Catalogue 13-564E Occasional c. 3

A compendium of methods of error evaluation in censuses and surveys

Statistics Canada Statistique Canada

STATISTICS CANADA

Statistical Services Field

A COMPENDIUM OF METHODS OF ERROR EVALUATION IN CENSUSES AND SURVEYS

This compendium was assembled within the Methodology Divisions of the Statistical Services Field, Statistics Canada by a team consisting of J.-F. Gosselin, B. N. Chinnappa, P. D. Ghangurde and J. Tourigny.

Published by Authority of

The Minister of Industry, Trade and Commerce

Version française de cette publication disponible sur demande (nº 13-564F au catalogue).

November 1978
6-0000-501

Price: 70 cents

Ottawa

TABLE OF CONTENTS

TABLE OF CONTENTS – Concluded

I. INTRODUCTION

1.1 Objectives

It is well known that all data, from whatever source, are subject to errors. This is true whether the data are derived from a complete enumeration of the population, from a sample survey, or from administrative records. It is true whether the data are obtained using direct observation or measurement, using interviews, or by any other collection method. The potential sources of errors include the failure to cover the population of interest, misunderstanding of questions, errors in recording answers, measurements or observations, errors in coding and processing the data, sampling errors, and others.

Generally, the interpretation of data requires some knowledge about their quality. This is essential for determining whether phenomena observed in the data are real or whether they are largely the results of either the inherent variability in the data or deficiencies in the methods of collecting and assembling the data.

Measures of data quality are also very important to designers of censuses and surveys. This arises from the need to continually monitor, and if necessary improve, the quality of the data being produced, and to provide information that may lead to a better allocation of resources amongst various phases of a survey. This usually requires more detailed investigations to measure the level of error and identify the types and sources of errors.

This document is a compendium or catalogue of methods that may be used to assess the quality of data obtained from censuses or surveys and to identify types and sources of errors.

1.2 Approach

The general approach taken in this document is to present methods according to the types or sources of errors as follows:

(a) coverage;

(b) response and measurement;

(c) non-response;

(d) coding;

(e) data capture;

(f) edit and imputation;

(g) sampling and estimation.

Each type is discussed separately in Sections 2 to 8.

In each section, attempts have been made to define the type or source of error being investigated together with a brief discussion of known methods of controlling such errors. This is then followed by a series of write-ups each describing a specific method of measuring these errors.

Each method is presented with a statement of objectives, a brief description of the methodology and a general evaluation of the method in terms of its effectiveness and practical implications.

The following points should be noted:

(i) The methods described in this document are presented in a general context and are not directed to a particular survey. Hence, details about the survey methodology (such as sample design, estimation) have been left out. However, some references to specific applications of the methods have been included. These are by no means exhaustive.

(ii) The methods presented are oriented mainly towards the measurement of data quality or the identification of problems that lead to errors. This document does not direct itself to the description of methods that are used in a survey to control errors, or to adjust or improve survey estimates. These control methods are considered to be an integral part of the survey itself in the sense that they contribute to the quality of the final estimates. In addition, we have deliberately excluded methods that deal essentially with the assessment of the operational efficiency of a particular phase of a survey.

(iii) Although the methods are subdivided according to types or sources of errors it is recognized that the various phases of a survey are very closely interrelated and therefore errors cannot always be separated or attributed uniquely to a particular type or source.

1.3 Categories of Methods

Most methods presented in this document fall into one of the following four categories:

1. Analysis of data obtained from the survey itself.
2. The comparison of survey aggregate estimates with estimates obtained from an external source.
3. Independent repetition of a particular survey operation under the same general conditions.
4. Micro-comparison of survey data with another more reliable source.

Methods in categories 1 and 2 usually involve the analysis or comparison of fairly readily available data. These may not always lead to precise measures of data quality but rather provide a fairly general assessment of the magnitude of error and in some cases the actual types or sources of errors.

Category 3 methods are usually designed to measure the variability in a particular survey phase (e.g., response, coding) while category 4 methods are geared towards the measurement of bias (e.g., coverage, response) and the identification of specific types of error. Both of these methods require a specially designed study and/or the availability of an alternative more reliable source of data.

It should be noted that very similar methods are presented for different sources of error (e.g., category 2). Although these may appear repetitious, they are included for completeness since they may be used, depending on the context, to assess different types or sources of error.

References to applications of the various methods are included throughout the text though no attempt at completeness has been made. A more extensive set of references on non-sampling errors in surveys can be found in the bibliography prepared by Professor T. Dalenius which appears in Volume 45 (1977) of the International Statistical Review.

2. COVERAGE

The objectives of a survey usually define a target population for which estimates of characteristics are required. Coverage errors are introduced whenever the sampling frame (which could be made up of subframes in a multi-stage design) does not adequately represent the target population at the time of the survey. These errors include:

(a) units in the target population that are missing from the sampling frame (undercoverage);

(b) units in the sampling frame that are not in the target population (overcoverage);

(c) units in the target population that are included more than once in the sampling frame (duplication).

These errors could arise for a variety of reasons, e.g., definitional ambiguities, ambiguities in delineation of boundaries of an areal unit, incorrect listing procedures, or events (like births, deaths or movements of units) which make the frame out of date.

In the case where the target population is defined as a collection of units which can only be identified during the survey, coverage errors could arise because of errors at the data collection stage in reporting the characteristics that define these units. For example, if, in a household survey, the target population is persons 14 years and over, coverage errors can be introduced when information on all members of the household is not supplied, or when age is stated incorrectly for some members.

Coverage errors may be reduced by:

(i) defining the sampling units (at any sampling stage) that contain the reporting units, with unambiguous, easily recognizable and stable boundaries;

(ii) using multiple-frame techniques;

(iii) implementing procedures to update the sampling frame using either external sources (e.g., Post Office, National Revenue, Social Insurance Records) or information obtained from the survey itself;

(iv) devising improved listing procedures and controls.

Five methods of assessing coverage errors in censuses and surveys are identified and described in this section.

2.1 Comparison at the macro level with estimates based on an external source.

2.2 Evaluation of coverage error using a dual collection system approach.

2.3 Estimation of coverage error based on a micro-match with a reliable source of information.

2.4 Analysis of survey counts.

2.5 Chain method.

2.1 Comparison at the Macro Level with Estimates Based on an External Source

Objective

To measure the net coverage error.

Methodology

A list based on an external source of information which measures the same target population (often an earlier census) is transformed to meet the survey objectives and brought up-to-date, by taking into account births, deaths, amendments and migration movements. Then the external source estimates are compared with the corresponding survey weighted estimates. Additional coverage error information can be provided by consistency checks (e.g., a survey ratio of males-to-females may be compared with the corresponding ratio from the external source).

Evaluation

This method is relatively quick and cheap provided comparable external estimates are available. No further data collection is required. The external source estimates may be prepared in-house while collecting the survey data.

The shortcomings of the method result from the shortcomings in the data used to construct an up-to-date external source. The external source should be reliable and comparable. Often information on births and deaths (in the general sense, e.g., companies coming into and going out of business) is scarce while information on amendments (e.g., changes to type of business) requires a well structured maintenance function.

The method can measure the net coverage error relative to the external source after allowing for sampling and non-sampling errors in both estimates. In continuous surveys, a consistent trend in the coverage error will tend to confirm the presence of coverage bias. The method does not discern causes of the coverage error.

If the aggregated level selected for comparison is a subclass based on the characteristics of the unit (e.g., age-sex-province) the net coverage error cannot easily be separated from response error.

Reference

Gray, G.B., General Quality Report (Canadian

Labour Force Survey), Technical Paper, HSDS, Statistics Canada, June 1973.

2.2 Evaluation of Coverage Error Using a Dual Collection System Approach

Objectives

To measure the undercoverage error.

To identify possible causes of undercoverage.

Methodology

This method involves the use of an independent (though not necessarily complete) list of the target population to estimate undercoverage in the original survey frame.

An independent sample is selected from each source of information and a linkage operation is carried out to determine whether or not each selected unit is also included in the other source. Once the matching is completed the sample cases may be classified according to the following table:

Total target population	In the original survey frame	Not in the original survey frame	Total
In the independent source	A	B	A + B
Not in the independent source	C	D	C + D
Total	**A + C**	**B + D**	**N**

Estimates of A, B and C can be derived from the matching operation, but D cannot be estimated directly from the sample. However, under the assumption that coverage errors in the two sources are independent (i.e., there is no tendency to miss the same units), D may be estimated by:

$$\hat{D} = \frac{CxB}{A}$$

which will lead to an estimate of the undercoverage $(B + \hat{D})$ in the original survey frame.

Usually a survey is compared with a Post Enumeration Survey. A Census of Population may be matched with a household survey or with administrative records.

Evaluation

The quality of the estimates and the level of information on coverage errors obtained from this method will largely depend on the independence assumption underlying the method. Indeed, if coverage errors in the two sources tend to be correlated, then D will necessarily underestimate the actual number of units missing from both frames, thus underestimating the total undercoverage in the main survey frame. In such situations the method will only provide an estimate of the undercoverage in the original survey frame relative to the independent source.

Great care must also be exercised when performing the linkage operation since matching errors (i.e., the matching of different units or a failure to match a unit common to both sources) could seriously bias the results. This matching operation may in some cases be very difficult and costly, especially when the two sources are not compatible in terms of the information used in the matching.

Out of scope units in either source may also create problems in matching and in estimation.

To summarize, this method will provide a good means of obtaining direct information on undercoverage provided that the independence assumption is a reasonable one, and that the problems and costs associated with the matching operation are not too great. The following reference provides complete detail on this method.

Reference

Marks, E.S., Seltzer, W. and Krotki, K.J., *Population Growth Estimation: A Handbook of Vital Statistics Measurement,* The Population Council, New York, 1974.

2.3 Estimation of Coverage Error Based on a Micro-match with a Reliable Source of Information

Objectives

To measure the undercoverage rate.

To identify possible causes of undercoverage.

Methodology

A reliable external source of information, measuring the same target population is compared with the survey. A micro-match is performed in the following way:

1. selection of a sample of units from the external source;

2. identification of information on which the matching will be based;
3. one-way match of sample units with the survey frame.

An estimate of undercoverage error is based on the results of the match of sample units.

Evaluation

This method should provide good estimates of undercoverage provided the independent source of information is complete and can be updated in such a way that characteristics to be matched (e.g., name, address) are correct for the survey reference period (i.e., besides being more reliable, the external source must be comparable in space and time).

The method will produce an undercoverage error estimate relative to the external source. This estimate will be biased if errors are made in matching (particularly failure to match).

Since this method identifies individual units that were not in the survey frame, the causes of undercoverage can also be investigated.

The linkage and matching operation may in some cases be quite costly.

References

Brackstone, G.J. and Gosselin, J.-F., The 1971 Reverse Record Check, 1971 Census Evaluation Programme, Results Memorandum, CDN 71-E-23, Part 1, Statistics Canada, October 1974.

The Coverage of Housing in the 1970 Census, Evaluation and Research Program, 1970 Census of Population and Housing, PHC(E)-5, U.S. Department of Commerce, Bureau of the Census, October 1973.

2.4 Analysis of Survey Counts

Objectives

To indicate possible deterioration of the survey frame.

To monitor changes and shifts in the population.

Methodology

The method normally involves (i) an analysis of data obtained from the actual survey operations regarding deaths, duplication notices, out-of-scope units, ill-defined units, and changes in classification, and (ii) the comparison of estimates for two or more replicates designed to carry the same expected number of units and the investigation of any differences that are statistically significant.

While the analysis may provide an estimate of coverage error, both the analysis and the comparison may either indicate errors in listing, in frame maintenance, in sample selection and administration, or real growth or decline in the population.

Evaluation

The method may provide a general indication of possible frame or listing problems as well as changes in the population under study.

The method is suitable for large scale continuous surveys where coverage errors can occur due to the frame becoming out-of-date as a result of growth, decline, etc. It is also a sample control mechanism.

Reference

Lawes, M. and Newton, F.T., Paired Areas Analysis Sample Yield Check, Technical Report, HSDS, Statistics Canada, May 1975.

2.5 Chain Method

Objective

To measure coverage errors.

Methodology

The survey questionnaire carries a set of questions requesting the respondent to identify other units in the target population. For example, in a health survey, respondents can be asked to name other patients they know with the same illness; in a business survey, respondents can be asked to list business partners along with their Social Insurance numbers and to provide a list of Payroll Deduction Account numbers.

In the case of a repeated survey, this method also provides a means of updating the survey frame.

Evaluation

This method is most appropriate for rare populations. It is an inexpensive way of assessing the completeness of the sampling frame. However, although this method will provide an indication of the coverage of an existing list, it will not provide an unbiased estimate of coverage error unless one is willing to assume that the units identified by respondents are a random sample from the target population.

This method is helpful in improving the completeness of the sampling frame.

3. RESPONSE

A response error occurs whenever a characteristic is mis-reported in a census or a survey. This may occur whether the information is obtained through self-enumeration,[1] by interview, or by a measurement process.

Generally, response errors may result from a misunderstanding of concepts or questions, accidental or deliberate error, proxy responses or simply a lack of knowledge on the part of respondents or interviewers. It has been shown that both respondents and interviewers may contribute significantly to the total response error. These errors will affect the quality of survey results to the extent that they introduce additional variability and bias into the final survey estimates.

The methods that are used to reduce and control response errors include, (i) testing of survey instruments such as questionnaires, (ii) adequate training, supervision and control of survey operations, (iii) follow-up of a sub-sample of cases using more detailed probing questions with a view to adjusting for response bias, and (iv) detailed editing and imputation.

This section describes possible methods of measuring response errors. These have been presented in the framework of the usual "response variance – response bias" model. A detailed description of this model may be found in the general references.

The methods are:

3.1 Analysis of a sample of records with error or rare conditions.

3.2 Macro-comparison with external data.

3.3 In-depth re-interview.

3.4 Re-interview under the same conditions.

3.5 Controlled experiments.

3.6 Micro-comparison with external data.

General References

Bailar, B.A. and Dalenius, T., Estimating the Response Variance Components of the U.S. Bureau of the Census Survey Model, *Sankhya (B)*, Vol. 31, 1969.

Hansen, M.H., Hurwitz, W.N. and Bershad, M.A., Measurement of Errors in Censuses and Surveys, *Bulletin of the I.S.I.*, Vol. 38, Part 2, 1960.

Koch, G.G., An Alternative Approach to Multivariate Response Error Models for Sample Survey Data With Applications to Estimators Involving Subclass Means, *JASA*, Vol. 68, 1973.

[1] That is, when a respondent completes, by himself, a questionnaire which is returned by mail or collected by an enumerator.

3.1 Analysis of a Sample of Records with Error or Rare Conditions

Objective

To gather evidence on the type and level of response errors.

Methodology

A sample of cases that fail certain pre-specified conditions is selected. In most cases these conditions would be edit rules. However, these conditions may also be chosen in such a way that they select possible, but rare combinations of values (this is sometimes done as part of the survey operation).

The corresponding survey questionnaires or source documents are then examined to determine whether the failure results from a processing error (coding, data capture, etc.) or whether the particular data configuration actually exists on the questionnaire. In the latter case, other data on the questionnaire are examined in an attempt to explain the error or rare condition.

Evaluation

This method will not generally provide measures of data quality such as variance or bias. It may however provide very useful information on the type, source, and frequency of response and processing errors that affect the quality of final survey estimates. This method is fairly cheap and uses readily available data.

3.2 Macro-comparisons with External Data

Objective

To obtain a general indication of extent of response errors.

Methodology

This method involves the comparison of survey aggregate estimates with independent external data with a view to obtaining a general indication of the level and direction of response errors.

Evaluation

The level of information that may be gathered from this method will generally depend on the quality and comparability of the actual external data being used, in terms of coverage, response, etc. Indeed the external data may themselves be subject to error. Also, they may have been gathered using different definitions, concepts or methodologies, they may suffer from a time-lag bias, and it may be difficult or impossible to make compensating adjustments. Hence the deficiencies may in some cases make this method difficult to apply. However, when these disadvantages are not too great, macro-comparisons may give a general indication of the level and direction of response errors.

This method will not generally provide detailed information on the types and sources of response errors. It is relatively cheap and fast.

Reference

Kempster, A.J., Background Information on the 1971 Census Labour Force Data, Population and Housing Research Memorandum, 71-EC-4, Census Field, Statistics Canada, November 1973.

3.3 In-depth Re-interview

Objectives

To measure response bias.

To gather evidence on the type or source of response errors.

Methodology

This method involves the re-enumeration of a random sample of respondents to the original survey under "ideal" conditions. This entails the use of better enumerators, trained to deal with problem cases, and the use of more detailed probing questions in order to obtain more accurate and complete responses.

These are then compared to the original survey responses and estimates of reliability are obtained by weighting these sample data to the population level.

Evaluation

Re-enumeration studies generally provide a fairly good means of producing data on the extent and type of response errors in surveys given that it is possible to obtain more reliable data at the re-interview stage. In certain circumstances they may also provide a suitable vehicle for other types of evaluation studies, such as checks on coverage or procedures.

One of the problems with such studies lies with the difficulty in obtaining more reliable data during the re-interview. This will obviously depend on the type of data being evaluated and the respondent's capability and willingness to provide more detailed information. In addition, such projects are usually very expensive since they involve extensive and extremely careful field work in most cases. They also place an extra burden on the respondents. Hence they are probably practical only in censuses and large surveys.

Reference

Garton, B.E., Analysis of Census Data on Type of Dwelling, Internal Report, Statistics Canada, August 1974.

3.4 Re-interview under the Same Conditions (with or without Reconciliation)

Objectives

To measure the extent of response errors.

To gather evidence on the type or source of response errors.

Methodology

This involves the re-interview of a sample of respondents to the original survey. Attempts are made to duplicate the general conditions of the original survey in terms of questions, interview procedures, etc. The two sets of responses are then compared to obtain estimates of response variance and some of its components.

This type of study may be extended to include reconciliation in the field, that is, the interviewer compares the two responses and asks further questions whenever a discrepancy is found. This will lead to estimates of response bias in addition to the estimates of response variance.

Evaluation

The difficulty in implementing this type of study lies in the achievement of independence and identical survey conditions. A re-interview under very similar conditions may be possible but independence may be lost due to respondents recalling their previous responses. This may be reduced by increasing the time lag between the two interviews, however recall problems may then begin to influence the response variance in the re-interview. Also, if carried out with reconciliation, the estimates of response bias will not generally reflect the "consistently wrong" responses.

In spite of these disadvantages, these studies generally produce important information on the types and sources of response errors.

Such studies are usually quite expensive since they involve extensive and extremely careful fieldwork. In addition, they place an extra burden on respondents.

References

Quality Control and Measurement of Non-sampling Error in the Health Interview Survey, NCHS, Vital and Health Statistics, Series 2, No. 54, U.S. Department of Health, Education and Welfare, March 1973.

Tremblay, V., Singh, M.P. and Clavel, L., Methodology of the Labour Force Survey Re-interview Program, *Survey Methodology Journal*, Vol. 2, No. 1, Statistics Canada, 1976.

3.5 Controlled Experiments

Objective

To measure components of the mean square error.

Methodology

This method involves carrying out controlled experiments in an attempt to measure some or all components of the mean square response error. Depending on the objectives of the experiment (i.e., which response error parameters are being measured), these experiments may involve:

(i) re-interview under similar conditions (see method 3.4);

(ii) interpenetration at various levels (e.g., of enumerator assignments, of training assignments);

(iii) in-depth re-interview (see method 3.3);

(iv) interviewer variance study.

These may also involve the testing of alternative methodologies or questionnaires.

Evaluation

These studies are normally carried out during the development phase of censuses and large surveys but may also be used as an evaluation tool to assess and evaluate the types and sources of response errors. They require very careful fieldwork and therefore these studies tend to be expensive. Depending on the type of controls used they may also suffer from some of the disadvantages described for methods 3.3 and 3.4.

Reference

Brackstone, G.J., and Hill, C.J., The Estimation of Total Variance in the 1976 Census, *Survey Methodology Journal*, Vol. 2, No. 2, Statistics Canada, 1976.

3.6 Micro-comparison with External Data

Objective

To assess the level of response errors.

Methodology

This method involves the comparison of individual responses obtained from a census or a survey with similar data obtained from an independent source. This would usually be carried out on a sample (or subsample) basis. The independent sources used are generally administrative records or in some cases another survey.

These studies are usually carried out in three steps:

1. the selection of a sample of cases;
2. the linkage between the two frames for the selected units;
3. the comparison of the data for each matched unit.

The results are then weighted to the population level and data quality measures obtained.

Evaluation

This method is obviously restricted to cases where an independent source of data exists. In a sample survey, the sample involved in such a study would usually be selected from the actual survey itself. In the case of a census, the sample may be selected from either source, provided the external source is fairly complete.

There are often identification problems associated with linkage of selected units between the two frames.

Finally, the interpretation of the results of micro-comparisons will depend on the accuracy and comparability of the data obtained from the independent source. If the data can be regarded as accurate and comparable, then such studies would provide estimates of response bias, otherwise the results will be more limited in the sense that they may only provide indications of the types and sources of response errors.

References

Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match, Evaluation and Research Program, 1970 Census of Population and Housing, PHC(E)-11, U.S. Department of Commerce, Bureau of the Census, January 1975.

Gosselin, J.-F., DC-7 Evaluation of 1971 Reporting on Demographic and Social Characteristics, 1971 Census Evaluation Programme, Internal Report, Statistics Canada, April 1974.

4. NON-RESPONSE

Non-response occurs when information that is required for a survey unit is missing. This could happen because the unit cannot be contacted, because the unit is unable to provide the information requested, or because the unit refuses to co-operate in the survey. Data that were collected but were missing at the tabulation stage because the questionnaire was lost or because the answers could not be deciphered, and responses that are not usable because the wrong unit was surveyed, are also generally considered as non-responses. Further, data that are rejected because of poor quality may also be considered as contributing to non-response if they cannot be corrected.

Only units that belong to the population being studied can contribute to non-response. If there is no response to a filter question used to identify units in the population of interest it is difficult to determine whether subsequent non-response on the questionnaire is real non-response or whether the unit is simply out-of-scope.

The degree of non-response for a unit can vary from partial non-response, i.e., non-response to one or some of the questions, to total non-response to all the questions asked in the questionnaire or interview.

Whenever non-response occurs the total error of estimates obtained from the survey could increase because of non-response bias (to the extent that the non-respondents differ from respondents with respect to the characteristics being estimated) and because of the larger sampling variance that stems from the reduction in effective sample size. The magnitude of these errors depends on the proportion and type of non-respondents and on the attempts made in the estimation procedure to compensate for non-response.

Attention to the problems of non-response during survey design and careful control of the different stages of the survey operation will help in reducing the level of non-response.

There is considerable literature on methods of handling and reducing non-response in surveys and some examples are given in the general references at the end of this section. Despite application of such methods some non-response is inevitable in most censuses and surveys. Among the techniques available for reducing the impact of non-response on survey estimates are the following:

(i) call-backs or follow-ups of all or a sample of non-respondents;

(ii) substitution of non-respondent units in the field by interviewing "similar" units (use of this technique should be carefully controlled by the field operation managers);

(iii) substitution at the processing stage by using data from similar units, by using estimates from similar groups of units, by updating previous records using longitudinal imputation rules (in repeated periodic surveys), or, in the case of partial non-response, by using relationships established between the missing data and available data.

This section is concerned with methods of estimating and evaluating the impact of non-response errors on survey estimates after any non-response reduction or compensation techniques that form part of the survey system have been applied. The following types of method will be described:

4.1 Analysis of characteristics of non-respondents using internal survey data.

4.2 Comparison of characteristics inferred for non-respondents with data from external sources.

General References

Dalenius, T., The Problem of Not-at-homes, *Statistisk Tidskrift*, Vol. 4, 1955.

Deming, W.E., On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Response and the Bias of Non-response, *JASA*, Vol. 48, 1953.

Durbin, J., Non-response and Call-backs in Surveys, *Bulletin of the I.S.I.*, Vol. 34, Part 2, 1954.

Hansen, M.H. and Hurwitz, W.N., The Problem of Non-response in Sample Surveys, *JASA*, Vol. 41, 1946.

Kish, L and Hess, I., A Replacement Procedure for Reducing the Bias of Non-response, *The American Statistician*, Vol. 13, No. 4, 1959.

Politz, A. and Simmons, W., An Attempt to Get the "Not-at-homes" into the Sample without Call-backs, *JASA*, Vol. 44, 1949.

Politz, A. and Simmons, W., Note on An Attempt to Get "Not-at-homes" into the Sample without Call-backs, *JASA*, Vol. 45, 1950.

Zarkovich, S.S., *Quality of Statistical Data*, Rome: FAO, 1966.

4.1 Analysis of Characteristics of Non-respondents Using Internal Survey Data

Objective

To identify questions and types of survey unit with high non-response rates and to study the impact of methods used to adjust for non-response on survey estimates.

Methodology

Estimates of the non-response rates for individual questions or for sets of questions are obtained by taking the ratio of the number of units that should have responded to the question(s) but did not, to the number of units that should have responded. These are unbiased estimates of the non-response rates if the survey is a complete enumeration or if it is a sample survey with a self-weighting design. If the sample design is not self-weighting, unbiased estimates of non-response rates are obtained by taking the same ratio but using sums of weights rather than simple counts.

Non-response rates can be analysed in several ways:

(i) by comparing rates between questions (for partial non-response);

(ii) by comparing rates between different groups of units where the groups are defined in terms of characteristics known for all members of the survey frame (e.g., comparisons between provinces or between age-sex groups if age and sex are available in the survey frame), or equivalently by comparing profiles of respondents and non-respondents in terms of certain characteristics;

(iii) by studying the components of non-response by reason: not-at-home, unable to respond, refusal, data lost, etc.

The impact of different techniques for adjusting for non-response may be studied by comparing the survey estimates obtained using different assumptions about non-respondents. In surveys that involve several stages of field collection such as an initial contact attempt followed by several follow-up attempts on non-contacts, it may be possible to calculate survey estimates based on data from all units responding before a certain stage, and to compare these estimates for different stages. An extrapolation of these estimates to a complete response situation could be compared to the actual survey estimates produced.

Evaluation

This method is fairly cheap since it does not involve any follow-ups in the field. However, it does require that certain information be recorded during the survey operations (e.g., reasons for non-response, stage at which response was obtained).

The analysis of response rates by reason will provide information on where to take steps to reduce or control non-response in future surveys.

The analysis of response rates by characteristics is limited by the data that are available for all sample members either in the sampling frame or by observation during field collection. This analysis will help to identify sub-groups of the population requiring special attention in future surveys, and to identify sub-groups for whom current data may be less accurate.

The study of the effect of different assumptions about non-respondents will help in assessing the reliability of the estimates obtained using the particular assumption about non-respondents adopted in the survey.

4.2 Comparison of Characteristics Inferred for Non-respondents with Data from External Sources

Objective

To determine the characteristics of non-respondents and to measure non-response bias.

Methodology

This method involves determining from a source external to the survey the characteristics of units remaining as non-respondents at the end of the survey. Sources which might be used are:

(i) existing sets of administrative or other records;

(ii) data obtained in other surveys;

(iii) data obtained in a special follow-up study after the survey.

Often the cost of obtaining and matching external data will restrict this type of study to a sample of non-respondents in the survey.

The analysis of these data could either be at the individual survey unit level by comparing the values obtained from the external source with those that were imputed for non-respondents in the survey, or at the aggregate level by recalculating the survey estimates using the data for non-respondents obtained from the external source.

Evaluation

This is a fairly costly method since it involves detailed matching of non-respondents with an external source. It becomes even more expensive if variation (iii) that requires a special follow-up survey is undertaken.

It is rarely possible to obtain all data for a complete sample of non-respondents so one is often left with a small set of hard-core non-respondents about whom nothing is known. However, as long as this hard-core set is not too large, this method will supply useful estimates of the magnitude of non-response bias in survey estimates.

5. CODING

A coding operation in a survey or census will be defined as the operation where data on questionnaires or source documents are transformed into a format which is suitable for input to the data capture operation. This often involves the assignment of codes for "write-in" entries but may also be a fairly straightforward transcription operation.

It is recognized that errors may be introduced during a coding operation. The actual level of error would usually depend on factors such as:

(i) the complexity of the coding scheme used;

(ii) the amount of interpretation or judgement required to perform the coding;

(iii) training and experience of coders;

(iv) the quality of the information on which the coding is based.

In large coding operations quality control procedures are often used. These do not eliminate all errors but rather, ensure that the overall percentage of error is not greater than a pre-specified level. Records on rejection rate and error rate over time for each coder are usually kept with a view to identifying those who may require further training.

This section describes possible methods of evaluating coding errors. These are studied under the same error model used for response errors in Section 3. It should be noted that whenever quality control is used, it is considered to be part of the actual survey operation. Therefore the methods described here attempt to measure the actual outgoing quality, taking into account the effect of the quality control operation.

The following evaluation methods are presented:

5.1 Analysis of quality control data.

5.2 Measurement of coder variance using quality control data.

5.3 Expert re-coding.

5.4 Re-coding under the same general conditions.

5.1 Analysis of Quality Control Data

Objective

To assess the outgoing quality of the coding operation.

Methodology

When quality control techniques are used for coding operations, detailed records are usually kept giving error rates, number of batches rejected, etc. This method consists of analysing these results in order to assess the outgoing quality of the coding operation and the effectiveness of the quality control plan.

This would usually involve calculating and analysing parameters such as:

(i) the average ingoing quality;

(ii) the average outgoing quality;

(iii) error rates per coder/noter;

(iv) proportion of batches accepted/rejected.

These would normally be analysed over time.

Evaluation

This method provides a fairly inexpensive means of evaluating the overall quality of the coding operation when quality control techniques are used. Unless individual records are examined, it will not give information on the type and source of coding error.

5.2 Measurement of Coder Variance Using Quality Control Data

Objective

To estimate the coder variance and its components.

Methodology

This method involves the use of data obtained from the quality control operations to evaluate coding variation.

A sample of coders is selected and a comparison is made between the original coder's code and the noter's or quality control clerk's code for cases selected in the quality control sample. Using this as a basis, estimates of simple, correlated and total coder variance are derived.

Evaluation

This method is appropriate when the quality control system uses independent verification. If the noting and the coding procedures are carried out separately, the independence of the two operations is quite well preserved. However, the general conditions of the two operations may be different in some cases. For example, the coding and the noting may be done on different types of forms in different environments, etc. These factors would normally have a relatively minor effect on estimates. However, there may exist differentials in the average quality of the work of coders and noters. This may potentially be more serious than the above and may lead to biased estimates of coder variance depending on the magnitude of these differentials.

The actual cost of carrying out this project is fairly small since it only involves the gathering and analysis of readily available data.

5.3 Expert Re-coding

Objective

To estimate the coding bias.

Methodology

This study has two variations each involving some form of expert re-coding.

(i) A sample of cases for which coding was required is selected. Each selected case is then examined by an expert who determines whether or not the original code was correct, and re-codes each case found in error. These data are then used to obtain the required estimates.

(ii) A sample of cases for which coding was required is selected. The actual re-coding is carried out in two phases. Specially trained clerks first re-code each case in the sample (which may or may not be done independently). The new codes are then compared with the original survey code and all disagreements are examined by an expert to determine which code is in error. A sub-sample of the remaining cases is also examined to check the possibility of both the coders and re-coders making similar coding errors. The sample results are then weighted properly to derive estimates of coding bias.

Evaluation

Most expert re-coding studies of type (i) deal with a fairly small number of cases mainly due to limited expert resources. The number of errors identified by such a procedure will necessarily be small, which implies that estimates can only be derived for very broad categories. Method (ii) is essentially a means of minimizing the limitations of the first method. The initial re-coding can be done on quite a large number of cases and is essentially a means of identifying cases that are likely to be in error. This method, however, involves more clerical resources.

5.4 Re-coding under the Same General Conditions

Objective

To estimate the coder variance and some of its components.

Methodology

This involves the re-coding of a sample of cases under the same general conditions as the original coding operation, i.e., same training, instructions, etc. The re-coding would usually be carried out by a different coder without access to the original code to ensure independence.

The two codes are then compared for each selected case and estimates of coder variance and its components are derived.

Evaluation

Given that the same general conditions and independence of the two operations are preserved, this study should provide valid estimates of coder variance and its components. A detailed analysis of the results would also provide very useful information on the type and sources of coding errors.

Reference

Effect of Coders, Evaluation and Research Program, 1960 Censuses of Population and Housing, Series ER 60, No. 9, U.S. Department of Commerce, Bureau of the Census, January 1972.

6. DATA CAPTURE

The data capture operation in a census or survey consists of converting the data received on questionnaires (e.g., respondent answers) or coding forms to a machine readable format. This conversion can be accomplished through manual intervention (e.g., keypunch, key-edit, and/or type and scan) or by directly reading the data, optically, with equipment such as FOSDIC (Film Optical Sensing Device for Input to Computer) or OCR (Optical Character Recognition) reader. The captured data then pass through subsequent processing stages of editing, imputation, weighting and tabulation.

Errors can be introduced into the data using both kinds of data capture methods but the types of error introduced will be different. The overwhelming majority of errors introduced in manual intervention methods are operator errors with relatively few errors occurring due to machine malfunction, whereas the errors which usually occur in optically based methods are related to reading failures resulting from unrecognizable characters, substitutions (e.g., interpreting a 4 as a 9, etc.), or erroneously reading an uncoded circle or mark as a coded one. The error rates associated with key-punch, key-edit or type and scan operations depend on various factors such as the operators' training and experience, the complexity of alphanumeric data on the questionnaire or coding form, and the extent of inspection used. On the other hand the direct read (optical) error rates depend primarily on the quality of coding on the incoming documents.

In most surveys using key-punch and key-edit operations, quality control inspection by sample verification is used to ensure that the overall percentage of cards or records with errors is not greater than a prespecified level. Generally in sample verification, a random sample is selected from each lot of cards or records and is verified by an operator called a "verifier". A verified card or record is called a "defective" if it has either one or more errors in transcription from identification or data fields, or an error pertaining to the card or record itself (e.g., duplication or omission). If the number of defectives in the sample from a lot exceeds a certain acceptance number, the lot is rejected and completely verified. The sampling plan (i.e., sample size and acceptance number) is determined for a given lot size, operators' error rate per card or record, and the desired level of outgoing quality.

In the case of monthly or other periodic surveys using key-punch or key-edit, information on operators' error rates from previous months is used in devising sampling plans which ensure that the prespecified outgoing quality is continually met with the minimum resources. Smaller volume surveys using manual intervention methods for data capture usually incorporate complete verification which guarantees a high quality output.

Data capture errors associated with optical methods are generally controlled by any one or combination of the following:

(i) quality control of printed forms to ensure their machine readability;

(ii) by the use of on line editing and correction (where this facility exists);

(iii) by reading specially prepared dummy documents, marks or data to ensure that the equipment is functioning correctly;

(iv) by the use of check digits, hash totals and redundancy checks on key fields to identify possible substitutions during data capture.

It should be noted that whenever quality control is used, it is considered to be part of the actual survey operation. That is, the evaluation methods described here attempt to measure the actual outgoing quality, taking into account the effect of the quality control operation.

A minimal evaluation of data capture errors for a survey operation that has a quality control inspection program would consist of evaluating the quality of inspected lots by using the error rate data available and the various parameters associated with the sampling plans. Other special studies can also be undertaken to evaluate data capture errors. These studies would be necessary for surveys that utilize the direct read data capture methodology or those that use key-punch, key-edit or type and scan and do not incorporate any type of verification. The evaluation methods for data capture errors can thus be identified as follows:

6.1 Analysis of quality control data.

6.2 Verification using error free data.

General References

Dodge, H.F. and Romig, H.D., *Sampling Inspection Tables*, Second Edition, New York: John Wiley and Sons, Inc., 1959.

Duncan, A.J., *Quality Control and Industrial Statistics*, Third Edition, Chapter XVI, Rectifying Inspection for Lot-by-lot Sampling, Homewood, Illinois: Richard D. Irwin, Inc., 1965.

Weatherill, G.B., *Sampling Inspection and Quality Control*, London: Methuen, 1969.

6.1 Analysis of Quality Control Data

Objective

To estimate the average outgoing quality (AOQ) of the data capture operation.

Methodology

When quality control techniques are used in data capture, detailed records are usually kept giving lot size, sample size, acceptance number, average error rate, number of batches rejected, etc. This method consists of analysing these data to estimate the average percentage defectives in the sample verified lots.

Evaluation

The estimate of AOQ obtained by this method is appropriate only if there are no verification errors (i.e., the error detection, correction and recording functions are performed according to instruction). This method provides a fairly inexpensive means of evaluating the overall quality of the capture operation.

The data on error rates can be used to alter the level of inspection by changing sampling plans. The detailed error rates by type and by operator can also give some insight into the error generating mechanism of the data capture operation and this can be used to improve the training of operators.

References

Ghangurde, P.D., Review of Quality Control Methods Used for Data Capture in Key-Punch Section, Technical Report, Statistics Canada, September 1974.

Minton, G., Inspection and Correction Error in Data Processing, *JASA*, Vol. 65, 1969.

6.2 Verification Using Error Free Data

Objective

To estimate the data capture error rate by independent verification.

Methodology

A random sample of records is selected and the data are then recaptured independently in a more controlled and error free environment than the one used for the original data capture. The recaptured sample data are then matched against the originally captured data. The errors observed in the original records are used to estimate error rates.

Evaluation

This is generally the only type of data capture evaluation that can be applied to surveys that utilize the direct read data capture methodology but it can also be used to evaluate error rates for surveys using the key-punch, key-edit or type and scan processes which do not incorporate any verification.

This method will provide an accurate estimate of the data capture error rate to the extent that errors in the recaptured data can be eliminated or minimized. The estimates of the error rates could be used to determine the acceptance sampling plans for future surveys with the same questions and codes, to improve questionnaire or document design and in data capture training.

Reference

March, M.J., QD-4 Data Capture Evaluation – 1974 Census Test, Internal Report, Statistics Canada November 1975.

7. EDIT AND IMPUTATION

As seen in earlier sections, errors can enter the data in a variety of ways during the survey operation and various checks and controls to screen for such errors can be built into many stages of a survey (e.g., field edit, quality control of coding or data capture). In spite of these checks and controls the data reaching the final stages will still contain some errors. Editing and imputation is the final filter in the survey process and attempts to catch the more obvious errors that have escaped the quality checks and controls incorporated into the earlier stages of collection and processing.

The edit procedure usually consists of:

(i) checking each field of every record to ascertain whether it contains a valid code or entry;

(ii) checking codes or entries in certain predetermined combinations of field to ascertain whether codes or entries are consistent with one another.

These checks are formally written in terms of edit rules or decision rules. The edit rules are formed using many types of information such as known constraints on values of variables in fields and historical experience about the range of values that variables are likely to take in the particular subject matter area.

The imputation procedure consists of changing values in some of the fields in records which failed the edit rules with a view to ensuring that the resultant data records satisfy all edit rules.

There are a number of ways in which the edit and imputation procedure of a survey can be done depending on time, budget and computer or personnel resources available. The edit rules and imputation procedure can be written in terms of flow charts or decision tables and both editing and imputation can be done manually, though for large surveys the operation will generally be automated. A systematic approach to automatic edit and imputation has been used in Statistics Canada for Census processing and some other major surveys. The computer system developed for this purpose, called CAN-EDIT, uses a particular method of hot-deck imputation for qualitative data.

The general approach to the evaluation of edit and imputation is somewhat different from that taken in previous sections. This stems from the fact that edit and imputation essentially attempts to reduce gross errors, and to remedy some of the defects introduced at earlier stages of the survey process.

An edit and imputation procedure may "correct" the data too much or too little depending on the actual set of edits and on the imputation method used. For example, edits are often constructed in such a way as to screen out very unlikely cases as opposed to impossible cases. In such situations cases failing edit will not only indicate errors introduced at earlier stages but may also include possible extreme values or outliers.

Similarly, when a particular imputation method is adopted, one is implicitly making an assumption about the nature of the errors in the data. For instance, when a hot-deck approach is used (constrained or unconstrained), one assumes that errors are essentially randomly distributed (in all survey units or within subgroups) which may or may not be the case depending on the specific application.

There is therefore a need generally to evaluate the reasonableness of the assumptions underlying an edit and imputation procedure with a view to measuring its effect on the overall quality of survey data. This however is often very difficult to achieve without the use of external evidence.

As a first step towards the evaluation of edit and imputation, one may analyse the frequency of edit failures and the net effect of edit and imputation on overall estimates. If particular problem areas are identified, this analysis may be supplemented by detailed investigation of a sample of edit failures. The methods described are:

7.1 Analysis of edit failure rates.

7.2 Analysis of the effect of edit and imputation.

7.3 Follow-up of a sample of edit failures.

General References

Fellegi, I.P. and Holt, D., A Systematic Approach to Automatic Edit and Imputation, *JASA*, Vol. 71, 1976.

Freund, R.J. and Hartley, H.O., A Procedure for Automatic Data Editing, *JASA*, Vol. 62, 1967.

Naus, J.I., *Data Quality Control and Editing*, New York: Marcel Dekker, Inc., 1975.

7.1 Analysis of Edit Failure Rates

Objective

To assess quality of data and appropriateness of edit rules.

Methodology

This method consists of analysing edit failure rates for each edit rule for each field in the record. The records of any manual or computerized edit and imputation system can give information required for such an analysis. The particular fields which show excessive failure rates and particular codes in these fields which cause these failures can be detected from the analysis. These particular fields could be more error prone due to questionnaire design or coding procedures, or the analysis could lead to detection of impractical edit rules.

Evaluation

The method is a fairly cheap and practical way to compare error proneness of various fields and to assess the impact of edit rules on the survey data. However, it can only provide general information on the probable cause of high edit failures.

7.2 Analysis of the Effect of Edit and Imputation

Objective

To analyse the effect of the edit and imputation procedure on the data passing through the edit and imputation system.

Methodology

The records which have failed one or more edit rules in one or more fields are input to an imputation system which imputes valid codes in these fields. The frequency distributions of variables in various fields for the pre- and post-imputation stages can be compared to evaluate the net effect of edit and imputation on survey estimates. This may be done graphically or by some goodness of fit tests. In cases where hot-deck imputation is used the variability due to the imputation procedure may be assessed by repeating the imputation procedure on a sample of edit failed records.

Evaluation

This method provides an explicit measurement of the effect of edit and imputation on the data. However, unless external evidence is available, it does not provide direct information on whether or not the changes in the data resulted in improved data quality.

7.3 Follow-up of a Sample of Edit Failure

Objective

To identify the sources of errors detected by the edit procedure.

To assess the validity of the assumptions underlying the edit and imputation procedure and their impact on the quality of data.

Methodology

A random sample of edit failed records is selected and the original survey schedules or questionnaires corresponding to these records are checked to determine whether the edit failure was due to a processing error (coding or data capture) or to a potential response error.

In some cases, it may be possible to detect the source or type of response error by using other information on the questionnaires. If this is not possible, this method may be extended to contacting the respondent to obtain a correct response or to verify that the reported information was correct in its original form.

Evaluation

Although this method is listed in the section on Edit and Imputation, it may in fact provide information about all the previous four sources of error (i.e., response, non-response, coding, data capture). It could be combined with methods 3.1 and 4.2.

The follow-up to the questionnaire provides a fairly cheap method of assessing the source and type of errors detected by the edit and imputation procedure. It may also indicate some problems in questionnaire design, coding and data capture procedures.

However, edit failures are quite often due to response errors or non-response, which cannot be explained by the examination of the original survey questionnaire. In such situations, a follow-up with the respondent will generally provide the only means of testing the validity of the edit procedures and the reasonableness of the imputation procedure. This would also provide very useful data that can be used in developing models for dealing with conflicts and imputation for non-response. However, this follow-up is likely to be costly and may involve problems such as response burden and recall bias.

8. SAMPLING AND ESTIMATION

Sampling error occurs whenever survey results are based on a sample of units from a survey frame. It is fundamentally different from the types of error described previously since it is present by design, can be controlled by the sampling technique, and is usually measurable from the sample data itself in the case of random sampling.

The sampling deviation is defined as the difference between the estimate of a population parameter obtained from a sample and the value of that parameter for the frame population obtained using essentially the same survey and measurement techniques. This difference is a function of the sample size, the variation within the population, and the sampling and estimation methods used in arriving at the estimate. Sampling error is usually assessed in terms of sampling variance which is the average value, over all possible samples, of the square of the sampling deviation. Obviously there is no sampling error in complete enumeration surveys.

Control of sampling error is achieved by increasing the sample size and/or using the information available in the sampling frame in the best possible manner to evolve efficient sampling techniques (using appropriate stratification, allocation, clustering, probabilities of selection, etc.) and improved estimation procedures. These are of course subject to constraints of cost and feasibility. Administrative control of the operational steps in selecting the sample is also essential to ensure that the sample selection has been correctly carried out.

The following methods of estimating and evaluating sampling errors are discussed in this section:

8.1 Comparison of the distribution of units in specified classes as estimated from a sample survey with that from a complete enumeration survey.

8.2 Estimation of sampling error using standard formulae.

8.3 Estimation of sampling error using replicated samples.

8.4 Estimation of sampling error using error charts of models.

8.5 Evaluation of the efficiency of the sample design and estimation procedures.

8.1 Comparison of the Distribution of Units in Specified Classes as Estimated from a Sample Survey with that from a Complete Enumeration Survey

Objective

To test whether the sample represents the population with respect to various distributions thus providing an overall indication of sampling error.

Methodology

The estimated distribution of units in the specified classes is tested against that for the complete enumeration survey using a χ^2 test. For a self-weighting design the distribution of the sampled units is compared with that of the complete enumeration survey.

Evaluation

This method is often used to test if the sample represents the sampling frame from which it was selected with respect to that distribution. Unless the data on the characteristics used to classify the units into the specified classes are collected using the same concepts and definitions in both the sample survey and complete enumeration survey, problems of comparability may arise.

An example of this method is the comparison of the age-sex distribution of units in the 1 in 3 sample from the Census as compared with that in the complete Census. For such large samples, rejection of the hypothesis that the sample has the same distribution may indicate the presence of errors in the selection of the sample or non-sampling errors in the classification.

8.2 Estimation of Sampling Error Using Standard Formulae

Objective

To obtain an estimate of the sampling error.

Methodology

For most sample designs and estimation procedures used, it is possible to derive formulae for the standard unbiased (or approximately unbiased) estimates of the sampling error from the observations on the ultimate sampled units. This method consists of using these formulae.

Evaluation

These estimates of sampling error will also include certain components of response error. Except for simple sampling and estimation procedures, these formulae could be complicated (e.g., in complex multiphase, multistage, clustered designs with varying probabilities of selection at each stage). Calculating the estimate of the sampling error for each estimate that is output for a survey could be a costly and time consuming effort, although not necessarily large relative to the total cost of the survey. When computers are used, much of the cost and effort will be in planning the calculation and developing the necessary computer systems rather that in their actual execution.

8.3 Estimation of Sampling Error Using Replicated Samples

Objective

To obtain an estimate of the sampling error.

Methodology

This method involves the use of techniques such as replicated samples built into the sample design, pseudo-replicated samples that are superimposed on the design, balanced repeated replicates or jack-knife techniques. These techniques provide estimates of sampling errors, however complicated the sample design and estimation procedures used may in fact be.

Evaluation

These methods are considerably simpler than method 8.2 and can be applied to all sample designs and estimation procedures. If the replication covers interviewing and processing operations as well, the techniques provide estimates of the sampling errors and part of the non-sampling errors (response and processing variance). Some of these estimates are biased and they tend to have larger sampling variances than the direct estimates discussed in method 8.2. But for some complex samples these may be the only feasible methods for estimating sampling errors.

However, building the replicates into the sample design involves some loss of efficiency in the estimation of the parameters of interest (i.e., an increase in their sampling error).

The cost of these methods at the field and processing stages depends on the level at which replication is done. At the field stage, if it is desired to allot each replicate to a different interviewer or interviewing team (so as to measure some of the non-sampling error) this technique involves more travel cost, depending of course on the level at which such replication is done. At the processing stage, for most designs, estimation of sampling error using method 8.2 is costlier than this method since the former needs computation of the variance components from each of the sampling stages.

8.4 Estimation of Sampling Error Using Error Charts or Models

Objective

To obtain an estimate of the sampling error.

Methodology

Based on the pattern of sampling errors calculated for various estimates for different effective sample sizes using method 8.2 or 8.3 either from the survey itself, from past surveys (if the survey is one of a continuing series), or from other similar surveys, it is often possible to "type" estimates and build approximate models (often based on design effects) or to draw nomograms or charts that allow one to calculate or read off the approximate sampling error for any given estimate.

Evaluation

This method is the cheapest, once the model or chart has been developed. Considerable investigation and checking may be involved in developing these models or charts. Their usefulness should be kept under continuous surveillance to check if the patterns have changed. These methods are particularly applicable and useful in continuing surveys and in multipurpose surveys with large numbers of variables and tabulations.

Reference

Gonzalez, M.E., Ogus, J.L., Shapiro, G. and Tepping, B.J., Standards for Discussion and Presentation of Errors in Survey and Census Data, *JASA*, Vol. 70, Part 2, 1975.

8.5 Evaluation of the Efficiency of the Sample Design and Estimation Procedures

Objective

To compare different sampling techniques and estimation procedures in an attempt to evaluate the methods used and find the best combination for a particular sample survey.

Methodology

Without going into details, this involves studying the different steps in the development of a sample design and estimation procedure and comparing alternative methods that could be used at each step. Some examples of the steps are: stratification, allocation, number of stages of selection, clustering and probabilities of selection of sampling units, use of self-weighting designs, use of unweighted estimators and of different sets of weights for the sample observations. Studies on components of variance, efficiency of stratification and design effects, estimates of biases and mean square errors of alternative estimators are some of the techniques that could be used.

Evaluation

This detailed evaluation is based on data available from past surveys and is generally not possible at the start of a new survey except to the extent that studies of other similar surveys may help. These methods, where possible, usually precede the development of a sample design but they also could be part of the ongoing evaluation and development of any continuing survey.

References

Fellegi, I.P., Gray, G.B., and Platek, R., The Redesign of the Canadian Labour Force Survey, *Proceedings of the Social Statistics Section, American Statistical Association*, 1964.

Fellegi, I.P., and Gray, G.B., Sampling Errors in Periodic Surveys, *Proceedings of the Social Statistics Section, American Statistical Association*, 1971.

Gray, G.B., Components of Variance Model in Multi-stage Stratified Samples, *Survey Methodology Journal*, Vol. 1, No. 1, Statistics Canada, 1975.

Gray, G.B., Stratification Index – Methodology and Analysis, *Survey Methodology Journal*, Vol. 2, No. 2, Statistics Canada, 1976.

Lawes, M., A Comparison of Some Binomial Factors for the Labour Force Survey, *Survey Methodology Journal*, Vol. 1, No. 1, Statistics Canada, 1975.

Tepping, B.J., Variance Estimation in Complex Surveys, *Proceedings of the Social Statistics Section, American Statistical Association*, 1968.