas. The check census required interviewing all households within a probability sample of postal delivery routes. The mail carriers did the interviewing and identified and sorted the voluntary mail returns. They then provided separate counts for each postal route, including the sample postal routes. This then gave an independent variable to use in the estimation, one of the earliest demonstrations of ratio estimation. The results of the check census were convincing on the usefulness of sampling. However, the entire effort was remarkable in many ways:

- the effects of nonresponse from a voluntary census were anticipated;
- the use of ratio estimation; and
- the speedy results.

Hansen, in an interview in Statistical Science [10], reports that the registration took place the week of November 20, 1937; that the household canvas was done during the week of December 4, 1937; and preliminary results became available on New Year's Eve, 1937. I don't think the Census Bureau could beat that record now.

Hansen attributes the success of the 1937 enumerative check census as a demonstration of the use of sampling as key in gaining acceptance within the Bureau. Before then, Bureau staff believed that complete coverage was necessary and that sampling would discredit the Bureau. The success of the study helped gain the acceptance of sampling in the 1940 census, the first census in which some questions were asked of only a sample, not the entire population. Unfortunately, in the last few months, some at Census have dragged out the old chestnut about needing to do the vacant delete check on a 100% basis because a census has less error than a survey [6]. Let's just assume that was a temporary aberration caused by litigation.

A great deal of the theory of sampling was developed in conjunction with the Labor Force Survey. The Works Progress Administration (WPA) sponsored a survey to measure unemployment. In 1942, when the WPA was abolished, the survey was moved to the Census Bureau. The sampling procedures were evaluated and many improvements were made. Several important contributions to sampling theory came from that revision. Some of the sampling principles introduced into the 1942 revision were: enlarged primary sampling units, sampling with probabilities proportionate to a measure of size, and area substratification. These principles were discussed in a 1943 paper by Hansen and Hurwitz in the Annals of Mathematical Statistics [7]. Rereading this paper, "On The Theory of Sampling From Finite Populations," always provides new insights. The article seems to be the first published by Federal employees on the topic of sampling of finite populations. Though the concepts had been discussed by others, the extension of theory was new. Also, a hallmark of Hansen and Hurwitz, the results were discussed in a series of practical comparisons highlighting the advantages of the recommended procedures.

Improvements in the Labor Force Survey continued over the years. Composite estimation, using the system of sample rotation to improve the estimates, was introduced. The Current Population Survey, as the Labor Force Survey is now called, has undoubtedly led the way throughout the world in setting the standards for a labor force survey.

Surveys of business establishments presented new sampling problems, also undertaken by the Statistical Research Division. The attitude frequently encountered was that sampling might be all right with relatively homogenous populations such as people but they would not work with highly skewed populations such as businesses. Working with the acknowledged skewness of the population, the sampling group stratified the retail stores by size. The largest stores were necessarily included in the sample, and the smaller businesses were sampled with probability proportionate to a measure of size.

It was also apparent that businesses came into being and died frequently. A static sample would not be able to capture this turnover. Therefore, an area sample to provide estimates for new stores was incorporated. The Monthly Retail Trade Survey has seen many innovations, but these basic cornerstones remain. The Retail Trade survey also makes use of composite estimation to provide more precise estimates.

Many other instances of sampling innovations could be mentioned. Many descriptions are given, and the theory and practical applications are described in the 1953 book Sample Survey Methods and Theory in two volumes, by Hansen, Hurwitz, and Madow [9]. Though the illustrations are seriously outdated, the books still provide more practical sampling applications than any other books I know of. I only regret that they were never updated.

3. NON-SAMPLING ERROR

Another major advance in sample surveys and censuses was to look beyond sampling error to try to control the errors arising from other sources, such as the interviewers, processors, questionnaires, and so forth. Hansen and Hurwitz moved in that direction before the 1950 Census, incorporating many experimental studies in the census designed to estimate the effect of measurement errors in the census. Total survey error became a strong focus at the Census Bureau. The measurement and control of non-sampling errors became a regular feature of Census Bureau work.

An impetus to this nonsampling error work was the recognition that measurement errors could have a much stronger effect on data than sampling errors, especially at larger levels of aggregation. Hansen, Hurwitz and Bershad [8] developed an integrated model for censuses and surveys that explicitly incorporated sampling error, response error, and bias. The response error component contained what are now known as a simple response variance and a correlated response variance. The simple response variance reflects the basic trial-to-trial variability that arises from differences in respondent reporting, different respondents, different interviewers, and the like. The term has also been generalized to include the variance that arises from trial-to-trial variability in coding. The correlated response variance refers to the variance that arises from a factor that pushes responses into a certain pattern. The most studied factor is that of the interviewer. By having certain expectations or from experience interviewing at a few households, the interviewer can push responses into certain categories. We see wide variability among interviewers working in the same areas on nonresponse rates, on questions about educational attainment, and many other items.

This model was first tested in the 1950 census and was a major factor in the decision to move from an "enumerator census" where an interviewer went to every household, asked the questions, and recorded the answers, to a "mail census," where the questionnaires are sent to every household and householders are asked to fill out the forms and return them by mail. Experiments in the 1960 and 1970 censuses show a large reduction in this variance component when self-enumeration is used [11, 12].

In addition, Hansen and Hurwitz encouraged work on coverage error. The Census Bureau has invested a large amount of time in investigating the effects of coverage error, both in censuses and surveys. After the 1950 census, using a model developed by Ansley Coale at Princeton University, the Census Bureau was able to measure the amount of undercounting in the decennial census at the national level, by age, race, and sex. This method, known as demographic analysis, showed that there was a differential undercount that affected blacks much more severely than whites [4]. In addition, the Census Bureau started development of a post-enumeration survey to learn more about the uncounted population. At first, the Bureau relied on a "do-it-better" approach, but in recent years has turned to a "do-it-again" approach. This latter emphasis will be used in the 1990 census. Similarly, coverage losses in surveys spurred work on ratio estimation procedures that would dampen the effect. Most Bureau household surveys use those procedures.

The Bureau of the Census is now well-known for its work on measurement error. In addition to work on response error and coverage, it has encouraged work on time-in-sample biases that affect the estimate from surveys in which respondents are contacted more than once.

The labor force survey, in which respondents are kept in sample four successive months, dropped for eight months, and then contacted for four additional months, has been carefully studied. Bailar [1] showed the difference between the higher estimates of employment and unemployment for those in sample for the first time and those in sample for later times. These differences affect the levels of employment and unemployment, though probably not the estimates of month-to-month change.

These are only a few examples of the work begun at the Census Bureau on measurement errors. Now work is carried on at all the statistical agencies.

4. SEASONAL ADJUSTMENT

The history of seasonal adjustment in the government began with the efforts of Julius Shiskin when he was at the Census Bureau. He was responsible for introducing computerized seasonal adjustment. Now the X-11 method is used around the world.

According to Julie Shiskin, in the 1950's the Federal agencies were under pressure from the

Council of Economic Advisors to produce seasonally-adjusted time series. The Census Bureau got the first electronic computer dedicated to data processing, the UNIVAC I, in 1953 and Julie heard a lot about how difficult it was to program from Eli Marks who was in his car pool. It dawned on Julie that the computer could be used for making the seasonal adjustments, so he checked with a computer technician and found that it would take 1 minute to do a 10-year series. Of course, it takes less than that now.

Seasonal adjustment is still somewhat of an art form, since the X-11 program provides so many options, and the analyst can choose among them. However, there was skepticism at the beginning of this computerization about whether a machine could do what a skilled technician could. Julie decided to challenge the Federal Reserve Board. He proposed that they take any series and spend as much time as they wanted adjusting it. Then he would run the same series through the computer. Both series would be plotted and given, without identification of who did the adjustment, to a small, very distinguished group at the Federal Research Board who would judge the results. The result was a unanimous decision that the computer method was superior.

The government now seasonally adjusts thousands of time series annually. Model-based methods, because of computer limitations, seemed impractical for many years. Also, new seasonal adjustment factors were developed every year, based on historical experience. For example, a factor to be used in the computation of the seasonally adjusted figures for July would be developed in December of the preceding year. No new data based on more recent events were allowed to influence the adjustment. This made sense when it took several days to prepare punch cards and run the series. But within the last ten years, that method received more criticism, and the method of concurrent seasonal adjustment was promoted. The time series staff at the Census Bureau, led by David Findley, did a thorough investigation of the merits of concurrent seasonal adjustment on Census Bureau series, and led the way for the adoption of that method by the Bureau.

The time series staff has also asked some very key questions that are central to seasonal adjustment. First, what kind of standard exists to judge whether or not a series should be seasonally adjusted? Second, given that there are several methods for adjusting time series, how do you evaluate the different methods? In a key paper, Bell and Hillmer [2] question the need for seasonal adjustment if series can be adequately modeled. They also describe some criteria for evaluating seasonal adjustments. I must be quick to point out that the Census Bureau is not the only government agency that has done ground-breaking work in this area. In fact, one very useful accomplishment of the time series staff at the Census Bureau is to hold regular meetings of interested and involved experts throughout the government. Thus, people at the Federal Reserve Board, Bureau of Labor Statistics, Energy Information Administration, and the Bureau of Economic Analysis, to name only a few, all participate and keep up-to-date on new developments. Estella Dagum at Statistics Canada has led many very successful efforts, including the development of the X-11 ARIMA method.

5. DISCLOSURE AVOIDANCE

Whether or not one agrees with the Census Bureau on its policies about keeping data confidential, one must agree that the Bureau has promoted disclosure avoidance techniques to protect data. Disclosure avoidance is an attempt to protect the answers of individual respondents. It has long been a problem in censuses, but is also a problem in surveys, especially surveys that are longitudinal in nature or where records exist that could be linked to the survey results.

Disclosure avoidance problems in the population censuses focus on disclosures that would occur from the publication of very small frequencies. These small numbers lead to the potential identification of single respondents or small groups of respondents. In addition, zeros in cells may also lead to disclosure. Disclosure in frequency tables is usually defined in terms of a threshold rule that states that disclosure occurs if, given any tabulation cell X, one can infer that the number of respondents in X is less than a predetermined threshold value. In 1980 decennial census publications this predetermined threshold value was defined separately for households and persons.

Methods for controlling disclosure in frequency count tables fall into three categories: suppressing all values, perturbing cell values, and replacing numeric cell values by intervals. Cell suppression insures that numeric values are not given and that inferences cannot be derived from manipulation of linear relationships between unpublished and published cell values. Data perturbation means adding or subtracting a small amount from most cell values so that inferences regarding the tabulated values cannot be made with certainty. The third method, replacing point estimates by intervals, is not useful for many data users for cross-classifications.

Cell suppression was the main technique used by the Census Bureau through 1980. Additive restraints along rows and columns of the table generate a series of linear constraints. Once the primary disclosures have been suppressed, mathematical programming is used as a disclosure audit on the table. Though this method was used on an ad hoc basis for years, Cox and his colleagues at the Census Bureau derived the mathematical underpinnings [3] and showed how complex cell suppression actually was.

Data perturbation methods, including random rounding, have been developed and used in the United Kingdom, Sweden, and Canada. All of these methods depend on adding or subtracting a small value, sometimes zero, from table cells, with a specified probability.

For data such as sales, value, inventory, and financial information from manufacturing and retail establishments, the Census Bureau is concerned about being able to identify the amounts from respondents. If a competitor reviews a tabulation and subtracts the amount for his firm, the amount for another respondent may be identified. Cell suppression techniques are used. The so-called (n, k)-rule, which states that X is a disclosure cell if a fixed number of respondents n account for more than a fixed percentage of k of the total cell value, belongs to a class of cell dominance rules, all of which are additive.

Disclosure-avoidance work is going on all over the world, primarily in government offices. No doubt this reflects the fact that these offices have serious problems that have been pushed to the fore by the demand for microdata.

6. A LOOK TO THE FUTURE

All four areas presented so far have relied on the development of mathematical models. Sampling, of course, relies on randomization methods, but the control of total survey error led to the formulation of a survey error model, first described by Hansen, Hurwitz, and Bershad [8]. That model and the experiments used to estimate the parameters were the basis for many policy decisions on the conduct of censuses and surveys.

Time series models are used widely around the world, replacing empirical methods such as the X-11. Researchers are now urging that time series methods become integrated with survey estimation methods to produce more accurate results. It will be interesting to observe how or whether this melding will take place.

Another area of active modeling within government agencies is to produce small-area data. Data are often collected for larger areas of aggregation, such as states, and then data needs are expressed for smaller areas, such as counties. Conferences have been held comparing and evaluating different techniques for producing small-area data. The Census Bureau used empirical methods to develop population estimates during decade. Several models were explored as part of the undercount research at the Census Bureau, and much was learned about the problem.

Ad hoc methods for editing and imputation are now being carefully scrutinized and mathematical models are being developed. We shall undoubtedly see more modeling of this type in the future.

Thus, the future, as I see it, will be a further expansion of models. This is not to denigrate the empirical methods used now. Statisticians have always recognized that theory and practice go hand in hand. Empirical methods that seem to work lead to modeling and theoretical developments that are tempered by practical experience. The government agencies have many fascinating statistical problems that will lead the way, as they have in the past, in certain areas of statistical methodology.

REFERENCES

[1] Bailar, Barbara A. "The Effects of Rotation Group Bias on Estimates from Panel Surveys," Journal of the American Statistical Association, vol. 70, March 1975, pp. 23-30.

[2] Bell, William R. and Hillmer, Steven C. "Issues Involved with the Seasonal Adjustment of Economic Time Series." Journal of Business and Economic Statistics, vol. 2, no. 4, October 1984, pp. 291-317.

[3] Causey, Beverley E., Cox, Lawrence H., and Ernst, Lawrence R. "Applications of Transportation Theory to Statistical Problems." Journal of the American Statistical Association, vol. 80, no. 392, December 1985, pp. 903-909.

[4] Citro, Constance F. and Cohen, Michael L., (eds.). The Bicentennial Census, National Academy Press, Washington, D.C., 1985.

[5] Duncan, Joseph and Shelton, William. Revolution in United States Government Statistics, U.S. Government Printing Office, 1978.

[6] Hansen, Morris H. "Some History and Reminiscences on Survey Sampling." Statistical Science, vol. 2, no. 2, pp. 180-190, 1987.

[7] Hansen, Morris H. and Hurwitz, William N. "On the Theory of Sampling from Finite Population." Annals of Mathematical Statistics, vol. 14, 1943, pp. 333-362.

[8] Hansen, Morris H., Hurwitz, William N., and Bershad, Max A. "Measurement Errors in Censuses and Surveys." Proceedings of the International Statistical Institute, vol. 38, 1961, pp. 358-374.

[9] Hansen, Morris H., Hurwitz, William N., and Madow, William. Sample Survey Methods and Theory, vols. 1 and 2, New York: John Wiley and Sons, 1953.

[10] Olkin, Ingram. "A Conversation with Morris Hansen." Statistical Science, vol. 2, no. 2, pp. 162-179, 1987.

[11] U.S. Bureau of the Census, Series ER 60, No. 7, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: The Effect of Interviewers and Crew Leaders, 1968.

[12] U.S. Bureau of the Census, PHC(E) Series, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1970.