ESTIMATING INTERVIEWER EFFECTS
IN SAMPLE SURVEYS
SOME CONTRIBUTIONS

Peter Lundquist

Department of Statistics
Stockholm University
2006
Doctoral Dissertation
Department of Statistics
Stockholm University
SE-106 91 Stockholm
Sweden

Abstract
This thesis focuses on measurement errors that could be ascribed to the interviewers. To study interviewer variability a measurement error model is formulated which makes a clear distinction between three sources of randomness: the sample selection, interviewer assignment, and interviewing.

In the first paper the variance of the observed sample mean is derived, and it is seen how this variance depends on parameters of the measurement error model and on the number of interviewers. An estimator of the interviewer variance, which is seen to be unbiased, and a biased intra-interviewer correlation estimator are suggested. In a simulation study it is seen that the simulation variance of the interviewer variance estimator increases for both high and low interviewer assignments and seems to have a minimum somewhere in between.

The second paper presents an expression of the variance of the observed sample mean under stratified random sampling. Two possible estimators of the variance of the mean are considered, one of which has a slight positive bias, the other a negative bias, which can be large. Two different estimators of the interviewer variance are studied. Only one of them makes it possible to find a reasonable estimate of the intra-interviewer correlation.

In the third paper an expression for the variance of the interviewer variance estimator is derived. This result may prove useful in designing future studies of interviewer variance. For a large population it will be possible to use an approximate variance, irrespective of the underlying distribution of the unknown true values. But it is still necessary to know the kurtosis for the interviewer effect. It is demonstrated that a high kurtosis will increase the variance of the estimated interviewer variance.

The fourth paper deals with some issues in planning and analyzing an interviewer variance study. Three problems are considered: (i) Determining the number of interviewers and the appropriate size of the interviewer assignments; (ii) Finding the probability of negative estimates of the interviewer variance; (iii) Testing for interviewer variance.

Key words: Response variance; survey nonsampling error; interviewer effects; interpenetration; power.
Contents

Part I: Introduction
Interviewer effects in sample surveys

Part II: Papers included
1. Estimating interviewer variance under a measurement error model for continuous survey data (co-authored with Jan Wretman)
2. Estimating interviewer variance under a measurement error model for continuous survey data – Stratified sampling
3. The variance of the interviewer variance estimator under an error model for continuous data
4. Issues in designing an interviewer variance study under an error model for continuous data
INTERVIEWER EFFECTS IN SAMPLE SURVEYS

1. The concept of interviewer variability

This thesis is about surveys where data are collected by interviewers. In such surveys, an interviewer’s work consists of more than just reading the questions to the respondent. He or she has also to trace the respondents, get in contact with them, find out if they are eligible for the survey, persuade them to participate if they are reluctant, clarify the meaning of survey questions when necessary, ask probing questions, and finally record the answers in an appropriate way.

It is evident that the way interviewers carry out their work will have an effect on the quality of the estimates resulting from the survey. Interviewers are human beings, and there will always be individual differences among them. It is unavoidable that interviewers contribute, more or less, to the total volume of errors in survey data. Efficient survey planning aims, among other things, at reducing the errors caused by imperfections in interviewer performance.

For planning purposes it is important to have an idea of how the size of the interviewer errors depends on various factors. Special interviewer studies can be helpful in providing such knowledge. Evaluation of interviewer errors is also of importance for another purpose, namely to give information on survey quality to users of statistics.

A recent discussion of the interviewer’s role in a survey, the type of errors that can be ascribed to them, factors that may affect the magnitude of the errors, and methods for evaluating and controlling interviewer errors, is given by Biemer and Lyberg (2003). A model of the interview process, with associated interviewer and respondent cognitive processes, is given by Japec (2005).

In this thesis, the focus is on the variability among interviewers when it comes to generating errors in data. By “error” we mean that there is a difference between the true value of a respondent and the value reported by the interviewer. Part of this error is assumed to depend in a systematic way on the specific interviewer. The model to be used in the thesis essentially says that for each interviewed element,

\[
\text{Element’s observed value} = \text{Element’s true value} + \text{Interviewer effect} + \text{Random error},
\]

where the interviewer effect is assumed to be the same for all interviews done by the same interviewer, but may vary among interviewers. It is a sort of systematic component, assumed to influence in the same way the responses from all respondents in the interviewer’s assignment. By contrast, the random error varies from interview to interview, even when done by the same interviewer. This type of ANOVA-like model has its origins in the article by Kish (1962).

Which are the real phenomena behind the assumed systematic interviewer effects in the model? For example, if survey instructions or definitions are unclear, they can be interpreted in different ways by different interviewers, and for each interviewer in a way that is consistent through all his or her interviews. Likewise, each interviewer may have his or her own way of handling ambiguous or unclear responses. Personal characteristics of an interviewer such as age, race, gender, social class, education, attitudes, and beliefs, may have a tendency to systematically influence the measurements made by the interviewer. Other external factors that could explain the interviewer effect are for example: the questionnaire, the mode of interview, methods for training, supervising, and monitoring the interviewer routines. Biemer and Lyberg (2003) discuss these things.

From now on, when we talk about studying interviewer variability we mean studying the variance of the interviewer effects over all interviewers. This variance will be called the interviewer variance, denoted \( s^2_i \). Another numerical measure connected with interviewer variability is the intra-
**interviewer correlation**, $\rho$, which can be interpreted as the ratio of the interviewer variance to the total variance.

The theoretical intra-interviewer correlation always takes a value between 0 and +1, and usually takes a value near zero. Groves (1989) made a compilation of estimated values of the intra-interviewer correlation from a number of studies reported in the literature, and it turned out that for telephone surveys the value was seldom higher than 0.02. However, the intra-interviewer correlation, in spite of its low numerical value, contributes to increasing the variance of the estimator of a population parameter in a way that may be devastating. For example, under simple random sampling, when intra-interviewer correlation is present, the variance of the sample mean is roughly multiplied by the factor $[1 + (m-1)\rho w]$, where $m$ is the average interviewer workload—that is, the average number of interviews per interviewer. This means that if the workloads are 40–50 interviews and the intra-interviewer correlation is 0.02, then the variance of the sample mean is increased by 80 to 100 percent.

### 2. Basic prerequisites and assumptions of the thesis

The thesis is written with the interviewer organization and other working conditions of a national statistical agency like Statistics Sweden in mind, and the results are assumed to be useful for future interviewer variability studies at Statistics Sweden. The results are considered to be applicable in telephone interview surveys of individuals or households, with a sample size larger than 1,000 and with fewer than 200 interviewers. The sampling designs considered are simple random sampling (**srs**) and stratified random sampling. These restrictions are motivated by Swedish survey conditions, characterized by reliable population registers covering the whole nation, together with a high telephone access rate.

In the case of simple random sampling of individuals we assume that interviewer assignments are constructed by a technique sometimes referred to as *interpenetration* and used originally by Mahalanobis (1946). In the actual context, interpenetration means that all interviewers get assignments of equal size and that each interviewer assignment is a random subsample taken from the original sample. Thus, if there are $n$ elements in the whole sample and $I$ interviewers, we randomly choose $m = n/I$ elements to be interviewed by interviewer no. 1, then $m$ elements from the remaining elements in the sample for interviewer no. 2, etc. In the case of stratified random sampling, the same kind of interpretation as above is assumed to be used within each stratum. However, the size of the interviewer workloads may differ among strata.

The measurement error model used is written as

$$y_k = \mu_k + b_i + \varepsilon_k,$$

where

- $y_k$ is the value reported for element $k$ when interviewed by interviewer $i$;
- $\mu_k$ is the true value for element $k$;
- $b_i$ is the (systematic) interviewer effect of interviewer $i$; and
- $\varepsilon_k$ is a remaining random error.

The $\mu_k$ are constants, while the $b_i$ and the $\varepsilon_k$ (and the $y_k$) are assumed to be random variables. The model is like an ANOVA model with a random effect $b_i$ and it is related to models used by, for example, Kish (1962), Hartley and Rao (1978), and Biemer and Trewin (1997).

Three sources of randomness are recognized, namely
The initial selection, by a probability sampling design, of a sample from the population.

The randomized construction of interviewer assignments.

The interview process.

Because of some ambiguity in the terminology among authors in this area, we will in Paper 1 spend some time defining basic concepts the way we will use them. Otherwise, the terminology will closely follow that of Wolter (1985) and Särndal, Swensson, and Wretman (1992).

3. Estimation problems

In Paper 1 of the thesis estimating the true population mean, \( \hat{m} = \frac{\sum \hat{y}_i}{N} \), is considered, using data from a simple random sample, assuming that the measurement error model above holds. The population mean, \( \mu \), is assumed to be estimated by the sample mean, \( \bar{y}_s = \frac{\sum y_k}{n} \), and the expected value of this estimator is derived. The variance of \( \bar{y}_s \) is also derived, and it is seen how this variance depends on parameters of the measurement error model and on the number of interviewers. Among other things it is noted that the variance of \( \bar{y}_s \) decreases as the number of interviewers increases (other things being equal). Two estimators of the variance of \( \bar{y}_s \) are suggested, neither of them unbiased. One of them, which has a small positive bias, is preferred to the other one which has a possibly large negative bias. It seems safer not to underestimate the variance of an estimator.

An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased.

An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased. An estimator of the intra-interviewer variance, \( s^2 \), is suggested, which is seen to be unbiased.

Paper 2 is mainly about the same problems as Paper 1, but now under another sampling design, stratified random sampling. The true population mean, \( \hat{m} \), is assumed to be estimated by the usual stratified sample mean, \( \hat{y}_s = \frac{\sum h=1 \cdots H W_h \sum s, y_k}{n} \), and the expected value of this estimator is derived. An expression for the variance of \( \hat{y}_s \) is obtained. By analogy with Paper 1, two possible estimators of the variance of \( \hat{y}_s \) are considered, one of which has a slight positive bias, the other a negative bias, which can be large.

Two different estimators of the interviewer variance are under consideration, both unbiased. However, only one of them makes it possible to find a reasonable estimate of the intra-interviewer correlation. Thus this estimator of the interviewer variance is the preferred one, and it is also seen in a simulation study to be more efficient than the other one.

In Paper 3 an expression for the variance of the interviewer variance estimator, \( \delta^2 \), is derived. This result may prove useful in designing future studies of interviewer variance. For a large population it will be possible to use an approximate variance, irrespective of the underlying distribution of the unknown true values. But it will still be necessary to know the kurtosis for the interviewer effect, \( b_h \). It is demonstrated that a high kurtosis will increase the variance of the estimated interviewer variance. Prior indications of non-normal interviewer effects should therefore be accounted for in the design of an interviewer variance study.
For large populations, with an srs design, and normally distributed errors, our variance of $\hat{s}_b^2$ approximates the variance of the variance component estimator in the classical ANOVA model with random effects (cf. Searle, Casella, and McCulloch 1992). This is true when our interviewer effect defined in Section 2 corresponds to the random factor in the classical random effect model, although our source of randomness is different.

4. Planning and analyzing an interviewer variance study

Paper 4 deals with some issues in planning and analyzing an interviewer variance study. Three problems are considered:

(i) Determining the number of interviewers and the appropriate size of the interviewer assignments.

(ii) Finding the probability of negative estimates of the interviewer variance.

(iii) Testing for interviewer variance.

For a given sample size, $n$, a problem is to find the number of interviewers, $I$, and the size of the interviewer assignment, $m$, so that the interviewer variance is estimated with high precision, under the constraint that $I \times m = n$. Under simple random sampling, it is demonstrated that the approximate relative variance of $\hat{s}_b^2$ (using the approximation from Paper 3) can be expressed explicitly as a function of three things: (i) the interviewer assignment size $m$, (ii) the intra-interviewer correlation $\rho_w$, and (iii) the kurtosis, $\beta_b$, of the distribution of the interviewer effects. (If the interviewer effects are normally distributed, the kurtosis term will vanish.) Using this approximate expression, it is possible to find the value of $m$ that will minimize the relative variance of $\hat{s}_b^2$ for given values of $\rho_w$ and $\beta_b$. Of course, in practice the values of $\rho_w$ and $\beta_b$ are unknown and have to be conjectured in one way or another. It may also turn out in practice that even the minimum value of the relative variance is too large, so that the intended study will not give any information of value. In that case, there are two possible options, either increase the size of the study (if the budget allows it) or call the whole thing off.

The interviewer variance estimator in this thesis may sometimes produce a negative estimate, which is unfortunate and difficult to handle, because the variance to be estimated is non-negative by definition. In Paper 4, under special distributional assumptions, an expression is derived for the probability of obtaining negative estimates. The theoretical result is confirmed in a simulation study, which also indicates that (i) for a given total sample size, the probability of a negative estimate decreases as the number of interviewers decreases (and the size of the interviewer assignment increases), and (ii) the probability of a negative estimate decreases as the total sample size increases.

Paper 4, finally, discusses how to test the null hypothesis that $s_b^2 = 0$ against the alternative hypothesis that $s_b^2 > 0$. Two tests are considered: the traditional $F$-test and a bootstrap test. The tests are compared in a simulation study, where it turns out that they perform roughly the same way, both for normal and gamma distributed interviewer effects.

In practice, when interviewer variance studies are carried out, they are often embedded in a regular survey. The purpose of the regular survey is to estimate a number of population parameters, while the purpose of the embedded interviewer study is to find out if answers to survey questions are affected by the interviewers. These two purposes are somewhat incompatible. It is usually impossible to find a combination of survey and experimental design that gives accurate estimates of both regular population parameters and interviewer variances. It is an important task to find an acceptable balance between the design of the interviewer variance study and the survey design.
5. Further research

The work on this thesis has raised some questions which need to be answered. Some of them are for a survey organization to consider, while others are of a more academic nature. The thesis will hopefully encourage Statistics Sweden to learn more about interviewer variability. Only one documented interviewer variance study exists at Statistics Sweden. Brorson (1981) studied the interviewer effects in the Survey of Living Conditions (a face-to-face survey). The study was criticized by Lindström (1981) because no randomized experimental design was used.

Some of the interviewers of Statistics Sweden work in a central telephone-group, while others work “in the field” (that is, do telephone interviewing from their homes). In many large telephone surveys, both the central group and the field interviewers are used to collect the data. The two types of interviewers have clearly different working situations and needs. There are also other things in the survey organization that indirectly could give rise to interviewer variability. For example a new computer-assisted system or new strategies to improve the survey participation may affect the interviewers in different ways. To improve the survey quality, a better understanding of important factors in the interviewer work must be obtained. This also calls for interviewer experiments in the future.

Full information about the distribution of interviewer effects is rarely available. In this thesis, normal and gamma distributed interviewer effects are studied. In practice, observed data are often used as a basis for distributional assumptions. Guidelines for using observed data in the calculations should be investigated. Graphical illustrations might be used, for example, as described by Korn and Graubard (1998).

To study the interviewer variability, one could use other types of measures than those described in this thesis. In most telephone surveys the same interviewer asks all the questions in the questionnaire. It is then possible to study interviewer effects for several questions simultaneously, using some multiple testing procedure (cf. Westfall and Young 1993).

As already mentioned, the thesis describes a small simulation study using a bootstrap technique in testing for interviewer variance. The purpose of such a resampling procedure is to handle non-normality in data. More research is needed on resampling methods in this area, especially in combination with stratification and nonresponse. Solutions might be developed using bootstrap and other resampling strategies given, for example, in Manly (1997) and Lunneborg (2000), together with findings in the sampling area given in Kovar, Rao, and Wu (1988) and Sitter (1992).

The ANOVA model is appropriate for continuous data. When study variables in a survey are categorical, which is often the case, the methods proposed in this thesis are no longer appropriate. Further research is needed, and a starting point could be the work by Stokes and Mulry (1987).
6. References


