ERRORS ASSOCIATED WITH THE FRAME

Judith T. Lessler, Research Triangle Institute

This paper summarizes some of the results of a review of the errors associated with the frame for a survey. All results in the original paper are not covered.

One of the goals of the taxonomy project* is to provide a listing of the terminology used to describe frame errors. As an illustration of the variation in terminology consider the definition of a frame.

1. Definitions of Frames

The definitions and descriptons of frames varied as to the number of concepts and/or operations the definition/description attempted to encompass. The names given to the concepts also varied. The concepts that the frame definitions attempted to encompass include the following:

1. The target population is a finite collection of N identifiable elements.
2. Sampling is done on some set of units, but this set is not necessarily the target population.
3. Some mechanism must exist for linking the target population and the set which is sampled.
4. In order to be able to collect information from elements, we must be able to locate the elements and delimit them from each other.
5. There is more than one type of linkage that can exist between the target elements and the sampled set. These linkages determine the type of sample design and estimation procedures that can be used in a survey.
6. Sample designs and estimation procedures vary as to their requirements for information about the population elements.

For example, two simple definitions of a frame are:

> Frame: The list of sampling units which make up the population. (Cochran, 1978)

> Frame: The list of units. (Zarkovich, 1966)

These definitions incorporate the concept of a finite collection of identifiable elements.

Other definitions include this concept, as well as, the idea that the sampling units are not necessarily equivalent to the population elements, such as,

> Frame: A list of sampling units. (Scheaffer, Mendenhall, and Olt, 1979)

> Sampling units are nonoverlapping collections of elements from the population.

Hansen, Huvwitz and Jabine (1963) add the third idea, i.e., that of a linkage mechanism by defining a list of sampling units, a target population of reporting units and rules of association which link the two sets.

> Rules of association: Those rules which allow establishment of a linkage between a selection of listed units with known probabilities to a selection of reporting units with known probabilities. (Hansen, Hurwitz, and Jabine, 1963).

They do not call this a frame, however.

In a roughly equivalent manner, Warwick and Lininger (1975) also consider these three concepts and do use them to define a frame. They first speak of lists:

> List: A list is an inventory of the units in a population or a subpopulation with a direct one-to-one correspondence between each item listed and the unit it represents.
>
> If the listing units are population units, it is "an elements list....the listing units may be groups of elements found in convenient, identifiable, and unambigious groups. -- in such cases the inventory would consist of a cluster list." (Warwick and Lininger, 1975)

Their definition of a frame then goes beyond these definitions of lists.

> Frame: A sampling frame consists of the materials and procedues used fully to account for the population when complete element lists are not available... It is basically the operational procedures and material used to account for the population in drawing the sample. (Warwick and Lininger, 1975).

Moser and Kalton (1971) also in their discription of a frame embody the notion that the frame should not be merely a list of units but "must contain sufficient details to ensure that each unit is identified with certainity; and for another it must contain the information required to enable the unit to be located." Szameitat and Schaffer (1963) define a frame using this same idea:

> Frame: All material which describes the components of the target population (or an adequate part of that population) in such a way that it is possible to determine in the course of the survey the individual components and to delimit them from other components.

Recognizing that there are several types of linkage that can exist between frame units and target elements, Dalenius (1974) first defines a population of objects {0}, and the frame is then defined as

> Frame: The frame is any material, device, etc., which is used to provide observational access to the population {0}.

A special class of frames is defined which consists of "finite sets {U} of units of same kind."

Three types of linkages between the frame units and the population objects are distinguished:

---

* RTI project 255U-1791 Taxonomy of Survey Errors, supported by Grant No. SOC-7804597 from the National Science Foundation.

One-to-one: Each frame unit is associated with one and only one population object.
Many-to-one: Several frame units are associated with a single population object.
One-to-many: A frame unit is associated with more than one object in the population.

Jessen (1968) gives a very extensive discussion of frames identifying several types of frames and six types of association between the frame units and the population elements. These six types of associations are discriminated by the types of linkage and by whether or not the number of linkages can be determined for elements in the sample.

Finally, several definitions of frames recognized that certain types of sample designs and estimation procedures require more information than mere identification of elements. Examples are stratified sampling, probability proportional to size sampling, and ratio estimation. As an example, Szameitat and Schaffer (1963) define a preferred frame as:

Preferred Frame: Of particular usefulness for a sample survey is a frame which not only provides a description of the individual components but also supplies additional information, such as size of the components or their inclusion into specific parts of the target population. If there are seveal frames of the same level available for a target population, that one will generally be used which contains the most exact information and can, at the same time, be easily applied.

After reading all of these definitions, I attempted the following definition which hopefully includes all of the ideas in the previous definitions:

Frame: The frame consists of the materials, procedures, and devices which identify, delimit, and allow access to the elements of the target population. The frame is composed of a finite set of units to which the probability sampling scheme is applied. Rules or mechanisms for linking the frame units to the population elements are an intergral part of a frame. The frame also includes any auxiliary information (measures of size, demographic information) that is used for (1) special sampling techniqus, such as, stratification and probability proportional to size sample selections; or for (2) special estimation techniques, such as, ratio or regression estimation. (Lessler, 1980).

2. Frame Errors

A wide variety of terminology for refering to frame errors was encountered in the review. The variety of errors associated with frames can be classified into 6 different types. They are:
(1) Population elements missing from the frame;
(2) Nonpopulation elements included in the frame;
(3) Population elements associated with the frame more than once;
(4) Failure to recognize that the frame units are clusters of elements;
(5) Incorrect auxillary information; and
(6) Information insufficient to locate target elements.

Some of the terms used to talk about this type of error are given in the Preliminary Taxonomy. They are discussed in Lessler (1979).

3. Models and Procedures for Measuring the Extent and Impact of Frame Errors. Special Procedures for Conducting Surveys in the Presence of Frame Errors.

In this section, we will review some of the ways for assessing frame errors. A distinction is made between measures of the extent of the error and measures of the impact of the error on survey statistics. All of the literature reviewed considered the impact of frame errors on two statistics, estimates of a total and a mean. Procedures for dealing with each of the 6 types of error are given in the main paper. Only one is discussed here.

3.1 Population Elements Missing From the Frame

This is probably the most serious frame error because it cannot be detected by examining either the frame or the sample. Two measures of the extent of undercoverage are to:
(1) the number of elements missing from the frame, and
(2) the proportion of elements missing from the frame.

Several measures of the impact of the missing elements were encountered. They are:
(1) net bias,
(2) relative bias,
(3) mean-square error, and
(4) the ratio of squared bias to mean square error.

For example, Kiranandana (1976) expresses the relative bias of the estimate of a population total and a population mean in terms of the proportion of elements missing from the frame and the ratio of the mean of the missing elements to that of the associated elements. If

W = proportion of elements not included in the frame,

and

$$r = \frac{\text{mean of missing elements}}{\text{mean of included elements}} ;$$

then

$$\text{relative bias (TOTAL)} = \frac{-Wr}{rW + [1-W]} ;$$

and

$$\text{relative bias (MEAN)} = \frac{W(1-r)}{rW + (1-W)} .$$

The key to measuring the effect of nonassociated target elements on estimates of the mean and total is to get an estimate of the total for the nonassociated elements. The extent of undercoverage can be measured by estimating the number of target elements missing from the frame.

Essentially, two different types of procedures were found for estimating the magnitude of the various error meaures. They are:
(1) External data procedures, and
(2) Data collection procedures.

In external data procedures, the results of the survey are compared with some data from outside

the survey. Examples are, direct comparison of two figures and the in-flow, out-flow technique in which the estimate is expected to equal the sums and differences of various quanities.

Another technique for detecting undercoverage is the reverse record check. In a reverse record check, target elements that should be included in the frame are identified from a set of records. The frame is then checked to see if they are indeed included. It is interesting to note that the use of this method can require a certain type of sample design. If this method is used in a multistage sample survey, it requires the use of compact clusters (Zarkovich and Krane, 1965) in which all elements that are clustered within a certain area, time period, etc. are included in the sample. This is because an element identified in the records must be slated to be in the sample, or we would not know whether it was missing from the survey becuase of noncoverage or because of not being selected in the sample.

Data collection procedures include the use of an independent relisting and the linking or predecessor/sucessor method.

Several methods exist for conducting a survey with an incomplete frame. They include:

(1) Redefining the target population to include only those elements associated with the frame;
(2) Using a linking procedure in which rules are made for linking non-associated elements to certain associated elements; and
(3) The use of multiple frames, either, overlapping or non-overlapping.

4. Preliminary Taxonomy for Frame Errors

To form the taxonomy, errors associated with the frame were first classified by type. Within each type the following is considered:

(1) the terminology employed,
(2) the models and measures for the extent of the error and the impact of the error,
(3) the procedures for estimating the values of the error measures, and
(4) methods for conducting surveys in the presence of error.

PRELIMINARY TAXONOMY

I. Frames

1. Definitions of Frames:

Frame: The list of sampling units which make up the population. (Cochran, 1978)

Frame: The list of units. (Zarkovich, 1966)

Frame: A list of sampling units. (Scheaffer, Mendenhall, and Olt, 1979)

List: When the elements of the population have been numbered or otherwise identified, the population together with its identification system is called a list. (Hansen, Hurwitz, and Madow, 1953)

Population defined by: A finite population is any well defined set or class containing a finite number of elements.

List: Nongeographically defined units for drawing a sample. (Hansen, Hurwitz, and Jabine, 1963)

Frame: A list of all the sampling units in the population. This list provides the basis for the selection and identification of units in the sample. (Sukhatme and Sukhatme, 1970)

Sampling units: The population is subdivided into a finite number of distinct and identifible units called sampling units.

Frame: A sampling frame consists of the materials and procedures used fully to account for the population when complete element lists are not available... It is basically the operational procedures and material used to account for the population in drawing the sample. (Warwick and Lininger, 1975).

List: A list is an inventory of the units in a population or a sub-population with a direct one-to-one correspondence between each item listed and the unit it represents.

If the listing units are population units, it is "an elements list." The listing units may be groups of elements found in convenient, identifiable, and unambigious groups. - - in such cases the inventory would consist of a cluster list. (Warwick and Lininger, 1975)

List: A finite collection of units labled $\ell_1, \ell_2, \ldots \ell_N$. (Hansen, Hurwitz, and Jabine, 1963)

Target Population: A finite collection of reporting units labeled $t_1, t_2, \ldots, t_N$. (Hansen, Hurwitz and Jabine, 1963)

Rules of Association: Those rules which allow establishment of a linkage between a selection of listed units with know probabilities to a selection of reporting units with known probabilities. (Hansen, Hurwitz, and Jabine, 1963).

Frame: The lists, indexes, maps or other population records from which the sample can be selected at each sampling stage. (Moser and Kalton, 1971)

Frame: Physical lists and procedures that can account for all the sampling units without the physical effort of actually listing them. (Kish, 1965)

Frame: The Frame consists of previously available descriptions of the objects or material related to the physical field in the form of maps, lists, directories, etc. from which sampling units maybe constructed and a set of sampling units selected; and also information on communications, transport, etc. which may be of value in improving the design for the choice of sampling units and in the formation of strata. (United Nations, 1964).

Frame: All material which describes the components of the target population (or an adequate part of that population) in such a way that it is possible to determine in the course of the survey the individual components and to delimint them from other components. (Szameitat and Schaffer, 1963)

Preferred Frame: Of particular usefulness for a sample survey is a frame which not only provides a description of the individual components but also supplies additional information, such as size of the

components or their inclusion into specific parts of the target population. If there are several frames of the same level available for a target population, that one will generally be used which contains the most exact information and can, at the same time, be easily applied. (Szameitat and Schaffer, 1963).

Frame: The frame is any material, device, etc., which is used to provide observational access to the population {0} (Dalenius, 1974).

Frame: Serial List/ A serially numbered listing of each element in the universe. (Jessen, 1978).

Frame type: Mixable objects/ A set of physical units each of which is associated with a population element and which may be mixed to simulate the process of randomization. (Jessen, 1978).

Frame type: Count Frame/ An ordered set of physical units in which a particular unit may be identified by counting through the set. (Jessen, 1978).

Frame type: Cluster Frame/ A listing of units in which each unit is associated with more than one population element. (Jessen, 1978).

Frame type: Area Frame/ A structure in which the sampling units are areas. Two subtypes are

a. grid frames in which the area contains a cluster of elements such as households, businesses, etc.; and
b. plot frames in which the elements are the areas.

Frame type: Noninteger Frame/ A frame in which the total number of population elements N associated with the frame is not pre-determined but which allows selection of a sample in which the units have equal probabilities of selection. These probabilities of selection are unknown, however. The numbers used in random digit dialing are an example of a noninteger frame. (Jessen, 1978).

Frame relationships:

The six types of association are:

(1) One-to-one correspondence.
(2) Simple cluster case, one-to-many.
(3) Many-to-one correspondence, which can be determined for the sample.
(4) Many-to-one correspondence, which can not be determined for the sample.
(5) Many-to-many correspondence, which can be determined for the sample.
(6) Many-to-many correspondence, which can not be determined for the sample. (Jessen, 1978).

2. Types of Frame Errors:

2.1 Population Elements Missing from the Frame

2.1.1 Terminology

NOT GIVEN DUE TO SPACE LIMITATIONS

2.1.2 Measures of Error

2.1.2.1 Extent of Undercoverage

a. The number of population elements missing from the frame
b. The proportion of the population elements excluded from the frame

2.1.1.1 Impact of Undercoverage on Estimates of Population Mean and Total

a. Net bias
b. Relative bias
c. Mean Square Error
d. Ratio of Squared bias to mean square error.

2.1.3 Procedures for Estimating Magnitude of Error Measures

2.1.3.1 Extent

2.1.3.1.1 External Data Procedures

a. Comparison with an external figure
b. In-flow, out-flow - Overall
c. In-flow, out-flow - Domains
d. Reverse record check

2.1.3.1.2 Data Collection Procedures

a. Quality Check
b. Linking procedure-Predecessor/ Sucessor Method

2.1.3.2 Impact

2.1.3.2.1 External Data Procedures

a. Comparison with an external figure
b. In-flow, out-flow - Overall
c. In-flow, out-flow - Domains
d. Reverse record check

2.1.3.2.2 Data Collection Procedures

a. Quality Check
b. Linking procedure - Predecessor/ Sucessor Method.

2.1.4 Methods for Conducting Surveys Using Incomplete Frames

2.1.4.1 Redefine Target Population
2.1.4.2 Linking Procedure
2.1.4.3 Multiple Frames

2.1.4.3.1 Novnoverlapping - Supplemental or Complementary Frames

2.1.4.3.2 Overlapping

a. Domain membership and domain sizes predetermined before sampling
b. Domain membership not predetermined, domain sizes known
c. Domain membership not predetermined, domain sizes not known.

2.2 Nonpopulation Elements Associated with Frame

2.2.1 Terminology

NOT GIVEN DUE TO SPACE LIMITATIONS

2.2.2 Measures of Error

2.2.2.1 Extent of Overcoverage

a. The number of nonpopulation elements associated with the frame.
b. The proportion of elements associated with the frame that are nontarget elements

2.2.2.2 Impact of Overcoverage on Estimates of Population Mean and Total

a. Net bias
b. Relative bias
c. Mean square error
d. Ratio of squared bias to mean square error

e. Relative efficiency of domain estimates to efficiency for a frame with no overcoverage.

2.2.3 Procedures for Measuring Magnitude of Error Measures

2.2.3.1 Nonpopulation members can be recognized from usual Sample Data

a. Domain estimate of number of nontarget elements
b. Domain estimates of target and nontarget population totals and means

2.2.3.2 Usual sample data does not reveal nontarget elements

2.2.3.2.1 Extent

2.2.3.2.1.1 Quality check approaches

a. Direct method
b. Indirect method
c. Sampling method

2.2.3.2.1.2 External data approaches

a. Comparison with an external figure
b. In-flow, out-flow - Overall
c. In-flow, out-flow - Domain

2.2.3.2.2 Impact

2.2.3.2.2.1 Quality check approaches

a. Domain analysis of already collected sample data using results on target/nontarget membership identified in 2.2.3.2.1.1 a, b, and c.

2.2.4 Methods for Conducting Surveys Using Frames With Overcoverage

2.2.4.1 Nonpopulation Elements Identified from Sample Data

a. Estimation using Theory of Domain Estimation

2.3 Simutaneous Investigation of Under and Overcoverage

2.3.1 Terminology

NOT GIVEN DUE TO SPACE LIMITATIONS

2.3.2 Measures of Error

2.3.2.1 Extent

a. Gross error of listing
b. Net error of listing
c. Percentage accurate listings

2.3.2.2 Impact of Coverage Errors on Estimates of Totals and Means

a. Net bias

2.3.3 Procedures for Measuring Magnitude of Error Measures

2.3.3.1 Quality Check Procedures

a. Single stage compact cluster design
b. Multi-stage compact cluster designs

2.3.4 Methods for Conducting Surveys Using a Frame with Under and Overcoverage

NONE

2.4 Population Elements Associated with the Frame More than Once

2.4.1 Terminology

NOT GIVEN DUE TO SPACE LIMITATIONS

2.4.2 Measures of Error

2.4.2.1 Extent

a. Number of population elements with multiplicity γ
b. Average multiplicity
c. Sum of multiplicities.

2.4.2.2 Impact of Multiplicity on Estimates of Totals and Means

a. Bias
b. Relative bias
c. Mean square error

2.4.3 Procedures for Measuring Magnitude of Error Measures

2.4.3.1 Extent

a. Estimate of number of population elements with two listings when maximum multiplicity = 2.

2.4.3.2 Impact

NONE

2.4.4 Methods for Conducting Surveys Using Frames with Multiplicity

2.4.4.1 Elimination of Multiplicity

a. Removal of by sorting frame units
b. Definition of a unique counting rule
c. Redefinition of target population

2.4.4.2 Adjusting for Multiplicity

2.4.4.2.1 Sampling Procedures

a. Additional stage of sampling at unit level
b. Additional stage of sampling at group level

2.4.4.2.2 Weighting procedures

a. Weighting by inverse of multiplicies
b. Weighting by inverse of multiplicaties under alternate counting rules
c. Weighting by probability of selection--requires matching within sample
d. Priority counting rules and weighting by probability of being in sample -requires matching within sample

2.5 Frames with Unrecognized Clustering

2.5.1 Terminology

NONE

2.5.2 Measures of Error

NONE

2.5.3 Procedures for Estimating Magnitude of Error Measures

NONE

2.5.4 Methods for Conducting Surveys Using Frames with Unrecognized Clustering

a. Reweight by cluster size

2.6 Incorrect Auxiliary Information

2.6.1 Terminology

NOT GIVEN DUE TO SPACE LIMITATIONS

2.6.2 Measures of Error

2.6.2.1 Extent

a. Number of elements incorrectly assigned to a group

2.6.2.2 Impact

a. Relative efficiency of domain estimates taken to estimates in which domains are strata.

2.6.4 Methods for Conducting Surveys with Incorrect Auxiliary Information

NONE

2.7 Frames with Nonaccessible Population Elements

2.7.1 Terminology
NOT GIVEN DUE TO SPACE LIMITATIONS

2.7.2 Error Measures
Same as 2.1.2

2.7.3 Procedures for Estimating Magnitude of Error Measures

2.7.3.1 Extent

a. Examination of Frame

2.7.3.2 Same as 2.1.3.2 Data collection procedures require some nonframe mechanism for obtaining the data

2.7.4 Methods for Conducting Surveys Using Frames with Nonaccessable Elements
Same as 2.1.4.

REFERENCES

1. Birnbaum, Z.W. and Monroe G. Sirken (1965) "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, PHS Publication No. 1000, Ser. 2, No. 11, National Center for Health Statistics, Washington, D.C., U. S. Government Printing Office.
2. Chervy, G. (1949) "Control of a General Census by Means of an Area Sampling Method," Journal of the American Statistical Association, Vol. 44, pp 373-379.
3. Cochran, W.G. (1963), Sampling Techniques, J. Wiley and Sons, Inc. New York. (1977)
4. Dalenius, Tore (1974) The Ends and Means of Total Survey Design, mimeographed monograph.
5. Deighton, Richard E., et al. (1977) Gossary of Nonsampling Error Terms, mineographed copy of a draft report prepared for the Executive Office of the President, Office of Management and Budget, Federal Committee on Statistical Methodology, Subcommittee on Nonsampling Errors, August, 1977.
6. Eckler, A.R. and L. Pritzker, (1951) "Measuring the Accuracy of Enumerative Surveys," Bulletin of the International Statistical Institute, Vol. 33, Part 4, pp 7-24.
7. Hansen, Morris H., William N. Hurwitz; and Thomas B. Jabine, (1963) "The Use of Imperfect Lists for Probability Sampling at the U.S. Bureau of the Census." Bulletin of the International Statistical Institute, Vol. 40, Book 1, Ottawa, pp. 497-517.
8. Hansen, M. H., W. N. Hurwitz, and W. G. Madow, (1953) Sample Survey Methods and Theory, Volume I ., Methods and Applications, J. Wiley and Sons, Inc., New York.
9. Hartley, H. O. (1962) "Multiple Frame Surveys," Proceedings of the Social Statistics Section, American Statistical Association, pp. 203-206.
10. Jessen, Raymond J. (1978) Statistical Survey Techniques, John Wiley and Sons, Inc., New York.
11. Kiranandana, Suchada Seridhoranakul (1976) Imperfect Frames in Sample Surveys, Ph.D. thesis, Harvard University, Cambridge, Mass.
12. Kish, Leslie (1965), Survey Sampling, J. Wiley and Sons, Inc., New York.
13. Lessler, Judith T. (1979), Frame Errors, Phase I Taxonomy Report, Project RTI/1791/00-01I.
14. Marks, E. S., P. W. Maudlin, and H. Nisselson (1953), "The Post-enumeration Survey of the 1950 Census/A Case History in Survey Design," Journal of the American Statistical Association, Vol. 48, pp. 220-243.
15. Moser, C.A. and G. Kalton (1972), Survey Methods in Social Investigation, Basic Books, Inc., New York.
16. Scheaffer, Rchard L., William Mendenhall, and Lyman Olt (1979), Elementary Survey Sampling, Duxbury Press, North Scituate, Mass.
17. Sirken, Monroe G. (1970) "Household Surveys with Multiplicity," Journal of the American Statistical Association, Vol. 65, pp. 257-66.
18. Sirken, Monroe G. (1972) "Stratified Sample Surveys with Multiplicity" Journal of the American Statistical Association, Vol. 67, pp. 224-227.
19. Sukhatme, P. V., and B. V. Sukhatme (1970) Sampling Theory of Surveys with Applications, Iowa State University Press.
20. Szmeitat, Klaus and Karl-August Schäffer (1963) "Imperfect Frames in Statistics and the Consequences of Their Use in Sampling," Bulletin of the International Statistical Institute, Vol. 40, pp. 517-44.
21. United Nations (1964), Recommendations for the Preparation of Sample Survey Reports (Provisional Issue), Statistical Papers, Series C, No. 1, Rev. 2, Dept. of Economic and Social Affairs, Statistical Office of the United Nations, New York.
22. Vogel, Federic A. (1975) "Surveys With Overlapping Frame-Problems in Application," Proceedings of the Social Statistics Section, American Statistical Association, pp. 694-699.
23. Warwick, Donald P. and Charles A. Lininger, (1975), The Sample Survey: Theory and Practice, McGraw-Hill Bok Company, New York.
24. Zarkovich, S. S. (1965), "Sampling Methods in the Yugoslav 1953 Census of Population," Journal of the American Statistical Association, Vol. 50, pp. 720-737.
25. Zarkovich, S. S. (1956), "Some Remarks on Coverage Checks in Population Censuses," Population Studies, Vol. 9, pp. 721-175.
26. Zarkovich, S. S. and J. Krane (1965), "Some Efficient Uses of Compact Cluster Sampling," Paper prepared for the 35th Session of the International Statistical Institute, Belgrade.
27. Zarkovich, S. S. (1966), Quality of Statistical Data, Food and Agriculture Organization of the United Nations, Rome.