COMPONENTS OF RESPONSE VARIANCE

FOR CLUSTER SAMPLES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

MIDDLE EAST TECHNICAL UNIVERSITY

BY

DENIZ AKDEMIR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF

MASTER OF SCIENCE

IN

DEPARTMENT OF STATISTICS

JUNE 2003

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Statistics

Prof. Dr. H. Öztas Ayhan
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Statistics.

Prof. Dr. H. Öztas Ayhan
Supervisor

Examining Committee Members

Prof. Dr. H. Öztas Ayhan

Doç.Dr. Aysen Akkaya

Doç. Dr. Bilgehan Güven

Y. Doç. Dr. Inci Batmaz

Y. Doç. Dr. M. Qamar Islam

ABSTRACT

COMPONENTS OF RESPONSE VARIANCE FOR CLUSTER SAMPLES

Akdemir, Deniz

M.S., Department of Statistics

Supervisor: Prof. Dr. H. Öztas Ayhan

June 2003, 109 pages

Measures of data quality are important for the evaluation and improvement of survey design and procedures. A detailed investigation of the sources, magnitude and impact of errors is necessary to identify how survey design and procedures may be improved and how resources allocated more efficiently among various aspects of the survey operation. A major part of this thesis is devoted to the overview of statistical theory and methods for measuring the contribution of response variability to the overall error of a survey.

A very common practice in surveys is to select groups (clusters) of elements together instead of independent selection of elements. In practice cluster samples tend to produce higher sampling variance for statistics than element samples of the same size. Their frequent use stems from the desirable cost features that they have.

Most data collection and sample designs involve some overlapping between interviewer workload and the sampling units (clusters). For those cases, a proportion of the measurement variance, which is due to interviewers, is reflected to some degree in the sampling variance calculations.

The prime purpose in this thesis is to determine a variance formula that decomposes the total variance into sampling and measurement variance components for two commonly used data collection and sample designs. Once such a decomposition is obtained, determining an optimum allocation in existence of measurement errors would be possible.

Keywords: Measurement Errors, Response Errors, Interviewer Variance, Cluster Sampling, Simple Response Variance, Correlated Response Variance

ÖZ

KÜME ÖRNEKLEMELERINDE YANIT VARYANSININ BILESENLERI

Akdemir, Deniz

Master, Istatistik Bölümü

Tez Yöneticisi: Prof. Dr. H. Öztas Ayhan

Haziran 2003, 109 sayfa

Arastirmalarda, sonuçlarin degerlendirilebilmesi ve daha sonraki arastirmalarin tasarim ve islemlerinin gelistirilebilmesi açisindan veri kalitesi ölçütleri oldukça önemli bir yere sahiptir. Hatalarin kaynaklari, büyüklükleri ve etkilerinin ayrintili bir biçimde incelenmesi, arastirmanin tasarim ve islemlerinin ne sekilde iyilestirilebilecegi ve kaynaklarin birçok farkli kullanim için ne sekilde dagitilacagi hakkinda bilgi verir. Bu tezin önemli bir kismi arastirmalarda meydana gelen yanit degiskenliginin payinin ölçülmesine yarayacak yöntemleri ve istatistik teorisini incelemeye ayrilmistir.

Arastirmalarda yaygin olarak kullanilan bir yöntem örneklem elemanlarin tek tek ve bagimsiz sekilde seçilmesindense kümeler halinde seçilmesine dayanir. Uygulamada küme örneklemelerinin varyansi çogunlukla ayni büyüklükteki eleman örneklemesininkinden daha büyüktür. Küme örneklemesinin yaygin olarak kullanilma nedeni sahip oldugu uygun maliyet özelliklerine dayanir.

Anketörlerin is yükleri ve kümeler arasinda çogunlukla bir örtüsme görülür. Bu gibi durumlarda anketör varyansinin bir kismi örnekleme varyansi tarafindan ölçülür. Bu tezin temel amaci bu gibi durumlar için yanit varyansini bilesenlerine ayiran bir model gelistirmektir.

Anahtar Kelimeler: Ölçme Hatasi, Yanit Hatalari, Anketör Varyansi, Küme Örneklemesi, Ölçme Varyansi.

ACKNOWLEDGMENTS

I wish to express my deepest appreciation to Prof. Dr. H. Öztas AYHAN for his guidance and insight throughout the research.

It is also my duty to express my thanks to my parents and my friends who supported me in many ways during the preparation of my thesis.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

All statistical data, from whatever source and whatever the manner of their collection, are potentially subject to errors of various types. Even a complete census of all known members of a population is subject to errors.

Knowledge about data quality is required for their proper use and interpretation. This knowledge is essential in determining whether and with what degree of confidence the patterns observed in the results are real, and not merely products of the variability and deficiency inherent in the data. Information on the nature and magnitude of errors can also be useful for making appropriate corrections to the data or adjustments in their interpretation.

A survey attempts to acquire knowledge by observing the population and making quantitative statements about aggregated and disaggregated population characteristics. Surveys consist of a number of survey operations. Each phase of the operations affects the quality of survey estimates, and with each phase we can associate sources of errors in the estimate. A survey error refers to deviations of obtained results from those which are true reflections of population values.

1.1. Survey operations associated with survey errors

a. Sample selection: This phase consists of the execution of a predetermined sampling design using a suitable sampling frame. The sample size necessary to obtain the desired precision is determined. Errors in estimates associated with this phase are; frame errors of which undercoverage is particularly serious and sampling error which arises because a sample not the whole population, is observed.

b. Data Collection: There is a predetermined measurement plan with a specified mode of data collection (personal interview, telephone interview, mail questionnaire, or other). The field work is organised, interviewers are selected, and interviewer assignments are determined. Data are collected, according to the measurement plan, for the elements in the sample. Errors in estimates resulting from this phase include;

i. Measurement errors, where the respondent gives (intentionally or unintentionally) incorrect answers, the interviewer misunderstands or records incorrectly, the interviewer influences the responses, the questionnaire is misinterpreted, etc.

ii. Errors due to nonresponse (i.e., missing observations).

c. Data Processing: During this phase collected data are prepared for estimation and analysis. It includes the following elements; coding and data entry, editing, renewed contact with respondents to get clarification if necessary, imputation. Errors in estimates associated with this phase include transcription error (keying errors), coding errors, error in imputed values, errors introduced by or not corrected by edit.

d. Estimation and Analysis: This phase entails the calculation of survey estimates according to the specified point estimator formula, with appropriate use of auxiliary information and adjustment for nonresponse, as well as a calculation of measures of precision in the estimates (e.g., variance estimate, coefficient of variation of the estimate, confidence interval). Statistical analyses may be carried out, such as comparison of subgroups of the population, correlation and regression analyses, etc. All error from phases (a) to (c) above will affect the point estimates, and they should ideally be accounted for in the calculation of the measures of precision.Dissemination of Results and Postsurvey Evaluation

e. Dissemination of Results and Postsurvey Evaluation: This phase includes the publication of the survey results, including a general declaration of the conditions surrounding the survey. This declaration often follows a set of specified guidelines for quality declaration, which traditionally include two major categories: sampling and nonsampling errors.

1.2. Objectives of this Study

A major aim of this thesis is to provide an overview of statistical theory and methods for measuring the contribution of response variability to the overall error of a survey. The first chapters of the thesis are devoted to overview of the measurement error theory. Some concepts and ideas related to the subject are reviewed concisely. Various available measurement error models are described. Those chapters, which are meant to provide an overview of measurement error theory, are not meant to give a complete treatment of the subject. They are meant as an introduction to the second aim of this thesis.

The second major aim of this thesis is determining a variance formula that decomposes the total variance into sampling and measurement variance components for data collection and sample designs which involve a degree of correspondence between interviewer workload and the sampling units (clusters). This data collection and sample designs need a specific variance decomposition model. Assigning only one interviewer for each cluster generates a different variance then assigning interviewers randomly to sampling units. A different assignment and sample design is often used to reduce the costs of the survey but it also changes the variance structure. Utilising a general measurement error model for a survey with interviewers given by Lessler and Kalsbeek (1992) the variance decomposition is obtained for the assignment and sample design which involves all cluster elements for chosen clusters being observed.

A random sub-sample of cluster elements, considered next, complicates the precision error of the cluster mean estimator, introducing a sampling error within the clusters. Within a cluster two kinds of error will be present: sampling error and measurement error made by the interviewers. These two errors have been assumed to be independent, and a linear additive model is used to illustrate their total effect. This second data collection and sample design involved further complexities but the proper use of the model for the first data collection and sample design made the solution possible.

This thesis examines the languages of measurement errors in Chapter 2. In Chapter 3, empirical estimation of survey measurement errors are covered. Sources of measurement error issues are examined in the Chapter 4. Survey costs subject is an important aspect of the survey operation, which is covered in Chapter 5. In Chapter 6, measurement error models is covered. Decomposition of the survey error is discussed in Chapter 7. Finally, cluster sampling and variance decomposition proposals are followed by the conclusion of this thesis.

CHAPTER 2

LANGUAGES OF MEASUREMENT ERROR

The field of measurement of survey error components has evolved through the somewhat independent, and uncoordinated, contributions of researchers trained as statisticians, psychologists, political scientists, and sociologists. Therefore, it lacks a common language and a common set of principles for evaluating new ideas. According to Groves (1989) at least three major languages of error appear to be applied to survey data. They are associated with three different academic disciplines and illustrate the consequences of groups addressing similar problems in isolation of one another. The three disciplines are statistics (especially statistical sampling theory), psychology (especially psychometric test and measurement theory), and economics (especially econometrics). Although other disciplines use survey data (e.g., sociology and political science), they appear to employ languages similar to one of those three.

2.1. Measurement Error Terminology in Survey Statistics

The total error of a survey statistic is labeled *the mean squared error*; it is the sum of all variable errors and all biases. Another common conceptual structure labels the total survey error of a survey statistic, by *the root mean square error* (Kish, 1965); which is the square root of the mean squared error.

Bias of a statistic is a systematic error that affects the statistic in all implementations of a survey design; in that sense it is a constant error. Bias of a survey can not be measured from within the survey. Its estimation involves validating information from sources that are external to the survey.

A *variable error*, measured by the *variance* of a statistic, arises because achieved values differ over the units (e.g., sampled persons, interviewers used, questions asked) that are the sources of the errors. The concept of variable errors inherently requires the possibility of repeating the survey, with changes of units in the replications.

Variable errors and biases are connected; bias is the part of error common to all implementations of the survey design, and variable error is the part that is specific to each trial. A survey design defines the fixed properties of the data collection over all possible implementations.

Define the mean square error as

$$\text{Mean Square Error} = \text{Variance} + \text{Bias}^2$$
$$E_{s,t,i,a}(\overline{y}_{s,t,i,a} - \overline{m})^2 = E_{s,t,i,a}(\overline{y}_{s,t,i,a} - \overline{y}_{....})^2 + (\overline{y}_{....} - \overline{m})^2 \qquad \textbf{(2.1.1)}$$

where $E_{s,t,i,a}()$ denotes the expectation over all samples, s, given a sample design; all trials, t; all sets of interviewers, i, chosen for the study; and all assignment patterns, a, of interviewers to sample persons; $\overline{y}_{s,t,i,a}$ denotes the mean over respondents in the s-th sample, t-th trial, i-th set of interviewers, a-th assignment pattern of interviewers to sample persons, for y, the survey measure of the variable m in the target population; $\overline{y}_{....}$ denotes the expected value of $\overline{y}_{s,t,i,a}$ over all samples of respondents, all trials, all sets of interviewers, and all assignment patterns; $\overline{m}$ denotes the mean of the target population for true values on variable m. The bias of the mean is

$$Bias(\overline{y}_{s,t,i,a}) = (\overline{y}_{....} - \overline{m}) \ . \qquad \textbf{(2.1.2)}$$

The variance of the mean is

$$Var(\overline{y}_{s,t,i,a}) = E_{s,t,i,a}(\overline{y}_{s,t,i,a} - \overline{y}_{....})^2 \ . \qquad \textbf{(2.1.3)}$$

2.1.1. Sample Selection

Observational errors concern the accuracy of measurement at the level of individual units enumerated in the survey. These arise from the fact that what is measured on the units included in the survey can depart from the actual (true) values for those units. Observational errors are deviations of the answers of respondents from their true values on the measure. Observational errors center on substantive content of the survey: definition of the survey objectives and questions; ability and willingness of the respondent to provide the information sought; the quality of data collection, coding editing, processing etc.

2.1.2. Errors of nonobservation

Errors of observation concerns generalizability from the units observed to the target population, includes sampling variability and various biases associated with sample selection and implementation, such as coverage, selection and non-response errors. These are errors in the process of extrapolation from the particular units enumerated to the entire study population for which estimates or inferences are required. These center on the process of sample design and implementation, and include errors of coverage, sample selection, sample implementation and non-response, as well as sampling errors and estimation bias.

The above categorisation is based on operational considerations, and in a sense is more fundamental than the distinction usually made between sampling and non-sampling errors. In the survey statistics terminology *sampling errors* are viewed as the errors emerging because of the sampling procedure. *Nonsampling errors* are often thought of being due to mistakes and deficiencies during the development and execution of the survey procedures.

Each group of errors may be further classified in as much detail as possible to identify specific sources of error, so as to facilitate their assessment and control:

a. Observational errors

a.1. Conceptual errors

- errors in basic concepts, definitions, and classifications
- errors in putting them into practice (questionnaire design, interviewers training and instructions)

a.2. Response errors

- response bias
- simple response variance
- correlated response variance

a.3. Processing errors

- editing errors

- coding errors
- data entry errors
- programming errors

b. Non-observational errors

b.1. Coverage and related errors

- omissions
- incorrect boundaries
- outdated lists
- sample selection errors

b.2. Non-response

- refusals
- inaccessible
- not-at-homes, etc.

b.3. Sampling error

- sampling variance
- estimation bias

There are two main alternative views on survey error within the survey statistics field. The simpler view on the survey error is based on the assumption that the only source of variation in survey results comes from measuring different subsets of the population. Thus, sampling variance is the only variable error. This view is taken in most standard statistical sampling theory. According to this view the variance given in (2.1.3) contains only the sampling variance,

$$E_{t,i,a}(E_s(\overline{y}_{s,t,i,a} - \overline{y}_{.,t,i,a})^2), \qquad (2.1.4)$$

where $\bar{y}_{.t,i,a}$ is the expected value of $\bar{y}_{s,t,i,a}$ over all samples of respondents, s, given a sampling design.

According to Groves (1989) a more elaborated view of survey error held by some survey statisticians comes from those interested in *total survey error*. Underlying this perspective is the notion that the survey at hand is only one of an infinite number of possible trials or replications of the survey design. Respondents are assumed to vary in their answers to a survey question over trials, leading to *simple response variance* (Hansen, Hurwitz and Pritzker, 1964). The interviewer is often treated as a source of error in this perspective, and is most often conceptualized as a source of variable error.

The variable effects that interviewers have on respondent answers are sometimes labeled *correlated response variance* in this perspective (Bailey, Moore and Bailar, 1978). Measurement bias or response bias refers to systematic errors that have a discernible pattern compared to the "true response". The response bias of an estimate will not be reflected in the variance of a sample statistic; its effect, if it can be estimated, will be reflected in the mean squared error. The simple response variance is defined as

$$E_{s,i,a}(E_t(\bar{y}_{s,t,i,a} - \bar{y}_{s,i,a})^2) \ . \qquad (2.1.5)$$

2.2. Measurement Error Terminology in Psychological Measurement

Groves (1989) states that when moving from survey statistics to psychometrics, the most important change is the notion of an unobservable characteristic the researcher is attempting to measure with a survey indicator (i.e., a question). In contrast, within survey statistics, the measurement problem lies in the operationalization of the question (indicator, in psychometric terms). The psychometrician, typically dealing with attitudinal states, is more comfortable labelling the underlying characteristic (construct, in psychometric terms) as unobservable, something that can only be approximated with any applied measurement.

There are two influential measurement models. In the first, classical true score theory, all observational errors are viewed as joint characteristics of a particular measure and the person to whom it is administered. In such measurement the expected value (over repeated administrations) of an indicator is the true value it is attempting to measure. That is, there is no measurement bias possible, only variable error over repeated administrations. Although classical true scores provide the basis for much of the language of errors in psychometrics, it is found to be overly restrictive for most survey applications.

An additional change when moving to the field of psychometric measurement is the explicit use of models as part of the definition of errors. That is, error terms are defined assuming certain characteristics of the measurement apply. In this perspective expectations of the measures are taken over trials of administration of the measurement of a person. That is, each asking of a question is one sample from an infinite population (of trials) of such askings. The *propensity distribution* describes the variability over trials of the error for the particular person. Under the classical true score assumption the mean of that distribution is zero. When there is interest in a population of persons, the expected value of the indicator is taken both over the many propensity distributions of the persons in the population and the different persons.

The true score m_j on the construct m, of a person, j, on the indicator g is defined as the expected value of the observed score; that is,

$$m_j = E_t(y_{gjt}), \tag{2.2.1}$$

where y_{gjt} denotes the response to indicator g on the t-th trial for the j-th person; and the expectation $E_t(.)$ is with respect to the propensity distribution over trials of the indicator's administration for the j-th person. The model for measurement is

$$\begin{array}{l}\text{Response = True Score + Error}\\ y_{gjt} = m_j + e_{gjt},\end{array} \tag{2.2.2}$$

where e_{gjt} is the error for the g-th indicator committed by the j-th person on the t-th trial.

Two terms in the psychometric perspective, validity and reliability, are frequently used to label two kinds of variable errors. The notion of theoretical validity, sometimes called construct validity, is used to mean the correlation between the true score and the respondents answer over trials. The measure is taken to be one of an extensible set of indicators of the construct. It is not equated with the construct it attempts to measure or it is not considered to define the construct itself. This is in contrast with strict operationism, in which each construct is defined in terms of a narrowly specified set of operations.

Theoretical validity of the g-th indicator, for the population of which the j-th person is a member, is

$$\frac{\text{Covariance of Indicator and True Score}}{\text{(Standard Deviation of Indicator)(Standard Deviation of True Score)}}$$

$$\frac{E_{jt}[(y_{gjt}-\bar{y}_{g..})(m_j-\bar{m})]}{\sqrt{E_{jt}(y_{gjt}-\bar{y}_{g..})^2}\sqrt{E_{jt}(m_j-\bar{m})^2}}=r_{ym}, \tag{2.2.3}$$

where r_{ym} is the correlation between the true scores and observed values over trials and persons in the population; $\bar{y}_{g..}$ is the mean over persons and trials of observed scores; and $\bar{m}$ denotes the mean over persons of true values.

The other error concept used in psychometrics is reliability, the ratio of the true score variance to the observed score variance. Variance refers to variability over persons in the population and over trials within a person.

$$\begin{aligned}\text{Index of Reliability} &= \frac{\text{Variance of True Score}}{\text{Variance of Indicator}}\\ &= \frac{E_{jt}(m_j-\bar{m})^2}{E_{jt}(y_{gjt}-\bar{y}_{g..})^2}=\frac{s_m^2}{s_y^2}=r_y,\end{aligned} \tag{2.2.4}$$

where s_m^2 is the variance of the true scores across the population and trials; s_y^2 is the variance of the observed scores across the population; r_y is the index of reliability. With this definition of reliability, it can be noted that the concept is not defined for measurements on a single person, only on a population of persons and reliability has a value specific to that population.

Validity and reliability can be assessed only with multiple indicators. Bohrnstedt (1983) makes the distinction between theoretical validity, which is defined on a single indicator, and empirical validity, an estimation of theoretical validity that can be implemented only with another measure of the same construct. Sometimes criterion validity is used to denote that the other measure is assumed to be measured without any variable error. Empirical or criterion validity of y_1 in relation to y_2, where y_1 and y_2 are two indicators of m, is given by the correlation of y_1 and y_2 over trials and persons in the population, $r_{y_1 y_2}$.

2.3. Measurement Error Terminology in Econometrics

In the field of econometrics the terminology for errors arise mostly through the language of the general linear model. The observations analysed are viewed to be a collection of events from a random process. In this respect the term measurement error model is used to denote a regression model, either linear or non-linear, where at least one of the covariates or predictors is observed with error. If m_j denotes the value of the covariate for the j-th sample unit, then m_j is unobserved, instead we observe $y_j = f(m_j, d_j)$ where d_j is known as the measurement error. The observed (or indicator) variable is assumed to be associated to the unobserved (or latent) variable via the function f. The form of this function defines the different types of measurement error models.

Within the class of error models , there are two variants: *classical additive error models*, and *error calibration models*. The classical additive error model establishes that the observed variable y_t on the $t-th$ trial is an unbiased measure of m. That is

$$y_t = m + d_t, \tag{2.3.1}$$

where d_t is a random variable with mean 0 and variance s_d^2.

We talk about error calibration models when the observed variable is a biased measurement of the variable of interest. In this case a regression model to associate the two variables is

$$y_t = a_0 + a_1 m + d_t \, , \tag{2.3.2}$$

where as before $E_t(d_t) = 0$, but now $E_t(y_t) = a_0 + a_1 m$. If information about the relationship (2.3.2) is available, the measurement y_t can be calibrated by using $a_1^{-1}(y_t - a_0)$.

CHAPTER 3

EMPIRICAL ESTIMATION OF SURVEY MEASUREMENT ERROR

This section describes some techniques for evaluating and controlling measurement error in surveys. The methods discussed are

a. Reinterview Studies

b. Multiple Indicators Studies

c. Record Check Studies

d. Cognitive Studies

3.1. Reinterview Studies

A reinterview (replicated measurement on the same unit in interview surveys) is a new interview which repeats all or part of the questions of the original interview. When implementing reinterview methodology, there are two underlying assumptions:

a. The reinterview is independent of the first interview,

b. The original interview and the reinterview either use the same mode of data collection and are conducted under the same general conditions or the reinterview and reconciliation provide "true" values.

Reinterview studies requiring two sets of measurements on the sample or part of it have been implemented since the early days of sample surveys

(Mahalanobis, 1946). There are three major purposes for conducting reinterview studies:

a. Estimation of simple response variance or reliability: A reinterview will permit the partitioning of the observed variability of responses into the sampling variance and the simple response variance. A reinterview used to measure either simple response variance or reliability must be an independent replication of the original interview. Independence is threatened, however, by conditioning, which occurs when respondents remember their first answer during the reinterview.

b. Estimation of the response bias: Theoretically, the measurement of response bias requires the existence of data from which the true value may be estimated; however, often these data do not exist. In practice, reinterview programs frequently estimate a measure of response bias by including a process known as reconciliation. This is when the respondent is asked to reconcile answers that differed between the original and the reinterview. Reconciliation can occur during or at the end of the reinterview or in a separate, third contact.

c. Evaluation of the field work: Reinterview studies can be used to identify interviewers who are falsifying data, and who misunderstand the survey procedures and require additional training. The different purposes for which reinterviews may be used necessitate different methodologies and thus dictate different reinterview designs. Forsman and Schreiner (1991) describe four basic reinterview designs. Two focus on evaluating interview performance (one of which was specifically developed to detect interviewer falsification), and two on estimating measurement error components of the interview data (one estimating simple response variance and reliability and the other estimating response bias.

Forsman and Schreiner (1991) explain that each basic design is characterized by the following six factors:

a. The method of reinterview sample selection. The reinterview sample can be a onestage sample of respondents, households, or clusters of households (such a cluster may consist of, e.g., four neighbouring households). The reinterview sample can also be a two-stage sample, where the original interviewers are primary sampling units, and respondents (or households or clusters) within interviewers are secondary sampling units (ssu). Such a two-stage sample permits a proper allocation of ssu's over interviewers.

b. The choice of reinterviewers. The reinterviewers can be selected from the same pool of interviewers as the original interviewers. They may also be selected from among the most experienced interviewers in this pool. A third option is to select the reinterviewers from a group of supervisors.

c. The choice of respondent. The respondent can be the same as in the original interview; he or she can be chosen according to the same procedure as in the original interview ("original respondent rule"); the respondent might be the most knowledgeable person in the household, or each person could respond for himself or herself ("self-response").

d. The design of the reinterview questionnaire. The reinterview questionnaire may be exactly the same as the original questionnaire, or may contain a subset of the original questions. To achieve "true" values, the reinterview questionnaire may contain probing questions.

e. Whether or not to conduct reconciliation. When the responses obtained during the reinterview differ from those obtained in the original interview the differences are evaluated through a process called reconciliation. During reconciliation the respondent is provided with the information received in both interviews and asked to determine what is the correct information.

f. The choice of mode. The choice is between telephone and face to face interviews. If the purpose of the reinterview is to estimate response variance or reliability the questions are repeated exactly, the responses are not reconciled, and the mode is the same as in the original interview. When estimating bias, however, the purpose is to obtain the "true" response. Here, the reinterview design should include the most experienced interviewers and supervisors. Likewise, reinterviews designed to measure response bias should target the most knowledgeable respondent, not necessarily the original respondent. If estimating response bias, the questions can be modified to elicit more accurate responses, reconciliation is used, and the mode of data collection need not be the same as the original interview.

3.2. Multiple Indicators Studies

Groves (1989) describes multiple indicators studies as another approach that uses replicated measures to estimate measurement error, but it uses multiple measurements of the same characteristic in a single survey. In this approach measurement error associated with a particular method of data collection and/or a particular question can be assessed. Measurement error, here, is defined as a component of variance in the observed value of indicators, not corresponding to variability in the true values of the underlying measures.

3.3. Record Check Studies

Record check studies are used to estimate response bias. As described in the section on reinterview studies, the measurement of response bias theoretically requires the existence of data from which the true value may be estimated. When these data do not exist, reinterview studies frequently use reconciliation. When these data do exist and are available, record check studies are possible. Such a study generally assumes that information contained in the records is without error, that is, the records contain the true values on the survey variables.

Groves (1989) describes three kinds of record check study designs: the reverse record check study, the forward record check study, and the full design record check study. The different designs are based in part on the relation of the survey sample to the external source of data providing the comparisons.

In the reverse record check study, which Groves also refers to as the retrospective design, the researcher goes back to the records which were the source of the sample to check the survey responses. That is, the survey sample is drawn from a record file considered to contain accurate data on a trait or characteristic under study, and the survey includes some questions on information already in the records. The survey data are compared with the record data to estimate measurement error.

The weakness of reverse record check studies is that they cannot by themselves measure errors of overreporting (falsely reporting an event). They can only measure what portion of the records sample correspond to events reported in the survey and whether the characteristics of the events are the same on the records as in the survey report.

In a forward record check study, the researcher obtains the survey data first and then moves to new sources of record data for the validity evaluation. Thus, in this design, the sample is drawn from a separate frame. Once the survey responses have been collected, the researcher searches for relevant records containing information on the respondents and makes comparisons. Some surveys may be designed to include questions asking about where records containing similar information on the sample person can be found.

Forward record check studies work well for measuring overreports in a survey, but they are not commonly used. They generally entail contacting several different record-keeping agencies and may require asking the respondents for permission to access their record files from the different agencies. They are also limited in their measurement of underreporting:

The full design record check study combines features of the reverse and forward record check designs. The survey sample comes from a frame covering all persons of the population (reverse record check design) and researchers seek records from all sources relevant to those persons (forward record check design). Thus, researchers measure survey errors associated both with underreporting and overreporting by comparing all records corresponding to the respondent. However, this design requires a data base that covers all persons in the target population and all events corresponding to those persons.

All validity evaluation designs share three limitations. As mentioned earlier, there is the assumption that the record systems do not contain errors of coverage, nonresponse, or missing data. Second, it is also assumed that the individual records are complete and accurate, without any measurement errors. The third limitation involves matching errors (difficulties matching respondent survey records with the administrative records) and these could affect the estimation of measurement errors.

3.4. Cognitive Studies

Forsyth and Lessler (1991) contend that "if we are to understand the sources of survey measurement error and find ways of reducing it, we must understand how errors arise during the question-answering process. This will allow us to develop better questions that will yield more accurate answers. The primary objective of cognitive laboratory research methods is not to merely study the response process, but through careful analysis to identify questioning strategies that will yield more accurate answers". As Nolin and Chandler (1996) explain, the methods of cognitive research can be used to increase understanding of the ways that respondents comprehend survey instructions and questions, recall requested information, and respond to the influence of word and question order.

Cognitive research draws on three different literatures: research in cognitive psychology on memory and judgment, research in social psychology on influences against accurate reporting, and evidence from survey methodology research regarding response errors in surveys. Literature in survey methodology concentrates on models of measurement of response errors, rather than on explaining their presence. For example, survey methodology has documented response errors and identified respondent groups and response tasks that are more prone to these errors.

Theories of cognitive psychology have been applied to survey measurement to gain insight into how the respondent's attributes and actions may affect the quality of survey data. These theories focus on how people encode information in their memories and how they retrieve it later. Social psychological literature, on the other hand, emphasizes the influences on communication of answers to survey questions.

Researchers generally agree on five stages of action relevant to survey measurement error:

a. Encoding of information: how the respondent obtains, processes, and stores information in memory

b. Comprehension: how the respondent assigns meaning to the interviewer's question

c. Retrieval: how the respondent searches for memories of events or knowledge relevant to the question

d. Judgement of appropriate answer: how the respondent chooses from alternative responses to the question

e. Communication: How the respondent answers through all the other personal characteristics and social norms that might be relevant (Groves, 1989)

Beyond acceptance of these five stages, cognitive research takes different paths. Forsyth and Lessler (1991) conducted a literature review of cognitive research methods used to study the survey question-answering process and discussed the topic with others who have conducted cognitive research. They concluded that no guidelines were available for choosing one cognitive research method over another. While a number of response models have been developed, there is yet little consensus on how the models are implemented.

Oksenberg and Cannell (1977) and Tourangeau (1984) models assumed a basic sequence that respondents followed when answering a question, but there is no consensus on the procedural details of these methods. Forsyth and Lessler "believe that this lack of consensus may be due, in part, to a lack of theoretical and empirical work that explores how methodological details can affect cognitive laboratory results" (Forsyth and Lessler, 1991). Nonetheless, they offer a summary of four general sets of methods that have been implemented: Expert evaluation methods, Expanded interview methods, Targeted methods, Group methods.

All of these methods provide more information about the question-answering process than can be obtained through simply asking the survey questions and recording the answers. The methods differ according to their timing and the amount of control the researcher has over what is observed. The task timing may be either concurrent, immediately after the respondent answers the questions, delayed, or unrelated. Either the respondent decides what information will be observed, as in concurrent think-aloud interviews, or response data are independently processed by the researcher as in behaviour coding. All cognitive laboratory methods are basically qualitative studies even though some of the methods do collect quantitative information.

CHAPTER 4

SOURCES OF MEASUREMENT ERRORS

Measurement error comes from four primary sources (Biemer et al. 1991). These four sources are the elements that comprise data collection. While generally these sources are addressed separately, they can also interact.

These are:

a. Questionnaire: The questionnaire is the presentation of the request for information.

b. Data Collection Method: The data collection mode is how the questionnaire is delivered or presented.

c. Interviewer: The interviewer is the deliverer of the questionnaire.

d. Respondent: The respondent is the recipient of the request for information.

4.1. Questionnaire Effects

The questionnaire is designed to communicate with the respondent in an unambiguous manner. It represents the survey designer's request for information. Questionnaires to be compared may differ in question wording, question order, response categories, and so on. If an independent data source were available, then results from the two questionnaire versions could be compared to the external data source to determine the "best" version. Otherwise, the result from the two groups could be compared to each other to determine the extent of any differences in reporting. As another variation, the same group of respondents can be asked similar versions of the same questions at a different point of time, but the questions asked must be those for which answers are expected to remain the same over time.

a. Specification problems: At the survey planning stage, error can occur because the data specification is inadequate or inconsistent with what the survey requires. Specification problems can occur due to poorly worded questionnaires and survey instructions, or may occur due to the difficulty of measuring the desired concept. These problems exist because of inadequate specifications of uses and needs, concepts, and individual data elements.

b. Question wording: The questionnaire designer attempts to carefully word questions so s/he will communicate unambiguously. The designer wants the respondent to interpret the question as the designer would interpret the question. Words, phrases, and items used in questionnaires are subject to the same likelihood of misunderstanding as any form of communication. The questionnaire designer may not have a clear formulation of the concept s/he is trying to measure. Even if s/he has a clear concept, it may not be clearly represented in the question. And, even if the concept is clear and faithfully reproduced, the respondent may not interpret the request as intended.

c. Length of the questions: The questionnaire designer is faced with the dilemma of keeping questions short and simple while assuring sufficient information is provided to respondents so they are able to answer a question accurately and completely. Longer questions may provide more information or cues to help the respondent remember and more time to think about the information being requested. The effect of question length may be measured if an independent source of data is available by randomly assigning sample units to one of two groups, one receiving a "short" version of the questions and the other group receiving the "long" version of the questions. Responses for each group can then be compared with the "known" values for these questions.

d. Length of the questionnaire: A questionnaire of excessive length can cause errors resulting from fatigue or boredom of the respondent or the interviewer. Length of the questionnaire may also be related to nonresponse error, discussed briefly in chapter 2. If an independent data source is available, the impact of questionnaire length may be tested using a designed experiment. In this experiment, the questions are split into two halves. The question sets appear in reverse order on the two questionnaires.

e. Order of questions: Question order can affect the responses when it affects recall or creates confusion. Asking questions may affect how respondents answer later questions, especially in attitude and opinion surveys, where researchers have observed effects of the question order. Respondents may also use information from previous items about what selected terms mean to help answer subsequent items. The effect of question order can be assessed by administering alternate forms of a questionnaire to random samples.

f. Open and closed formats: Question formats in which respondents are asked to respond using a specified set of options (closed format) may yield different responses than when respondents are not given categories (open format). The closed format may remind respondents of something they may not have otherwise remembered to include. The response options to a question cue the respondent as to the level or type of responses considered appropriate.

g. Questionnaire format: For the self-administered questionnaire the design and layout of the instrument may help or hinder accurate response. The threat is that a poor design may confuse respondents, lead to a misunderstanding of skip patterns, fatigue respondents, or contribute to their misinterpretation of questions and instructions. Cognitive research methods provide information to asses the design and format of questionnaires.

4.2. Data Collection Mode Effects

Various methods or modes are available for collecting data for a survey. Lyberg and Kasprzyk (1991) present an overview of different data collection methods along with the sources of measurement error for these methods.

a. Face-to-face interviewing: Face-to-face interviewing is the mode in which an interviewer administers a structured questionnaire to respondents. Using a paper questionnaire, the interviewer completes the questionnaire by asking questions of the respondent. Although this method is generally expensive it does allow a more complex interview to be conducted. This mode also allows the use of a wide variety of visual aids to help the respondent answer the questions.

One problem for face-to-face interviewing is the effect of interviewers on respondents' answers to questions, resulting in increases to the variances of survey estimates. Another possible source of measurement error is the presence of other household members who may affect the respondent's answers. In situations where multiple respondents are required to complete a questionnaire, the interaction of the group of respondents can cause differences in the reported values. This is especially true for topics viewed as sensitive by the respondents. Measurement error may also occur because respondents are reluctant to report socially undesirable traits or acts.

b. Telephone interviewing: This mode is very similar to face-to-face interviewing except interviews are conducted over the telephone rather than in person. Telephone interviewing is usually less expensive and interviews often proceed more rapidly. However, this mode also provides less flexibility. This mode can be conducted from the interviewers' homes or from centralised telephone facilities. Centralised telephone interviewing makes it possible to monitor interviewers' performance and provide immediate feedback. Since the interviewer plays a central role in telephone interviewing as well, the sources of measurement error are very similar to those in face-to-face interviewing although the anonymity of the interviewer may improve reporting on sensitive topics by providing adequate "distance" between interviewer and respondent.

c. Self-administered surveys: Any survey technique that requires the respondent to complete the questionnaire him/herself is referred to as a self-administered survey. The most common ways of distributing these surveys are through the use of mail, fax, newspapers/magazines, and increasingly the internet, or through the place of purchase of a good or service (hotel, restaurant, store).

A considerable advantage of the self-administered survey is the potential anonymity of the respondent, which can lead to more truthful or valid responses. Also, the questionnaire can be filled out at the convenience of the respondent. Since there is no interviewer, interviewer error or bias is eliminated. The cost of reaching a geographically dispersed sample is more reasonable for most forms of self-administered surveys than for personal or telephone surveys, although mail surveys are not necessarily cheap. In most forms of self-administered surveys, there is no control over who actually fills out the questionnaire. Also, the respondent may very well read part or the entire questionnaire before filling it out, thus potentially biasing his/her responses.

Self administered mail surveys are the most commonly used data collection mode for economic surveys. In mail surveys, the questionnaires are mailed to the ultimate sampling. The respondents complete and mail back the questionnaire. Mail surveys have different sources of measurement error than face-to-face and telephone interviewing. Self administered mail surveys have no interviewer effects and less risk of “social desirability” effects. However, this mode is more susceptible to misreading and misinterpretation of questions and instructions by the respondents. Good questionnaire design and formatting are essential to reduce the possibility of these problems.

d. Diary surveys: Diary surveys are usually conducted for topics that require detailed behaviour reporting over a period of time. The respondent uses the diary to enter information about events soon after they occur to avoid recall errors. Interviewers are usually needed to contact the respondent to deliver the diary, gain the respondent’s co-operation and explain the data recording procedures, and then again to collect the diary and, if it is not completed, to assist the respondent in completing the diary.

Lyberg and Kasprzyk (1991) identify a number of sources of measurement error for this mode such as, respondents giving insufficient attention to recording events and then failing to record events when fresh in their memories; the structure and complexity of the diary can present significant practical difficulties for the respondent; and respondents may change their behaviour as a result of using a diary.

e. Direct observation: Direct observation is a method of data collection where the interviewer collects data by direct observation using his/her senses (vision, hearing, touching, testing) or physical measurement devices. This method is used in many disciplines. An inaccurate counter, a faulty scale, or poorly calibrated equipment may cause measurement errors.

Measurement errors may be introduced by observers in ways similar to the errors introduced by interviewers; for example, observers may misunderstand concepts and misperceive the information to be recorded, and may change their pattern of recording information over time because of complacency or fatigue.

f. Mixed data collection mode: Two or more modes of data collection are used for some surveys to save money, improve coverage, improve response rates, or to reduce measurement errors.

4.3. Interviewer Effects

Because of individual differences, each interviewer handles the survey situation in a different way, that is, in asking questions, probing and recording answers, or interacting with the respondent, some interviewers appear to obtain different responses from others. Interviewers may not ask questions exactly as worded, follow skip patterns correctly or probe for answers nondirectively. They may not follow directions exactly, either purposefully or because those directions have not been made clear enough. Interviewers may vary their inflection, tone of voice, etc. without even knowing it.

To the extent these errors are large and systematic, a bias, as measured in the mean squared error of the estimate, will result and this is called the interviewer effect. Another potential source of interviewer effects is respondent reaction to characteristics of the interviewer, such as age, race, sex, or to attitudes or expectations of the interviewer.

a. Interviewer characteristics: Groves (1989) reviewed a number of studies and concluded, in general, demographic effects appear to apply when the measurements are related to the characteristics but not otherwise. That is, there may be an effect based on the race of the interviewer if the questions asked were related to race. Other interviewer factors may also play a role in interviewer-produced error, such as voice characteristics and interviewing expectations.

Three different means to control interviewer errors are: training, supervision or monitoring, and workload manipulation. Standardisation of the measurement process especially as it relates to interviewers' tasks leads to a decrease in interviewer effects. One way to accomplish standardisation is through a training program of sufficient length to cover interview skills and techniques as well as information on the specific survey

Supervision and performance monitoring are essential ingredients of a quality control system. Developing good supervisory practices is essential because the supervisors are often the first level at which problems are recognised or corrected. Supervisors can help interviewers understand their job better, provide additional training, and assure that workload does not impact the quality of the work. Reinterview programs and field observations are conducted to evaluate individual interviewer performance. Observations in the field are conducted using extensive coding lists or detailed observers' guides where the supervisor or monitor checks whether the procedures are properly followed.

A third way to control interviewer effects is to change the average workload; interviewer variance increases as average workload increases. The issue is to find the optimal average workload. Optimal workload as a function of interviewer hiring and training costs, interview costs, and size of intra-interviewer correlation.

b. Correlated interviewer variance: In the early 1960's attention turned to estimating the size of the interviewer effect and three different approaches were suggested (Hansen, Hurwitz, and Bershad (1961), Kish (1962), and Fellegi (1964)). Even apparently small interviewer intraclass correlations can produce important losses in the precision of survey statistics. For practical and economic considerations, each interviewer usually has a large workload. An interviewer who is contributing a systematic bias will thus affect the results obtained from several respondents and the effect on the variance is large.

4.4. Respondent Effects

Respondents may contribute to error in measurement by failing to provide accurate responses. Groves (1989) indicates that both traditional models of the interview process and the cognitive science perspectives on survey response identify the following five sequential stages in the formation and provision of answers by survey respondents:

a. Encoding of information: involves the process of forming memories or retaining knowledge.

b. Comprehension of the survey question: involves knowledge of the words and phrases used for the question as well as the respondent's impression of the purpose of the survey, the context and form of the question, and the interviewer's behaviour in asking the question.

c. Retrieval of information from memory: involves the respondent's attempt to search her/his memory for relevant information.

d. Judgement of appropriate answer: involves the respondent's choosing from the alternative responses to a question based on the information that was retrieved.

e. Communication of the response: involves the consideration of influences on accurate reporting that occur after the respondent has retrieved the relevant information as well as the respondent's ability to articulate the response.

There are many aspects of the survey process that can affect the quality of the respondent's answers resulting from this five-stage process.

a. Respondent rules: One survey factor related to the response process is the respondent rules. For surveys collecting information for the sample unit, the specific respondent's knowledge about the answers to the questions may vary among the different eligible respondents. Surveys collecting information for individuals within the sample unit (e.g., persons within households, employees within businesses, and students and teachers within schools) may use self-reporting or proxy reporting. Self versus proxy reporting differences vary by subject matter.

b. Questions: The respondent's comprehension of a question is affected by the wording and complexity of the question, and the design of the questionnaire. The respondent's ability to recall the correct answer is affected by the type of question asked and by the difficulty of the task in determining the answer. The respondent's willingness to provide the correct answer to questions is affected by the type of question being asked, by the difficulty of the task in determining the answer, and by the respondent's view concerning the social desirability of the responses.

c. Interviewers: The respondent's comprehension of the question is affected by the interviewer's visual clues as well as audio cues. The interviewer reads the question incorrectly, does not follow the appropriate skip pattern, misunderstands or misapplies the questionnaire, or records the wrong answer.

d. Recall period: The longer the time period between an event and the survey the more likely it is respondents will have difficulty remembering the activity the question is asking about. Survey designers need to identify the recall period that minimises the total mean squared error in terms of the sampling error and possible biases.

e. Telescoping: Telescoping occurs when respondents report occurrences within the recall period when they actually occurred outside the recall period.

f. Timing of the interview: The timing of an interview can also impact respondent error. Interviews soon after the end of a business cycle, tax preparation, or other reporting period may improve recall, while interviews during busy times may result in rushed responses.

CHAPTER 5

SURVEY COSTS

In survey work, one generally seeks to use a sample design that has two properties: a satisfactory level of information capacity, and costs that are consistent with available budgets and that make reasonably efficient use of resources. Information capacity is generally measured by the variances of the estimators of selected population quantities that are considered to be of principal interest.

The costs of survey activities often act as limiting influences on efforts to reduce survey errors. A classical problem in survey research is how to optimize sample design with respect to variance and cost. Survey costs and errors are reflections of each other; increasing one reduces the other.

Determining an optimum allocation requires assumptions about the variance of survey estimators and about the nature of survey costs. The variance model can be derived explicitly, depending on the type of design and the population value being estimated from the sample. Furthermore, estimates of the important parameters of the variance model are easily estimable and can be obtained from published reports.

However, unlike the variance model, which can be mathematically derived given the statistical implications of the sampling design, identification of the functional form of the cost model is a less rigorous process. The model reflecting survey costs is largely dependent on how one views the survey protocol and the amount of complexity one allows in its formulation.

The ideal cost model should have three characteristics: First, it must realistically represent the way in which costs are incurred in an actual survey operation. Second, the formulation should be simple enough so that the optimum solution is tractable. Third, unit costs, which constitute the parameters of the cost model, should be sufficiently straightforward in interpretation so that they can be easily understood by operations staff to develop useful estimates for calculating optimum allocations.

The selection of design parameters is based on an examination of costs and on an understanding of the error structure. To solve the optimisation problem cost model needs to be developed which contains terms that are also present in the error model. Each of the units which acts to improve the quality of the survey statistics also brings with it a cost.

In addition to the matter of choosing an appropriate functional form for a cost model, one is faced with the problem of obtaining good estimates of unit costs, the parameters of the model that is chosen. By combining the cost models and error structures, an optimisation problem is posed. This optimisation problem involves one of the following optimisation criteria:

a. Minimise the total cost for required variance,

b. Minimise the variance for a given total cost,

c. Minimize the product of the variance and the total cost.

To the extent that the variance models are good approximations of reality, the method of negotiating the total cost of a survey ensures that the analyses that can be supported by the information content of the resulting survey data base are consistent with pre-survey expectations. Some questions have been asked about the practicality of cost and error modelling at the survey design stage. Fellegi and Sunter (1974) offer a set of criticisms of attempts to address multiple sources of error and cost to guide design decisions:

a. There are practical constrains to feasible alternatives open to the researcher.

b. Major alternative survey designs do not present themselves within a fixed budget.

c. Components of cost function may not be continuous, over the whole range of possible designs. The discontinuities in the cost models imply that partial derivatives do not exist.

d. In a complex design the error reduction functions will be complex.

e. Terms in error function may interact in some unknown way.

f. Important interaction may exist between different surveys.

g. Major surveys are seldom designed to collect only one item of information. A single optimum design for a multipurpose survey may not be identified.

h. A survey is seldom designed to measure variables at a single level of aggregation; subclass statistics are also important.

i. The time constraint of the survey may inject another set of considerations very much related to the balance between different sources of error.

j. The method spends part of budget to obtain data on costs errors of components of the design, yet it offers no guidance on how much money to spent on those evaluation activities.

Most of the time the cost-return (return in terms of precision) relationship for each error type will be such that the marginal contribution of expenses made on improvement of the processes incurring errors of this type decrease as the total expense made on these processes increase.

Consider the cost model that separates costs of a cluster from costs of each sample element in the cluster (Groves, 1989):

$$\begin{gathered}\text{Total cost} = \text{ Fixed cost } + \text{ Cluster costs } + \text{ Element costs} \\ \mathrm{C} = \mathrm{C}_0 + C_a a + C_b b \ ; \end{gathered} \qquad (5.1.1)$$

where C_0 denotes the fixed costs of doing the survey, independent of the number of sample clusters or sample elements per cluster; C_a denotes the cost of selecting, and locating of each cluster, independent of the number of sample elements for each cluster; a denotes the number of sample clusters; C_b denotes the cost of selecting, contacting, and interviewing a single sample element from a cluster; b is the number of sample elements per cluster.

For the sampling error of the estimated population mean and the cost model specified above, the optimal number of elements per cluster would be

$$b_{opt} = \sqrt{\frac{C_a(1-roh)}{C_b roh}} \ . \qquad \textbf{(5.1.2)}$$

That is, large numbers of sample elements should be taken from clusters that exhibit internal homogeneity on the survey variable, small cluster sizes should be taken with low homogeneities.

CHAPTER 6

MEASUREMENT ERROR MODELS

Data that are collected from individuals by personal interview are known to be subject to response error. Response errors, sometimes called measurement errors, have long been recognised as one of the major problems in surveys. The effect of response errors can be quite severe in statistical data analysis.

In defining the concept of error it is necessary to postulate a true value. It is generally assumed that a true value of characteristics under study exists for each individual. Hansen, et al. (1953) suggest three criteria for the definition of the true value for an individual:

a. It must be uniquely defined;

b. It should be defined in such a manner that the purposes of the survey are met; and

c. It should be defined la terms of operations which can be carried through, even though it might be difficult or expensive to perform the operations.

For a situation in which survey response for a given individual can be considered as coming from a population of conceptual responses for that individual, it may be appropriate to define the individual true value as the expected response obtained under certain well-defined survey conditions.

Individual true value is a useful ideal at which to aim and the consideration of departures from its value is helpful in assessing the methods by which we obtain information.

The basic approach to the analysis of the individual response errors depends on an understanding of the measurement process and the way in which the conditions under which the survey is carried out may affect the results of the survey. It is useful to distinguish between two components of response error. One can define an *expected survey value* as the expected value under the *essential survey conditions*. The difference between this value and the true value is the *response bias*. In addition to this there are random fluctuations about the expected value. These variable errors contribute to the response error, in the form of response variance. The *response variance* is a measure of variability between different responses on different trials.

Cochran (1968) gave a short description of the experiments conducted by Pearson (1902) in a review paper on measurement errors. From these experiments, Pearson (1902) observes that

a. The mean errors differed significantly from zero;

b. For a given measurer, the size of the bias varied throughout the series of trials .

c. The errors were not, in general, normally distributed; and

d. The errors of two apparently independent observers in measuring the same quantity were positively correlated.

The measurement of response errors requires that they be represented by a mathematical model. A number of alternative models have been proposed, often to accommodate special situations. The variations in the response error models which have been developed depend upon the survey itself. Survey factors which must be considered by the model formulation include the existence of, or ability to obtain, "correct" values for units in the survey, the complexity of estimation given the sample design, the ability to make re-measurements under reasonably fixed conditions, one of the most difficult conditions to achieve, the ability to randomize work assignments, budget constraints for these costly measurement studies.

Cochran (1968) reviews the various types of mathematical models to represent errors of measurement. In his discussion, the following models are mentioned:

Let y_{jt} denote the recorded measurement on the t-th trial for the j-th unit in the sample (j=1,2,….,n), and the symbol m_j denotes the correct or true measurement. The error of measurement on the j-th unit on the t-th trial is $d_{jt} = y_{jt} - m_j$. d_{jt} is called the individual response error. The subscript t will refer to the t-th trial or repeated measurement

$$y_{jt} = m_j + d_{jt}, \qquad (6.1.1)$$

where, both y_{jt} and d_{jt} have a frequency distribution for each member of the population. m_j is assumed fixed for any specific member of the population.

The simplest model is one in which

$$E(d_{jt} \mid j) = 0;\; E(d_{jt}^2 \mid j) = s^2;\; E(d_{jt}, d_{jt'}) = 0 \;\left(\mathrm{t} \neq \mathrm{t}'\right);$$
$$E(d_{jt}, d_{j't} \mid j, j') = 0 \;\left(\mathrm{j} \neq \mathrm{j}'\right) . \tag{6.1.2}$$

In the above model the errors assume zero mean and constant variance, they are uncorrelated with the true values, with one another on different units, and on different trials for the same unit. It is possible to make a modification by assuming $E(d_{jt}^2 \mid j) = s_{d_j}^2$, meaning that measurements on different units involve differing precision.

Incorporating an overall bias of amount a in the measurement process into the model we get

$$y_{jt} = m_j + a + d_{jt} \; . \tag{6.1.3}$$

The next stage is to introduce a variable bias term a_j, and to make the additional assumption that a_j ‘s are uncorrelated with the true values m_j;

$$\begin{aligned} &y_{jt} = m_j + a_j + d_{jt}; \\ &E(a_j, m_j) = 0; \\ &E(d_{jt} \mid j) = 0;\; E(d_{jt}^2 \mid j) = s_{d_j}^2; \\ &E(d_{jt}, d_{jt'}) = 0 \;\left(\mathrm{t} \neq \mathrm{t}'\right); \\ &E(d_{jt}, d_{j't} \mid j, j') = 0 \;\left(\mathrm{j} \neq \mathrm{j}'\right) . \end{aligned} \tag{6.1.4}$$

It is also possible to assume that a_j and m_j are correlated;

$$\begin{aligned} &y_{jt} = m_j + a_j + d_{jt}; \\ &E(a_j, m_j) \neq 0; \\ &E(d_{jt} \mid j) = 0;\; E(d_{jt}^2 \mid j) = s_{d_j}^2; \\ &E(d_{jt}, d_{jt'}) = 0 \;\left(\mathrm{t} \neq \mathrm{t}'\right); \\ &E(d_{jt}, d_{j't} \mid j, j') = 0 \;\left(\mathrm{j} \neq \mathrm{j}'\right) . \end{aligned} \tag{6.1.5}$$

It is often convenient to combine the terms m_j and a_j by writing $m'_j = m_j + a_j$, since most of the times no feasible method of measuring the true value is available.

Another modification of the model (6.1.4) involves the situations where the relation between the variable bias a_j and the true value m_j can be expressed as a linear regression of a_j on m_j with regression coefficient g ;

$$y_{jt} = a + bm_j + a_j + d_{jt},\ E(a_j) = 0,\ Cov(a_j, m_j) = 0\ , \qquad (6.1.6)$$

where $b = 1 + g$.

With binomial data y takes only the values 0 and 1. Hansen, Hurwitz and Bershad (1961) presented the consequences of model (6.1.4) in case of Binomial data and also for interval data. It is customary to study the effect of measurement errors in the context of estimating a population total (or a mean). They presented their response model in the context of estimation of the proportion $(p = \frac{1}{N}\sum_{j=1}^{N} m_j)$ of individuals that belong to a given class of a finite population.

An observation on the j-th unit in the t-th trial is designated by y_{jt} .

$y_{jt} = 1$, if the j-th unit is assigned to the particular class under consideration on the t-th trial

$= 0$, otherwise.

Let, m_j denote the correct or true measurement for the j-th unit, and a_j be the variable bias term. The measurement error model that will be considered is of the following form:

$$y_{jt} = m_j + a_j + d_{jt};\ E(a_j, m_j) = 0;$$
$$E(d_{jt} \mid j) = 0;\ E(d_{jt}^2 \mid j) = s_{d_j}^2;\ E(d_{jt}, d_{jt'}) = 0 \ (\mathrm{t} \neq \mathrm{t}');$$
$$E(d_{jt}, d_{j't} \mid j, j') = 0 \ (\mathrm{j} \neq \mathrm{j}') .$$

For a unit for which $m_j = 1$ if $y_{jt} = 1$, there is no measurement error for this unit on this trial. Let $y_{jt} = 0$ on a certain proportion q_j of trials. q_j is the probability of misclassification for this unit. Thus, in the model (6.1.4) with $m_j = 1$, we have $a_j = -q_j$, and d_{jt} is a binomial variates with variance $s_j^2 = q_j(1-q_j)$. Similarly, for a unit for which $m_j = 0$ with probability of misclassification for this unit is f_j, we have $a_j = f_j$, and d_{jt} is a binomial variable with variance $s_j^2 = f_j(1-f_j)$.

$$\begin{aligned} P &= p\{1 - E(q_j \mid m_j = 1)\} + (1-p)E(f_j \mid m_j = 0) \\ &= p(1-q) + qf, \end{aligned} \tag{6.1.7}$$

where $q = E(q_j \mid m_j = 1)$ is called the probability of false negative, and $f = E(f_j \mid m_j = 0)$ is called the probability of false positive. Also, y_{jt} is binomially distributed.

The sample proportion $p_t = \frac{1}{n}\sum_{j=1}^{n} y_{jt}$ is distributed like the mean of a binomial sample of size n with parameter P. Here, p_t is a biased estimate of p, and the bias amounts to $-pq + (1-p)f$. Variance of p_t is $\frac{P(1-P)}{n}$, and variance of p is $\frac{p(1-p)}{n}$. Errors of measurement cause an increase in variance only if P is nearer $\frac{1}{2}$ than p.

The expected value of p_t is the average value taken over all possible trials including all possible samples and all possible responses under the general conditions:

$$P = E(p_t) = E(\frac{1}{n}\sum_{j=1}^{n} y_{jt}) \ . \qquad (6.1.8)$$

Bias of p_t is,

$$B_{p_t} = E(p_t - p) = P - p \ . \qquad (6.1.9)$$

Variance of p_t is,

$$s^2_{p_t} = E(p_t - P)^2 \ . \qquad (6.1.10)$$

MSE of p_t is,

$$MSE_{p_t} = E(p_t - p)^2 = s^2_{p_t} + B^2_{p_t} \ . \qquad (6.1.11)$$

The expected value of the observation on the j-th unit is,

$$m'_j = m_j + a_j = E_j(y_{jt}) \ . \qquad (6.1.12)$$

The response deviation (the difference between the observed value of the j-th unit on the t-th trial and the expectation of observation on the t-th unit) is

$$d_{jt} = y_{jt} - m'_j \ . \qquad (6.1.13)$$

For each element the response can be expressed as

$$y_{jt} = m_j + \underbrace{(m'_j - m_j)}_{a_j} + \underbrace{(y_{jt} - m'_j)}_{d_{jt}}, \qquad (6.1.14)$$

where m_j is the individual true value, $a_j = (m'_j - m_j)$ is the individual response bias, and $(y_{jt} - m'_j)$ is the individual response deviation. The true value does not effect the response variance, but only the response bias.

Similarly, the estimator obtained from the survey can also be divided into components as

$$p_t = p + (p_t - P) + (P - p), \qquad (6.1.15)$$

where p is the true population proportion, $(P - p)$ is the response bias, and $(p_t - P)$ is the response deviation. The response deviation consists of fluctuations about the expected value and produces the total variance.

The total variance of the survey is,

$$\begin{aligned} s^2_{p_t} &= E(p_t - P)^2 \\ &= E(p_t - \overline{m'} + \overline{m'} - P)^2 \\ &= \underbrace{E(p_t - \overline{m'})^2}_{\text{response variance: } s^2_{d_t}} + 2\underbrace{E(p_t - \overline{m'})(\overline{m'} - P)}_{Cov(d_t, p)} + \underbrace{E(\overline{m'} - P)^2}_{\text{sampling variance of } p_t : s^2_p}, \end{aligned} \qquad (6.1.16)$$

where,

$$\overline{m'} = \frac{1}{n}\sum_{j=1}^{n} m'_j . \qquad (6.1.17)$$

$E(p_t - \overline{m'})^2$ is defined as the response variance contribution to the total variance of p_t .

$$
\begin{aligned}
s^2_{\bar{d}_t} &= E(p_t - \overline{m'})^2 \\
&= E(\frac{1}{n}\sum_{j=1}^{n} y_{jt} - \frac{1}{n}\sum_{j=1}^{n} m'_j)^2 \\
&= E[\frac{1}{n}\sum_{j=1}^{n} (y_{jt} - m'_j)]^2 \qquad (6.1.18) \\
&= E(\frac{1}{n}\sum_{j=1}^{n} d_{jt})^2 \\
&= E(\bar{d}_t)^2 = Var(\bar{d}_t) = s^2_{\bar{d}_t} \ .
\end{aligned}
$$

$2E(p_t - \overline{m'})(\overline{m'} - P)$ is twice the covariance of $\bar{d}_t$ and $\overline{m'}$, the covariance between response and sampling deviations. Koch (1973) calls this component as the interaction variance. When $P = \overline{m'}$ (when $n = N$ or repetitions are defined on a fixed sample) this covariance term becomes zero.

$$E(p_t - \overline{m'})(\overline{m'} - P) = Cov(\bar{d}_t, \overline{m'}) \ . \qquad (6.1.19)$$

$E(\overline{m'}) = P$ where the expectation is taken over all possible samples (and trials). $E(p_t - \overline{m'}) = E(\bar{d}_t)$ and the expectation of $\bar{d}_t$ over all possible samples and all possible trials equal to 0. $E(\bar{d}_t - 0)(\overline{m'} - P)$ taken over all possible samples and all possible trials equals to $Cov(\bar{d}_t, \overline{m'})$. This term is excluded in Hansen, Hurwitz and Bershad (1961) discussion.

$s^2_p = E(\overline{m'} - P)^2$ is the sampling variance of p_t. This variance is only due to sampling. For simple random sampling with replacement s^2_p is

$$s^2_p = \frac{s^2_{m'}}{n} , \qquad (6.1.20)$$

where $s^2_{m'}$ is the population variance of the m'_j.

In case of a complete census or a simple random sample of units of analysis, the response variance, $s^2_{\bar{d}_t}$, can be restated in the following form:

$$\begin{aligned} s^2_{\bar{d}_t} &= E(\bar{d}_t^{\,2}) \\ &= \frac{1}{n} s^2_d + \frac{n-1}{n} r s^2_d \\ &= \frac{1}{n} s^2_d [1 + r(n-1)], \end{aligned} \tag{6.1.21}$$

where

$$s^2_d = E(d^2_{jt}) = \frac{1}{N}\sum_{j=1}^{N} E_t(d^2_{jt}), \tag{6.1.22}$$

is the simple response variance, and

$$r = \frac{E(d_{jt}d_{kt})}{s^2_d} \text{ (for j} \neq \text{k)}, \tag{6.1.23}$$

is the intraclass correlation among the response deviations in a trial.

$$\begin{aligned} s^2_{\bar{d}_t} &= E(\bar{d}_t^{\,2}) \\ &= E\left(\frac{1}{n}\sum_{j=1}^{n} d_{jt}\right)^2 \\ &= \frac{1}{n^2} E\left(\sum_{j=1}^{n} d_{jt}\right)^2 \\ &= \frac{1}{n^2}\sum_{j=1}^{n} E(d_{jt}^{\,2}) + \frac{1}{n^2}\sum_{j}^{n}\sum_{k}^{n} E(d_{jt}d_{kt}), \text{ for j} \neq \text{k} \\ &= \frac{n}{n^2} E(d_{jt}^{\,2}) + \frac{n(n-1)}{n^2} E(d_{jt}d_{kt}), \text{ for j} \neq \text{k}; \end{aligned} \tag{\textbf{6.1.24}}$$

since $E(d_{jt}d_{kt}) = r s^2_d$,

$$\begin{aligned} s^2_{\bar{d}_t} &= \frac{1}{n} s^2_d + \frac{(n-1)}{n} r s^2_d \\ &= \frac{1}{n} s^2_d [1 + r(n-1)]. \end{aligned} \tag{6.1.25}$$

From the examination of (6.1.21), one can see that the possible impact of even a very small intraclass correlation can be substantial when the sample size is large. In case of continuous variates, the consequences of the model (6.1.4) would be similar to the consequences presented for Binomial data.

CHAPTER 7

DECOMPOSITION OF THE SURVEY ERROR

As it was stated before the total variability of estimates obtained from a survey is the sum of the sampling variability and the non-sampling variability. O'Muircheartaigh (1982) partitions the total variance of estimators into four components, each of which has a different implication for the survey design.

7.1. Simple Sampling Variance

Consider the following model:

$$y_{jt} = m_j + d_{jt}, \qquad (7.1.1)$$

where, as before, y_{jt} denotes the observation on the j-th unit on t-th trial, m_j denotes the true value for element j, and d_{jt} denotes the variable response error (or response deviation) obtained for element j at trial t. The response biases are excluded from this model. The specification of the model involves the specification of the distribution of d_{jt}. Suppose that we want to estimate the population mean

$$\overline{m} = \frac{1}{N}\sum_{j=1}^{N} m_j \ . \qquad (7.1.2)$$

The sample mean of the observations is

$$\overline{y}_t = \frac{1}{n}\sum_{j=1}^{n} y_{jt} \ . \qquad (7.1.3)$$

Sampling error is an error of nonobservation. Survey estimates are subject to sampling error because not all members of the population are measured. The particular units which happen to be selected into a particular sample depends on chance, the possible outcomes being determined by the procedures specified in the sample design. This means that, even if the required information on every selected unit is obtained entirely without error, the results from the sample are subject to a degree of uncertainty due to these chance factors affecting the selection of units. Sampling variance is a measure of this uncertainty.

Sampling variance of a survey statistic, y , can be described as average squared deviations of individual sample values of the statistics and its own average value:

$$E[y_r - E(y_r)]^2 , \tag{7.1.4}$$

where y_r is a sample statistic on the r-th distinct sample of the sampling design and $E(y_r)$ is the expected value of y_r over all samples of the given design. The sampling variance is thus a feature of a distribution over all possible samples that could be drawn with a particular design. Each observation in that sampling distribution is the result of one sample of the given design.

There are three types of distribution that should be kept conceptually distinct when considering sampling error. The first is the distribution of characteristic to be measured in the survey in the population. Population distributions of elements form this first kind of distribution; they have N points, for each N elements in the population. The second type of distribution is the sample distribution. It mimics the corresponding population distribution, but it is based on smaller number of elements. The third type of distribution is the sampling distribution of a sample statistic. Sampling error concerns the variability of values of statistics over different samples that could be drawn.

Sampling variance is the variability of a statistic over all possible samples using the same design, but the majority of surveys are conducted once, using only one sample. When probability samples are drawn with two or more independent selections, the sampling variance of many statistics can be estimated from only implementation of the design. A probability sample is one for which all members of the population have a known, nonzero chance of selection.

To illustrate the estimation of sampling variance of a survey statistic from a single probability sample, consider the case of a simple random sample with replacement of size n. The statistics of interest is the mean for a variable, y. If the true values, m_j, were observed in this survey, the only variability in the estimator would arise from the fact that only a sample from the population is observed. The variance of the sample mean $\overline{y}_t$ (assuming m_j is observed for all j in the sample) will be

$$Var(\overline{y}_t) = \frac{s_m^2}{n}, \tag{7.1.5}$$

where $s_m^2 = \frac{1}{N}\sum_{j=1}^{N}(m_j - \overline{m})^2$ is the population element variance of the true values. In case of simple random sample without replacement the variance formula for the sample mean is complicated by the finite population correction $(1 - f')$, where $f' = \frac{n-1}{N-1}$.

The population element variance, s_m^2, is a property of the population that was sampled. It is not a property of the sample design. The sample designs in practice are rarely a simple random sample, it is however possible to obtain a good estimate of s_m^2 in such situations. In practice an acceptable approximation can be obtained by treating the sample observations as though they had arisen from a simple random sample.

7.2. Simple Response Variance

Another source of variation of the estimator is caused by the response deviations, d_{jt}. If we assume that the response deviations are not correlated with the true values or with each other, the model (7.1.1) becomes

$$\begin{aligned} & y_{jt} = m_j + d_{jt}; \\ & E(d_{jt} \mid j) = 0; \\ & E(d_{jt}^2 \mid j) = s_{d_j}^2 = s_d^2; \\ & E(d_{jt}, d_{j't'}) = 0 \ \left(\text{for } t \neq t' \text{and } j \neq j'\right) . \end{aligned} \tag{7.2.1}$$

The component of the variance contributed by these uncorrelated response deviations is

$$Var(\bar{d}_t) = \frac{s_d^2}{n} . \tag{7.2.2}$$

The variance component in (7.2.2) is called the simple response variance and it is a function of the sizes of response deviations and the sample size. The sum of simple sampling variance, and simple response variance can be called the simple total variance (s_{stv}^2). This is the variance of the mean of the sample of size n from the population when the response deviations are uncorrelated.

$$s_{stv}^2 = \frac{1}{n}(s_m^2 + s_d^2) \ . \qquad (7.2.3)$$

The simple total variance is estimated by taking the observed variance of the observations. Ignoring the finite population correction factor

$$E(\frac{s^2}{n}) = s_{stv}^2 = \frac{1}{n}(s_m^2 + s_d^2) \ , \qquad (7.2.4)$$

where

$$s^2 = \frac{\sum_{j=1}^{n}(y_{jt} - \overline{y}_t)^2}{n-1} \ . \qquad (7.2.5)$$

Hansen, Hurwitz and Pritzker (1964) defined the *index of inconsistency* as the ratio of the simple response variance to the total variance of individual response; that is

$$I = \frac{s_d^2}{s_m^2 + s_d^2} \ . \qquad (7.2.6)$$

Index of inconsistency (I) is a relative measure of random response variability and is defined as the ratio of simple response variance to simple total variance per element.

The estimation procedure for the simple response variance and index of inconsistency where we have two observations on the same units obtained in two independent trials is simple. The simple response variance can be estimated by the gross difference rate (GDR), where GDR is

$$GDR = \sum_{j=1}^{n} \frac{(y_{j1} - y_{j2})^2}{n} \ . \qquad (7.2.7)$$

Thus, the gross difference rate is the average squared difference between the original interview and the reinterview responses.

Observing that the difference of response deviations in the two trials is the difference between the two y_{jt} values observed. That is,

$$(d_{j1} - d_{j2}) = (y_{j1} - y_{j2}). \qquad (7.2.8)$$

Its variance is expressed simply as

$$s^2_{d12} = \sum_j^n \frac{(y_{j1} - y_{j2})^2}{n} = \sum_j^n \frac{(d_{j1} - d_{j2})^2}{n} = GDR. \qquad (7.2.9)$$

Since by definition, with r_{12} as the correlation between the two trials

$$s^2_{d12} = s^2_{d1} + s^2_{d2} - 2r_{12} \cdot s_{d1} \cdot s_{d2}, \qquad (7.2.10)$$

and furthermore, since the two variances in trials under the same conditions are also the same, we have

$$s^2_{d12} = 2s^2_d \cdot (1 - r_{12}); \text{ with } s_{d1} = s_{d2} = s_d. \qquad (7.2.11)$$

Finally, simple response variance can be estimated from the above on the assumption that the correlation r_{12} is zero, that is the two trials are independent

$$s^2_d = \frac{1}{2} s^2_{d12}; \text{ with } ?_{12} = 0. \qquad (7.2.12)$$

That is, it is one-half the mean squared deviation between values on the same units obtained in the two independent trials.

The assumption of independence between repetitions of the survey is usually not valid, because the second measurement is often influenced by the first: the respondent and/or the interviewer may remember and try to be consistent with the response given earlier. This tends to make r_{12} positive, and hence the independence assumption to result in underestimation of response variance.

With binomial data, the estimators are often presented in a very simple table showing the original and reinterview estimates (or counts if the design is simple random sampling), which the information on the cross tabulation is given in Table 1:

Table 1.: *Cross Tabulation of responses for Original Interview and Reinterview*

Original Interview				
Reinterview		Number of cases with characteristics	Number of cases without Characteristics	
	Number of cases with characteristics	a	b	a + b
	Number of cases without Characteristics	c	d	c + d
Total		a + c	b + d	n = a + b + c + d

For tables formatted in this fashion, the GDR takes a very simple form:

$$GDR = \frac{b+c}{n} \quad . \qquad (7.2.13)$$

The GDR is the proportion of cases that were reported differently in the original and reinterview surveys. It is equal to the proportion of cases reported as having a characteristic in the original interview but not having it in the reinterview, plus the proportion of cases reported as not having the characteristic in the original interview but having it in the reinterview. Similarly, from the table, the index of inconsistency also takes on a very simple form:

$$I = \frac{b+c}{2np(1-p)}, \text{ where } p \text{ is } \frac{a+c}{n} . \qquad (7.2.14)$$

7.3. Correlated Sampling Variance

Although sampling error is partly a function of variability in the population studied, the sampling error in a statistic is under control of sample designer. The sample design features which are most important in this regard are:

a. Stratification: Stratification is the sorting of the population into separate subgroups (strata) prior to selection. Each element of the population belongs to one and only one stratum. After groups are identified, separate samples are selected from each group. Stratification tends to reduce sampling error.

Consider a population of N units. If H different strata are constructed, let N_h be the number of elements in the h-th stratum, where $h = 1,2,...,H$. In this case the population mean would be

$$\overline{Y} = \frac{1}{N}\sum_{h=1}^{H}\frac{N_h}{N}\overline{Y_h} = \sum_{h=1}^{H}W_h\overline{Y_h} , \qquad (7.3.1)$$

where $W_h = \frac{N_h}{N}$ is the proportion of the population in the h-th stratum, and $\overline{Y_h}$ is the mean of the h-th stratum. If simple random samples were drawn from each of the H strata separately, then one estimator of the sample mean is

$$\overline{y} = \sum_{h=1}^{H}W_h\overline{y}_h . \qquad (7.3.2)$$

The sampling variance of the sample mean is

$$\sum_{h=1}^{H} \frac{1}{n_h} W_h^2 s_h^2 , \qquad (7.3.3)$$

where W_h is the proportion of the population in the h-th stratum; s_h^2 is the population element variance in the h-th stratum; H is the total number of strata; and n_h is the sample size in h-th stratum.

b. Assignment of probabilities of selection to different kinds of elements in the population: A simple random sample assigns equal probabilities of selection to each element in the population. Sometimes there are practical reasons to depart from this design. The costs of measuring some members of the population may be very high, lower probabilities of selection is often assigned to these members of the population. There may be a desire to study a subgroup of the population intensively, with smaller sampling errors. The members of such a subgroup may be assigned higher probabilities of selection than other members of the population. In some infrequent cases, there may be prior information about the within-strata variability on survey measures. In such cases oversampling the strata with higher element variances can reduce the sampling error in the estimator relative to a design using equal probabilities of selection the same total sample size.

c. Clustering: Sometimes selection of groups (clusters) of elements together instead of independent selection of separate elements is preferred. Cluster sampling involves selecting a sample in a number of stages. The units in the population are grouped into convenient, usually naturally occurring clusters. These clusters are non-overlapping, well-defined groups which usually represent geographic areas.

In practice, cluster samples tend to produce higher sampling errors for statistics than element samples of the same size. The loss of precision arises because most natural groupings of persons contain persons who are similar to one and another on the variables that are measured. Despite the loss of precision for survey statistics from cluster samples, the reason for the frequent use of cluster samples is the desirable cost features they are likely to have. Generally, Cluster samples cost less than element samples.

The sampling variance of the estimated population mean for a cluster sample is inflated by two factors: the correlation of values among persons in the same clusters, and number of sample elements chosen from a cluster.

$$Var(\bar{y}) = \frac{s^2[1 + roh(b-1)]}{n}, \qquad (7.3.4)$$

where s^2 is the population element variance; roh is the intracluster correlation coefficient; b is the number of sample persons chosen from each cluster; and n is the total number of persons in the entire sample.

The intracluster correlation measures covariation of pairs of persons in the same cluster, calculated by deviations from the overall mean. If elements in the same cluster have similar deviations from the population mean, then roh will be positive, and the sampling variance of the estimated mean from a cluster sample will be inflated over that from an element sample of the same size.

d. Sample size: The fourth feature in the control of the survey designer is the sample size itself. Sample size has an impact on sampling variance as a function of the number of independent selection at each stage of the sample and the relative within and between unit variability at each stage.

In practice simple random samples are very rarely used. Most sample designs are stratified multistage designs and the sampling variance of such designs is normally greater than the sampling variance of a simple random sample of the same size. The term design effect is used to describe the variance of sample estimates for a particular sample design relative to the corresponding variance of a simple random sample with the same sample size.

The concept of *design effect* was popularised by Kish (1965) to deal with complex sample designs involving stratification and clustering. Stratification generally leads to a gain in efficiency over simple random sampling, but clustering usually leads to deterioration in the efficiency of the estimate due to positive intracluster correlation among the subunits in the clusters. In order to determine the total effect of any complex design on the sampling variance in comparison to the alternative simple random sampling, one calculates a ratio of variances associated with an estimate, namely

$$deff = \frac{\text{sampling variance of a complex sample}}{\text{sampling variance of a simple random sample}} . \qquad (7.3.5)$$

This ratio is called the design effect $deff$ of the sampling design for the estimate which is based on the same sample size. This ratio measures the overall efficiency of the sampling design and the estimation procedure utilised to develop the estimate.

In cluster samples, the ratio is typically larger than one, expressing the losses due to clustering. If subsamples of size b are selected randomly from equal clusters, the design effect is

$$deff = [1 + roh(b-1)], \qquad (7.3.6)$$

where roh, the coefficient of intraclass correlation, is a measure of the homogeneity within clusters.

In case of unequal clusters, the design effect can be approximated by

$$deff = [1 + roh(\bar{b} - 1)], \qquad (7.3.7)$$

where $\bar{b} = \frac{\sum_{a=1}^{a} b_i}{a}$ is the average number of individuals interviewed in each cluster. The intraclass correlation coefficient gives an indication of relative similarity of individuals within a cluster compared to the similarity of individuals in the population as a whole. The more similar individuals are to one and another within a cluster, the larger the value of roh will be.

roh takes on values within the interval $[-\frac{1}{b-1}, 1]$. The highest possible value of roh means all elements comprising any cluster have the same value. The lowest possible value of roh indicates that there is zero variance between cluster means. If the variable is distributed completely at random among clusters roh takes on the value zero, and the design effect becomes unity. Generally, roh tends to be greater than zero, and even a relatively small positive roh can have a large effect on the variance if the average cluster size is large.

The sampling variance formula in (7.1.5) underestimates the total sampling variance in the presence of intraclass correlation within clusters. A more realistic sampling variance formula can be obtained by incorporating the design effect into the analysis. Again, if we assume that only the true values, m_j, were observed in a survey:

$$Var(\overline{y}_t) = \frac{s_m^2}{n}[1 + roh(\overline{b} - 1)], \qquad (7.3.8)$$

where $s_m^2 = \frac{1}{N}\sum_{j=1}^{N}(m_j - \overline{m})^2$ is the population variance of the true values as before. The increase in the variance over simple sampling variance given by (7.1.5) is

$$\frac{s_m^2}{n}[roh(\overline{b} - 1)], \qquad (7.3.9)$$

and may be called the correlated sampling variance.

7.4. Correlated Response Variance

A mathematical model which assumes independent responses of all individuals will not represent a survey which uses interviewers unless the interviewer is assumed to have no influence on the response. If we assign at random a different interviewer to each individual, the effect of the interviewer on the responses would be uncorrelated for any two obtained responses. However, a given interviewer usually obtains responses for a number of individuals, and often the errors made by a particular interviewer are correlated.

The important contributions to response variance are likely to arise from the factors involving correlated response deviations. The analysis of response deviations is complicated when we want to take into account of the possible correlations among the response deviations. The simple model in

(7.2.1) can be modified to take the possibility of such correlations, namely the correlated response deviations.

The simple response variance is based on independence assumption, while correlated response variance is based on dependence in these models. The correlated response variance reflects the part of total response variance due to a common influence on a group of respondents. It reflects the correlations among response deviations of different units in a given sample and a given trial. It has been long recognised that the results obtained by the same interviewer on different sampling units may be positively correlated, thus the correlated response variance is often interpreted as the interviewer effect. In order to analyse the effect of interviewer variance consider the following model

$$y_{ijt} = m_j + d_{ijt}, \qquad (7.4.1)$$

where m_j is the true value for the j-th individual, and d_{ijt} represents the response deviation for the i-th interviewer on the j-th unit. We make the following assumptions on the model:

$$\begin{aligned} &E(d_{ijt}) = 0 \text{ and } Var(d_{ijt}) = s_d^2, \text{ for all j;} \\ &Cov(d_{ijt}, d_{i'j't}) = r_1 s_d^2, \text{ for } i = i'; \qquad (7.4.2) \\ &Cov(d_{ijt}, d_{i'j't}) = r_2 s_d^2, \text{ for } i \neq i'. \end{aligned}$$

In (7.4.2), r_1 denotes the intra-interviewer correlation coefficient. It represents the ratio of the correlation between the response deviations for individuals interviewed by the same interviewer to the simple response variance

$$r_1 = \frac{Cov(d_{ijt}, d_{i'j't})}{s_d^2}, \text{ for } i = i'. \qquad (7.4.3)$$

r_2, in (7.4.2), denotes the between interviewer correlation coefficient. It represents the ratio of the correlation between the response deviations for individuals interviewed by the different interviewers to the simple response variance

$$r_2 = \frac{Cov(d_{ijt}, d_{i'j't})}{s_d^2}, \text{ for } i \neq i'. \qquad (7.4.4)$$

In case of a simple random sample of size n, where there are k interviewers each obtaining m randomly assigned interviews $(n = km)$, the contribution of the response deviations to the total variance of the sample mean will be

$$\begin{aligned}
Var(\overline{d_t}) &= Var(\frac{1}{km}\sum_{i=1}^{k}\sum_{j=1}^{m} d_{ijt}) \\
&= \frac{1}{n^2}\{Var(\sum_{i=1}^{k}\sum_{j=1}^{m} d_{ijt})\} \\
&= \frac{1}{n^2}\{nVar(d_{ijt}) + km(m-1)\underbrace{Cov(d_{ijt}, d_{i'j't})}_{\text{for } i=i'} + m^2k(k-1)\underbrace{Cov(d_{ijt}, d_{i'j't})}_{\text{for } i\neq i'}\} \qquad (7.4.5) \\
&= \frac{1}{n^2}\{ns_d^2 + km(m-1)s_d^2 r_1 + m^2k(k-1)s_d^2 r_2\} \\
&= \frac{s_d^2}{n}\{1 + r_1(m-1) + r_2 m(k-1)\}.
\end{aligned}$$

Typically, r_2 will be negligibly small; ignoring r_2, (7.4.5) becomes

$$Var(\overline{d_t}) = \frac{s_d^2}{n}[1 + r_1(m-1)]. \qquad (7.4.6)$$

When the interviewer workloads is not constant, substituting the average interviewer workload, $\overline{m} = \frac{n}{k}$, for m in the above formula provides a good approximation. The increase in the variance over simple response variance is given by

$$\frac{s_d^2}{n}[1+r_1(m-1)]-\frac{s_d^2}{n}=\frac{s_d^2}{n}[r_1(m-1)], \qquad (7.4.7)$$

and is called the correlated response variance. One can observe from (7.4.7) that, as the size of the average interviewer workload increases, the effect of correlated response variance becomes greater.

Consider a two stage sampling plan, where the first stage involves simple random selection of a clusters from a total of A clusters in the population. On the second stage, from each selected cluster of size B, b subunits are selected by simple random sampling again. The observation on the chosen $n = ab$ subunits are made by k interviewers each having a workload of m subunits ($n = km$). The total variance can be written as

$$Var(\overline{y})=\frac{s_m^2}{n}+\frac{s_d^2}{n}+\frac{s_m^2}{n}[1+roh(b-1)]+\frac{s_d^2}{n}[1+r_1(m-1)]. \qquad (7.4.8)$$

The implications of each of these components are different in terms of survey design and execution. The total variance given by (7.4.8) cannot be eliminated completely but can be controlled to a certain amount. Each error type can be reduced by improving the measurement processes that cause these errors. Improvement of processes involves costs.

The simple sampling variance can be effected only by changing the sample size. The sample size can be increased by incurring additional costs, increased sample size increases the precision. The correlated sampling variance can be modified by the choice of sample design. The intracluster correlation coefficient is determined by the choice of clusters for the design: the more homogeneous the clusters the larger the clustering effect. The average subsample size within the selected clusters is the other determining factor, and for a given sample size depends on the number of clusters included in the sample.

A similar argument goes with the simple response variance and the correlated response variance. The contribution of simple response variance can be effected only by changing the sample size. The correlated response variance, however, can be modified by the choice of sample design. Assuming that the quality of the interviewers is not effected by increasing their number, reduction of the average interviewer workload will decrease the effect of the correlated response variance on the total variance. Employing additional interviewers in a survey will have a positive effect on the total precision; reducing the effect of correlated interviewer variance on the total measurement variance.

7.5. Estimates of Response Variance

The method of measuring response variance involves formulating a response error model, postulating that the survey is repeatable under some fixed set of identical conditions, and measuring the components of variability among the repetitions. There are two alternative methods to obtain approximate estimates of the response variance or of the specified components of the response variance, although none of these methods provide unbiased or consistent estimates of them (Groves, 1989).

a. The Replication method: One way of estimating the response variance under a set of conditions is to replicate the survey procedure on the same sample. Suppose that a simple random sample of size n is selected from the population; Assume that the survey is taken twice. Let $\overline{y}_1$ be the sample mean obtained in the first trial, and $\overline{y}_2$ be the sample mean obtained in the second trial. We might use

$$\frac{(\overline{y}_1 - \overline{y}_2)^2}{2}, \tag{7.4.9}$$

as an estimate of the simple response variance. The expected value of equation (7.4.9) is approximately,

$$\frac{s^2_{\overline{d}_1} + s^2_{\overline{d}_2} - 2r_{\overline{d}_1\overline{d}_2} s_{\overline{d}_1} s_{\overline{d}_2}}{2}, \tag{7.4.10}$$

if, in fact m_{1j} and m_{2j} are approximately the same for all j. Thus if $s^2_{\overline{d}_1} = s^2_{\overline{d}_2}$ and the correlation term, $r_{\overline{d}_1\overline{d}_2}$ is zero equation (7.4.9) is an unbiased estimate of the simple response variance, with one degree of freedom. The number of degrees of freedom can be increased by increasing the number of replications.

The principal disadvantages of the replication method lies in the necessity of making the assumption that the correlation is zero. If this correlation is positive then (7.4.9) will be an underestimate of the simple response variance by a factor of $(1 - r_{\overline{d}_1\overline{d}_2})$ on the average. Another limitation is that the second or subsequent trials have to be conducted at a later point in time, and therefore to the extent that the change in time changes the essential survey conditions of the subsequent trials, the differences between m_{1j} and m_{2j} will increase.

b. The method of interpenetrating samples: In order to estimate the correlated response variance due to interviewers the survey design must be modified. Ideally, to assess the impact of interviewers on correlated measurement errors, we would use a design in which subsamples of the sample segments are randomly assigned to interviewers so that no systematic difference between the workloads of the interviewers can contaminate the comparison of the results of the interviewers. This procedure of random allocation of workloads to interviewers is called interpenetration and is due to Mahalanobis (1946).

Suppose that a simple random sample of size $n = km$ is selected from the population; the sample is partitioned into k equal subsamples of size m - $(s_1, s_2, ..., s_k)$. Each subsample is allocated to a single interviewer. The label (i, j) is used to indicate that individual j belongs to the workload of interviewer i.

From the data we can calculate two linearly independent sums of squares: the between-interviewers sums of squares (C), and the within-interviewer sum of squares (F).

$$C = \frac{m}{k-1}\sum_{i=1}^{k}(\bar{y}_i - \bar{y})^2 , \qquad F = \frac{1}{k(m-1)}\sum_{i=1}^{k}\sum_{j\in s_i}(y_{ij} - \bar{y}_i)^2 . \qquad (7.4.11)$$

The expected values of the mean squares, C and F, will be

$$E(C) = s_m^2 + s_d^2\{1 + (m-1)r_1 - m(k-1)r_2\} , \qquad E(F) = s_m^2 + s_d^2(1 - r_1) . \qquad (7.4.12)$$

Since r_2 can generally be assumed to be small relative to r_1, we can use $\frac{1}{m}(C - F)$ as a possible estimator of $s_d^2 r_1$. As an negligibly unbiased estimator of the total simple variance, $\frac{1}{m}(C-F)+F$, can be used. $\frac{1}{m}(C-F)+F$ will underestimate the total simple variance, $s_m^2+s_d^2$, by an amount $s_d^2 r_2$. Interpenetrating subsamples provide a valid estimate of the total variance of an estimated total in the presence of measurement errors. But such designs are not often used due to cost and operational considerations.

On a response model which was similar to that of Hansen, Hurwitz and Bershad (1961), Fellegi (1964) used a sampling design involving both interpenetration and replication. A sample design which involves both re-enumeration and random allocation of respondents makes it possible to estimate both the simple and correlated response deviations.

CHAPTER 8

CLUSTER SAMPLING AND VARIANCE DECOMPOSITION

The similarity of the formulas for correlated response variance in (7.4.7) and the correlated sampling variance in (7.3.9) stems up from the fact that the sampling design caused by the interviewer assignments induces a type of clustering effect. Each interviewer's workload generates a cluster.

Consider the case where, a cluster sample is drawn and one interviewer is assigned at random to each sample cluster. Each interviewer completes all interviews in the assigned cluster. The interviewer assignment and the sample cluster will be completely equal in this case, and the traditional cluster sample standard error computation will reflect both sampling and response error variance associated with different interviewers. For example, with large-cluster sampling where one or more interviewers will work in only a single primary unit, the traditional sampling error estimate of the total variance will fully reflect the measurement error variance contribution associated with the interviewers to the total variance.

However, most data collection and sample designs do not lead to complete correspondence between the interviewer workload and the sampling units. For those cases, only a proportion of interviewer variance is reflected in the sampling variance calculations. Sampling error calculations reflect some portion of the interviewer variance, but not all.

8.1. Cluster sampling

Suppose that from a population of A clusters of equal size, a sample clusters are selected with equal probability. In the selected clusters, all B elements are included in the sample which contains $aB = n$ elements. The equal probability selection of any of the $N = AB$ population elements is

$$f = \frac{a}{A}\frac{B}{B} = \frac{a}{A} = \frac{n}{N} \quad . \tag{8.1.1}$$

The sample mean of n elements in the sample serves to estimate the population mean. It is also the mean of the a cluster means:

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j = \frac{1}{aB}\sum_{a=1}^{a}\sum_{b=1}^{B} y_{ab} = \frac{1}{a}\sum_{a=1}^{a}\bar{y}_a \quad . \tag{8.1.2}$$

In order to analyse the effect of interviewer variance for this case, let y_{abit} be the measurement made by the i-th interviewer on the b-th element in cluster a; the index t is used to denote that y_{abit} is a random variable. Also consider a finite population of interviewers indexed by $i = 1,2,...,K$. Following Lessler and Kalsbeek (1992);

Let,

$$U_a = \begin{cases} 1, \text{if the } a\text{-th cluster is selected for the sample} \\ 0, \text{ otherwise} \end{cases}$$

$$V_i = \begin{cases} 1, \text{if the i-th interviewer is selected for the survey} \\ 0, \text{ otherwise} \end{cases}$$

$$C_{ai} = \begin{cases} 1, \text{if the i-th interviewer is assigned to the } a\text{-th cluster} \\ 0, \text{ otherwise .} \end{cases}$$

Now assume each interviewer is assigned at random to c clusters. Each of the selected $k = a/c$ interviewers completes all interviews in the assigned clusters.

The sample mean of n elements in the sample, given in (8.1.2), can be rewritten as

$$\bar{y}_t = \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}(V_i U_a C_{ai}\frac{1}{B}\sum_{b=1}^{B} y_{abit}) \ . \qquad (8.1.3)$$

Let $\bar{y}_{ait}$ denote $\frac{1}{B}\sum_{b=1}^{B} y_{abit}$, then $\bar{y}_t$ in (8.1.3) can be written as

$$\bar{y}_t = \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}(V_i U_a C_{ai}\bar{y}_{ait}) \ . \qquad (8.1.4)$$

As before, let y_{abit} be written as the sum of two components

$$y_{abit} = m_{ab} + d_{abit} \, , \qquad (8.1.5)$$

where the first component denotes the individual true value for the b -th element of the a -th cluster, and the second term denotes the individual response error for the b -th element of the a -th cluster on the observation by the i -th interviewer on trial t. By replacing y_{abit} with the equivalent sum we can rewrite $\bar{y}_{ait}$ as

$$\begin{aligned} \bar{y}_{ait} &= \frac{1}{B}\sum_{b=1}^{B}(m_{ab} + d_{abit}) \\ &= \frac{1}{B}\sum_{b=1}^{B} m_{ab} + \frac{1}{B}\sum_{b=1}^{B} d_{abit} \qquad (8.1.6) \\ &= \bar{m}_a + \bar{d}_{ait} \ . \end{aligned}$$

Assuming $E_t(d_{abit}) = 0$, the expected value of an observation on b -th element of the a -th cluster is

$$\begin{aligned} E_t(y_{abit}) &= E_t(m_{ab} + d_{abit}) \\ &= \underbrace{E_t(m_{ab})}_{m_{ab}} + \underbrace{E_t(d_{abit})}_{0} \qquad (8.1.7) \\ &= m_{ab} \ . \end{aligned}$$

Since $E_t(d_{abit}) = 0$, $E_t(\overline{d}_{ait}) = E_t(\frac{1}{B}\sum_{b=1}^{B} d_{abit}) = \sum_{b=1}^{B} E_t(d_{abit}) = 0$ for all a, so we can rewrite the expected value of $\overline{y}_{ait}$ as

$$
\begin{aligned}
E(\overline{y}_{ait}) &= E_t(\overline{y}_{ait} \mid U_a V_i C_{ai} = 1) \\
&= E_t(\overline{m}_a \mid U_a V_i C_{ai} = 1) + E_t(\overline{d}_{ait} \mid U_a V_i C_{ai} = 1) \\
&= E_t(\overline{m}_a \mid U_a V_i C_{ai} = 1) \\
&= \overline{m}_a \ .
\end{aligned}
\tag{8.1.8}
$$

We can define the sample mean of the expected cluster means for the a clusters included in the sample as

$$
\overline{m} = \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai} \overline{m}_a \ . \tag{8.1.9}
$$

The expected value of $\overline{y}_t$ is

$$
\begin{aligned}
E(\overline{y}_t) &= E_s E_{t|s}(\overline{y}_t) \\
&= E_s E_{t|s}[\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}(U_a V_i C_{ai} \overline{y}_{ait})] \\
&= \frac{1}{a} E_s[\sum_{i=1}^{K}\sum_{a=1}^{A}(U_a V_i C_{ai} E_t(\overline{y}_{ait} \mid V_i U_a C_{ia} = 1)] \\
&= \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} E_s(V_i U_a C_{ai}) \overline{m}_a \ .
\end{aligned}
\tag{8.1.10}
$$

Since we assume single random selection of clusters and interviewers without replacement, and the assignment of clusters to interviewers is also without replacement so that each cluster is assigned to a single interviewer

$$
\begin{aligned}
&\Pr(U_a = 1) = \frac{a}{A} \ , \\
&\Pr(V_i = 1) = \frac{k}{K} \ , \\
&\Pr(C_{ai} = 1 \mid U_a = 1,\ V_i = 1) = \frac{c}{a};
\end{aligned}
\tag{8.1.11}
$$

where $c = a/k$ is the number of clusters assigned to each interviewer. Thus,

$$E_s(V_i U_a C_{ia}) = \frac{a}{A}\frac{k}{K}\frac{c}{a} = \frac{a}{AK} \ . \tag{8.1.12}$$

$E(\overline{y}_t)$ becomes

$$\begin{aligned} E(\overline{y}_t) &= \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} E_s(V_i U_a C_{ai})\overline{m}_a \\ &= \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} \frac{a}{AK}\overline{m}_a \\ &= \frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A} \overline{m}_a \\ &= \overline{\overline{m}} \ . \end{aligned} \tag{8.1.13}$$

Under the assumption of no interaction between sampling and measurement errors, we can write the variance of $\overline{y}_t$ as the sum of a measurement and a sampling variance component

$$Var(\overline{y}_t) = E(\overline{y}_t - \overline{m})^2 + E\,(\overline{m} - \overline{\overline{m}})^2 \ . \tag{8.1.14}$$

Expanding the first term in (8.1.14), we get

$$\begin{aligned} E(\overline{y}_t - \overline{m})^2 &= E\,(\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\overline{y}_{ait} - \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\overline{m}_a)^2 \\ &= E[\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}(\overline{y}_{ait} - \overline{m}_a)]^2 \\ &= \frac{1}{a^2}\{\sum_{i=1}^{K}\sum_{a=1}^{A} E(V_i U_a C_{ai}) E_t[(\overline{d}_{ait})^2 \,|\, V_i U_a C_{ai} = 1]\} \\ &+ \frac{1}{a^2}\{\sum_{i=1}^{K}\sum_{a\neq a'}^{A} E(U_a U_{a'} V_i C_{ai} C_{a'i}) E_t[(\overline{d}_{ait})(\overline{d}_{a'it}) \,|\, U_a U_{a'} V_i C_{ia} C_{ia'} = 1\} \\ &+ \frac{1}{a^2}\{\sum_{i\neq i'}^{K}\sum_{a=1}^{A} E(U_a V_i V_{i'} C_{ai} C_{ai'}) E_t[(\overline{d}_{ait})(\overline{d}_{ai't}) \,|\, U_a V_i V_{i'} C_{ai} C_{ai'} = 1\} \\ &+ \frac{1}{a^2}\{\sum_{i\neq i'}^{K}\sum_{a\neq a'}^{A} E(U_a U_{a'} V_i V_{i'} C_{ai} C_{a'i'}) E_t[(\overline{d}_{ait})(\overline{d}_{a'i't}) \,|\, U_a U_{a'} V_i V_{i'} C_{ai} C_{a'i'} = 1\} \ . \end{aligned} \tag{8.1.15}$$

Under the above assumptions about the selection of clusters and interviewers and assignment of clusters to interviewers, we have

$$
\begin{aligned}
&\Pr(U_a = 1) = \frac{a}{A}, \\
&\Pr(U_a = 1,\ U_{a'} = 1) = \frac{a}{A}\frac{a-1}{A-1}, \\
&\Pr(V_i = 1) = \frac{k}{K}, \\
&\Pr(V_i = 1,\ \mathrm{V}_{i'} = 1) = \frac{k}{K}\frac{k-1}{K-1}, \\
&\Pr(C_{ai} = 1 \,|\, U_a = 1,\ V_i = 1) = \frac{c}{a}, \\
&\Pr(C_{ai} = 1,\ C_{a'i} = 1 \,|\, U_a U_{a'} = 1,\ V_i = 1) = \frac{c}{a}\frac{c-1}{a-1}, \\
&\Pr(C_{ai} = 1,\ C_{a'i'} = 1 \,|\, U_a U_{a'} = 1,\ V_i V_{i'} = 1) = \frac{c}{a}\frac{c}{a-1}, \\
&\Pr(C_{ai} = 1,\ C_{ai'} = 1 \,|\, U_a = 1,\ V_i V_{i'} = 1) = 0\ .
\end{aligned}
\tag{8.1.16}
$$

Thus, the expression in (8.1.15) becomes

$$
\begin{aligned}
E(\overline{y}_t - \overline{m})^2 = \frac{1}{a}\Big[&\frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A} Var_t(\overline{d}_{ait}) \\
&+ \frac{(c-1)}{A(A-1)K}\sum_{i=1}^{K}\sum_{a \neq a'}^{A} Cov_t(\overline{d}_{ait}, \overline{d}_{a'it}) \\
&+ \frac{c(k-1)}{A(A-1)K(K-1)}\sum_{i \neq i'}^{K}\sum_{a \neq a'}^{A} Cov_t(\overline{d}_{ait}, \overline{d}_{a'i't})\Big]\ .
\end{aligned}
\tag{8.1.17}
$$

The first term of (8.1.17) is the variance of the measurements on a-th cluster mean by the i-th interviewer given that both a-th cluster and i-th interviewer are selected for the survey and a-th cluster is assigned to i-th interviewer. The second term is the measurement covariance between the cluster means measured by the same interviewer. Finally, the last term in the expression, is the between cluster between interviewer covariance.

If we further make the following assumptions on the distribution of d_{abit};

$$
\begin{aligned}
&Var_t(d_{abit}) = s_d^2, \text{ for all } a,\ b, \text{ i;}\\
&Cov_t(d_{abit}, d_{ab'it}) = r_1 s_d^2, \text{ for all } a, \text{ i, and all } b \neq b';\\
&Cov_t(d_{abit}, d_{a'bit}) = r_2 s_d^2, \text{ for all i, and all } b \text{ in } a \neq a';\\
&Cov_t(d_{abit}, d_{a'b'i't}) = r_3 s_d^2, \text{ for all } b \text{ where } i \neq i',\ a \neq a',
\end{aligned}
\qquad (8.1.18)
$$

where $r_1 s_d^2$ is the intra-interviewer intra-cluster covariance of the individual response deviations; $r_2 s_d^2$ is the intra-interviewer between-cluster covariance of the individual response deviations; and $r_3 s_d^2$ is the between-interviewer between-cluster covariance of the individual response deviations, the variances and covariances in expression (8.1.17) becomes

$$
\begin{aligned}
&Var_t(\bar{d}_{ait}) = Var_t(\frac{1}{B}\sum_{b=1}^{B} d_{abit}) = \frac{1}{B^2} Var_t(\sum_{b=1}^{B} d_{abit})\\
&= \frac{1}{B^2}[B Var_t(d_{abit}) + B(B-1) Cov_t(d_{abit}, d_{ab'it})]\\
&= \frac{1}{B}[s_d^2 + (B-1) r_1 s_d^2]\\
&= \frac{s_d^2}{B}[1 + (B-1) r_1];\\
&Cov_t(\bar{d}_{ait}, \bar{d}_{a'it}) = Cov_t(\frac{1}{B}\sum_{b=1}^{B} d_{abit},\ \frac{1}{B}\sum_{b=1}^{B} d_{a'bit})\\
&= \frac{1}{B^2} Cov_t(\sum_{b=1}^{B} d_{abit},\ \sum_{b=1}^{B} d_{a'bit}) = \frac{1}{B^2} B^2 r_2 s_d^2 = r_2 s_d^2;\\
&Cov_t(\bar{d}_{ait}, \bar{d}_{a'i't}) = Cov_t(\frac{1}{B}\sum_{b=1}^{B} d_{abit},\ \frac{1}{B}\sum_{b=1}^{B} d_{a'bi't})\\
&= \frac{1}{B^2} Cov_t(\sum_{b=1}^{B} d_{abit},\ \sum_{b=1}^{B} d_{a'bi't}) = \frac{1}{B^2} B^2 r_3 s_d^2 = r_3 s_d^2\ .
\end{aligned}
\qquad (8.1.19)
$$

The expression in (8.1.17) then becomes

$$
\begin{aligned}
E(\bar{y}_t - \bar{m})^2 &= \frac{1}{a}\Big[\frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A}\frac{1}{B}\{s_d^2 + (B-1)r_1 s_d^2\} \\
&+ \frac{c-1}{A(A-1)K}\sum_{i=1}^{K}\sum_{a\neq a'}^{A} r_2 s_d^2 \\
&+ \frac{c(k-1)}{A(A-1)K(K-1)}\sum_{i\neq i'}^{K}\sum_{a\neq a'}^{A} r_3 s_d^2\Big] \\
&= \frac{1}{a}\Big[\frac{1}{B}\{s_d^2 + (B-1)r_1 s_d^2\} + (c-1)r_2 s_d^2 + c(k-1)r_3 s_d^2\Big] .
\end{aligned}
\tag{8.1.20}
$$

Similarly, expanding the second term in (8.1.14), we get

$$
\begin{aligned}
E(\bar{m} - \bar{\bar{m}})^2 &= E\Big(\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\bar{m}_a - \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\bar{\bar{m}}\Big)^2 \\
&= E\Big[\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}(\bar{m}_a - \bar{\bar{m}})\Big]^2 \\
&= \frac{1}{a^2}\Big[\sum_{i=1}^{K}\sum_{a=1}^{A} E(V_i U_a C_{ai})(\bar{m}_a - \bar{\bar{m}})^2 \\
&+ \sum_{i=1}^{K}\sum_{a\neq a'}^{A} E(U_a U_{a'} V_i C_{ai} C_{a'i})(\bar{m}_a - \bar{\bar{m}})(\bar{m}_{a'} - \bar{\bar{m}}) \\
&+ \sum_{i=1}^{K}\sum_{a\neq a'}^{A} E(U_a U_{a'} V_i V_{i'} C_{ai} C_{a'i'})(\bar{m}_a - \bar{\bar{m}})(\bar{m}_{a'} - \bar{\bar{m}})\Big] .
\end{aligned}
\tag{8.1.21}
$$

The above expression can be rewritten as

$$
\begin{aligned}
E(\bar{m} - \bar{\bar{m}})^2 &= \frac{1}{a}\Big[\frac{1}{A}\sum_{a=1}^{A}(\bar{m}_a - \bar{\bar{m}})^2 \\
&+ \frac{a-1}{A(A-1)}\sum_{a\neq a'}^{A}(\bar{m}_a - \bar{\bar{m}})(\bar{m}_{a'} - \bar{\bar{m}})\Big] .
\end{aligned}
\tag{8.1.22}
$$

Letting

$$
\begin{aligned}
Var(\bar{m}_a) &= \frac{1}{A}\sum_{a=1}^{A}(\bar{m}_a - \bar{\bar{m}})^2, \\
Cov(\bar{m}_a, \bar{m}_{a'}) &= \frac{1}{A(A-1)}\sum_{a\neq a'}^{A}(\bar{m}_a - \bar{\bar{m}})(\bar{m}_{a'} - \bar{\bar{m}}), \text{ for } a \neq a' ,
\end{aligned}
\tag{8.1.23}
$$

we can rewrite (8.1.22) as

$$E(\overline{m}-\overline{\overline{m}})^2=\frac{1}{a}[Var(\overline{m}_a)+(a-1)Cov(\overline{m}_a,\overline{m}_{a'})] \,. \qquad (8.1.24)$$

8.2. Subsampling Within Clusters

Introducing a random sub-sample of cluster elements, will complicate both error formulas: When the sampling fraction is less than one, from the true values of the randomly selected cluster elements only, we can not achieve perfect accuracy in estimating the true value of the cluster parameters. When sub-sampling within clusters, we can estimate $\overline{m}_a$ with some precision error not equal to zero. The precision error of $\overline{y}_{ait}$ will be, then, complicated further by a sampling error within the cluster. Error was only due to measurement before, now, we must also consider the sampling error within clusters.

Assuming that errors resulting from measurement and sampling are uncorrelated, true value of an element within cluster is uncorrelated with error of interviewer, and selection of sample elements from clusters is simple random sample without replacement and independent of true value of elements; we can use a single linear additive model to conceptualise these two kinds of errors. The sub-sampling fraction is $f_a=\frac{b}{B}$. b selected elements from each selected cluster is observed by a randomly assigned interviewer.

Let, y_{abit} denote the observation for the b-th element of the a-th cluster on the observation by the i-th interviewer on trial t, and $\overline{y}_{asit}$ denote the observation on the s-th sample from the a-th cluster; by the i-th interviewer, on the t-th trial. The error due to measurement is still viewed as being caused by the interviewer. As before, y_{abit} is the sum of two components; however, $\bar{y}_{asit}$ is the sum of three components:

$$\begin{aligned} y_{abit} &= m_{ab} + d_{abit} \ , \\ \bar{y}_{asit} &= \bar{m}_a + h_{as} + \bar{d}_{ait} \ , \end{aligned} \tag{8.2.1}$$

where m_{ab} is the true value of the b -th element of the a -th cluster, d_{abit} is the b -th element of the a -th cluster on the observation by the i-th interviewer on trial t, $\bar{m}_a$ is the true mean of the a -th cluster, h_{as} is the error due to a selection of the s-th sample not the whole cluster, $\bar{d}_{ait}$ is the mean of the measurement errors made on the b elements of the cluster.

Taking the expectation of $\bar{y}_{asit}$ involves two kinds of expectations: the expectation taken over different trials, and the expectation taken over different subsets of elements within the cluster. Since it has been assumed that interviewer errors are uncorrelated with the true value of elements and the inclusion of different elements in the sample, $\bar{d}_{ait}$ is only due to interviewer. The same is true with h_{as}, this error is only due to sampling. Expectation of $h_{as} + \bar{d}_{ait}$ over all samples and trials will be the sum of each components expectations. Assuming that both expectations are zero, the expectation of $\bar{y}_{asit}$ is simply,

$$E_{st}(\bar{y}_{asit}) = \bar{m}_a \ . \tag{8.2.2}$$

With fixed a, taking the expectation over either s or t will require conceptualising one error term as variable and the other as fixed:

$$\begin{aligned} &E_{s|t}(h_{as} + \overline{d}_{ait}) = E_{s|t}(h_{as}) + \overline{d}_{ait} = \overline{d}_{ait}; \\ &E_t E_{s|t}(h_{as} + \overline{d}_{ait}) = E_t E_{s|t}(h_{as}) + E_t(\overline{d}_{ait}) = 0; \\ &E_{t|s}(h_{as} + \overline{d}_{ait}) = h_{as} + E_{t|s}(\overline{d}_{ait}) = h_{as}; \\ &E_s E_{t|s}(h_{as} + \overline{d}_{ait}) = E_s(h_{as}) + E_s E_{t|s}(\overline{d}_{ait}) = 0 \ . \end{aligned} \quad \textbf{(8.2.3)}$$

So, with the above assumptions, the order of taking the expectations leads to identical results. Since, in most real life situations, the identification of the sample units is done before the measurements made on those sampled units, when inferring on the population parameters, the ordering of the expectations will follow the reverse order here, first over trials and then over samples. Doing the other way shall not give different results.

Variance of the cluster mean estimator can be decomposed into components: variance resulting from sampling within cluster and variance resulting from the interviewer error. For any cluster, the sampling variance within the cluster and the measurement variance for the interviewer assigned to this cluster will be summed to give the total variance of the estimator of that cluster.

$$Var_{st}(h_{as} + \overline{d}_{ait}) = Var_s(h_{as}) + Var_t(\overline{d}_{ait}) \ . \quad (8.2.4)$$

Let $\overline{y}_t$ be an estimate of the population mean obtained at t-th trial

$$\overline{y}_t = \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}(V_i U_a C_{ai}\,\overline{y}_{asit}) \ . \quad (8.2.5)$$

The expected value of $\bar{y}_t$ is

$$
\begin{aligned}
E(\bar{y}_t) &= E_s E_{t|s}(\bar{y}_t) \\
&= E_s E_{t|s}[\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}(U_a V_i C_{ai}\bar{y}_{asit})] \\
&= \frac{1}{a}E_s[\sum_{i=1}^{K}\sum_{a=1}^{A}U_a V_i C_{ai} E_{t|s}(\bar{y}_{asit})] \\
&= \frac{1}{a}E_s[\sum_{i=1}^{K}\sum_{a=1}^{A}(U_a V_i C_{ai} E_t(\bar{m}_a + h_{as} + \bar{d}_{ait} | U_a V_i C_{ia} = 1)] \\
&= \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}E_s(V_i U_a C_{ai})E_s(\bar{m}_a + h_{as}) \\
&= \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A}E_s(V_i U_a C_{ai})\bar{m}_a \\
&= \frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A}\bar{m}_a \\
&= \bar{\bar{m}}\ .
\end{aligned}
\qquad (8.2.6)
$$

Under the assumption of no interaction between sampling and measurement errors, we can write the variance of $\bar{y}_{st}$ as the sum of a measurement and a sampling variance component

$$Var(\bar{y}_{st}) = E(\bar{y}_t - \bar{m})^2 + E(\bar{m} - \bar{\bar{m}})^2\ . \qquad (8.2.7)$$

Expanding the first term in (8.2.7), we get

$$E(\bar{y}_t - \bar{m})^2 = E_s E_{t|s}(\bar{y}_t - \bar{m})^2$$

$$= E_s E_{t|s}\Big(\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\bar{y}_{asit} - \frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}\bar{m}_a\Big)^2$$

$$= E_s E_{t|s}\Big[\frac{1}{a}\sum_{i=1}^{K}\sum_{a=1}^{A} V_i U_a C_{ai}(\bar{y}_{asit} - \bar{m}_a)\Big]^2$$

$$= \frac{1}{a^2}\{\sum_{i=1}^{K}\sum_{a=1}^{A} E(V_i U_a C_{ai}) E_{st}[h_{as}^2 + \bar{d}_{ait}^2 + 2h_{as}\bar{d}_{ait} \,|\, V_i U_a C_{ai} = 1]\} \qquad (8.2.8)$$

$$+ \frac{1}{a^2}\{\sum_{i=1}^{K}\sum_{a \neq a'}^{A} E(U_a U_{a'} V_i C_{ai} C_{a'i}) E_{st}[h_{as} h_{a's} + h_{as}\bar{d}_{a'it} + \bar{d}_{ait} h_{a's} + \bar{d}_{ait}\bar{d}_{a'it} \,|\, U_a U_{a'} V_i C_{ai} C_{a'i} = 1\}$$

$$+ \frac{1}{a^2}\{\sum_{i \neq i'}^{K}\sum_{a=1}^{A} E(U_a V_i V_{i'} C_{ai} C_{ai'}) E_{st}[h_{as}^2 + h_{as}\bar{d}_{ai't} + \bar{d}_{ait} h_{as} + \bar{d}_{ait}\bar{d}_{ai't} \,|\, U_a V_i V_{i'} C_{ai} C_{ai'} = 1\}$$

$$+ \frac{1}{a^2}\{\sum_{i \neq i'}^{K}\sum_{a \neq a'}^{A} E(U_a U_{a'} V_i V_{i'} C_{ai} C_{a'i'}) E_{st}[h_{as} h_{a's} + h_{as}\bar{d}_{a'i't} + \bar{d}_{ait} h_{a's} + \bar{d}_{ait}\bar{d}_{a'i't} \,|\, U_a U_{a'} V_i V_{i'} C_{ai} C_{a'i'} = 1\}\,.$$

Since we assumed independence of sampling and measurement errors, the expectation of the cross terms involving both h_{as} and $\bar{d}_{ait}$ will disappear. Taking the necessary expectations the equation in (8.2.8) simplifies further:

$$E(\bar{y}_t - \bar{m})^2 = \frac{1}{a}\{\frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A}[Var_s(h_{as}) + Var(\bar{d}_{ait})]$$

$$+ \frac{c-1}{AK(A-1)}\sum_{i=1}^{K}\sum_{a \neq a'}^{A}[Cov_s(h_{as}, h_{a's}) + Cov_t(\bar{d}_{ait}, \bar{d}_{a'it})] \qquad (8.2.9)$$

$$+ \frac{c(k-1)}{A(A-1)K(K-1)}\sum_{i \neq i'}^{K}\sum_{a \neq a'}^{A}[Cov_s(h_{as}, h_{a's}) + Cov_t(\bar{d}_{ait}, \bar{d}_{a'i't})]\}\,.$$

If we let all variance and covariance terms in the above sum be same for different units:

$$
\begin{aligned}
&Var_s(h_{as}) = s_h^2 \text{ for all } a\,;\\
&Var_t(\bar{d}_{ait}) = Var_t\Big(\frac{1}{b}\sum_{b=1}^{b} d_{abit}\Big)\\
&= \frac{1}{b^2}[bVar_t(d_{abit}) + b(b-1)Cov_t(d_{abit}, d_{ab'it})]\\
&= \frac{s_d^2}{b}[1 + (b-1)r_{d_1}] \text{ for all } a\,, \text{ and all } i\,; \qquad (8.2.10)\\
&Cov_s(h_{as}, h_{a's}) = r_{s_1} s_h^2 \text{ for all } a \neq a'\,;\\
&Cov_t(\bar{d}_{ait}, \bar{d}_{a'it}) = r_{d_2} s_d^2 \text{ for all i for } a \neq a'\,;\\
&Cov_t(\bar{d}_{ait}, \bar{d}_{a'i't}) = r_{d_3} s_d^2 \text{ for all } i \neq i', and\ a \neq a'\,,
\end{aligned}
$$

where s_h^2 is the within-cluster sampling variance of the cluster mean estimator; s_d^2 is the interviewer variance of the individual response deviations; $r_{s_1} s_h^2$ is the between-cluster sampling covariance of the cluster mean estimator; $r_{d_1} s_d^2$ is the intra-interviewer intra-cluster covariance of the individual response deviations; $r_{d_2} s_d^2$ is the intra-interviewer between-cluster covariance of the individual response deviations; and $r_{d_3} s_d^2$ is the between-interviewer between-cluster covariance of the individual response deviations, the expression in (8.2.10) becomes

$$
\begin{aligned}
E(\bar{y}_t - \bar{m})^2 &= \frac{1}{a}\Big\{\frac{1}{AK}\sum_{i=1}^{K}\sum_{a=1}^{A}[s_h^2 + \{\frac{s_d^2}{b}[1+(b-1)r_{d_1}\}]\\
&+ \frac{c-1}{KA(A-1)}\sum_{i=1}^{K}\sum_{a\neq a'}^{A}[r_{s_1}s_h^2 + r_{d_2}s_d^2]\\
&+ \frac{c(k-1)}{K(K-1)A(A-1)}\sum_{i\neq i'}^{K}\sum_{a\neq a'}^{A}[r_{s_1}s_h^2 + r_{d_3}s_d^2]\Big\} \qquad (8.2.11)\\
&= \frac{1}{a}\{s_h^2 + \frac{s_d^2}{b}[1+(b-1)r_{d_1}] + (a-1)r_{s_1}s_h^2 + (c-1)r_{d_2}s_d^2 + c(k-1)r_{d_3}s_d^2\}\,.
\end{aligned}
$$

The second part of the equation (8.2.7) is calculated as in the first data collection and sample design in the previous section of this chapter; we get

$$E(\bar{m}-\bar{\bar{m}})^2 = \frac{1}{a}[Var(\bar{m}_a)+(a-1)Cov(\bar{m}_a,\bar{m}_{a'})] \ . \qquad (8.2.12)$$

8.3. Proposed Models

Most data collection and sample designs need a specific variance decomposition model. The aim of this chapter was to build a variance decomposition model which will adequately reflect the variance generating processes involved by the specific data collection and sample design considered.

The literature on this subject usually considers basic data collection and sample designs. Complex designs are usually considered when they are actually needed and these models are generally not suitable for generalising in a direct fashion, because they often involve the particular data collection and sample design.

When a cluster sample is drawn this is often done to reduce the costs of the data collection process and thus the assignment of interviewers to clusters is made to get advantage of the cost features involved with cluster samples. An assignment scheme which randomly assigns interviewers to actual sampling units is no more preferable. The cost model in (5.1.1) explains the cost features of a cluster design (Groves, 1989):

Total cost= Fixed cost + Cluster costs + Element costs

$$\mathrm{C}=\mathrm{C}_0 + C_a a + C_b b$$

where C_0 denotes the fixed costs of doing the survey, independent of the number of sample clusters or sample elements per cluster; C_a denotes the cost of selecting, and locating of each cluster, independent of the number of sample elements for each cluster; a denotes the number of sample clusters; C_b denotes the cost of selecting, contacting, and interviewing a single sample element from a cluster; b is the number of sample elements per cluster. The model explains why a different assignment scheme which is not random may reduce costs.

Assigning only one interviewer for each cluster generates a different variance then assigning interviewers randomly to sampling units. A different assignment and sample design is often used to reduce the costs of the survey but it also changes the variance structure. The sections 1 and 2 of this chapter deal with decomposition of the variance of two similar sample designs and assignment schemes. The first covered design is a special case of the second one. However, here the aim was moving slightly from the theory built for simple random sampling designs and simple assignment schemes to a more complex one; and the first design provides a step between.

Lessler and Kalsbeek (1992) provide a variance decomposition model which takes into account of the measurement errors generated by the interviewers. Their model is provided for a case where the sample design is simple random sample and inclusion and assignment of interviewers is done similarly in a random fashion. From a pool of interviewers, an interviewer is selected randomly for the interview and assigned randomly to a sampling unit included in the sample. The model by Lessler and Kalsbeek (1992) has been chosen as a starting point since it involved indicator functions which made it suitable for taking into account of the changes in the variance structure as a result of modifications in the assignment and sampling design.

The first design proposed in this chapter is only slightly different from the one used by Lessler and Kalsbeek (1992). In this paper, their model has been used at a different level: the clusters are thought of as creating a different level then the sampling units. The model proposed by Lessler and Kalsbeek (1992) is used in this paper to capture the variance structure at the cluster level; in other words, here the clusters take the place of the sampling units in the model proposed by Lessler and Kalsbeek (1992). In the first model all cluster elements for chosen clusters are observed. Selection of all cluster elements in the sample simplifies the solution; the mean obtained from the observation of all units in the cluster does not involve sampling error but only measurement error. Making the necessary calculations one arrives at the following total variance formula:

$$
\begin{aligned}
Var(\bar{y}_t) &= E(\bar{y}_t - \bar{m})^2 + E(\bar{m} - \bar{\bar{m}})^2 \\
&= \frac{1}{a}\left[\frac{1}{B}\{s_d^2 + (B-1)\, r_1 s_d^2\} + (c-1)\, r_2 s_d^2 + c(k-1)\, r_3 s_d^2\right] \qquad (8.3.1) \\
&+ \frac{1}{a}[Var(\bar{m}_a) + (a-1)Cov(\bar{m}_a, \bar{m}_{a'})] \; .
\end{aligned}
$$

A random sub-sample of cluster elements is considered in the next section. Sampling error within the cluster complicates the precision error of the cluster mean estimator. Here, within a cluster two kinds of error are present: sampling error and measurement error made by the interviewers. These two errors have been assumed to be independent, and a linear additive model is used to illustrate their total effect. At the sampling unit level there is only measurement error, but when we move from sampling units to clusters sampling error must also be considered.

$$y_{abit} = m_{ab} + d_{abit} \ ,$$
$$\bar{y}_{asit} = \bar{m}_{a} + h_{as} + \bar{d}_{ait} \ ,$$

where m_{ab} is the true value of the b -th element of the a -th cluster, d_{abit} is the b -th element of the a -th cluster on the observation by the i -th interviewer on trial t, $\bar{m}_{a}$ is the true mean of the a -th cluster, h_{as} is the error due to a selection of the s -th sample not the whole cluster, $\bar{d}_{ait}$ is the mean of the measurement errors made on the b elements of the cluster.

Making the necessary calculations one arrives at the following total variance formula:

$$\begin{aligned} Var(\bar{y}_t) &= E(\bar{y}_t - \bar{m})^2 + E(\bar{m} - \bar{\bar{m}})^2 \\ &= \frac{1}{a}\{s_h^2 + \frac{s_d^2}{b}[1+(b-1)r_{d_1}] + (a-1)r_{s_1}s_h^2 + (c-1)r_{d_2}s_d^2 + c(k-1)r_{d_3}s_d^2\} \\ &+ \frac{1}{a}[Var(\bar{m}_a) + (a-1)Cov(\bar{m}_a, \bar{m}_{a'})] \ , \end{aligned} \qquad (8.3.2)$$

One can observe that letting $b = B$ in (8.3.2) gives the same result as in (8.3.1) since s_h^2 will be zero for $b = B$.

CHAPTER 9

CONCLUSIONS

The basic objective of a survey is to provide information on the basis of survey variables, the measurement error problem should be studied with this aim in mind. The survey variables can be measured with a certain precision, and the knowledge on this precision is very important for both utilisation of the information and arriving at more precision. Trying to understand the mechanism of our measurements and their variance structure will surely be fruitful.

In this study the concentration was on response variance components. Literature on the subject has been reviewed and major contributions have been presented. The different perspectives of various disciplines on measurement error is considered in Chapter 2. Three major languages of error which appear to be applied to survey data are overviewed. They are associated with three different academic disciplines and illustrate the consequences of groups addressing similar problems in isolation of one another. The three disciplines are statistics (especially statistical sampling theory), psychology (especially psychometric test and measurement theory), and economics (especially econometrics). In Chapter 3, some techniques for evaluating and controlling measurement error in surveys are discussed. The methods discussed are reinterview studies, multiple indicators studies, record check studies, and. cognitive studies. Four sources of measurement errors (questionnaire, data collection method, interviewer, and respondent) which are the elements that comprise data collection are considered in Chapter 4. The costs of survey activities often act as limiting influences on efforts to reduce survey errors. This subject is considered in Chapter 5. The measurement of response errors requires that they be represented by a mathematical model. A number of alternative models have been considered in Chapter 6. Decomposition of the variance by O'Muircheartaigh (1982) is discussed in chapter 7. O'Muircheartaigh (1982) partitions the total variance of estimators into four components, each of which has a different implication for the survey design: simple and correlated response error, simple and correlated sampling error.

A lot has been done by the survey theorists on interviewers' contribution to the measurement error. Correlated response variance component is used to capture the correlated effect of interviewers' contribution to the total variance. Most data collection and sample designs need a specific variance decomposition model. As data collection and sample design methods move from simple to complex, isolating the interviewer contribution from the total error gets more complicated.

When a cluster sample is drawn this is often done to reduce the costs of the data collection process and thus the assignment of interviewers to clusters is made to get advantage of the cost features involved with cluster samples. Assigning only one interviewer for each cluster generates a different variance then assigning interviewers randomly to sampling units. A different assignment and sample design is often used to reduce the costs of the survey but it also changes the variance structure.

The Sections 1 and 2 of Chapter 8 deal with decomposition of the variance of two similar sample designs and assignment schemes. The first covered design is a special case of the second one. However, here the aim was moving slightly from the theory built for simple random sampling designs and simple assignment schemes to a more complex one; and the first design provides a step between.

In the first model, all cluster elements for chosen clusters are observed. Selection of all cluster elements in the sample simplifies the solution; the mean obtained from the observation of all units in the cluster does not involve sampling error but only measurement error. Utilising a general measurement error model for a survey with interviewers given by Lessler and Kalsbeek (1992) the variance decomposition is obtained.

A random sub-sample of cluster elements is considered in the next section. Sampling error within the cluster complicates the precision error of the cluster mean estimator. Within a cluster two kinds of error are present: sampling error and measurement error made by the interviewers. These two errors have been assumed to be independent, and a linear additive model is used to illustrate their total effect. This second data collection and sample design involved further complexities but the proper use of the model for the first data collection and sample design made the solution possible.

However, first of all, the identifiability and the estimability of the parameters of the model have to be worked out for the estimation of unknown parameters of the models. For a future study this may be considered as a subject. Also, for a future recommendation we propose the use of the decomposition models obtained in this thesis in determining an optimum allocation in existence of measurement errors.

REFERENCES

Bailey, L., Moore, T. and Bailar, B. 1978. "An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample," *Journal of the American Statistical Association*, 73, 1: 16-23.

Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N. and Sudman, S. 1991. *Measurement Errors in Surveys*. New York: John Wiley & Sons, Inc.

Biemer, P. and Stokes, S. 1991. "Approaches to the Modeling of Measurement Error," in P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.), *Measurement Error in Surveys*. 487-516. New York: John Wiley & Sons, Inc.

Bohrnstedt, G. 1983. "Measurement," in P. Rossi, R. Wright, and A. Anderson (eds.), *Handbook of Survey Research.* 70-122. New York: Academic Press.

Cochran, W. G. 1968 "Errors of Measurement in Statistics," *Technometrics*, 10: 637-666.

Fellegi, I. 1964. "Response Variance and Its Estimation," *Journal of the American Statistical Association,* 59: 1016-1041.

Fellegi, I., and Sunter A. 1974 "Balance Between Different Sources of Survey Errors-Some Canadian Experiences," *Sankhya,* Series C, Vol. 36, Pt. 3: 119-142.

Forsman, G. and Schreiner, I. 1991. "The Design and Analysis of Reinterview: An Overview," in P. Biemer, R. Groves, L. Lyberg, N. Mathiewetz, and S. Sudman (eds.), *Measurement Error in Surveys*. 279-302. New York: John Wiley & Sons, Inc.

Forsyth, B. and Lessler, J. 1991. "Cognitive Laboratory Methods: A Taxonomy," in P. Biemer, R. Groves, L. Lyberg, N. Mathiewetz, and S. Sudman (eds.), *Measurement Error in Surveys*. 393-418. New York: John Wiley & Sons, Inc.

Groves, R. 1989. *Survey Errors and Survey Costs.* New York: John Wiley & Sons, Inc.

Hansen, M., Hurwitz, W. and Bershad, M. 1961. “Measurement Errors in Censuses and Surveys,” *Bulletin of International Statistical Institute,* 38, 359-374.

Hansen, M., Hurwitz, W. and Madow, W. 1953. *Sample Survey Methods and Theory*. New York: John Wiley & Sons, Inc.

Hansen, M., Hurwitz, W. and Pritzker, L. 1964. “The Estimation and Interpretation of Gross Differences and the Simple Response Variance,” in C. Rao (ed.), *Contributions to Statistics*. Calcutta: Statistical Publishing Society.

Kish, L. 1962. “Studies of Interviewer Variance for Attitudinal Variables,” *Journal of the American Statistical Association*, 49: 520-538.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons, Inc.

Koch, G. 1973. “An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimators Involving Subclass Means,” *Journal of the American Statistical Association,* 68: 906-913.

Lessler, J. and Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York: John Wiley and Sons, Inc.

Lyberg L., and Kasprzyk. 1991. “Data Collection Methods and Measurement Error:An Overview,” in P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.), *Measurement Error in Surveys*. 487-516. New York: John Wiley & Sons, Inc.

Mahalanobis, P. 1946. “Recent Experiments in Statistical Sampling in the Indian Statistical Institute,” *Journal of the Royal Statistical Society,* 100, 109: 325-378.

Nolin, M. and Chandler, K. 1996. *Use of Cognitive Laboratories and Recorded Interviews in the National Household Education Survey*. (NCES 96-332). Washington, D.C.: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Oksenberg, L. and Cannell, C. 1977. “Some Factors Underlying the Validity of Self- Report,” *Bulletin of the International Statistical Institute*: 325-46l.

O’Muircheartaigh, C. 1982. “Methodology of the Response Errors Project,” *World Fertility Survey, Scientific Report 28*.

O'Muircheartaigh, C. and Marckward, A. 1980. "An Assessment of the Reliability of World Fertility Study Data," *Proceedings of the World Fertility Survey Conference*, 3: 305-379.

Tourangeau, R. 1984. "Cognitive Sciences and Survey Methods," in T. Jabine, E. Loftus, M. Straf, J. Tanur, and R. Tourangeau (eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, D.C.: National Academy of Science.