Survey Methodology, June 1990
Vol. 16, No. 1, pp. 95–98
Statistics Canada

Some Developments of Sampling Techniques and their Use in Official Statistics in Sweden

TORE DALENIUS and CARL-ERIK SÄRNDAL[1]

In this paper we present some important features of the history of sample surveys in Sweden, and we comment on related developments of sampling techniques (methods and theory) in official statistics. The account is organized into three periods as follows: (i) before 1900; (ii) 1900-1950; and (iii) after 1950. The emphasis is on the third period.

I. THE PERIOD BEFORE 1900

1. A summary view. As described in Dalenius (1957), there was a noticeable resistance against sample surveys in traditional fields of official statistics, especially among statisticians in leading positions. Sample surveys were considered justified primarily in cases where circumstances did not admit *total* surveys. In other fields there were, however, signs of appreciation, as illustrated in the next section.

2. Two classic illustrations. In the 1820's, the area of meadowland in Sweden was estimated using the following technique. For each county separately, the ratio of meadow acreage to arable land was computed for a sample of farms. This ratio was then applied to the total arable land acreage of the county, for which a separate estimate was available. And in 1830, the proposal was made by an official in a forestry board to estimate the volume of timber in a forest by means of a "strip survey method".

II. THE PERIOD 1900-1950

3. The main features. The potential of sample surveys in official statistics was slowly being understood. To the extent that sample surveys were used during this period, the design typically called for systematic sampling, whenever this was operationally feasible. In many applications, the sampling fraction was 1/10 or 1/5. In the 1940's, a major factor favouring total surveys was the war-time economy with its regulations and rationing. This influence, which lasted roughly until the end of that decade, was however counteracted by the introduction of Gallup polls into Sweden and especially by the spectacular accuracy of the Gallup Institute's forecast of the 1944 election. In particular, these trends were followed with interest by official statisticians.

4. The 1911 Forest Survey in Värmland. The essential feature of the design was that the volume of timber was measured on sample plots along 10 meter wide strips covering the area of Värmland. It is worth noting that the "representative characteristics" of the survey were analysed by means of probability theory.

---

[1] Tore Dalenius, Brown University, Carl-Erik Särndal, Université de Montréal.
The circumstances did not permit the authors to discuss the contents of this paper with representatives of Statistics Sweden.

5. The 1911 Housing Survey in Göteborg. This survey was carried out by the municipal statistical office in Göteborg. The selection of the sample of apartments was based on an urn scheme. Each building in Göteborg was represented by a slip with identification data. The slips were thoroughly mixed in an urn and a 20% sample of slips was selected. The motive behind the scheme was to avoid that the survey be criticized for using a biased sample. The urn scheme was described by the person in charge of the survey as the only method "which can be called representative".

6. The 1935-36 Partial Population Census. This sample census used an elaborate scheme of controlled selection. The results from this census played a decisive role in an intense debate in Sweden concerning a "population crisis" which was feared as a result of low birth rates at the time.

III. THE PERIOD AFTER 1950

7. The beginnings of a new era. The greatly improved international communications after the end of World War II contributed to making the statistical community in Sweden aware of the recent advancements in sample survey theory, methods, and applications in the United States and India, to mention two of the leading countries. The new developments were studied and discussed, for example, at the conference of the Scandinavian statisticians in Helsinki in 1949. Statisticians were proud to be able to "talk sample survey methods"; to be sure, in some cases this ability was limited to knowledge of certain technical terms, notably "stratification". Mention should also be made of the influence exercised by the United Nations and affiliated agencies such as the Food and Agriculture Organization. In the following we give some examples of sample surveys and related developments of methods and theory. For cases dating to the early 1950's, details are found in Dalenius (1957).

8. The 1950 sample inventory of acreages and livestock. In the 1930's, sample surveys were used to estimate acreages of various crops and animal stocks. These surveys were referred to as "representative counts". They were based on nonprobability selection of farms. The aim, which however was not achieved, was to select 1/10 of the farms in each of several size-groups into which the farms had been divided. In the 1940's, these surveys were carried out on a total basis. A decision was made for the 1950 survey to return to sampling. The design that was suggested and largely implemented for the 1950 survey represented a partial break with the classical tradition of selecting every tenth unit. While the total sample size was fixed by the government authorities to be 1/10 of the total number of farms in the target population, the new design called for stratifying the farms by size groups based on acreage and using minimum variance allocation, which implied a selection of relatively speaking more large farms than small farms. It is interesting to note that the government authorities responsible for assessing the design felt it necessary to consult the U.N. Subcommission on Statistical Sampling about the appropriateness of the drastic deviation from the "every tenth unit rule". The Subcommission wholeheartedly endorsed the design. Consequently it was accepted in principle. The design provided considerable opportunity for research. In fact, three contributions to the theory of stratified sampling emerged, namely, (i) how best to divide a population into L strata; (ii) the best choice of the number of strata; and (iii) sample allocation to the strata for estimation of several parameters. The suggested design also called for addressing the problem of "measurement errors" in the acreage, and a special calibration survey was proposed. However, the authorities rejected this proposal.

9. Yield estimation. During World War II, the yield of various crops was estimated using data collected by "eye estimates" of the yield per unit area. By 1950, it was realized that this data collection method could be seriously biased. In the beginning of the 1950's, time was ripe for considering a different approach, namely, crop estimation based on harvesting sample plots, referred to as "objective crop estimation". Accordingly, a pilot study was carried out to test the use of this approach. The outcome of the test was convincing. From then on, the "objective" method has been used. As part of the pilot survey design, a scheme was developed for without replacement selection of $n = 2$ farms from a stratum with probabilities proportional to size, as discussed in Dalenius (1953). The scheme called for dividing each stratum at random into two parts, and selecting one farm from each part.

10. Developments relating to nonsampling errors. In the early 1950's, the problem of nonresponse received considerable attention in Sweden as in other countries. Surveys with 20-30% nonresponse were not unusual. This generated a vivid and sometimes heated debate in the statistical community about the distortion of the estimates. For a while, the statisticians seemed to have the problem under their control. The public concern about invasion of privacy has lately changed this picture; nonresponse has again become a serious problem. In the last 15 years, several contributions were made in the area of control of nonsampling errors. The problem of "evasive answer bias", to use the term introduced by S. Warner in connection with randomized response, was addressed in Swensson (1976). And Lyberg (1981) successully tackled the problem of controlling the coding operation in a population census or in a survey with interviews.

11. Respondent burden. In recent years there has been a growing concern about respondent burden and its negative effects on response rates. For example, the target population in many business surveys is the same, rather limited population. The problem can be alleviated by special sample selection techniques. The SAMU system for business surveys at Statistics Sweden permits "negative coordination" of samples, in the sense that samples without overlap can be selected with the technique known as JALES. To each unit in the sampling frame, a uniformly distributed random number is attached. This number stays with the unit, and is used in the selection of samples over time.

12. Modeling in combination with traditional probability sampling principles. Since the 1950's, the methodology for surveys had closely followed the strong probability sampling tradition established by Neyman and by Hansen and his co-workers in the United States. However, sometimes modeling is necessary in surveys when the traditional probability sampling theory is not sufficient. Since the 1970's the use of modeling in surveys has been explored. The book Foundations of Inference in Survey Sampling by Cassel, Särndal and Wretman (1977) exposed the new trends. Also, a number of papers by these and other Swedish authors showed how models may assist in inference from surveys. In recent years, methodologists at Statistics Sweden have shown unusual openness to incorporating modeling in the making of survey estimates. An early example where design-based and model-based ideas were combined is the "Öresund survey" for measuring traffic flow between Sweden and Denmark. The design is discussed in Cassel (1978). Some surveys are now designed with the aid of modeling assumptions, as in the work force survey described in Lundström (1987) and in an ongoing project of restructuring of the business survey sector.

13. Safeguarding privacy in surveys. In the last two decades, the general public has become increasingly concerned about invasion of privacy in connection with surveys, including population censuses, carried out by Statistics Sweden. As a result, there has been a trend

towards increasing nonresponse rates in some surveys. Several measures have been taken to deal with the problem: (i) Statistics Sweden has adopted the Ethical Declaration of the International Statistical Institute (1986); a translation of that declaration was distributed to all employees; (ii) In 1987, Statistics Sweden held an international conference which focused on policy issues (as distinguished from "techniques"); the discussions at the conference are summarized in Statistics Sweden (1987); (iii) Statistics Sweden has promoted the development of new safeguards for privacy in its surveys and has taken active steps to apply them. A review is given in Dalenius (1988). Of special interest are papers by Block and Olsson (1976), who describe a measure for the identifying power of quasi-identifiers, and Cassel (1976), who discussed probability-based disclosure.

14. Specific events. The increasing appreciation of sample surveys since around 1950 led to the creation of the Survey Research Center at Statistics Sweden in 1953. A similar interpretation may be given to the establishment of a professorship in "statistics, especially official statistics" at the University of Stockholm in 1965. Also, professorships in survey methodology were recently created at Statistics Sweden.

REFERENCES

BLOCK, H., and OLSSON, L. (1976). Bakvägsidentifiering. (Backwards identification) *Statistisk Tidskrift*, 1976, 135-144.

CASSEL, C.M., SÄRNDAL, C.E., and WRETMAN, J.H. (1977). *Foundations of Inference in Survey Sampling*. New York: Wiley.

CASSEL, C.M. (1978). Probability based disclosure. In Dalenius, T., and Klevmarken, A. (eds.), *Proceedings of a symposium on personal integrity and the need for data in the social sciences*. Swedish Council for Social Science Research, Stockholm, 189-193.

CASSEL, C.M. (1978). On errors in the predictions with logit models. Technical report, Statistics Sweden.

DALENIUS, T. (1953). Något om metoder för objektiva skördeberäkningar. (About methods for objective crop estimation.) *Kungliga Lantbruksakademiens Tidskrift*, 92, 99-118.

DALENIUS, T. (1957). *Sampling in Sweden. Contributions to the Methods and Theory of Sample Survey Practice*. Stockholm: Almqvist and Wiksell.

DALENIUS, T. (1988). Controlling Invasion of Privacy in Surveys. Statistics Sweden.

INTERNATIONAL STATISTICAL INSTITUTE (1986). Declaration of Professional Ethics. *International Statistical Review*, 54, 227-242.

LUNDSTRÖM, S. (1987). Utveckling av estimatorer för skattning av antal förvärvsarbetande i olika arbetstidsklasser inom små redovisningsgrupper. R&D Report, U/STM 40, Statistics Sweden.

LYBERG, L. (1981). Control of the coding operation in statistical investigations. Urval no. 13, Statistics Sweden.

STATISTICS SWEDEN (1987). Statistics and Privacy: Future Access to Data for Official Statistics – Cooperation or Distrust? Statistics Sweden.

SWENSSON, B. (1977). Survey measurement of sensitive attributes. Ph.D. Thesis, Department of Statistics, University of Stockholm.