SCB STATISTISKA CENTRALBYRÅN

1983-09-01

PROMEMORIOR FRÅN P/STM

NR 10

# CURRENT SURVEY RESEARCH STATISTICS SWEDEN

AV LARS LYBERG, BENGT SWENSSON OCH
JAN HÅKAN WRETMAN

INLEDNING

TILL

**Promemorior från P/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1978-1986. – Nr 1-24.**

**Efterföljare:**

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

Promemorior från P/STM 1983:10. Current survey research at Statistics Sweden / Lars Lyberg m.fl.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-PM-PSTM-1983-10

SCB STATISTISKA CENTRALBYRÅN

1983-09-01

PROMEMORIOR FRÅN P/STM

NR 10

# CURRENT SURVEY RESEARCH STATISTICS SWEDEN

AV LARS LYBERG, BENGT SWENSSON OCH
JAN HÅKAN WRETMAN

# Current Survey Research at Statistics Sweden[1]

by

Lars Lyberg, Bengt Swensson and Jan Håkan Wretman

Statistics Sweden

CONTENTS



---

[1] Invited paper to be presented at the ASA meeting in Toronto, August 1983.

# Current Survey Research at Statistics Sweden

by

Lars Lyberg, Bengt Swensson and Jan Håkan Wretman

Statistics Sweden

## 1. Introduction

As laid down by the Government, Statistics Sweden is the central authority for the production of governmental statistics in Sweden and our agency is responsible for coordinating the production of such statistics.

Among other things Statistics Sweden has been commissioned to conduct - on behalf of central and local authorities - statistical surveys, to carry out data processing and to provide an advisory service on statistics. Furthermore, Statistics Sweden must keep certain registers, such as the Central Register of Enterprises, the Register of the Total Population and the Register of Farms. Statistics Sweden must ascertain that statistical considerations are not disregarded when other authorities collect data for administrative purposes. Commissions are also undertaken for private persons, corporations and organizations.

The actual statistical production is performed by three different departments; namely Area Statistics, Enterprise Statistics, and Statistics on Individuals. Four functional departments, Planning and Coordination (including the Statistical Research Unit), Central Administration, Systems and Information, and Operations, serve the subject-matter departments in different ways.

The authors of this paper work at the Statistical Research Unit (SRU) where research is carried out within two main projects: (1) Sampling and estimation and (2) Quality. The paper reflects the work currently carried out by the SRU. However, statistical research is also conducted in many other units of our agency. In the Department for Statistics on Individuals, for instance, there are two research groups working with statistical methods and measurement methods. The authors estimate that within Statistics Sweden approximately 25 persons work at least part-time with statistical research. This number should be related to the approximately 200 surveys carried out every year. Thus, impulses for new developments are by no means lacking. On the other hand, by necessity many fields must be neglected.

The present paper gives an overview of the work conducted within the main projects of SRU. The topics discussed are: pilot survey design, sampling from skew distributions, sampling on two occasions from a finite changing population, small domain estimation, estimation of regression coefficients for domains, measures for reducing nonresponse rates, inference in the presence of varying response probabilities, randomized response, control of interviewer performance, computer-assisted interviewing, automated coding, risk assessment, standards for quality presentation, and selected problems treated outside the SRU.

## 2. Pilot survey design

Efficient survey design calls for information on at least the following:

- Data on variability and data explaining this variability.

- Data on costs for different survey operations.

- Information on the characteristics of different measurement designs with respect to errors and administrative fitness.

- Information for the choice of sampling unit (elements, clusters etc.), and for the choice of a sampling strategy (sampling design and estimator), i.e., information on available auxiliary variables and their relation to the survey variables.

- Information on different processing systems (optical character recognition vs key punching, dependent vs independent verification of coding, manual coding vs automated coding, manual vs computerized editing, etc) and their impact on survey results.

- Information on administrative resources available.

Information on these and other aspects might be obtained from:

i) earlier surveys of the same or similar kind;

ii) evaluation studies;

iii) expert guesswork;

iv) pilot surveys, more or less informal pretests, different kinds of experiments, and method studies.

Often, information generated from i-iii is insufficient as a means for efficient survey design. A pilot survey or a sequence of pretests must be carried out. Pilot surveys and pretests, as is the case with producer-oriented evaluation studies, aim at improving a forthcoming main survey. The difference between piloting and evaluation is that the former is carried out prior to the main survey and generally stands by itself. Furthermore, in pilot surveys different designs can be put to a test while an evaluation study concerns a given design. Naturally there are surveys which are hybrids, for instance when built-in experiments are conducted in main surveys.

Textbooks on survey sampling only deal superficially with pilot surveys. In some of them the subject is mentioned en passant. Others discuss the matter more thoroughly but most of them still on a couple of pages only. There is no obvious reason for this. Possibly, pilot surveys are characterized as special cases of regular surveys and should be designed as such. It is also possible that the techniques used in pilot surveys, for instance experimental designs, are considered standard statistical methodology.

The emphasis differs between textbooks. A common view is that pilot surveys are recommended prior to large surveys where unknown factors might be important. The purpose is to estimate quantitative data (variances and costs) and qualitative data (for instance how different procedures work in practice). Pilot surveys might also be conducted in order to discover weak spots. Sometimes such efforts are referred to as "dress rehearsals."

Several authors urge that the utility of pilot surveys must be critically examined. If the purpose is to estimate quantitative data it is important to conduct a pilot survey of considerable size resulting in rather precise estimates. If precise estimates are not obtained such a survey is superfluous.

Since Statistics Sweden is the agency responsible for most of the governmental statistics in Sweden, many pilot surveys are carried out. However, we have a strong feeling that some of these could have been designed more efficiently or, in some cases, should not have been designed at all. A quick review of some pilot surveys conducted during the seventies seems to confirm our view.

- The pilot survey goals are often obscure.
- Variances and costs are seldom investigated.
- Inference is often replaced by intuition when for instance only a few interviewers are used or when only a few clusters are investigated. One consequence is that nonresponse rates are hard to predict.
- In some instances only one "treatment" is tested; i.e. only one type of questionnaire is studied, only one type of coding control is investigated etc. In those cases it is never found whether better or worse alternatives exist.

- Often indications of problems are given but one cannot always define which problems are the most serious ones.
- The pilot survey reports are seldom very specific when it comes to suggestions concerning the design of the main survey.
- Many pilot surveys are often in the nature of "warming-up" surveys of questionable utility.

Of course we have also found pilot surveys which seem efficiently designed and some approaches are promising and might be developed in a more formalized way.

Consultations with statisticians in USA, Canada, United Kingdom and Australia show that these problems have been recognized by other agencies as well. Some useful references are Jabine (1981), United Nations (1981, 1982), Hunt el al (1982), Moser and Kalton (1972), Brewer et al (1977), Brackstone (1976), and Dunnel and Martin (1982).

One area that has been examined thoroughly over the years, though, is the development of questionnaires. The literature on that particular topic in survey design is extensive. Some recent references are US Bureau of the Census (1983), Sudman and Bradburn (1982) and Kalton and Schuman (1982).

We are now working on a manual for the design of pilot surveys and pretests. Some fields to be discussed are:

a) The problem with many and/or competing goals.
A pilot survey is often multi-purpose and the purposes are sometimes competing as well. For instance studies of problems concerning the development of a questionnaire might call for a design involving a few primary sampling units to allow for meetings with interviewers etc. Reliable estimates of nonresponse rates would probably demand a nation-wide sample. One possible approach might be to draw one large sample and use subsamples of this for different extensive studies (for instance analysis of the questionnaire). The large sample could be used for estimates of nonresponse rates etc.

b) Should the samples be randomized or not?
Accurate estimates of variability, costs, nonresponse rates, etc, require random samples. It is, however, almost inevitable that the sampled population often has to be restricted compared to the one to be used in the main survey. The restrictions are sometimes a function of the administrative resources available. For instance, it happens that we choose primary sampling units where the interviewers have time to work with the pilot survey. Naturally, such strategies cause inferential problems. If randomization among first stage units is impossible or impractical it may be more advantageous to choose primary sampling units from "problem" areas and "non-problem" areas to obtain some kind of upper and lower limits of the characteristics under study.

c) The design of experiments.
The value of experimental studies should be stressed. Especially this is typical of experiments imbedded in ongoing surveys. Such experiments can be very profitable.

d) Studies of variability and cost.
Such studies are very rare at Statistics Sweden and at other agencies as well.

e) Sequential studies.
If flexibility is possible, sequential studies could be worth while. We have come across some examples of such studies. The results from one small study have been used as input to another. Small studies can give indications. Sometimes the indications are strong and may lead to alterations in the next study of the sequence. Sequential studies can be looked upon as an alternative to experiments when the "treatments" are difficult to define from the beginning.

f) Computer-assisted telephone interviewing (CATI).
The development of CATI-procedures makes it easier to put questionnaires to a test. Questions can be changed overnight without any printing. The same goes for question order, instructions, etc.

g) Reporting of results from pilot surveys.
This is an area where improvement is needed. We have come across some nice examples, though, of good and reliable reporting.

## 3. Sampling from skew distributions

The problem of estimating characteristics of a population with a very skew distribution occurs frequently in different guises in survey practice. In business sample surveys it is a major problem although stratified sampling improves the situation to some extent. In such surveys the problem is often - somewhat dubiously - referred to as an "outlier problem." Another guise of the same problem occurs when applying estimators for domains of study, where most of the observations are set to zero.

In these situations the standard practice of calculating confidence intervals based on the normal approximation is doubtful. Two questions arise:

i) When does the normal approximation apply?

ii) What should be done when the normal approximation is not good enough?

Our efforts so far have been concentrated on the first question. Mathematical statisticians have approached this problem as one of establishing convergence rates of sampling distributions to the normal distribution and making remainder term estimates for simple random sampling from finite populations. The attempts to reach exact theoretical results which are easy to apply have more or less failed. Therefore we have tried another approach based on empirical, numerical investigations.

With respect to the first question, W.G. Cochran (1977) gives a crude rule of thumb which he considers applicable to simple random sampling in certain situations when estimating a population mean or total. This rule is

$$n > 25\, G_1^2,$$

where n is the sample size necessary for application of the normal approximation and $G_1$ is Fisher's measure of population skewness:

$$G_1 = \frac{\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^3}{\left[\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^2\right]^{3/2}} .$$

Cochran states that the rule is designed so that a 95 % confidence probability statement will be wrong not more than 6 % of the time. However, he gives no reference or theoretical argument to underpin the rule. Anyway, the rule has occasionally worked well in cases where it has been empirically tested.

This fact has inspired us to make a more systematic investigation of rules of the Cochran type, i.e. rules such as

$$n > K_\alpha G^2$$

where G is a measure of population skewness and $K_\alpha$ is a constant, such that a nominal 95 % confidence interval covers the true value at least $\alpha$ % of the time.

We have studied a very simplified case, namely dichotomous populations consisting of ones and zeros. By calculating the actual coverage probability for a large number of different population sizes and degrees of skewness, a relatively stable relation between n and $G^2$ was observed for a fixed coverage probability. However, $G_1$ is not a suitable measure of skewness for our purpose. A better measure is

$$G_2 = \frac{\frac{1}{N}\sum_{j=1}^{N} |x_j - \bar{x}|^3}{\left[\frac{1}{N}\sum_{j=1}^{N} (x_j - \bar{x})^2\right]^{3/2}} ,$$

which could be used over the whole range of degrees of skewness and which, for large values, is approximately equal to $G_1$.

Using $G_2$, the following combinations of $K_\alpha$ and $\alpha$-values were obtained.

| $\alpha$ | $K_\alpha$ |
|---|---|
| 85 % | 3 |
| 90 % | 5 |
| 93 % | 11 |
| 94 % | 20 |
| 94.5 % | 40 |

These values were also tried for more diversified finite populations based on known parametric distributions (beta, Weibull, log-normal and power-function). By drawing simple random samples from these finite populations we found that the combinations above were approximately valid for these cases also. We therefore believe that a rule based on these values is sound survey practice.

A more detailed description of these problems is found in Dalén (1983).

Let us give some examples of other problems currently dealt with in this field.

- The rules given above assume that we have access to reliable information on $G_2$. Such information rarely exists, though, and therefore there is a need to use a sample estimate $\hat{G}_2$. However, it is not sufficiently known to what extent such an estimate could replace $G_2$. The sample analogue of $G_2$, for example, is a consistent but biased estimator with a typically large variance in skew distributions. This problem is now being studied theoretically as well as empirically.

- Attempts are made to find working rules for other sampling strategies.

- For qualitative data there are alternative and more exact ways to calculate confidence intervals. Properties of these alternatives are currently investigated. The aim is to make suggestions as to which one to use in different situations.

## 4. Sampling on two occasions from a finite changing population

The problems of efficient estimation of totals and means as well as changes between totals and means from one time to another are often thoroughly examined in standard texts for very simple designs when the population is assumed to be infinite and composed of the same units at times $t_1$ and $t_2$. See, e.g., Cochran (1977) and Raj (1968). At the SRU one line of research aims at creating efficient strategies for estimating parameters of change and levels when the population is finite and changeable, that is, units may have joined or left the population between $t_1$ and $t_2$.

The initial work concentrated on the following specific situation.

The changeable population is called U at $t_1$ and U' at $t_2$ (with size N and N' respectively), while the variable under study is called x at $t_1$ and y at $t_2$. At time $t_2$ it is assumed that $N_1$ units have left U and $N_2$ units have joined it - the corresponding subpopulations are called $U_1$ and $U_2$, respectively. The new population U' contains

$N' = N_{12} + N_2$ units, where $N_{12}$ is the number of units in $U_{12}$ (the intersection of U and U').

At time $t_1$ a probability sample s of fixed size n is drawn. At time $t_2$ this sample is made up of two subsamples, viz. $s_1$ of size $n_1$ from $U_1$ and $s_{12}$ of size $n_{12}$ from $U_{12}$, where $n_1 + n_{12} = n$.

At time $t_2$ three probability samples are selected: $s_{12m}$ (m = matched) of size $n_{12m}$ from $s_{12}$, $s_{12u}$ (u = unmatched) of size $n_{12u}$ from $U_{12} - s_{12}$, and $s_2$ of size $n_2$ from $U_2$. The total sample at $t_2$ is denoted s' and its size n'.

Forsman and Garås (1982 a,b) considered the case when

(a) s, $s_{12m}$, $s_{12u}$ and $s_2$ are simple random samples drawn without replacement

(b) the variances of x and y in $U_{12}$ are equal,

(c) $n_{12u} = n_{12} - n_{12m}$,

(d) $n_{12}$ is fixed.

For given $\theta = n_{12u}/n_{12}$, they gave minimum variance unbiased estimators (restricted to a certain linear class) and their variances for estimation of change and levels. They also determined optimal values of θ.

Current work concentrates on

(i) variance estimation,
(ii) the use of auxiliary information z through regression estimation,
(iii) alternative sample designs,
(iv) sampling on more than two occasions.

## 5. Small domain estimation

In Sweden, as in many other countries, there is a growing demand for small domain statistics (administrative regions as well as other types of domains). Nation-wide samples, however, are likely to give poor estimates for small domains, because the sample may contain only a few observations from any given domain. Methods to overcome this difficulty by utilizing auxiliary information have recently begun to come into focus throughout the world.

Since in Sweden quite a lot of computerized registers are available,it seems that conditions would be favorable for producing small domain statistics through a combined use of registers and sample survey data. Endeavours in this direction are, however, still in their infancy, and considerable methodological research is needed, before any ideas can be put into practice. The SRU is presently involved in three methodological studies on small domain estimation with application to household statistics.

The first study was concerned with the so-called SPREE (Structure Preserving Estimate) methods suggested by Purcell (1979). Briefly, these methods work by iteratively adjusting cell frequencies of a contingency table to agree with known marginal totals.

The problem was to estimate, for each municipality (=domain) d, the cell frequencies $N_{hid}$ of a two-dimensional contingency table, where $N_{hid}$ denotes the number of households in municipality d belonging to class h with respect to size of dwelling, and to class i with respect to size of household. The study concentrated on the model-bias of the methods, and no samling aspects were involved. Marginal frequencies $N_{h.d} = \sum_i N_{hid}$ and $N_{.id} = \sum_h N_{hid}$ were assumed known exactly for all municipalities in the population, as well as $N_{hi.} = \sum_d N_{hid}$. The population was a miniature population of 28 municipalities, with data taken from the 1975 Census of Population and Housing. Data from the 1970 census were also used to generate starting-points for the iterative procedure. The resulting estimates $\hat{N}_{hid}$ were compared with the known true values $N_{hid}$. Some findings of the study are reported in Statistics Sweden (1983 a).

The second study, which is now underway, is concerned with estimating, for each municipality (=domain) d, the quantity $N_{.id}$, which here denotes the number of people in municipality d belonging to class i with respect to cohabitational status. Here, an associated dimension to be utilized is given by the classification (indexed by h) of people with respect to sex and age in combination, which makes $N_{.id} = \sum_h N_{hid}$. Estimates $\hat{N}_{hi.}$ for the whole population of municipalities are assumed available, and frequencies $N_{h.d} = \sum_i N_{hid}$ and $N'_{hid}$ are assumed known, where $N'_{hid}$ denotes the value from the latest census.

A Monte Carlo study is presently being accomplished with the purpose of comparing alternative methods for estimating $N_{.id}$ by means of sample data combined with the auxiliary information just described. The methods to be considered includes: Poststratified estimation, Synthetic estimation, Generalized regression estimation (Särndal 1981), and SPREE estimation.

A third study is being designed in order to study the generalized regression method for estimating the number of unemployed people in municipality d, denoted $N_{.id}$ (where class i is the class of unemployed people). The auxiliary information to be utilized refers to a classification of people (as in the second study) by sex and age in combination.

## 6. Estimation of regression coefficients for domains

The research described in this section links up with the small domain estimation described in Section 5 and with the generalized regression approach proposed by Cassel et al (1976) and Särndal (1982). The question that initiated the work was: Given sample survey data from a population divided into many domains, how should we make inference, domain by domain and in the standard design-based (randomization theory) fashion, about finite population characteristics describing the relationship between a criterion variable y and explanatory variables $x_1, \ldots, x_r$, such as regression slopes?

The possible shortage of observations in any given domain poses a difficulty which might be overcome by exploiting auxiliary information. To this end, Elvers et al. (1983) propose two general methods, which are also explored by means of Monte Carlo experiments. They consider the case with several explanatory variables and estimation of multiple (possibly weighted) regression coefficients. Here, however, we limit ourselves to describing the ideas for the simple case when there is only one explanatory variable, and when the population characteristic to be estimated is a simple (ordinary least squares) regression slope within a domain.

Let U be a finite population of labeled units, $U = \{1,\ldots, k,\ldots,N\}$, divided into D nonoverlapping domains $U_{d.}$ of size $N_{d.}$, where $d = 1,\ldots, D$, and $\sum_{d=1}^{D} N_{d.} = N$. Let (for $k = 1,\ldots,N$) $y_k$ and $x_k$ denote the values of unit k with respect to the criterion variable and the explanatory variable, respectively.

The population characteristic to be estimated is, for $d = 1,\ldots, D$, the least squares regression slope

$$B_d = \frac{N_{d.} \sum_{U_{d.}} x_k y_k - (\sum_{U_{d.}} x_k)(\sum_{U_{d.}} y_k)}{N_{d.} \sum_{U_{d.}} x_k^2 - (\sum_{U_{d.}} x_k)^2}$$

$$= \frac{N_{d.} T_{dxy} - T_{dx} T_{dy}}{N_{d.} T_{dxx} - T_{dx}^2} ,$$

where $T_{dx} = \sum_{U_{d.}} x_k$, $T_{dy} = \sum_{U_{d.}} y_k$, $T_{dxy} = \sum_{U_{d.}} x_k y_k$, and $T_{dxx} = \sum_{U_{d.}} x_k^2$. ($\sum_A$ denotes sum over k in the set A.) Thus, the problem here is to estimate finite population characteristics, not "superpopulation" model parameters.

A probability sample s of fixed size n is drawn from U by a sampling design p(s) with inclusion probabilities $\pi_k > 0$ ($k = 1,\ldots, N$). The part of s that happens to fall within domain $U_{d.}$ is denoted $s_{d.}$ of random size $n_{d.}$ ($\sum_{d=1}^{D} n_{d.} = n$). Data ($y_k$, $x_k$) are observed for all $k \in s$.

The problem of estimating $B_d$ is considered as the problem of estimating the sums $T_{dx}$, $T_{dy}$, $T_{dxy}$, and $T_{dxx}$. In addition to the method of straightforward $\pi$-inverse weighting of sampled units (here called the C method;

C for Common) two other estimation methods are examined (the P and F methods; P for Product variable, and F for First order variable), which utilize auxiliary information of a special kind. The randomization theory approach is adhered to inasmuch as (1) adjustment for varying inclusion probabilities is carried out through $\pi$-inverse weighting of sampled units, and (2) the design variance is used to assess the efficiency of the estimators.

With the C method, $B_d$ is estimated by

$$\hat{B}_{dC} = \frac{\hat{N}_{d.}\,\hat{T}_{dxyC} - \hat{T}_{dxC}\,\hat{T}_{dyC}}{\hat{N}_{d.}\,\hat{T}_{dxxC} - \hat{T}^2_{dxC}} ,$$

where

$$\hat{N}_{d.} = \sum_{s_{d.}} \frac{1}{\pi_k} , \quad \hat{T}_{dxC} = \sum_{s_{d.}} \frac{x_k}{\pi_k} , \quad \hat{T}_{dxyC} = \sum_{s_{d.}} \frac{x_k y_k}{\pi_k} ,$$

and analogously for $\hat{T}_{dyC}$ and $\hat{T}_{dxxC}$.

The P and F methods both try to exploit the homogeneity gained by an a priori known grouping of the population units. The population is assumed to be divided into G mutually exclusive groups $U_{.g}$ of size $N_{.g}$, which cut across the domains $U_{d.}$ ($g = 1,\ldots, G$; $\sum_{g=1}^{G} N_{.g} = N$). Domains crossed with groups divide the population into DG cells $U_{dg}$ of size $N_{dg}$ ($\sum_{d=1}^{D} \sum_{g=1}^{G} N_{dg} = N$). These $N_{dg}$'s are the auxiliary quantities that must be known, from censuses or other sources, in order to make the P and F methods work.

The P method uses the principle of generalized regression estimation under the assumption that the groups ($g=1,\ldots, G$) explain the variation in both the x and the y variable, as well as in the product variables xy and $x^2$. The resulting estimator of $B_d$ is

$$\hat{B}_{dP} = \frac{N_{d.}\,\hat{T}_{dxyP} - \hat{T}_{dxP}\,\hat{T}_{dyP}}{N_{d.}\,\hat{T}_{dxxP} - \hat{T}^2_{dxP}} ,$$

where

$$\hat{T}_{dxP} = \hat{T}_{dxC} + \textstyle\sum_{g=1}^{G} \tilde{x}_{s_{.g}} \,(N_{dg} - \hat{N}_{dg}) ,$$

$$\hat{T}_{dxyP} = \hat{T}_{dxyC} + \textstyle\sum_{g=1}^{G} \widetilde{xy}_{s_{.g}} \,(N_{dg} - \hat{N}_{dg})$$

and analogously for $\hat{T}_{dyP}$ and $\hat{T}_{dxxP}$ . Here,

$$\hat{N}_{dg} = \textstyle\sum_{s_{dg}} \frac{1}{\pi_k} , \quad \tilde{x}_{s_{.g}} = (\textstyle\sum_{s_{.g}} \frac{k}{\pi_k})/(\textstyle\sum_{s_{.g}} \frac{1}{\pi_k}) ,$$

$$\widetilde{xy}_{s_{.g}} = (\textstyle\sum_{s_{.g}} \frac{x_k y_k}{\pi_k})/(\textstyle\sum_{s_{.g}} \frac{1}{\pi_k})$$

(and analogously for $\tilde{y}_{s_{.g}}$ and $\widetilde{x^2}_{s_{.g}}$ ).

The F method differs from the P method only in the estimation of $T_{dxy}$ and $T_{dxx}$. To obtain the estimator $\hat{B}_{dF}$, the quantities $\widetilde{xy}_{s_{.g}}$ and $\tilde{x}^2_{s_{.g}}$ in the expressions for $\hat{T}_{dxyP}$ and $\hat{T}_{dxxP}$, respectively, should be replaced by $\tilde{x}_{s_{.g}}\,\tilde{y}_{s_{.g}}$ and $\tilde{x}^2_{s_{.g}}$ .

Elvers et al. (1983) also give methods for estimating the design variances $V_p(\hat{B}_{dI})$ (I = C, P, F). This makes it possible to calculate confidence intervals

$$\hat{B}_{dI} \pm z_{1-\alpha/2} \{\hat{V}_p(\hat{B}_{dI})\}^{1/2} \quad (I = C, P, F) ,$$

$z_{1-\alpha/2}$ being the normal score, although the actual coverage probability will most likely differ from the $100(1-\alpha)$ % confidence level that was aimed at.

Monte Carlo experiments were designed to assess the variance reductions realizable by the P and F methods over the simple C method. Three populations, called REAL, ART1, and ART2, were used. REAL consisted of real data on 1202 Swedish households divided into D = 24 domains by Sweden's major administrative regions ("län") and into G = 5 groups by size and age characteristics of the household; x and y were, respectively, disposable household income and taxable household income. These groups were a rather weak explanatory factor for x as well as for y, so a priori the structure of REAL does not strongly favor the P and F methods. The artificial populations ART1 and ART2 were therefore created to have the same cell frequencies as REAL, but a smaller within group variance, relative to the between group variance, in x as well as in y. ART2 was created to provoke a situation where extremely large efficiency gains are expected from the P and F methods. The domains varied in size from 20 % to about 1 % of the total population.

1000 repeated simple random samples of size n = 300 were drawn. For each sample and each domain, $\hat{B}_{dI}, \hat{V}_p(\hat{B}_{dI})$, and the confidence interval $\hat{B}_{dI} \pm 1,96\{\hat{V}_p(\hat{B}_{dI})\}^{1/2}$ were calculated by each of the methods: I = C,P,F. Summary statistics for the 1000 repetitions were calculated, including (for each method separately) mean and variance of the 1000 $\hat{B}_d$-values, mean of the 1000 $\hat{V}$-values, and coverage rate of the 1000 confidence intervals (i.e., the percentage of the 1000 samples which gave a confidence interval covering the true value $B_d$).

The tree methods C, F, and P shared the following features: (1) The mean of the 1000 $\hat{B}_d$-values differed but little from the true value $B_d$; (2)The variance of the 1000 estimates agreed well with the average estimated variances in the larger domains, but differed markedly in some of the smaller ones; (3) The achieved coverage rates were close to (but always somewhat short of) the nominal 95 % in the larger domains, but considerably less in the smallest domains.

The following emerged in the comparison of the three methods: (4) The P and F methods performed very similarly for all three populations, in terms of variance as well as coverage rate; (5) The variance reductions realized by the P and F methods over the C method were modest for most domains in the REAL population (0 % - 30 %), strong in virtually all domains in the ART1 population (20 % - 60 %), and dramatically large in all domains in the ART2 population (over 90 %).

## 7. Measures for reducing nonresponse rates

During the 70's Statistics Sweden noted an increase in nonresponse rates but the problem seemed to be restricted to surveys of individuals and households. It is a general fact that all subject-matter departments have been forced to resort to increased "salesmanship" to cope with the growing resistance among respondents, but in our surveys of individuals and households these efforts did not prove adequate.

Statistics Sweden carries out a number of continuous surveys of individuals and households. The nonresponse level for these surveys is high. The total nonresponse rate is for the Labor Force Survey 6-7 %, for the Survey of Consumer Buying Expectations 14-15 %, for the Survey of Living Conditions 14-15 %, for the Household Income Survey 12-13 % and for the Survey of Household Energy Consumption 14-15 %.

Generally speaking, we noted an increasing rate until 1976 with levels higher than those reported above. Over the last five years the situation has stabilized and the rates have been decreasing slightly for most surveys. Due to our efficient tracking procedures the nonresponse problem is mainly one of refusal. Those not-at-home play an important role, though, in surveys with a tight time schedule such as the Labor Force Survey.

The causes of this still irritating nonresponse situation are not quite clear. However, starting with the 1970 Census of Population, the media have taken an interest in aims and methods for surveys and censuses resulting in public debates on invasion of privacy. Furthermore, Sweden's Data Act imposes restrictions the activities of Statistics Sweden. For instance, it must be explicitly stressed in an advance letter that respondent cooperation is voluntary if the survey is a non-mandatory one. Now and then, Statistics Sweden suffers from highly adverse publicity. It happens that leading newspapers propagate for stalling and non-cooperation among potential respondents. Such incidents are rare, though.

The situation around 1974-75 called for special action by Statistics Sweden. The challenge was met by starting a nonresponse project comprising three subprojects: (a) Information to media and respondents, (b) Forms and strategies for data collection, and (c) Statistical methods and techniques. Here some of these efforts are briefly presented. W'thin project (a) the following is worth mentioning:

- A public relation manual. A PR manual has been compiled. It is intended for use by the survey designers, and it covers various aspects of the relationship between survey designers and respondents as well as between survey designers and media.

- The image of Statistics Sweden. Statistics Sweden continuously tries to improve its relations to respondents and media. The protection of privacy is an important task for society according to a survey on privacy and confidentiality carried out in 1976 (Wärneryd (1977)). It seems, although it cannot be formally verified, as if it still is. The signs speak for themselves. Media take a great interest in privacy and survey policy matters, and people avail themselves of the right to obtain file excerpts more now than a few years ago. Furthermore Sweden has a very active Data Inspection Board. As a matter of fact Statistics Sweden has been forced to create a special service unit for Data Act matters. Therefore, we have decided not to fight media when it comes to adverse publicity, since such battles are likely to be lost, but rather to provide media and the public with good information

- Refusal surveys. Statistics Sweden has conducted two small exploratory interview surveys on respondent reasons for refusal in the Survey of Living Conditions and the Labor Force Survey, reported in Bergman et al (1978). Of course, there emerged a great number of such reasons, often interacting ones, and the refusers turned out to be a very heterogeneous group with respect to personality characteristics and living conditions.

Within project (b) the following is worth mentioning:

- Incentives. According to the Data Act each individual has the right to demand and receive a print out from his own data file stored at Statistics Sweden. Many people avail themselves of this right. It has been suggested that offering file excerpts to respondents should be put to a test in order to find out if such an incentive could influence the response rate. A split-plot experiment was conducted where one half of the sample received the usual advance letter while the experimental group got a more informal information material and a chance to request

survey results and/or their own file excerpts. As expected, many (50%) made the request. The nonresponse rate was 20.5 % in this group compared with 23.3 % in the control group. One might conclude that this effort had a positive effect on the nonresponse rate. It is impossible, though, due to deficiencies in the experimental design, to distinguish which parts of the effort that contributed most to the result obtained.
Money is seldom used in our surveys as a means to gain respondent cooperation. It is used in the intermittent Household Expenditure Survey because of the extremely heavy respondent burden in that survey involving 2 or 4 weeks of diary keeping. Experimental pretests have shown that this kind of incentive has small but quite perceptible positive effects on the nonresponse rate. However, it is usually too costly to be considered a standard means for reducing nonresponse rates in our surveys.

- Brief reminders. Experiments with different call back procedures have been conducted in some mail surveys. The experiments revealed similar results. (1) It is possible to get a faster inflow of questionnaires by sending out a brief reminder a few days after the initial mailing. (2) The brief reminder is a post-card which is sent to all respondents whether they have answered the initial questionnaire or not. Thus costs can be reduced compared with conventional call back. (3) An intensive call back procedure does not have any negative effects on later follow up attempts. (4) Unfortunately, brief reminders do not decrease the final nonresponse rate compared with conventional call back procedures.

- The interviewer organization. When studying nonresponse problems at Statistics Sweden we have found that the lack of a formal quality control program for interviewer performance may contribute to the variability between individual interviewer nonresponse rates. Naturally these rates vary between rural and urban areas, but they also differ considerably within limited geographical areas, differences that must be related to the interviewers themselves and their different attitudes and approaches to interviewing. The nonresponse rates for some interviewers are very high. In Table 1 the average annual nonresponse rates in the 1979 Labor Force Survey are given for the "best" and the "worst" interviewer in each of our 25 counties. The same general pattern is found in other surveys as well.

Table 1 1979 average nonresponse rates for interviewers in the Labor Force Survey: the "best" and the "worst" interviewer and "interviewer quartiles" (%).

| County | "Best" | "Worst" | "Best" quartile | "Worst" quartile | No of inter-viewers |
|---|---|---|---|---|---|
| 1 | 3 | 17 | 5 | 14 | 19 |
| 2 | 3 | 17 | 5 | 15 | 16 |
| 3 | 2 | 12 | 3 | 11 | 8 |
| 4 | 4 | 12 | 4 | 11 | 8 |
| 5 | 2 | 14 | 2 | 11 | 15 |
| 6 | 2 | 9 | 2 | 7 | 12 |
| 7 | 1 | 4 | 2 | 4 | 7 |
| 8 | 1 | 6 | 1 | 5 | 8 |
| 9 | 2 | 3 | - | - | 3 |
| 10 | 2 | 4 | 2 | 4 | 6 |
| 11 | 1 | 11 | 1 | 8 | 12 |
| 12 | 2 | 14 | 4 | 12 | 23 |
| 13 | 2 | 6 | 2 | 5 | 8 |
| 14 | 1 | 16 | 2 | 14 | 23 |
| 15 | 1 | 8 | 1 | 6 | 12 |
| 16 | 2 | 9 | 2 | 7 | 9 |
| 17 | 0* | 13 | 1 | 9 | 12 |
| 18 | 1 | 4 | 2 | 3 | 8 |
| 19 | 1 | 7 | 2 | 6 | 12 |
| 20 | 1 | 3 | 1 | 3 | 9 |
| 21 | 1 | 4 | 2 | 3 | 11 |
| 22 | 1 | 7 | 2 | 6 | 11 |
| 23 | 0* | 8 | 1 | 7 | 6 |
| 24 | 1 | 7 | 2 | 6 | 8 |
| 25 | 1 | 11 | 2 | 9 | 9 |

* Zero nonresponse means that the rate is less than .5 %.

As can be seen from Table 1 the range in nonresponse rates between the "best" and the "worst" interviewer is quite large in many counties. If we investigate the difference between the "best"and the "worst" interviewer quartiles, the interquartile range, we find a similar pattern.

Thus we have one group of interviewers with very low nonresponse rates and one with high rates. It is quite clear that part of the difference can be explained by factors beyond the interviewer's control.
Part of it, though, must be due to divergencies in utilizing the time schedule (some interviewers start working with the assignments too late), in tracking down respondents and in approaching respondents.

The present situation characterized by lack of formal control must be changed. Therefore we are trying to develop a more tight program for interviewer control (see Section 10 ). It will contain systems for continuous reinterviews, observations in the field, close follow-up of interviewer performance including comparisons with established quality standards and extended training and education.

Within project (c) the following is worth mentioning:

- Handbook on statistical methods. This handbook contains an overview of statistical methods and techniques available for coping with nonresponse problems. The English translation of the title of the book is "You can count on nonresponse" (which is supposed to be a crack joke in Swedish). The handbook covers topics such as subsampling among nonrespondents (with nonresponse in the subsample), weighting and imputation, item nonresponse and methods for evaluating the effect of nonresponse.

- The dividing up of questionnaires. A common reason for refusing to cooperate in surveys is lack of time. Many of these refusers are, at least in principle, prepared to cooperate, but they find it troublesome to be burdened by, say, a one hour or ninety minute interview. An obvious conclusion from this is that a reduction of the length of the questionnaire might have a positive effect on the nonresponse rate. Naturally, such a conclusion might lead to an initial limitation of the number of questions when the survey is planned. If a reasonable level of response burden is the result, everything is fine. If, on the other hand, the number of questions is still too large, we might consider the possibility of reducing the response burden by dividing up the questionnaire.

The situation may be illustrated by the following example:

Suppose that a survey aims at studying the distribution of 120 variables in a population. Suppose also that to each variable there is associated a specific question in the questionnaire. In a conventional survey design a random sample of n units is drawn from the population. Then all n units in the sample are supposed to answer all the 120 questions.

Now suppose that the 120 variables (and the corresponding questions) can be divided into three sets - x, y, and z - where for example the first set (x) contains 40 of the variables, the second set (y) 40 other variables and the third set (z) contains the remaining 40 variables. Suppose, finally, that the x-set contains the variables which are considered the most important according to some specified criterion.

In this situation two questionnaires, A and B, might be constructed. A contains the x- and y-sets and B contains the x- and z-sets. The sample of n units is now divided randomly into two groups. The A-questionnaire is administered to the first group of $n_1$ units and the B-questionnaire to the remaining $n_2$ units. Naturally, other sets and combinations of sets are conceivable.

Many interesting (and difficult) methodological problems occur in this procedure, for example: Which combination of sets is the most effective? How shall the sample be allocated optimally between the sets?

Some work has been done in this field. Hocking and Smith (1972) solved the following problem:

Suppose that the questionnaire consists of p=r+s questions. The r questions need a large sample if we want an acceptable precision, while the remaining s questions can be estimated from a smaller sample. For instance, $n_p$ units get a complete questionnaire while $n_r$ units get a reduced one comprising r questions. How, then, shall the sample sizes $n_p$ and $n_r$ be determined in order to obtain the precision aimed at and an acceptable reduction of the respondent burden?

In Hocking (1979) the results of Hocking and Smith were developed for the general cases of complete and reduced questionnaires applied to stratified sampling designs.

In most surveys it is of interest to estimate relationships between variables, e.g. through the estimation of cell frequencies for contingency tables. When dividing up a questionnaire it is therefore necessary to make sure that such relations can be estimated with sufficient precision. In the statistical literature different methods are described to estimate the value of a contingency table cell when supplementary information about one or two marginal distributions is available. The statistical properties of the raking ratio method in simple random sampling (SRS) were first investigated by Ireland and Kullback (1968). The method can be used in most sample designs even though their statistical properties in designs other than SRS are not known completely. Other methods have been described by Stephan (1942), Smith (1947), El-Badry and Stephan (1955) and Hocking and Oxspring (1971). In this context we also want to draw attention to methods for small domain estimation, discussed in Sections 5 and 6.

At Statistics Sweden we intend to penetrate this problem further.

- <u>Negatively coordinated samples</u>. Several thousand individuals might be selected for surveys carried out by Statistics Sweden at least twice within a twelve month period provided nothing is done to prevent it. As for prevention, Statistics Sweden has started to develop routines for negatively coordinated sampling. The ultimate aim is that no person should be included in a sample for an individual or household survey more than once during, say, a period of five years. The actual length of period is arbitrary, but should be based on the assumption that after such a period the respondent would not experience participation in an earlier survey as a burden.

For the moment, three surveys with heavy respondent burden, the Household Expenditure Survey, the Survey of Living Conditions and the Household Energy Consumption Survey, are negatively coordinated for five years. We are planning to gradually extend this coordination in order to include more surveys and hopefully also samples selected for commercial purposes such as samples sold to other private survey institutes.

- *A bibliography on nonresponse*. One by-product of the project work on nonresponse problems is a bibliography on nonresponse and related topics. The bibliography items are of different kinds. First, there are items whose title gives definite indication of a treatment of nonresponse problems. Secondly, there are items whose title contains a more indirect reference to nonresponse. Titles in this latter category contain key words such as "privacy", "respondent burden", "data collection methods" and "incentives". Thirdly, there are items where the type of indications mentioned do not exist, but where inclusion was nevertheless deemed important. The bibliography (Bogeström et al (1981)) contains about 1200 titles but is of course by no means to be considered complete.

## 8. Inference in the presence of varying response probabilities

The conceptual set-up for the nonresponse work on statistical methods and techniques described in the previous section was classical design-based sampling theory coupled with the most simple (non)response model, postulating a deterministic response behavior. In the words of Cochran (1977): "In the study of nonresponse it is convenient to think of the population as divided into two 'strata', the first consisting of all units for which measurements would be obtained if the units happened to fall in the sample, the second of the units for which no measurements would be obtained." The first stratum (the response stratum) thus consists of the units supposed to respond with probability 1, while the second stratum (the nonresponse stratum) consists of the units supposed to respond with probability 0.

The adoption of this simple response model places the survey statistician in a very uneasy position since she (he, for short) is supposed to provide reliable estimates for the entire population while classical design-based theory only allows valid inferences to the response stratum. If he wants to adhere to this response model, the collected data from the response stratum must be supported with some kind of model, for example a model relating the survey variable to available auxiliary information supposed to be valid for the response stratum as well as the nonresponse stratum. Since models might be misspecified, substantial (and unknown) bias may easily result for any sizeable nonresponse rate. Furthermore, different models may have to be used for different survey variables, resulting in highly increased data processing costs.

The above simple response model seems to have attracted much attention among survey statisticians during many years. It is somewhat surprising to find that many statisticians stick to a deterministic model, especially since nondeterministic models for response behavior have been known for a long time, e.g. Politz and Simmons (1949, 1950), Deming (1953). Furthermore, the simple response model certainly does not adequately reflect most survey situations. Cochran (op. cit) also admits: "This division into two distinct strata is, of course, an oversimplification. Chance plays a part in determining whether a unit is found and measured in a given number of attempts. In a more complete specification of the problem we would attach to each unit a probability representing the chance that it would be measured by a given field method if it fell in the sample."

During the last few years nondeterministic response models are more frequently found in the literature. At Statistics Sweden we are carrying out work to establish results for design-based and design-model-based inference in the presence of varying response probabilities. The work started while Carl-Erik Särndal was head of SRU, and he continues to cooperate with us on a consulting basis.

It is obviously not possible in this context to give a full account of the results attained hitherto. We will merely indicate a few points of departure and give an example.

Let $U = \{1, \ldots, k, \ldots, N\}$ be the finite population of size N. We want to estimate the total $T = \sum_1^N Y_k$ from a probability sample s, selected according to the design p(s) of fixed size n, characterized by the probabilities of inclusion $\pi_k = Pr(k \varepsilon s) > 0$ and $\pi_{k\ell} = Pr(k \text{ and } \ell \varepsilon s) > 0$.

Due to nonresponse we only observe $m < n$ objects, which thus form the response set $r \subset s$. Suppose the response mechanism is such that the objects respond independently of each other and of s. Let $q_k = Pr(k \varepsilon r | k \varepsilon s)$ (by assumption $> 0$) be the probability that object k responds $(k = 1, \ldots, N)$. If the response probabilities are known it is easy to give an unbiased estimator of T, and to determine its variance and an unbiased variance estimator. This will, of course, never be the case in a real-life application.

Now suppose that the population U is comprised of G groups, with $N_g$ (unknown) objects in group $U_g$, and let $q_k = \theta_g$ if $k \varepsilon U_g$ ($g = 1, \ldots, G$). These groups may be referred to as response homogeneity groups, since objects in the same group have the same response probability, while objects from different groups have different response probabilities. Further, suppose that we can correctly classify every object in s with respect to homogeneity group.

Let $s_g = s \cap U_g$, $r_g = r \cap U_g$, let $n_g$ be the number of objects in $s_g$, $m_g$ the number of objects in $r_g$, $\underset{\sim}{n} = (n_1, \ldots, n_G)$ and $\underset{\sim}{m} = (m_1, \ldots, m_G)$. Further, let

$$\pi_{k|s,\underset{\sim}{m}} = \Pr(k \varepsilon r | k \varepsilon s, \underset{\sim}{m})$$

and

$$\pi_{k\ell|s,\underset{\sim}{m}} = \Pr(k \text{ and } \ell \varepsilon r \,|\, k \text{ and } \ell \varepsilon s, \underset{\sim}{m})$$

Then, since r for $\underset{\sim}{n}$ and $\underset{\sim}{m}$ fixed is a stratified simple random without replacement sample from s,

$$\pi_{k|s,\underset{\sim}{m}} = \frac{m_g}{n_g} \quad \text{if } k \varepsilon r_g$$

and

$$\pi_{k\ell|s,\underset{\sim}{m}} = \begin{cases} \dfrac{m_g}{n_g} \dfrac{m_g - 1}{n_g - 1} & \text{if } k \text{ and } \ell \ \varepsilon \ r_g \\[2ex] \dfrac{m_g}{n_g} \dfrac{m_{g'}}{n_{g'}} & \text{if } k \text{ and } \ell \text{ belong to different groups} \end{cases}$$

Finally let $\pi_k^* = \pi_k \pi_{k|s,\underset{\sim}{m}}$ and $\pi_{k\ell}^* = \pi_{k\ell}\pi_{k\ell|s,\underset{\sim}{m}}$, and let the sample size n be such that $\Pr(n_g = 0) \simeq \Pr(m_g = 0) \tilde{=} 0$ ($g = 1, \ldots, G$). Then the population total T is unbiasedly estimated by

$$\hat{T}^* = \sum_{k \varepsilon r} Y_k / \pi_k^* \tag{8.1}$$

with variance

$$V(\hat{T}^*) = E_s \, E_{\underset{\sim}{m}} (\cdot) - T^2 , \tag{8.2}$$

where

(8.3) $$E_s E_{\underset{\sim}{m}} (\cdot) = E_s E_{\underset{\sim}{m}} \{ \sum_{k\varepsilon s} (\frac{Y_k}{\pi_k})^2 / \pi_{k|s,\underset{\sim}{m}} + \sum\sum_{\substack{k\neq\ell \\ \varepsilon s}} \frac{\pi_{k\ell|s,\underset{\sim}{m}}}{\pi_{k|s,\underset{\sim}{m}} \pi_{\ell|s,\underset{\sim}{m}}} \frac{Y_k}{\pi_k} \frac{Y_\ell}{\pi_\ell} \} .$$

If $\Pr(n_g \leqq 1) = \Pr(m_g \leqq 1) = 0$ $(g = 1,\ldots, G)$ an unbiased estimator of $V(T^*)$ is

(8.4) $$\hat{v}(T^*) = \sum_{k\varepsilon r} \frac{1-\pi_k^*}{\pi_k^{*2}} Y_k^2 + \sum\sum_{\substack{k\neq\ell \\ \varepsilon r}} (1 - \frac{\pi_k^* \pi_\ell^*}{\pi_{k\ell}^*}) \frac{Y_k}{\pi_k^*} \frac{Y_\ell}{\pi_\ell^*} .$$

Obviously, if the response model is correct, the above results hold for any survey variable, allowing a uniform treatment of all variables. The results, of course, are model dependent. However, this will be the case for any survey situation facing nonresponse. Thus, the survey statistician cannot normally hope to act in a completely model-free context,an impression he might achieve as a student of classical design-based sampling theory.

Even if we make a very careful choice of a, hopefully realistic, response model, there is a risk of specification errors. This implies that we perhaps should try to use models that connect the survey variables with known auxiliary information. However, in order not to make a situation that may seem shaky even worse, this modeling should not call for model-dependent, but design-model-based inference. Some such approaches (not including the problem of nonresponse, though) have lately been formulated, e.g. Särndal (1982).

At SRU we are at present, in collaboration with Carl-Erik Särndal, incorporating certain nonresponse theory with the general randomization theory approach discussed in Särndal (1982). We will here give a short account of some results.

Let the population, the design and the response mechanism be as described above. Let $\underset{\sim}{C}_N = (c_{kq})$ be an NxQ matrix of constants, and $\underset{\sim}{Y}_N = (Y_1,\ldots,Y_N)'$. We seek estimates of the Q linear functions $T_q = \sum_1^N c_{kq} Y_k$ $(q = 1,\ldots, Q)$ or, equivalently,of the Q-vector $\underset{\sim}{T} = (T_1,\ldots, T_Q)' = \underset{\sim}{C}_N' \underset{\sim}{Y}_N$.

The values $Y_k$ for $k\varepsilon r$ are observed; the values $Y_k$, $k\varepsilon$ U-r, remain unknown. However, auxiliary information is assumed available in the form of known R-vectors $\underset{\sim}{x}_1,\ldots, \underset{\sim}{x}_N$.

The estimator $\hat{\underset{\sim}{T}}{}^*$ proposed by (8.9) below uses the observed values $Y_k$ for $k\varepsilon r$, as well as fitted values $\hat{Y}_k = \underset{\sim}{x}_k' \hat{\underset{\sim}{B}}_r$ for $k = 1,\ldots, N$. Given by (8.7) below, $\hat{\underset{\sim}{B}}_r$ results from fitting the model $\xi$, such that the $Y_k$ are independent (throughout) and

$$E_\xi(Y_k) = \underset{\sim}{x}_k' \underset{\sim}{\beta} = \sum_{t=1}^{R} x_{kt}\beta_t; \quad V_\xi(Y_k) = \sigma_k^2 \tag{8.5}$$

to the m sample data points $(Y_k, \underset{\sim}{x}_k)$ for $k\varepsilon r$. The $\sigma_k^2$ may be known or unknown. If unknown, the $\sigma_k^2$ are assumed to have a structure which leaves $\hat{\underset{\sim}{B}}_r$, $\underset{\sim}{C}_N$ and $\underset{\sim}{G}_r$ below unaffected by unknowns.

Now (8.5) can also be written $E_\xi(\underset{\sim}{Y}_N) = \underset{\sim}{X}_N\underset{\sim}{\beta}$; $V_\xi(\underset{\sim}{Y}_N) = \underset{\sim}{\Sigma}_N$, where $\underset{\sim}{X}_N$ is the NxR matrix whose rows are $\underset{\sim}{x}_k'$ $(k = 1,\ldots, N)$, and $\underset{\sim}{\Sigma}_N$ = diag $\{\sigma_k^2 : k = 1,\ldots, N\}$. Moreover, let $\underset{\sim}{X}_r$, $\underset{\sim}{Y}_r$ and $\underset{\sim}{\Sigma}_r$ be the respective sample parts of $\underset{\sim}{X}_N$, $\underset{\sim}{Y}_N$ and $\underset{\sim}{\Sigma}_N$. Define also $\Pi_r^*$ = diag $\{\pi_k^*; k\varepsilon r\}$ and

$$\underset{\sim}{A}_{xr} = \underset{\sim}{X}_r' \underset{\sim}{\Sigma}_r^{-1} \underset{\sim}{\Pi}_r^{*-1} \underset{\sim}{X}_r \; ; \quad \underset{\sim}{A}_{Yr} = \underset{\sim}{X}_r' \underset{\sim}{\Sigma}_r^{-1} \underset{\sim}{\Pi}_r^{*-1} \underset{\sim}{Y}_r \; . \tag{8.6}$$

Introduce

$$\hat{\underset{\sim}{B}}_r = \underset{\sim}{A}_{xr}^{-1} \underset{\sim}{A}_{Yr} \tag{8.7}$$

which estimates the regression coefficient vector

$$\underset{\sim}{B} = (\underset{\sim}{X}_N' \underset{\sim}{\Sigma}_N^{-1} \underset{\sim}{X}_N)^{-1} \underset{\sim}{X}_N' \underset{\sim}{\Sigma}_N^{-1} \underset{\sim}{Y}_N \tag{8.8}$$

arising in the hypothetical generalized least squares 'census fit' of (8.5) to the N data points of the entire finite population.

Let $\hat{\underset{\sim}{Y}}_N = \underset{\sim}{X}_N \hat{\underset{\sim}{B}}_r$ be the N-vector of fitted Y-values $\hat{Y}_k = \underset{\sim}{x}_k' \hat{\underset{\sim}{B}}_r$, $k = 1,\ldots, N$. Let $\hat{\underset{\sim}{Y}}_r = \underset{\sim}{X}_r \hat{\underset{\sim}{B}}_r$ (mx1) and $\underset{\sim}{C}_r$ (mxQ) be the sample parts of $\hat{\underset{\sim}{Y}}_N$ and $\underset{\sim}{C}_N$, respectively.

As an estimator of T we propose

(8.9) $\hat{\underset{\sim}{T}}^{*} = \underset{\sim}{C}_r' \underset{\sim}{\Pi}_r^{*^{-1}} (\underset{\sim}{Y}_r - \hat{\underset{\sim}{Y}}_r) + \underset{\sim}{C}_N' \hat{\underset{\sim}{Y}}_N$

which - if the response model holds - is approximately design unbiased.

To emphasize $\hat{\underset{\sim}{T}}^{*}$'s linearity in $Y_k$, $k \varepsilon r$, we can, equivalently, write

(8.10) $\hat{\underset{\sim}{T}}^{*} = \underset{\sim}{G}_r' \underset{\sim}{\Pi}_r^{*^{-1}} \underset{\sim}{Y}_r$

where

(8.11) $\underset{\sim}{G}_r' = \underset{\sim}{C}_r' - \underset{\sim}{D}_{xr} \underset{\sim}{A}_{xr}^{-1} \underset{\sim}{X}_r' \underset{\sim}{\Sigma}_r^{-1}$

(8.12) $\underset{\sim}{D}_{xr} = \underset{\sim}{C}_r' \underset{\sim}{\Pi}_r^{*^{-1}} \underset{\sim}{X}_r - \underset{\sim}{C}_N' \underset{\sim}{X}_N$

and $\underset{\sim}{A}_{xr}$ is given by (8.6). In the q-th linear function, $Y_k$ receives the weight $g_{rkq} \pi_k^{*^{-1}}$, where $g_{rkq}$ is the kq-element of $\underset{\sim}{G}_r = (g_{rkq})$.

In Särndal (1982) rules of thumb for the estimated variance-covariance matrix in single and two stage sampling are given. They are not immediately applicable, since we in the present nonresponse context lose the fixed size n for the variable size m. We will give analogous rules for the nonresponse case for one stage sampling.

Define $z_{kq} = c_{kq} Y_k$. Then $T_q = \sum_1^N z_{kq}$ is estimated unbiasedly (assuming that the response model holds, and that m is large) by the sum $\sum_r z_{kq}/\pi_k^*$ (q = 1,..., Q), similar to a Horvitz-Thompson (HT) sum. The covariance between $\sum_r z_{kq}/\pi_k^*$ and $\sum_r z_{kp}/\pi_k^*$ (variance if q = p) is estimated unbiasedly (under conditions assumed to be satisfied) by

(8.13) $v_{qp} = HT_r (z_{kq}, z_{kp})$ ,

where

$$HT_r(z_{kq}, z_{kp}) = \sum_{k \varepsilon r} \frac{1-\pi_k^*}{\pi_k^{*2}} z_{kq} z_{kp} + \sum_{\substack{k \neq \ell \\ \varepsilon r}}\sum \left(1 - \frac{\pi_k^* \pi_\ell^*}{\pi_{k\ell}^*}\right) \frac{z_{kq}}{\pi_k^*} \frac{z_{\ell p}}{\pi_\ell^*}$$

which we call the HT-transform. When the model (8.5) is fitted with $\hat{\underset{\sim}{B}}_r$ given by (8.7), the residual arising for object k is denoted

$$e_k = Y_k - \underset{\sim}{x}_k' \hat{\underset{\sim}{B}}_r \tag{8.14}$$

Rule 1. TAY-estimator in single stage sampling.
Consider $v_{qp}$ given by (8.13). Replace $Y_k$ by the residual $e_k$ given by (8.14). That is, $z_{kq} = c_{kq} e_k$ replaces $\hat{z}_{kq} = c_{kq} \hat{Y}_k$, and analogously for $z_{kp}$. Call the result $v_{qp}$. The matrix $(v_{qp}) = V_{TAY}(T^*)$ defines the TAY estimator of the variance-covariance matrix $\hat{V}_p(\underset{\sim}{T}^*)$.

Rule 2. TAY2 estimator in single stage sampling.
Consider $v_{qp}$ given by (8.13). In addition to the replacement of $Y_k$ by $e_k$ in Rule 1, replace $c_{kq}$ by $g_{rkq}$ defined through (8.10). That is, $z_{kq} = g_{rkq} e_k$ replaces $\hat{z}_{kq} = c_{kq} \hat{Y}_k$ in (8.13), and analogously for $z_{kp}$. Call the result $\hat{v}_{qp}$. The matrix $(v_{qp}) = \hat{V}_{TAY2}(\underset{\sim}{T}^*)$ defines the TAY2 estimator of $V_p(\underset{\sim}{T}^*)$.

The above results are taken from Swensson (1983). A detailed report by Swensson and Särndal is under preparation.

In Lyberg (1983) a model to study nonresponse effects in competing risks analysis is proposed. (A competing risks model is a Markov chain with a continuous time parameter, one transient state (State 0) and some (finite) number K of absorbing states.) The response probabilities are assumed to depend on whether, and from which cause, decrement has occurred during an observation period with right censoring. The model has been used to study nonresponse effects on estimates of transition intensities in the 1981 Swedish Fertility Survey. Some empirical results from that survey are presented to give realistic estimates of the parameters in the model.

By means of the model, the nonresponse effects on technical bias, variances and variance estimators of occurrence/exposure rates (estimated intensities) are investigated. It is shown that the technical bias (i.e. the bias due to ratio estimation) is often insignificant compared with the standard error, which in turn can often be estimated in an approximately unbiased manner by the usual variance estimator even in the nonresponse situation.

The nonresponse bias of estimates of transition intensities and transition probabilities is also investigated. It is shown that the nonresponse bias may be very large, if the response probabilities for decrements and survivors differ much. Two methods to adjust for the nonresponse bias are investigated. Both require accurate estimates of the ratios between the response probabilities for decrements and survivors. If this requirement is not met, the adjustment methods may in fact increase the nonresponse bias.

## 9. Randomized response

Survey questions which are sensitive or highly personal (e.g. on tax cheating, drug use or sexual behavior) generate substantial nonresponse and/or untruthful reporting. To cope with measurement problems of this type Warner (1965) introduced the randomized response (RR) technique. Since then, numerous papers have been published on the subject - for example, more than 20 articles on RR techniques have appeared in JASA. By far, the majority of the papers deal with theoretical aspects - only a limited number of well-designed empirical validation studies have been reported.

In Sweden the only uses of the technique seem to be in a small pilot test on receipt of public relief reported in Eriksson (1973) and in a recent survey briefly described below. The survey is part of a research project carried out at the Department of Sociology, the University College of Örebro, in which the SRU is engaged as statistical consultant. The purpose of the project is twofold: (i) to estimate the extent of drug use (especially cannabis) among adolescents (15-16 years of age) in Örebro (a city with 120000 inhabitants); (ii) to compare different methods for collecting data on drug use, including the randomized response technique.

The survey population was defined as pupils belonging to the 50 class units (with an average size of approximately 25 pupils) in Örebro forming the 9th grade in the comprehensive school. The 50 class units were randomly divided into two groups, with 20 units in group 1 and 30 units in group 2. (This allocation was approximately optimal considering the number of skilled interviewers at disposal.) Pupils in group 1 had to answer an anonymous questionnaire at a homeroom session, while pupils in group 2 were individually interviewed, using the original

Warner RR technique for the two sensitive questions on drug use (only one of which will be illustrated).

The RR procedure was as follows.

The interviewer handed over a plastic cup containing two ordinary dice. The respondent was told to shake the cup with the two dice and then count the resulting total number of dots. The respondent was then instructed to give an answer, depending on the outcome, as follows.

I. <u>If the total number of dots is 4, 5, 6, 7, 8, 9 or 10</u>

If you have smoked hashish or marihuana at least once give the answer "A".

If you have never smoked hashish or marihuana give the answer "B".

II. <u>If the total number of dots is 2, 3, 11 or 12</u>

If you have never smoked hashish or marihuana give the answer "A".

It you have smoked hashish or marihuana at least once give the answer "B".

(Of course,precautions were taken so that the interviewer had no possibility whatever to reveal the true drug use status of the respondent. Also, a thorough instruction and a practical demonstration preceded the actual RR interview.)

At the time of writing this paper the field work has just been terminated and the data have not yet been analyzed in detail. Rough preliminary figures indicate a point estimate close to 6 % for anonymous questionnaires and a point estimate close to 8 % for the RR interviews concerning the frequency of cannabis use. A detailed report on the project is planned to appear at the end of 1983.

## 10. Control of interviewer performance

Statistics Sweden has a staff of approximately 260 interviewers at its disposal. Twelve years ago the number of interviewers was 570.Since then the interviewers have become unionized and a new labor market legislation has come into effect. This development has brought certain consequences. Firstly, unsatisfactory interviewer performance is no longer a valid reason for dismissing an interviewer. Secondly, it is the responsibility of Statistics Sweden to ensure that all its interviewers meet the specific quality and efficiency standards. Today, we have no formal system for control of individual interviewer performance, which is rather unusual by international standards.

Measures taken to deal with, say, nonresponse have been general, i.e., directed towards all the interviewers including those having low non-response rates. The interviewers, spread all over the country, have obtained similar training and education in spite of the fact that the conditions for a successful field work depend heavily on the geographical site. Field work problems are not the same in big cities as in rural areas.

During the last couple of years it has become obvious that there is a need for a more tight quality control program with respect to interviewing. The data collection operations for some of the surveys conducted at Statistics Canada and US Bureau of the Census have been studied in detail. Some of the control operations carried out by these agencies will be put to a test at Statistics Sweden, at least for some of our surveys of individuals and households. The long-term purpose is to design a quality control program suited to Swedish conditions, where we make use of the methodological advances from Canada and USA, but skip some of the corrective actions towards individual interviewers. Such corrective actions ('probation', different types of 'punishments') are neither possible nor desirable within our legislation framework.

The quality control development work is concerned with the following problem areas:

A. Questionnaire edits

It is important to study how interviewers fill out and edit their questionnaires.We investigate

- how interviewer editing compares to central staff editing,
- which types of editing statistics that are necessary,
- how to develop an efficient feedback operation (from staff to individual interviewers).

B Observation of field work

Field observation is a method that has been used occasionally at Statistics Sweden. The observer takes notes of the interviewer's performance including how the interview is conducted and how the survey forms are completed. The observer reviews the interviewer's household-by-household performance and discusses the interviewer's strong and weak points with an emphasis oncorrecting bad habits.

We know that observation of field work is an excellent method (see US Bureau of the Census (1983)) to check interviewer performance, question 'quality' and field work instructions. An important drawback is that field work observations are rather costly. The observation itself is costly. Often there are certain travel costs involved and it is very important that the observer is well prepared for his/her task. As a consequence the method could hardly be used very often.

We investigate

- when to conduct field observations,
- what the observer's part finally comes down to and how to develop this part,
- who should be observers (central staff personnel or other interviewers),
- how to develop the interviewer-observer interaction in a positive way,
- how to develop efficient interviewer feedback,
- how to carry out observations of telephone interviews.

## C. Interviewer's work at home

By means of field observations we know a lot about what happens in face-to-face interviews. We know much less about the interviewer's work in his/her own home. Since knowledge is lacking in this field we are worse off when it comes to education and supervising the interviewer staff.

Examples of work carried out at the interviewer's home are:

i) tracking down respondents,
ii) making interviewing appointments,
iii) questionnaire editing,
iv) sending information to respondents,
v) studying,
vi) making out bills.

The interviewer's work at home cannot be studied by regular observations. Perhaps some kind of diary keeping could be an alternative, possibly accompanied by interviewing the interviewers. Perhaps even some type of home work observation might be possible.

In fact, one minor study of this kind has already been carried out. People from the central staff visited some interviewers and studied the premises and the keeping of material. The participation in this study was voluntary on part of the interviewers.

## D. Reinterviews

Reinterviews are not very common at Statistics Sweden. We have no continuing program of reinterviews in our surveys. However, we would like to have one, aiming at

i) obtaining a measurement of the content error,
ii) checking the nonresponse classification,
iii) an evaluation of manuals, question wording and interviewer training.

We investigate

- how to choose a reinterview method depending on the purpose,
- how to conduct reinterviews (a reinterview manual),
- who should be a reinterviewer (an 'ordinary' interviewer or a supervisor),
- how to choose questions for reinterviews,
- how to deal with discrepancies between the original interview and the reinterview.

E. Production statistics

Quality control needs reliable data on interviewer performance. A new information system designed for the field work has been developed. Eventually we will for every interviewer or group of interviewers be able to directly read off data about nonresponse rates, assignments, reassignments, the flow of completed interviews etc.

We investigate

- if it is possible to incorporate data about questionnaire editing, results from reinterviews and field observations, and some measure of interviewer efficiency.

Data of this kind could be used in the feedback operation and in the general analyses of the field work operations.

## 11. Computer-assisted interviewing

Computer-assisted telephone interviewing, CATI, refers to the use of computer systems for telephone interviewing and related forms of data collection. As pointed out by Nicholls (1981), systems with these capabilities might more properly be called "computer assisted data collection" systems. However, the acronym "CATI" has become so accepted that it is used in almost all reports dealing with systems with these broader capabilities. The advantages associated with CATI are, for instance, increased possibilities to get a better data quality, data processing and questionnaire development become less time-consuming, pretests can be conducted more efficiently and interviewer training can be improved.

CATI has been around for a decade or so. Some relevant documents describing phases of its development are, for instance, Cannel et al (1982), Fink (1981), Groves (1980), Groves et al (1980), Nicholls (1978), Palit (1980),Rustemeyer et al (1978), Shanks et al (1981), Vigderhous (1979) and Shure and Meeker (1978). Recently statistical agencies in countries like the Netherlands, United Kingdom, Denmark and Sweden have become interested in using existing CATI facilities or developing similar systems of their own.

In Sweden the basic facilities for surveys are favorable compared to the situation in most other countries. We have within Statistics Sweden a continuously updated computerized register of the total population, containing, e.g., information on birth registration numbers and addresses. This register, or subsets of it, is regularly used as a sampling frame for individual and household surveys. It contains excellent information for conducting mail surveys and surveys involving a personal visit by the interviewer. When it comes to telephone interviewing, we are a little bit worse off, because a person's telephone number cannot be found automatically, once we have address and birth registration number. Interviewers have to track telephone numbers by means of local telephone directories. If the number cannot be found in the directory, the interviewers have to try other ways, e.g., directory enquiries, asking local social insurence offices, parish registration offices, or employers. If no telephone number can be found, the interviewer has to visit the person or household selected. We are still fortunate, though, in our telephone survey work because interviewers always know the identity of the respondent.The register serves them with the respondent's name, address, age, status and some other data as well. "Cold" interviews are never conducted at Statistics Sweden. For instance, we have no need for systems involving random digit dialling. Some further progress is possibly underway. The National Swedish Telephone Company is building a register where an individual's birth number and telephone number is connected. Such a register would be very useful for us. Perhaps, one should have these basic advantages in mind when discussing CATI in Sweden.

At Statistics Sweden we have been working with CATI since 1979. The work has not been especially extensive so far but rapir progress is now beeing made. Rather, it has been of the kind suggested above by Nicholls: computer-assisted data collection. The project started with an installation and test of a CATI system developed at UCLA. This was a "warming-up" procedure to get us aquainted with CATI facilities.

Since the main part of the interviewers at Statistics Sweden are scattered all over the country we are trying to develop a special CATI model designed for Swedish survey conditions. We recommend the use of micro-computer technology as a basis for a portable survey data collection equipment. At present the project works along two interacting lines:

- tests of available, handheld computers in survey data collection,

- specifications of requirements for a prototype equipment for survey data collection to be available on the market within a couple of years.

A "portable" solution aims at removing some of the common CATI limitations. One such obvious limitation is the fixed physical location associated with CATI use today (which is good for quality control purposes, though). Data collection often takes place in shops, the respondent's home, premises or even cornfields. In fact we would like to allow all kinds of data collection in our portable system. So far we have conducted one major pilot study concerning the collection of retail price index data. In this study a sample of stores were visited by interviewers who recorded current prices and other data for designated products. The micro-computers used in this study weighed approximately two pounds and the complete field equipment held in one hand (micro-computer, forms and pencil mounted on a tray) weighed one pound. After one day's practice interviewers became proficient users of the portable system and the pilot test went smoothly.

Existing portable micro-computers are not adapted to survey data collection, though. Therefore, we aim at constructing a handheld micro-computer to be used in centralized telephone survey work, in personal interviews, and in other kinds of data collection as well. All testing of software facilities will be done on our display consoles since we must be sure of the system specifications before we start producing the computer hardware. Our endeavours so far is described in Danielsson and Mårstad (1982a, b).

## 12. Automated coding

Manual coding is a major operation in such statistical studies as censuses of population, censuses of business and labor force surveys. The problems with coding are of different kinds. As with most other survey operations, coding is susceptible to errors. In some studies it is the most error-prone operation next to data collection. For some variables error frequencies at the 10 % level are not unusual. Furthermore coding is time-consuming and costly, difficult to control, and boring. To cope with these drawbacks, it appears inevitable to focus on the very basis of manual coding and to consider the possibilities offered by access to a computer for developing a basically new approach. This may be viewed as a natural extension of earlier uses of computers in the editing operations.

During the last decade we have conducted a series of experiments at Statistics Sweden in order to find out whether or not it is possible to automate the coding process. Some of these experiments have been so promising that we have dared to tackle some ongoing surveys with this technique. Swedish applications of automated coding are the coding of goods in the 1978 Household Expenditure Survey, occupation in the 1980 Census of Population, the Survey of Living Conditions and the Pupil Surveys, and, finally, book loans for the Swedish Author's Fund bonus disbursements. The endeavors so far is described in Lyberg (1981) and Andersson and Lyberg (1983). We distinguish four operations in a system for automated coding:

i) Construction of a computer-stored dictionary;
ii) Entering element descriptions into the computer;
iii) Matching and coding;
iv) Evaluation.

There are two general kinds of algorithms for automated coding: weighting algorithms and dictionary algorithms. Weighting algorithms assign weights to each word-code combination using information from a basic file. Dictionary algorithms look in a dictionary for words which imply specific code numbers.If a match occurs the element is coded.

Otherwise it is rejected and referred to manual coding. At Statistics Sweden we have worked with the dictionary approach only. Descriptions of efforts at the U S Bureau of the Census are found in Lakatos (1977), O'Reagan (1972), Knaus (1978) and Hellerman (1982).

The computer-stored dictionary replaces the nomenclature and the coding instructions used in manual coding. The creation of such a dictionary is a cumbersome task involving a lot of different operations. It can be constructed by man or by computer, but presumably a combination of the two is the most effective. At Statistics Sweden we have constructed such a computer program. When evaluating the efficiency of automated coding we use data on costs, coding degree (the proportion coded by the computer) and quality (the proportion correctly coded among those coded automatically). The following table shows some results for our Swedish applications.

| Survey | Dictionary size | Coding degree | Quality (error rate) | Cost savings | Variable |
|---|---|---|---|---|---|
| 1978 Household Expenditure Survey | 1500 to 4200 | 65 % | ≤ 1 % | Not estimated | Goods |
| 1980 Census of Population | 4000 to 12000 | 71,5 % | Not estimated | 200 000 US dollars | Occupation and socio-economic classification |
| Continuous Survey of Living Conditions | Parts of the Census dictionary | > 60 % | ≤ 1 % | Not estimated | Occupation and socio-economic classification |
| Continuous Pupil Surveys | Parts of the Census dictionary | 50 % | ≤ 1 % | Not estimated | Occupation |
| Continuous Surveys of Book Loans | 65000 | 80 % | 0 % | Not estimated | Authors and book titles |

The low error rates are the results of the dictionary construction principle. In most applications only unambiguous descriptions are permitted in the dictionary. The resulting error frequency is due to the fact that sometimes an "unambiguous" description turns out to be ambiguous. Generally, cost savings are not explicitly estimated. However, savings have been made in all the applications but the exact magnitude has not been calculated. Perhaps this is explained by the fact that a conventional large-scale manual coding operation is an impossible alternative in some of these surveys. Our labor market legislation makes it difficult to hire coding personnel for occasional efforts such as the coding in a census. When it comes to cutting work load peaks we have to rely on our permanent staff and automated coding in combination. In our experiments and applications we have used methods that are rather unsophisticated. The methodological development has probably suffered from the fact that those responsible for the survey have been satisfied with rather modest coding degrees around 65-70 % or even less. We certainly ought to strive for more efficient systems with coding degrees around, say, 90 %. This could be done by means of more sophisticated methods, for instance, using auxiliary information, but also by changing the codes. Perhaps it is not too preposterous to make changes in the codes in order to obtain a less costly coding. That option should certainly be considered more often in times of scarce financial resources.

So far, our strategy has been to put the variables easy to code to a test first. Now we have to proceed to the more difficult ones and make the dictionaries and the supporting routines more efficient.

## 13 Risk assessment

Problems concerning probability and risk assessment in connection with the production of energy have been treated by Statistics Sweden in comments to official investigations (see Statistics Sweden (1978, 1980, 1981)), and in a study prepared within the SRU (see Björk and Hagberg (1982)).

The public discussion on nuclear risks in Sweden has largely centered on the "Rasmussen report" (WASH-1400) and similar studies performed in Sweden. Statistics Sweden discusses the use of probability statements in risk and safety assessment and draws attention to the critical discussion of the Rasmussen report in the USA and in other countries.

Statistics Sweden notes that it seems possible to survey limited systems and thus get a certain guidance for practical measures with the aid of the event tree/fault tree methodology. However, calculations of this kind do not work in a meaningful way with regard to large socio-technical systems that presuppose an interaction between human beings and a complicated technology. The data presented must be regarded as pure fiction without prognostic value.

In the Rasmussen report "probabilities" of various kinds of damage in connection with the operation of nuclear power reactors are calculated by using a "probabilistic" model that gives a simplified picture of possible sequences of events. The model does not include the effects of changes during the life of a reactor (e.g. aging), unplanned human intervention, sabotage, war, etc. Different events and failures are assigned numerical "probabilities", based on various sources, such as statistics concerning nuclear reactors and other activities, handbooks, expert reports etc.

The event tree/fault tree technique used in the Rasmussen report is analogous to the decision tree technique used in operations analysis. The report refers to the textbook "Decision theory" by H. Raiffa, who declares himself a Bayesian and stresses the difference between the "judgmental probability" of a decision-maker and "a bona fide, objective, tangible, real-world, frequency-based probability".

In the Rasmussen report, the word "probability" has various meanings. Statistics Sweden stresses that the final results are presented in a frequentist form, while the basic material to a large extent consists of subjective "probabilities". In the public discussion of the report, the choice of causal factors included in the models has been questioned and criticized, as has the choice of parameter values and of mathematical methods for calculating point estimates and error bounds.

Sometimes, as exemplified below, misconceptions and pseudoarguments have been presented in support of the "credibility" or "realism" of calculations of the Rasmussen type.

The methodology of the Rasmussen report is applied in the West German "Birkhofer report" and in similar official Swedish studies concerning nuclear reactor risks. To a great extent, these studies have used the same model assumptions and quantitative assessments as the Rasmussen group. The relatively close agreement between the numerical results has been considered as indicative of the "realism" of the method. A similar logical error is to be found in the comparison made by N. Rasmussen (in response to the findings of the Lewis report) between the estimates of core melt frequency presented in the Rasmussen report and the corresponding estimates in a study by the subjectivistists G. Apostolakis and A. Mosleh (1979) who used results from the Rasmussen report as one of their inputs and who have declared that they had not intended to "produce a definitive distribution for the frequency of core melts" but rather to "demonstrate how an expert's opinion can be formally handled and what difficulties arise in doing so".

## 14. Standards for quality presentation of statistics

Tentative guidelines on quality presentation of statistics were issued by Statistics Sweden in 1979 (Statistics Sweden (1979)). The guidelines were concerned with quality presentation directed to the users of statistics, the purpose being to inform them of quality, applicability, and limitations of the statistics.

A distinction was made between quality presentations (intended for users in general) and technical reports (not covered by the guidelines, and intended for professionals such as the producer and those users who have statistical expert knowledge). The quality presentation should concentrate on the effect (preferably the net effect) of various sources of error on the final data.

The quality concept, relevance, and error sources were discussed at some length, and recommendations for quality presentation were put forward. In addition to the guidelines proper, the document also contains a set of annotated examples of quality presentations.

The guidelines were thoroughly discussed by government statisticians from several European countries and from Canada at meetings in 1981 and 1982 with the Conference of European Statisticians in Geneva.

After a period of launching the guidelines within Statistics Sweden, we can now observe that the quality of quality presentations have gradually improved, although there still remains a lot to be done until we reach a uniformly high level.

Following these experiences, a decision was taken by Statistics Sweden in 1983 on a more definitive policy for a user-oriented quality presentation of statistics (Statistics Sweden (1983 b)). The new policy is concerned with all kinds of statistical data released by Statistics Sweden. Thus, it is of a more general scope than the earlier guidelines, which were mainly applicable to survey data. It is stated, however, that the earlier more detailed guidelines are to be followed whenever applicable.

The main principle of the policy document is that "The producer of statistics has to inform the users of factors which are important for a correct interpretation of the statistics. The information should be accessible and easy to understand for the users, and in all respects formulated to meet their needs."

For the special case of primary dissemination of statistics, the policy document states that the quality presentation should cover (when applicable) the following topics:

(1) Definitions, including explanation of concepts used, definition of populations, objects, variables, and classifications.

(2) Methodology, including data collection methods, sampling and estimation methods (only a short description), and references to more extensive technical reports.

(3) Comparability over time, and with data from other sources.

(4) Accuracy, including important sources of error, such as coverage, response rates, sampling error, and measurement and processing error. The user will appreciate a statement on the overall accuracy of the statistics.

## 15. Survey research outside the SRU

From the presentation so far, it is clear that the survey research of the SRU is of a somewhat general nature, directed towards exploration and introduction of new ideas. Survey research and development more closely tied to specific surveys is mostly undertaken by the subject-matter departments themselves. These efforts are scattered, however, due to the shortage of methodology staffs within the subject-matter departments.Here is a list of some selected survey research projects currently dealt with in the subject-matter departments.

The Department of Area Statistics:
- Methodology for describing spatial variation.
- Classification of farms.
- Problems connected with the objective crop yield surveys.
- Trend analysis of environmental statistics.
- Use of sampling in environmental statistics.
- Remote sensing, e.g., using aerial photographs for studying things like urban expansion into agricultural land, and land use within urban areas.

The Department of Enterprise Statistics:
- Time series of economic data, including seasonal adjustment and forecasting.
- Methodology of sampling and price measurement for the Consumer Price Index
- Evaluation of computerized editing.

The Department of Statistics on Individuals and Households:

- Methodology for small area estimation (including generalized regression estimation) applied on income statistics.
- Estimating measures of income inequality.
- Applicability of statistical matching.
- Combining model aspects and sampling design aspects in demographic analysis.
- Use of Markov models for event-history analysis applied to register data and sample survey data. (See Lyberg (1983) and further references given there.)
- Classification by principal components in multiregional demographic analysis. (See Martinelle (1982)).
- Models for classification errors in panel data.
- Using sample survey data for private consumption studies.
- Longitudinal studies in the field of educational statistics.

ACKNOWLEDGMENT

The authors are grateful to Jörgen Dalén, Tomas Garås, Sixten Lundström and Jan Hagberg/Leif Björk for providing basic materials to Sections 3, 4, 5,and 13, respectively. We are also very grateful to Ms Elseliv Lindfors for efficient typewriting under a tight time schedule.

REFERENCES

Andersson, R. and Lyberg, L. (1983): Automated Coding at Statistics Sweden. Paper to be presented at the ASA meeting in Toronto, Canada, 1983.

Apostolakis, G. and Mosleh, A. (1979): Expert Opinion and Statistical Evidence: An Application to Reactor Core Melt Frequency. Nuclear Science and Engineering: 70, pp. 135 - 149.

Bergman, L.R., Hanve, R. and Rapp, J. (1978): Why Do Some People Refuse to Participate in Interview Surveys. Statistisk tidskrift, pp. 341 - 356.

Björk, L. and Hagberg, J. (1982): Some Considerations Concerning Risk Studies in the Field of Energy. Statistisk tidskrift, pp. 348 - 366, 395 - 397.

Bogeström, B., Larsson, M. and Lyberg, L. (1981): Bibliography on Non-Response and Related Topics. Statistics Sweden.

Brackstone, G.J. (1976): Drawing Inferences from Test Results. Memo, Statistics Canada.

Brewer, K.R.W., Foreman, E.K., Mellor, R.W. and Trewin, D.J. (1977): Use of Experimental Design and Population Modelling in Survey Sampling. Bulletin of the ISI, New Delhi.

Cannel, C.F., Groves, R.M., Magilavy, L.J., Mathiowetz, N.A. and Miller, P.V. (1982): Comparison of CATI and NonCATI Questionnaires. In Charles F. Cannell et al., An Experimental Comparison of Telephone and Personal Health Surveys. Unpublished research report of the Survey Research Center, University of Michigan, prepared for the National Center for Health Statistics. Ann Arbor, Mich.; Survey Research Center, University of Michigan, April, 1982. Volume 1. Section III, pp. 43 - 70.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1976): Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. Biometrika 63, pp. 615 - 620.

Cochran, W.G. (1977): Sampling Techniques, Wiley, New York, 3rd ed.

Dalén, J. (1983): How Large Must the Sample Size Be? Nominal Confidence Levels Versus Actual Coverage Probabilities in Simple Random Sampling. Promemorior från SCB, No 8. Memo, Statistics Sweden.

Danielsson, L. and Mårstad, P.-A. (1982a): Towards Development of a Computer Assisted Survey Data Collection System. DBM-PM 82:35. Memo, Statistics Sweden.

Danielsson, L. and Mårstad, P.-A. (1982b): Statistical Data Collection with Handheld Microcomputers - A Test in Consumer Price Index. DBM-PM 82:41. Memo, Statistics Sweden.

Deming, W.E. (1953): On a Probability Mechanism to Attain an Economic Balance between the Resultant Error of Non-Response and the Bias of Non-Response. Journal of the American Statistical Association, 48, pp. 743 - 772.

Dunnel, K. and Martin, J. (1982): Piloting - Purposes and Evaluation. Survey Methodology Bulletin, Social Survey Division, Office of Population Censuses and Surveys, London, No 14, pp. 36 - 41.

El-Badry, M.A. and Stephan, F.F. (1955): On Adjusting Sample Tabulations to Census Counts. Journal of the American Statistical Association, 50, pp. 738 - 762.

Elvers, E., Särndal, C.E., Wretman, J.H. and Örnberg, G. (1983): Regression Analysis and Ratio Analysis for Domains, a Randomization Theory Approach. Unpublished report.

Eriksson, S. (1973): Randomized Interviews for Sensitive Questions. Ph.D. Thesis, Institute of Statistics, University of Gothenburg, Sweden.

Flink, J.C. (1981): Quality Improvement and Time Savings Attributed to CATI - Reflections on 11 Years Experience. Paper presented at the 36th Annual AAPOR Conference, Buck Hill Falls, PA. Radnor, Pa.: Chilton Research Service, May 29, 1981.

Forsman, G. and Garås, T. (1982a): Optimal Estimation of Change in Sample Surveys. Memo, Statistics Sweden.

Forsman, G. and Garås, T. (1982b): Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Memo, Statistics Sweden. (In Swedish)

Groves, R.M. (1980): Computer Assisted Telephone Interviewing and the Future of Survey Research. Paper presented at the Conference on Computer Assisted Telephone Interviewing, March 1980, Berkeley, California.

Groves, R.M., Berry, M. and Mathiowetz, N. (1980): Some Impacts of Computer - Assisted Telephone Interviewing on Survey Methods. American Statistical Association. Proceedings of the Section on Survey Research Methods, pp. 519 - 524.

Hellerman, E. (1982): Overview of the Hellerman I&O Coding System. Memo, U.S. Bureau of the Census.

Hocking, R. (1979): The Design and Analysis of Sample Surveys with Incomplete Data: Reduction of Respondent Burden. Paper presented at the symposium of incomplete data: Preliminary proceedings, Washington Hilton *Hotel on August 10-11, 1979*, Washington D.C.

Hocking, R. and Oxspring, H.H. (1971): Maximum Likelihood Estimation with Incomplete Multinomial Data. Journal of the American Statistical Association, 66, pp. 65 - 70.

Hocking, R. and Smith, W.B. (1972): Optimum Incomplete Multi-Normal Samples. Technometrics 14, pp. 299 - 307.

Hunt, S.D., Sparkman, R.D.J. and Wilcox, J.B. (1982): The Pretest in Survey Research: Issues and Preliminary Findings. Journal of Marketing Research, Vol XIX, pp. 269 - 273.

Ireland, C.T. and Kullback, S. (1968): Contingency Tables with Given Marginals. Biometrika, 55, pp. 179 - 188.

Jabine, T.B. (1981): Guidelines and Recommendations for Experimental and Pilot Survey Activities in Connection with the Inter-American Household Survey Program. Inter-American Statistical Institute.

Kalton, G. and Schuman, H. (1982): The Effect of the Question on Survey Responses: A Review. Journal of the Royal Statistical Society, Series A, Part 1, pp. 42 - 73.

Knaus, R. (1978): Automated Industry Coding - an Artificial Intelligence Approach. Memo, U.S. Bureau of the Census.

Lakatos, E. (1977): Automated I&O Coding. Memo, U.S. Bureau of the Census.

Lyberg, I. (1983): Nonresponse Effects on Survey Estimates in the Analysis of Competing Exponential Risks. Stockholm Research Reports in Demography, No 12, Department of Statistics, University of Stockholm.

Lyberg, L. (1981): Control of the Coding Operation in Statistical Investigations - Some Contributions. Ph D thesis. URVAL No 13, Statistics Sweden.

Martinelle, S. (1982): Classification by Principal Components - A Tool for Multiregional Demographic Analysis. Statistical Review No 5, pp. 333 - 346.

Moser, C. and Kalton, G. (1972): Survey Methods in Social Investigation. New York: Basic Books.

Nicholls, W.L. II (1978): Experiences with CATI in a Large-Scale Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 9 - 17.

O'Reagan, R.T. (1972): Computer-Assigned Codes from Verbal Responses. Communications from the ACM, Vol 15, No 6, pp. 455 - 459.

Palit, C.D. (1980): A Microcomputer Based Computer Assisted Interviewing System. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 516 - 518.

Politz, A.N. and Simmons, W.R. (1949, 1950): An Attempt to Get the "Not at Homes" into the Sample without Callbacks. Journal of the American Statistical Association, No 44, pp. 9 - 31 and No 45, pp. 136 - 137.

Purcell, N.J. (1979): Efficient Estimation for Small Domains: A Categorical Data Analysis Approach. Unpublished Ph.D. thesis, University of Michigan, Ann Arbor, Michigan.

Raj, Des (1968): Sampling Theory. New York: McGraw-Hill.

Rustemeyer, A., Shure, G.H., Rogers, M.S. and Meeker, R.J. (1978): Computer-Assisted Telephone Interviewing: Design Considerations. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 1 - 8.

Shanks, J.M., Nicholls, W.L. II and Freeman, H.E. (1981): The California Disability Survey: Design and Execution of a Computer-Assisted Telephone Study. Sociological Methods & Research. Vol 10, No 2, pp. 123 - 140.

Shure, G.H. and Meeker, R.J. (1978): A Minicomputer System for Multiperson Computer-Assisted Telephone Interviewing. Behavior Research Methods & Instrumentation, Vol 1, No 2, pp. 196 - 202.

Smith, J.H. (1947): Estimation of Linear Functions of Cell Proportions. The Annals of Mathematical Statistics, 18, pp. 231 - 254.

Statistics Sweden (1978): Comments on the Report of the Swedish Energy Commission "Energy - health, environment and safety risks", Statistisk tidskrift, pp. 472 - 481.

Statistics Sweden (1979): Riktlinjer för kvalitetsredovisning av statistik (Guidelines for presentation of the quality of statistics). (In Swedish). Reports on Statistical Coordination 1979:8. Stockholm. Statistics Sweden.

Statistics Sweden (1980): Comments on the report from the National Institute of Radiation Protection concerning emergency planning against nuclear power accidents. Statistisk tidskrift, pp.343 - 353.

Statistics Sweden (1981): Comments on the report by the Swedish Government Committee on Nuclear Reactor Safety and the proposals on safety increasing measure by the Swedish Nuclear Power Inspection. Statistisk tidskrift, pp. 56 - 59.

Statistics Sweden (1983a): Nya metoder för folk- och bostadsräkningar. Detaljrapport. (New Methods for Population and Housing Censuses. Detailed report.) (In Swedish). Stockholm: Statistics Sweden.

Statistics Sweden (1983b): Policy för användarorienterad kvalitetsredovisning av statistik. (Policy for a User-Oriented Presentation of the Quality of Statistics.) (In Swedish). Reports on Statistical Coordination 1983:1. Stockholm: Statistics Sweden.

Stephan, F.F. (1942): An Iterative Method of Adjusting Sample Frequency Tables when Expected Marginal Totals are Known. The Annals of Mathematical Statistics, pp. 166 - 178.

Sudman, S. and Bradburn, N.M. (1982): Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass Publishers.

Svensson, B. (1983): Generaliserad regressionsestimation av linjära funktioner vid bortfall, II: Okända svarssannolikheter. Ytterligare specialfall. Memo, Statistics Sweden. (In Swedish)

Särndal, C.E. (1981): Frameworks for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for Nonresponse. Invited paper to the ISI meeting in Buenos Aires.

Särndal, C.E. (1982): Implications of Survey Design for Generalized Regression Estimation of Linear Functions. Journal of Statistical Planning and Inference 7, pp. 155 - 170.

Vigderhous, G. (1979): Optimizing the Time Schedules and Response Rates in Telephone Interviews - A Study of the Seasonal Patterns. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 256 - 261.

Warner, S.L. (1965): Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association 60, pp 63 - 69.

Wärneryd, B (1977): SCB och allmänheten. Resultat från en intervjuundersökning våren 1976. Memo, Statistics Sweden. (In Swedish).

United Nations (1981): National Household Survey Capability Programme. Handbook of Household Surveys.

United Nations (1982): National Household Survey Capability Programme.

U S Bureau of the Census (1983): Approaches to Developing Questionnaires. A Report by the Subcommittee on Questionnaire Design. Edited by T.J. De Maio. Draft.

Tidigare nummer av Promemorior från P/STM:

Nr

1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)

2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)

3 En presentation av Box-Jenkins metod för analys och prognoser av tidsserier. Åke Holmén (1979-12-20)

4 Handledning i AID-analys. Anders Norberg (1980-12-20)

5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)

6 Metoder för evalvering av noggrannheten i SCBs statistik, En översikt. Jörgen Dalén (1981-03-02)

7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)

8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)

9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan H Wretman och Göran Örnberg (1983-06-20)

Kvarvarande exemplar av ovanstående nummer kan rekvireras från SCB, P/STM, Elseliv Lindfors, 115 81 Stockholm, eller per telefon 08 14 05 60 ankn 4178.

Garnisonstryckeriet ★ Stockholm ★ 1983