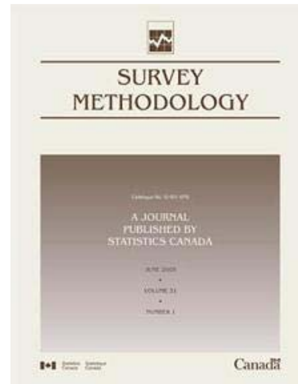
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Survey sampling theory over the twentieth century and its relation to computing technology

by D.R. Bellhouse



June 2000

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 11-20
Statistics Canada, Catalogue No. 12-001

Survey sampling theory over the twentieth century and its relation to computing technology

D.R. Bellhouse [1]

Abstract

Computation is an integral part of statistical analysis in general and survey sampling in particular. What kinds of analyses can be carried out will depend upon what kind of computational power is available. The general development of sampling theory is traced in connection with technological developments in computation. What is possible in theory is only practicable with the proper computing technology. At the same time new developments in technology can motivate new areas of theory to investigate. One hundred years ago, it was the requirements of statisticians that spurred on technological development. Although theoretical developments in sampling theory have often run ahead of computational capabilities, it is now the case that survey statisticians are now followers of computing technology that has been motivated by others instead of acting as the catalyst that leads to technological change.

Key Words: Analysis of survey data; Digital computers; Punch cards; Scientific programming; Statistical software; Survey data analysis; Survey estimation.

1. Introduction

There are several ways to approach the history of survey sampling. Two are very tempting, but will not be followed here. The first is to examine sampling in the context of the history of ideas – who formulated them and then how and why they are formulated, promoted, defended and discarded or supplanted. With respect to the personalities, it is not necessarily the one who espouses the idea first who is given prominence but the one who promotes it the best or the one who can best put the idea into practice. The approach of the history of ideas has been followed to a certain extent by Kruskal and Mosteller (1980) and Bellhouse (1988) who examined the progression of ideas beginning with the espousal of the representative method by Kaier (1897) over censuses combined with the use of randomization in surveys by Bowley (1906). The whole story of the debates over the foundations of sampling falls directly under this approach. From this debate, which was initiated by Godambe (1955), has emerged the continuing question of when to use models in sampling design and estimation. A second way to approach the history of sampling is to look at sampling theory as a branch of mathematics and then to fit this development into the general pattern of how research in mathematics evolves. Complicating this is that there are several approaches to how mathematics evolves, as discussed in Gillies (1992). One approach is to note that periodically there are results which seem to open up new areas of research while other areas become seemingly complete or “fished out” for new research ideas. Emerging areas of research often attract several talented researchers to work on these new problems and away from other potential research problems. This has its parallels in sampling. Hansen and Hurwitz (1943) obtained results on sampling with probability proportional to size and with replacement. Then Horvitz and Thompson (1952) extended this idea to sampling without replacement. The basic problem in unequal probability sampling without replacement is to find a sampling design that yields the desired inclusion probabilities. This resulted in several papers on the subject culminating in the review monograph by Brewer and Hanif (1983). Lately, very few papers are written to promote new without replacement sampling designs that result in inclusion probabilities proportional to a size variable. However, statistics and survey sampling cannot be equated to pure mathematics. Much of statistical research is motivated by practical problems in data interpretation and analysis not by abstract ideas.

In view of the explosion of technology over the 20th century, I chose another approach. This is to view the history of sampling over the 20th century as the history of the interplay between ideas that have been put into practice and computing technology that has defined the limits of practice or that has encouraged ideas for new developments in practice. The development of sampling methods may be categorized by the intersection of two strands: the use of surveys for descriptive and analytic purposes, and whether or not hypothetical models should be used.

2. Beginnings: The first half of the twentieth century

The first two major breakthroughs for survey sampling, one in the formulation of a statistical concept and the other in the development of technology, occurred at the end of the nineteenth century. Both breakthroughs faced some initial opposition or apathy, the idea more so than the technology,

1. D.R. Bellhouse, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, N6A 5B7, Canada.

but both prevailed and were developed further. These breakthroughs were: (1) Kaier's (1895/6, 1897, 1905) espousal of sampling through a "representative method" over attempts at complete enumeration for social surveys, and (2) the development of punch card machines for data processing by Hollerith (1894). Both breakthroughs were directly related to survey or census work. This was the first and last time that survey or census issues inspired major technological innovation. From then on, survey sampling has adapted itself to the available technology.

Kaier's idea was to get a sample that was an approximate miniature of the population. Through sampling, more detailed information could be obtained and more specialized studies could be carried out, all at a fraction of the cost of a census. The idea initially met with opposition and it took upwards of a decade for his ideas to be accepted.

The development of machinery by Herman Hollerith for data processing came directly out of the needs of the U.S. Bureau of the Census and the encouragement of the Bureau's Director of Vital Statistics, John Shaw Billings. The events that led to this development are described by Willcox (1926):

> "While the returns of the Tenth Census [1880] were being tabulated at Washington, Billings was walking with a companion through the office in which hundreds of clerks were engaged in laboriously transferring items of information from schedules to the record sheets by the slow and heart-breaking method of hand tallying. As they were watching the clerks he said to his companion, 'There ought to be some mechanical way of doing this job, something on the principle of the Jacquard loom, perhaps, whereby holes on a card regulate the pattern to be woven.' The seed fell on good ground. His companion was a young talented engineer in the office who first convinced himself that the idea was practicable and then that Billings had no desire to claim or use it. Thereafter he devoted the bulk of his life with great ultimate profit for himself and the world to ripening the invention and securing its adoption. I have no need to describe or eulogize Hollerith machines."

A full description of the development and use of these machines for surveys is given in Mandeville (1946). Hollerith's machine was applied to processing the 1890 U.S. census. While the 1880 census took over seven years to complete, the 1890 census was finished by early 1895. The Bureau used 180 tons of cards that were processed at a speed of 6,900 cards per 6½-hour day. Not only did the machine save time, it also significantly reduced tabulation errors. The punched card machine was used to process the 1891 Census of Canada, but it did not see early use in the censuses of the United Kingdom and the rest of the British Empire. It was felt that the level of detail required in these censuses did not justify the use of a Hollerith machine since the time saved by the machine would be balanced by the time taken to punch the card (Hooker 1894). In a paper on census taking Baines (1900) expressed a preference for manual over machine tabulation, especially when labour was cheap. Despite these initial misgivings, improvements to the machine continued and the use of the Hollerith machine for statistics became highly developed by mid-century. Hartley (1946) demonstrated the most sophisticated use of these punched card machines for statistical analysis. This included the calculation of moving averages and serial correlations as well as the solution of simultaneous equations on Hollerith machines.

After these near simultaneous and unrelated innovations in ideas and technology, theory ran ahead of practice for the next 50 or 60 years. Theoretical developments in sampling continued through the first half of the century. Out of discussions over the path to follow in the "representative method," Bowley (1926) put together a monograph describing all the known theoretical results in sampling under random selection and under purposive selection. In addition, he developed the theory for stratified sampling under proportional allocation. The triumph of randomization over purposive selection was due to Neyman (1934) who showed why randomization gave a more reasonable solution to sampling problems than purposive selection. Although not the first to do so, he also developed optimal allocation strategies for stratified sampling. Prior to the middle of the century the last major development, in terms of sampling design with accompanying estimates and variance estimates, was the concept of unequal probability sampling introduced by Hansen and Hurwitz (1943).

The practical implementation of these theoretical results was limited to relatively small-scale surveys. The analyses for most surveys used calculators, either electric ones such as those manufactured by Friden, Marchant or Monroe, or hand calculators operated by turning a crank such as the Brunsviga used by Pearson and the Millionaire used by Fisher. Since the labour in the analysis increased significantly with the sample size, standard errors were seldom calculated, and when calculated the correct formulas were seldom applied. Bowley (1936) describes a typical situation showing the infrequency of standard error calculations:

> "Tabulation is usually a dull and tedious job, but there is a certain interest in watching the entries accumulating in a cross table and seeing the gradual growth of continuity out of randomness. When the results take the form of a frequency curve, and especially if we have reason to expect a normal curve and find it, we have good reason to suppose that we have measured satisfactorily a real entity. Thus the distribution of price changes or their logarithms on a normal scale gives a great deal of support to the validity of an index number. In such cases the computation of standard error is reasonable."

Box and Thomas (1944) describe a survey of approximately 4,500 respondents stratified by the industry in which they worked. The standard errors, when presented, were calculated using the formula for simple random sampling. A decade later Deming (1956) noted:

> "Although the possibility of showing a valid standard error is by definition a feature of any probability sample, it is a fact that results of probability samples have too often appeared in the past without standard errors because of the sheer labor of computation."

It is within this context that Mahalanobis (1946) suggested the technique of interpenetrating subsamples. This technique, which Mahalanobis developed at the Indian Statistical Institute in the 1930's (Murthy 1967 and Deming 1956), is very simple: two or more independent subsamples are chosen according to the same sampling design. Then the variation between the subsample estimates of the population total provides an unbiased estimate of the variance of the final estimator of the total. Computationally, the method has distinct advantages in the punch card environment where sums are easier to obtain than variances. With interpenetrating subsamples the main computational effort is in finding the subsample estimates that are based on sums only. The Indian Statistical Institute obtained its first Hollerith machine in 1944. Prior to that time, tabulations and other calculations were done by hand. The Institute's Annual Report for 1945-46 published in *Sankhyā* shows the initial unease that always greets technological change and the eventual positive benefits to change. With respect to the introduction of these machines, the report states:

> "Contrary to apprehensions among certain sections of workers that the Hollerith machine would to a large extent eliminate manual computations, it was found that new and detailed studies which could not be formerly undertaken could now be handled without difficulty so that the demand for trained computers in the later stages was on the increase. In addition to routine projects undertaken from time to time, special studies such as mechanical solution of determinants, construction of tables, fitting of orthogonal polynomials, *etc*. were conducted."

In the United States, Deming (1956), for example, picked up on the general idea and put forward methods of replicated sampling. The U.S. Bureau of the Census used this method for variance estimation. At the Bureau this idea evolved into pseudo-replication, or eventually balanced repeated replication, for variance estimation (McCarthy 1969).

3. The advent of the digital computer

The initial development of the digital computer was for military purposes during the Second World War (Ceruzzi 1998). For some years after the war the military continued to play a central role in the advancement of computing. By the 1950's commercial uses were developed for the computer, and this is where sampling practice begins to catch up with sampling theory. The first generation of commercial computers included the UNIVAC followed by the IBM 700 series. These computers contained thousands of vacuum tubes as internal memory. The tubes for the IBM machine were about three inches in diameter and held 1,024 bits of information. The UNIVAC ran at 2.25 MHz and could carry out 465 multiplications per second. For both machines, data were input via punched cards and stored data was on magnetic tape rather than continued use of the punched cards. The 1961 census in the United Kingdom underscores the continuing central role of the military in computing at this point in time. The census was processed on an IBM 705 computer (Benjamin 1961). The computer belonged to the War Office and was used by the Royal Army Pay Corps. The census workers were able to use the computer when not in use by the army. Information was input via cards punched in one location and then taken to the computer in another location.

Although it was not at the forefront of the development of the computer as it had been with the Hollerith equipment, the U.S. Bureau of the Census was central in the initial commercial development of the digital computer. Not only did the Bureau receive the first UNIVAC that was produced, but also some of its employees participated in design decisions for its construction (Ceruzzi 1998 and Hansen 1987). The computer was delivered in March of 1951 and was used for processing the 1950 census. It ran 24 hours a day all week until the task was completed. Once the census work was completed, the computer was used for other censuses and surveys including the Current Population Survey. Technology was now catching up to theory; the computer was now used for better calculation of variance estimates. It also opened up new possibilities, in particular imputation of missing values. With respect to variance estimates Hansen, Hurwitz, Nisselson and Sternberg (1955) comment:

> "Until the acquisition of a high-speed electronic computer, the UNIVAC, extensive approximations were introduced into the estimates of variances to avoid computations that would be exceedingly time consuming with the available equipment. The availability of the UNIVAC makes it possible to avoid

most of these approximations. Even with the electronic computer, however, the work of making variance computations would be extremely heavy if variances were computed for all items directly. Approximate methods will continue to be used in the future, but they will be evaluated by more exact computations than have been feasible in the past."

Other statistical organizations followed but at a slower pace. The slow pace in Canada was perhaps due in part to the American experience. A 1956 report to the Dominion Statistician at the Dominion Bureau of Statistics in Canada on the subject of computing at the Bureau of the Census (reported and quoted by Worton 1998) states:

"Subject-matter people … are not entirely convinced that the UNIVAC system has given them the results which might be expected from a computer system. Undoubtedly UNIVAC has given a great deal of trouble – much of it probably not the fault of UNIVAC at all. Factors such as poor programming, inadequate analysis of the job, inexperienced operating staff, maintenance problems, and even friction between the three operating groups, *i.e.*, the subject matter staffs, the Central Operations Group, and the Central Electronics Unit are reflected in the performance of the UNIVAC system."

The Dominion Bureau of Statistics, now Statistics Canada, obtained its first computer in 1960, an IBM 705. The computer was used to process the 1961 census. As noted already, the British used an army-owned computer to process their 1961 census. In the late 1940's, Mahalanobis was on a list showing interest in obtaining one of the first UNIVACs (Ceruzzi 1998). However, the annual reports of the Indian Statistical Institute published in *Sankhyā* show that the Institute did not obtain a computer until 1956 at which time it received an HEC-2M.

Variance estimation for survey estimates of means, totals and proportions was now feasible for large-scale surveys. Widespread use of this technology now depended on two things – access to a computer, which was an expensive item to buy, and appropriate software to carry out the calculations.

4. Scientific programming

Certain kinds of research, and the application of these research results, are possible only with computing. These possibilities expand not only with the expansion in computing power, but also with easier access to the computer's power through programming languages or packaged programs. For several years the most popular scientific programming language was FORTRAN (FORmula TRANslation). This was introduced in 1957 by IBM for its 704 computer. Part of what popularized FORTRAN was the development of the WATFOR (WATerloo FORtran) compiler at the University of Waterloo in 1965. This popular compiler, which was used for teaching purposes, combined with the dominance of IBM in the marketplace made FORTRAN accessible to many students and subsequently to researchers (Ceruzzi 1998). In reporting on the development of his own computer programs for survey research, Yates (1973) shows how pervasive FORTRAN had become over the 1960's. Yates's programs for the computer at Rothamsted Experimental Station were originally written in the late 1950's with code specific to the computer they had. In the mid-1960's the code was written in Extended Mercury Autocode. By the end of the 1960's this code had to be translated into FORTRAN using a machine translator; otherwise it was not usable at any other computer location. The earliest use of FORTRAN in sampling that I can find is in Fan, Muller and Rezucha (1962). These three individuals, all of who worked at IBM, developed algorithms and accompanying FORTRAN code to select simple random samples by computer.

There were two different paths that were followed in the application of FORTRAN programming to survey sampling. One was among statistical agencies or survey research centres and the other was among individual academic researchers. The kind of work followed along each path is strongly correlated with the evolving power of the computer and the dominance of IBM (and hence FORTRAN) in the market. By the end of the 1960's, many institutions had new and more powerful mainframe computers, often one of the IBM 360 series that was originally announced in 1964. Moreover, the software (FORTRAN in particular) remained compatible with machine changes and upgrades, especially for machines in the IBM 360 series (Ceruzzi 1998). The Dominion Bureau of Statistics obtained its first IBM 360 in 1969, while for example the Universities of Manitoba, Toronto and Waterloo obtained their first machines in the years 1966-67 (Day 1971). At the agencies and research centres, various formulae and procedures necessary to survey design and analysis were computerized. For example, Fellegi, Gray and Platek (1967) report that when the Canadian Labour Force Survey was redesigned over 1964-65, sample selection by Fellegi's (1963) method of unequal probability sampling was coded into a FORTRAN routine. From the University of Michigan Survey Research Center, Kish and Frankel (1970) report that they had FORTRAN code for obtaining variance estimates for a variety of statistics including regression coefficients using balanced repeated replication. By the mid-1960's academic researchers began to use the computer via FORTRAN programming to study, numerically or empirically, the sampling theory that they or others had derived. One of the first was Sedransk (1965) who carried out some efficiency comparisons in FORTRAN on an IBM 7074 (marketed by IBM in 1964) for a double sampling scheme. In particular, efficiency comparisons were made between optimal values for the first and second phase sample sizes and an approximation to the optimal values. The computations involved taking expected values over a trinomial distribution in which several conditions had been imposed. The use of the computer here was to obtain a numerical comparison between exact methods and approximate ones. By the end of the decade a new kind of computer-based research process emerged. Rao and Bayless (1969) and Bayless and Rao (1970) compared several unequal probability sampling schemes by generating all possible samples and calculating the exact finite population mean square error for several real and constructed populations. It then became the norm to carry out extensive empirical studies on any newly proposed estimator or design.

The past 30 years have seen remarkable changes in computing technology. Modern computers are much faster, physically smaller and have much greater storage capacity. The steady increase in computing power and the availability of standard programming languages has allowed survey researchers to expand as well into survey data analysis. This technological change is reflected in developments in sampling theory for variance estimation. From the 1960's to the 1980's there were three basic computerized approaches to variance estimation of complex survey statistics: Taylor linearization (see Woodruff 1971, for early references to its usage), jackknife (first proposed in sampling by Durbin 1959) and balanced repeated replication (McCarthy 1969). The rise of computing power saw a new technique, Efron's (1982) bootstrap, for variance estimation. This new statistical technique, which was contemporaneous with the development of networked RISC (Reduced Instruction Set Computing) workstations running under a UNIX operating system, is highly computer intensive. Over the 1980's RISC workstations gradually replaced most mainframes in research organizations. Near the end of this transition away from mainframes, Rao and Wu (1987) extended bootstrap methodology to variance estimation for smooth statistics under stratified multistage designs.

The most recent software to have an effect on statistical research is the development of computer algebra packages. Although computer algebra has been in existence for some time, it is only in the last decade that it has progressed to the point that it is accessible to many researchers. With computer algebra many complex manipulations can be done automatically and much quicker than by hand and without risk of error. Similar to several other areas of statistics, many of the algebraic manipulations in sampling theory are related to algorithms that generate partitions. Based on the computer algorithms developed by Andrews and Stafford (1993) and Stafford and Andrews (1993), Stafford and Bellhouse (1997) have extended computer algebra techniques to survey sampling theory. Using their methodology, most of the results of so-called classical sampling theory, either existing in the literature or yet to be obtained, can be derived automatically.

5. Analysis of survey data

While steady and substantial progress had been made in research on problems of survey estimation or enumerative surveys over the 20th century, by 1970 little had been accomplished on the analytical aspects of surveys. The terms "enumerative" and "analytical" surveys were coined by Deming in 1950 (Deming 1953). In the same article he also gives a succinct definition:

> "Briefly, the enumerative question is how many? The analytic question is why? is there a difference between two classes, and if so, how big are the differences?"

There is an implication in this quotation that the purpose of analytical surveys was for comparisons of domain means. Certainly, throughout the 1960's the understanding of what constituted an analytical survey was often limited to this. Cochran (1963) states:

> "In an analytical survey, comparisons are made between different subgroups of a population, in order to discover whether differences exist among them that may enable us to form or to verify hypotheses about the forces at work in the population."

Yates (1960) also focused mainly on domain comparisons in his discussion of analytic surveys. He did, however, discuss regression analysis and the problem of attenuation, but not the problem of general survey weights. Skinner, Holt and Smith (1989) attribute the pioneering work in analytical surveys to social scientists, Paul Lazarsfeld in particular. I will use the theoretical development of regression analysis in complex surveys to illustrate these connections to social science, in this case economics.

One of the earliest studies to take into account the survey weights in regression analysis was by Klein and Morgan (1951). At the time both were at the University of Michigan; Morgan was in the Survey Research Center. At the outset of their paper they state:

> "The sample design, the methods of collecting the data, and underlying economic behavior will all contribute to the formulation of the model. The study of data collected in consumer surveys has convinced us that one cannot proceed simply by the application of conventional statistical methods in the estimation of economic relationships because of the existence of some basic difficulties which we classify as follows: (1) weighting of observations, (2) heteroscedasticity, (3) non-linearities, (4) the choice of alternative economic concepts, (5) errors of observation."

They addressed the first four "basic difficulties" but not the fifth. In their analysis of the approximately 2,300 responses to the Survey of Consumer Finances, which was a multistage sample, Klein and Morgan used the survey weights through weighted least squares estimation of the regression parameters but ignored the clustering effect when it came to variance estimation. They noted that in many cases the use of the survey weights had little effect on the estimates of the regression coefficient estimates but noted that there was a reduction in the estimated variance for the model error. Though Klein went elsewhere, Morgan remained at the Michigan Survey Research Center. Twenty years later, he and another (Lansing and Morgan 1971) gave an overview of the state of the art for the analysis of economic survey data. Not much had changed in terms of the incorporation of the survey design into the analysis. The same is true for other areas of social research; in many cases not even the survey weights were used. In the economics literature debate continued for at least twenty years over whether to use the survey weights in regression analysis; Porter (1973) has several references to this debate.

It was out of this milieu that Kish, who also worked at the Michigan Survey Research Center, initially put forward the concept of the design effect (Kish 1957), which is the measure of increase or decrease in variance over simple random sampling experienced in a survey with a design other than simple random sampling. Design effects have become central to many aspects of the analysis of complex survey data. With respect to regression analysis, Kish and Frankel (1970) studied the design effects in the estimation of regression coefficients. They used balanced repeated replication to obtain their variance estimates. It is not entirely clear in their presentation exactly what regression coefficients they were estimating. Later, the parameters were explicitly spelled out in Kish and Frankel (1974). Specifically, the finite population parameters are what would be obtained in least squares estimation of superpopulation regression parameters when the entire finite population is available. Estimation of these parameters has become one of the standard approaches to regression analysis from complex surveys. Fuller (1975), using Taylor approximations to the variances, put the whole inference process on a solid theoretical foundation by providing limit theorems for the estimates. In addition, he addressed the one problem that Klein and Morgan (1951) ignored: errors in the variables or measurement errors in the independent variables.

Konijn (1962) took a different approach to regression analysis. Under a cluster sampling design, he assumed different simple linear regression models within each cluster. The parameters of interest were weighted averages of regression parameters with the weights given by the cluster sizes. This approach is model-based in the sense that it is the model parameters that are of interest, not a finite population parameter. Konijn's approach was not followed for several years. However, there is now a substantial literature that has grown out of this model-based approach; Pfeffermann (1993) contains several references.

With regard to the social science origins of survey analysis, there were similar experiences in categorical data analysis. The sociological literature from the 1960's and on contains many examples of categorical data analysis ignoring the sampling design. After Rao and Scott (1981, 1984) developed contingency table and goodness of fit analyses for complex surveys, Rao and Thomas (1988) tried to promote this methodology among sociologists using a review article. A search through citation indexes shows that, although this work has had great impact in the statistical and medical literature, it has had little impact in the sociological literature. The reason for this may be due, in part, to lack of computer software. The most popular software among sociologists, which is SPSS, does not at the moment contain any routines for the analysis of complex survey data. This points to a wider problem: regression, categorical data analysis and other techniques that have been proposed for complex surveys are not widely practicable without the appropriate computer software. Fuller himself tried to respond to this need by developing a packaged program for survey data analysis (Hidiroglou, Fuller and Hickman 1980).

6. Statistical software for survey research

Frank Yates at Rothamsted Experimental Station was the first statistician to develop software for survey research. His work began in the late 1950's (Yates and Simpson 1960). Originally, programs were written that were specific to each survey. This evolved into a general-purpose program by the early 1960's (Simpson 1961). Although it was the first in the field and was available for many years, it never achieved widespread popularity. There are at least four reasons for its general lack of success, reasons that point to the success of other software developers.

(1) The package was not user friendly. In his obituary of Yates, Dyke (1995) made allusion to this fact. He says:

> "Yates believed that the analyst should understand the relevant theory, and so be ready to specify in exact detail what he wanted. Perhaps for this reason the program was not excessively easy to use! But its power and flexibility, and uncluttered clarity of its output were, and are, outstanding."

(2) It was too expensive for what it did and could not compete with cheaper competitors. Wolter (1985) lists a number of packages that were available in the mid-1980's. At the time the package was twice as expensive as SUDAAN but could do only tabulations, whereas SUDAAN had the additional capability of regression analysis and ratio estimation.

(3) Marketing is an important factor in the success of a product. Yates appeared to be more interested in tinkering with his product to improve it rather than investing time in marketing it.

(4) Other than a manual, by 1985 there was no technical support for the package.

Yates was not alone in having software that did not catch on. I had the same experience when I developed variance estimation software based on tree traversal algorithms (Bellhouse 1985). Other than the expense factor (mine was free), my package was a living example for the other three reasons why some software does not fly.

By the early 1970's there were over 40 packaged programs and routines, written mainly in FORTRAN, that would do statistical analyses (Schucany, Minton and Shannon 1972). Of these original packages only two have remained popular in the marketplace, SAS first released in 1970 and SPSS released in the late 1960's.

The survey software that has maintained predominance in the market for several years is SUDAAN developed by B.V. Shah of the Research Triangle Institute (Shah 1978 and 1984). It is marketed well and fully supported by its developer. It was originally accessed as a SAS procedure and has now become a stand-alone package. The tie with SAS was probably one of the reasons for its initial success. Those who were familiar with SAS could easily familiarize themselves with this new procedure, or equivalently the package, so that in a sense it was user friendly. Further, the package has continued to keep pace with survey research. The original program contained routines to calculate standard errors for survey estimates including means, totals, proportions and ratios. This was expanded to include regression analysis in the late 1970's when research on regression in complex surveys was under way. The program now contains routines for regression analysis, logistic regression, categorical data analysis and survival analysis. It has also kept pace with developments in computing machinery. Originally developed on a mainframe computer, the package is now available for use on a PC. It still maintains its links to SAS, although SAS currently has its own survey analysis procedures under development.

Currently, there are several other programs for survey analysis. The most popular among these programs, in addition to SUDAAN, are STATA and WesVarPC. While SUDAAN has been linked to SAS, the future development of WesVarPC, which was originally developed by the research corporation Westat, has been turned over to SPSS. Further, the survey routines in STATA are part of a larger statistical analysis package. As with mergers in the general business world, along with product and service integration, the future trend for survey data analysis packages is to become part of an omnibus statistical package. The development and maintenance of statistical packages, for survey research or for a wider context, is a time-consuming enterprise requiring a substantial capital investment. This can only be done by a well-financed organization.

SUDAAN, STATA and WesVarPC, along with the software packages GES from Statistics Canada and another named CLAN, have been recently reviewed and evaluated in Bergdahl, Black, Bowater, Chambers, Davies, Draper, Elvers, Full, Holmes, Lundqvist, Lundström, Nordberg, Perry, Pont, Prestwood, Richardson, Skinner, Smith, Underwood and Williams (1999). SUDAAN and STATA have also been evaluated by Cohen (1997). Among three of the packages reviewed (STATA, SUDAAN and WesVarPC), SUDAAN appears to have the most options. For example, Bergdahl *et al.* (1999) note that SUDAAN carries out variance estimation for complex statistics using any one of Taylor linearization, jackknife and balanced repeated replication. WesVarPC covers jackknife and balanced repeated replication, while STATA relies solely on Taylor linearization. So far, none of the packages does variance estimation using the bootstrap. It may just be a matter of time before this technology is incorporated into these packages. For some of its public use sample files, Statistics Canada provides bootstrap variance estimation procedures in SAS code. These procedures, however, are specific to the surveys in question.

7. Models in sampling

Models have come in and out of favour among sampling practitioners. Due to Neyman's (1934) pioneering work, the paradigm of randomization and the randomization distribution was paramount until the 1960's. However, the use of models did not disappear during the intervening years. Cochran (1946), for example, used models to study certain sampling designs and was able to conclude that systematic sampling was a good design to use under certain population structures. The 1960's debate over models arose out of the questioning of the foundations of sampling initiated by Godambe (1955). Since then the use of models has not only crept back in to sampling theory but has flourished substantially.

Since the 1960's the use of models in sampling has gone in several directions. At the same time, the practical and general use of models in survey estimation and analysis is only feasible with high speed computing and the appropriate software. In keeping with the theme I have been following here, I will take a very narrow approach to models by tying their usage to computing technology.

Several model-related methodologies have been computerized, either through the provision of numerical examples to illustrate the use of the methodology or through simulation studies to examine how the methodology works. At the present time there is only one model-related approach that has matured to the point where a general package program is available. This is the model-assisted approach that C.-E. Särndal has taken over several years resulting in generalized regression estimation or GREG. The bulk of the work is summarized in Särndal, Swensson and Wretman (1992). The work was initially motivated by the debates over the foundations of sampling. Under a model, a best, in some sense, estimator of a finite population parameter can be derived. Those on the side promoting randomization inference pointed out that when the model fails the associated estimator can perform very poorly. The solution propounded by Särndal was to obtain the estimate under the model and then to adapt it in such a way that it would remain consistent and perform adequately under the randomization distribution. It is an attempt to obtain the best of both worlds. Generalized regression estimation, as well as several other estimators, have been programmed into GES, a generalized estimation system developed at Statistics Canada. This SAS-based software is aimed at the descriptive side of surveys rather than the analytic and is described in Estevao, Hidiroglou and Särndal (1995). It is a package that could easily catch on under the right conditions.

8. Conclusions

Developments in sampling research are inextricably tied to computing and computational methods. Where research is headed will be guided, in part, by computer developments. What the immediate future holds for computing is greater speed and greater storage capacity so that packages can become bigger and more comprehensive. Generally acceptable practices in survey estimation and the analysis of survey data will be determined by the contents of generally available computer packages for survey sampling. On the research methodology side, new methodology will continue to be increasingly computer intensive. One other foreseeable development is the explosion of the internet. As a result of this explosion, several complete survey datasets are now easily available via the web. The extensive testing of new methodology on a variety of real surveys prior to publication of the methodology may soon become the norm.

Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

Andrews, D.F., and Stafford, J.E. (1993). Tools for symbolic computation of asymptotic expansions. *Journal of the Royal Statistical Society, B*, 55, 613-628.

Baines, J.A. (1900). On census-taking and its limitations. *Journal of the Royal Statistical Society*, 63, 41-71.

Bayless, D.L., and Rao, J.N.K. (1970). An empirical study of stabilities of estimators and variance estimators in unequal probability sampling (n = 3 or 4). *Journal of the American Statistical Association,* 65, 1645-1667.

Bellhouse, D.R. (1985). Computing methods for variance estimation in complex surveys. *Journal of Official Statistics*, 1, 323-329.

Bellhouse, D.R. (1988). A brief history of random sampling. *Handbook of Statistics*. (Eds. C.R. Rao and K.R. Krishnaiah) 6, 1-14. Amsterdam: North-Holland.

Benjamin, B. (1961). The 1961 census of population. *Incorporated Statistician*, 11, 130-143.

Bergdahl, M., Black, O., Bowater, R. Chambers, R., Davies, P., Draper, D., Elvers, E., Full, S., Holmes, D., Lundqvist, P., Lundström, S., Nordberg, L., Perry, J., Pont, M., Prestwood, M., Richardson, I., Skinner, C., Smith, P., Underwood, C. and Williams, M. (1999). *Model Quality Report in Business Statistics Volume II: Comparison of Variance Estimation Software and Methods*. London: Office of National Statistics.

Bowley, A.L. (1906). Address to the Economic and Statistics Section of the British Association for the Advancement of Science, York. *Journal of the Royal Statistical Society*, 69, 540-558.

Bowley, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin of the International Statistical Institute* 22 (1), 1-62.

Bowley, A.L. (1936). The application of sampling to economic and sociological problems. *Journal of the American Statistical Association*, 31, 464-480.

Box, K., and Thomas, G. (1944). The Wartime Social Survey. *Journal of the Royal Statistical Society*, 107, 151-189.

Brewer, K.R.W., and HANIF, M. (1983). *Sampling with Unequal Probabilities*, (Lecture Notes in Statistics, Volume 15). New York: Springer-Verlag.

Ceruzzi, P.E. (1998). *A History of Modern Computing*. Cambridge, Massachusetts: MIT Press.

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, 17, 164-177.

Cochran, W.G. (1963). *Sampling Techniques*, (2nd Edition). New York: John Wiley & Sons, Inc.

Cohen, S.B. (1997). An evaluation of alternative PC-based software packages developed for the analysis of complex survey data. *American Statistician*, 51, 285-292.

Day, N. (1971). *Canadian Computer Census* 1971. Toronto: Canadian Information Processing Society.

Deming, W.E. (1953). On the distinction between enumerative and analytic surveys. *Journal of the American Statistical Association*, 48, 244-255.

Deming, W.E. (1956). On simplifications of sampling designs through replication with equal probabilities and without stages. *Journal of the American Statistical Association*, 51, 24-53.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, 46, 477-480.

Dyke, G. (1995). Obituary: Frank Yates. *Journal of the Royal Statistical Society, A*, 158, 333-338.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Estevao, V., Hidiroglou, M.A. and Särndal, C.-E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. *Journal of Official Statistics*, 11, 181-204.

Fan, C.T., Muller, M.E. and Rezucha, I. (1962). Development of sampling plans by using sequential (item by item) selection techniques and digital computers. *Journal of the American Statistical Association*, 57, 387-402.

Fellegi, I.P. (1963). Sampling with varying probabilities and without replacement: rotating and non-rotating samples. *Journal of the American Statistical Association*, 58, 183-201.

Fellegi, I.P., Gray, G.B. and Platek, R. (1967). The new design of the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 62, 421-453.

Fuller, W.A. (1975). Regression analysis for sample survey. *Sankhyā C*, 37, 117-132.

Gillies, D. (1992). *Revolutions in Mathematics*. Oxford: Clarendon Press.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, B*, 17, 269-278.

Hansen, M.H. (1987). Some history and reminiscences on survey sampling. *Statistical Science*, 2, 180-190.

Hansen, M.H., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.

Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Sternberg, J. (1955). The redesign of the Current Population Survey. *Journal of the American Statistical Association*, 50, 701-719.

Hartley, H.O. (1946). The application of some commercial calculating machines to certain statistical calculations. Supplement to *Journal of the Royal Statistical Society*, 8, 154-183.

Hidiroglou, M.A., Fuller, W.A. and Hickman, R.D. (1980). *SUPER CARP*. Ames: Iowa State U.P.

Hollerith, H. (1894). The electrical tabulating machine. *Journal of the Royal Statistical Society*, 57, 678-689.

Hooker, R.H. (1894). Modes of census-taking in the British Dominions. *Journal of the Royal Statistical Society*, 57, 289-368.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Kaier, A.N. (1895/6). Observations et experiences concernant des dénombrements representatives. *Bulletin of the International Statistical Institute*, 9, 176-183.

Kaier, A.N. (1897). *The Representative Method of Statistical Surveys* (1976, English translation of the original Norwegian). Oslo: Central Bureau of Statistics of Norway.

Kaier, A.N. (1905). Untitled speech with discussion. *Bulletin of the International Statistical Institute*, 14, 119-134.

Kish, L. (1957). Confidence intervals for clustered samples. *American Sociological Review*, 22, 154-165.

Kish, L., and Frankel, M.R. (1970). Balance repeated replication for standard errors. *Journal of the American Statistical Association*, 65, 1071-1094.

Kish, L., and Frankel, M.R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society B*, 36, 1-37.

Klein, L.R., and Morgan, J.N. (1951). Results of alternative statistical treatments of sample survey data. *Journal of the American Statistical Association*, 46, 442-460.

Konijn, H.S. (1962). Regression analysis in sample surveys. *Journal of the American Statistical Association*, 57, 590-606.

Kruskal, W., and Mosteller, F. (1980). Representative sampling, IV: the history of the concept in statistics 1895 - 1939. *International Statistical Review*, 48, 169-195.

Lansing, J.B., and Morgan, J.N. (1971). *Economic Survey Methods*. Ann Arbor: Survey Research Center.

Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-378.

Mandeville, J.P. (1946). Improvements in methods of census taking and survey analysis. *Journal of the Royal Statistical Society*, 109, 111-129.

McCarthy, P.J. (1969). Pseudo-replication: half samples. *Review of the International Statistical Institute*, 37, 239-264.

Murthy, M.N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

Neyman, J. (1934). On the two different aspects of the representative method: stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-337.

Porter, R.D. (1973). On the use of survey sample weights in the linear model. *Annals of Economic and Social Measurement*, 2, 141-158.

Rao, J.N.K., and Bayless, D.L. (1969). An empirical study of stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum. *Journal of the American Statistical Association*, 64, 540-559.

Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, 76, 221-230.

Rao, J.N.K., and Scott, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, 12, 46-60.

Rao, J.N.K., and Thomas, D.R. (1988). The analysis of cross-classification data from complex sample surveys. *Sociology Methodology*, 18, 213-269.

Rao, J.N.K., and Wu, C.F.J. (1987). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 321-241.

Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Schucany, W.R., Minton, P.D. and Shannon, B.S. (1972). A survey of statistical packages. *Computing Surveys*, 4, 65-79.

Sedransk, J. (1965). A double sampling scheme for analytical surveys. *Journal of the American Statistical Association*, 60, 985-1004.

Shah, B.V. (1978). SUDAAN: Survey data analysis software. *Proceedings of the Statistical Computing Section, American Statistical Association*.

Shah, B.V. (1984). Software for survey data analysis. *American Statistician*, 38, 68-69.

Simpson, H.R. (1961). The analysis of survey data on an electronic computer. *Journal of the Royal Statistical Society, A*, 124, 219-226.

Skinner, C.J., Holt, D. and Smith, T.M.F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons, Inc.

Stafford, J.E., and Andrews, D.F. (1993). A symbolic algorithm for studying adjustments to the profile likelihood. *Biometrika*, 80, 715-730.

Stafford, J.E., and Bellhouse, D.R. (1997). A computer algebra for sample survey theory. *Survey Methodology*, 23, 3-10.

Willcox, W.F. (1926). The past and future developments of vital statistics in the United States I: John Shaw Billings and federal vital statistics. *Journal of the American Statistical Association*, 21, 257-266.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Woodruff, R.S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66, 411-414.

Worton, D.A. (1998). *The Dominion Bureau of Statistics: A History of Canada's Central Statistical Office and Its Antecedents*, 1841-1972. Montreal and Kingston: McGill- Queen's University Press.

Yates, F. (1960). *Sampling Methods for Censuses and Survey*, (3rd edition). London: Griffin.

Yates, F. (1973). The analysis of surveys on computers – features of the Rothamsted Survey Program. *Applied Statistics*, 22, 161-171.

Yates, F., and Simpson, H.R. (1960). A general program for the analysis of surveys. *Computer Journal*, 3, 136-140.