The Simplicity of List Sampling Weighting -- Is It Always So?

F. W. Immerman, L. L. Piper and Douglas J. Drummond, Research Triangle Institute

1. Introduction

The impression that most people gain from an introductory survey sampling course or text is that weighting a stratified simple random sample drawn from a list frame is a rather straightforward exercise. They may have learned that lists used to construct frames usually have some imperfections, that frame elements and analysis units are not always the same entities, and that most surveys have some degree of nonresponse, but they do not fully comprehend the ways in which these concepts are related to weighting. They are even less likely to appreciate that these concepts interact with each other, and that the interaction can make the calculation of weights a formidable task.

In this paper we will present illustrations of some of the difficulties which can be encountered in weighting survey data from such list samples. Rather than provide a broad overview, we will concentrate on problems associated with list frame surveys of businesses. In recent years surveys of this type have become increasingly common as regulatory agencies, marketing firms, and economic forecasters escalate their use of survey-based data. The potential for problems in weighting data from these surveys is high, and often not fully appreciated at the outset of the design effort. Our intent is to describe some of the difficulties that may be encountered, and to suggest ways in which similar difficulties can be avoided or minimized in future surveys. Much of the discussion will reflect our experiences with the National Urban Pesticide Applicator Survey (NUPAS), which RTI conducted and has been analyzing for the U.S. Environmental Protection Agency. A detailed description of the design and analysis of NUPAS can be found in Drummond et al. (in prep.).

2. Basic Concepts

2.1 Definitions

Given the context of a list frame, the following definitions can be made. A frame unit is a single element on the list frame used to select the sample. An analysis unit is the entity for which data is reported. In many surveys, the frame units and analysis units are identical, or have a one-to-many correspondence. In some surveys, however, the correspondence between frame units and reporting units is many-to-one or many-to-many. Such surveys are referred to as surveys with multiplicity (Sirken, 1970) or network surveys (Sirken, Graubard, and LaValley, 1978). Multiplicity may be defined as the number of linkages between a particular analysis unit and the sampling frame.

An example may help to clarify these concepts. Suppose a survey of pest control firms is conducted and that one such firm has five locations. Of these, four appear on the sampling frame. The location that is not on the frame might be absent because the list used to construct the frame was out of date, or excluded locations that didn't meet some criterion. Further suppose that any of the locations would, if selected, provide the desired data about the entire firm. In this case, the frame units are the four locations, and the analysis unit is the firm. Each frame unit associated with this firm has a multiplicity of four.

Sometimes multiplicity is a desirable characteristic (Birnbaum and Sirken, 1965; Sirken, 1970; Sirken, 1972; Lessler, 1981). In other cases it is an undesirable characteristic which either must be removed prior to final sample selection (Hansen, Hurwitz and Jabine, 1963; Kish, 1965), or must be accounted for during data weighting and analysis if unbiased estimates are to be obtained. In all cases, the mapping from frame units to analysis units has to be established. Once the mapping is known, if the multiplicity is not removed prior to sample selection, unbiased estimates of population totals can be obtained by calculating the overall inclusion probability of each analysis unit and using a Horvitz-Thompson estimator, or by applying counting rules involving the adjustment of the sampling weights (Birnbaum and Sirken, 1965; Sirken 1970; Sirken 1972).

2.2 A Simple Model

In greatly simplified terms, NUPAS was based on a stratified simple random sample selected from a list frame of pest control firms. For such a sample design, a general form for an unbiased estimator of the population total for a characteristic Y can be written as

$$\hat{Y}_{tot} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} y_{hi} I(e)_{hi} \qquad (1)$$

where h = indexing variable for strata
i = indexing variable for selected frame units within strata
n_h = the number of selected frame units in stratum h
y_{hi} = the observed value of characteristic Y for sample member hi
w_{hi} = the raw weight of sample member hi

$$I(e)_{hi} = \begin{cases} 1 \text{ if sample member is an eligible member of the target population} \\ 0 \text{ otherwise} \end{cases}$$

The eligibility indicator is included to acknowledge that in many surveys the frame used to select the sample is inefficient; that is, it contains elements that are not part of the target population. In the simple case in which a one-to-one or one-to-many correspondence exists between frame units and analysis units, and in which nonresponse does not occur, the raw weight for all sample members in stratum h is

$$w_{hi} = \frac{N_h}{n_h} ,$$

where N_h = the number of frame units in stratum h.

If there is multiplicity in the survey, raw weights can be adjusted in order to produce unbiased estimates of population totals. The

adjustment is performed for any frame unit with a multiplicity greater than one to reflect the fact that it's associated analysis unit had more than one way of being selected in the sample. Operationally, this adjustment is accomplished by dividing the raw weight by the frame unit's multiplicity. Formula (1) is thus modified to

$$\hat{Y}_{tot} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{w_{hi}}{k_{hi}} y_{hi} I(e)_{hi} , \qquad (2)$$

where

k_{hi} = multiplicity factor of sample member hi.

The theory which supports this procedure has been developed on two underlying assumptions. The first is that the multiplicity is a deterministic, rather than a probabilistic, factor. This means that the mapping between reporting units and frame units must be determinable without error. The second assumption is that all of the frame units linked to a particular analysis unit would provide identical data if sampled. If more than one is selected, each is retained in the sample.

If the survey suffers from nonresponse, further adjustment of the weights may be necessary if unbiased estimators are to be produced. This is especially important for estimators of totals. If industry profiles exist from prior surveys or censuses, it is often possible to adjust weights based on the data in these profiles. If no suitable industry profile is available, the weight adjustment factor may be similar in structure to the following:

$$\frac{\sum_{j=1}^{J} \frac{w_{gj}}{k_{gj}}}{\sum_{j=1}^{J} \frac{w_{gj}}{k_{gj}} I(r)_{gj}} , \qquad (3)$$

where

g = indexing variable for the weighting class
j = indexing variable for sample member within weighting class
w_{gj} = raw weight for sample member gj
k_{gj} = multiplicity factor for sample member gj

$$I(r)_{gj} = \begin{cases} 1 & \text{if sample member gj responded} \\ 0 & \text{otherwise} \end{cases}$$

Inspection of (3) reveals a dependence of the nonresponse adjustment on the multiplicity of both respondents and nonrespondents. This introduces a complication that is difficult to deal with. Multiplicity is often only directly determinable for respondents. Therefore, nonrespondent multiplicity has to be estimated. This obviously violates the assumption of deterministic multiplicity implicit in the use a multiplicity weight adjustment. Moreover, there is no way to obtain an unbiased estimate of the nonrespondent multiplicity. The estimation may be based on average respondent multiplicity, but to the degree that multiplicity is related to perceived burden, and therefore to nonresponse, average nonrespondent multiplicity may differ from average respondent multiplicity. To date a satisfactory way to deal with nonresponse in surveys with multiplicity has not been developed.

3. Practical Frame Unit Definition

3.1 Choosing a Level

Businesses come in all sizes and configurations, and have devised numerous ways of organizing in response to financial and management considerations. A firm may have multiple office locations, or may provide all services from one central location. Larger companies often have divisions or other hierarchical internal structures. Knowledge of the types of variation in business structure in the industry under study should ideally be obtained at the beginning of the design effort, and incorporated in the decision on what the level of the frame unit will be. Should the frame unit be an entire firm, a single location of that firm, or some subset of one of these? The choice of the level of the frame unit is both crucial and difficult. It impacts frame construction, questionnaire design, and subsequently effects the weighting task and the quality of data obtained by the survey.

The level chosen is often based primarily on the nature of the list or lists that will be used in frame construction. However, it is alsc important to try to assess the ability of potential sample members to provide data at the desired level. If some sample members are unwilling or unable to provide data at this level, the survey will suffer from increased nonresponse and/or respondent-induced multiplicity and inaccurate information.

Many business surveys ask for data which require respondents to refer to files or records. In these surveys, unwillingness or inability to provide data at the desired level is often due to the fact that the sample member's records are kept at some level other than the one requested. This may be true for all items in the questionnaire, or for only some of them. In general, businesses do not maintain summary profiles of their operations at lower levels than those needed for management, financial, or legal purposes. Different types of data may be summarized at different levels. For example, in a firm or location that services multiple states, sales may be tabulated at the state level because of tax requirements, whereas costs are summarized only at the firm or location level. A respondent who wishes to minimize the burden on himself and yet still provide data will often use the information in these summary profiles, regardless of how well they match the level requested. Other sample members will simply choose not to respond. The extent to which the record keeping systems of respondents are accomodated by the survey design and instrument is often a major factor in determining the quality and quantity of data obtained, and the ease with which the data is weighted and analyzed.

If the frame unit is below the level of the firm, there is potential for some businesses to be sampled multiple times. This can have several undesirable effects. It may create an impression of excessive burden, leading to nonresponse, or it may result in a decision to provide only firm-level data, regardless of the desired level, leading to multiplicity. The worst case occurs when the sampled component members of the firm chose to respond at different levels. For example, one location may respond with data only for itself, while another responds with data for the entire firm, while still another chooses not to respond at all. The best way to handle such a case is not clear. The data for the member that responded at the level of the firm could be discarded, thereby increasing nonresponse. Another option is to replicate it for the other two members and handle with a multiplicity factor, despite the obvious violation of the assumption of identical response for all frame units which is implicit in the use of multiplicity weight adjustments.

In industries which have been previously surveyed, assessment of potential sample member's willingness and ability to respond at a particular level can be based on prior experience. Assessment in fields not previously surveyed is more difficult. Consultation with subject matter specialists is a good first step. However, it's important to remember that most such specialists, while very knowledgeable about technical aspects of their field, may not be familiar with survey techniques, and may have a biased view of their profession. The obviously preferred option is a pilot study. While this may seem to be an expensive alternative, the resources seemingly saved by not having a pilot study may end up being spent several times over in order to clean the data obtained in the survey and properly weight it.

It is important to address some subtle distinctions during the frame unit definition process. Are franchises to be handled as independent firms, or as single locations of the franchising company? How are mergers, splits, or internal reorganizations that occur prior to data collection to be handled? What constitutes a "location" for the profession under study? Are areas that are strictly for production or storage of material to be counted as locations in service industries? What about corporate headquarters, which may have only administrative personnel? These sorts of questions should be answered prior to frame construction and instrument design.

3.2 Instrument Design Considerations

Once the intended frame unit level is decided upon, an instrument must be developed that will adequately convey to sample members the desired level at which data is to be reported. Any respondent misunderstanding of this level results in inaccurate data and may lead to multiplicity and coverage problems. Unfortunately, business terminology is not standardized, and tends to be ambiguous and vague, which increases the likelihood of misinterpretation, especially in mail surveys. Surveys which use personal enumeration or telephone interviews for data collection can overcome this to some extent through the use of well-trained interviewers, who can clarify for the respondent any misunderstood terms and concepts.

Regardless of mode of data collection, quality of instrument design and choice of frame unit level, the possibility always exists that some respondents will either intentionally or unintentionally provide data at a level other than the desired one. This should be anticipated during instrument development, and questions included to alert survey researchers when it has occurred, or when it may have occurred. The researchers will then have several options. They can recontact the respondent, ascertain at what level the data was reported, and, if necessary, attempt to obtain the data at the desired level. They may instead choose to transform the data provided to the appropriate level by using auxiliary information and a regression-based approach. For example, if data are desired for only one location of a firm, but the response is at the level of the entire firm, the proportion of the firm's sales or net worth associated with that location might be used to prorate the response. Another option is to use the data as is, correctly accounting for it's problem through the use of multiplicity weight adjustments. If this last option is likely to be used, the instrument should gather any information, such as names, addresses, and phone numbers, that will allow the mapping from the analysis unit back to the sampling frame to be established.

4. The Sampling Frame

4.1 The Ideal Case

The preceeding discussion implies that the ideal sampling frame for a business survey would be composed of frame units defined at a level easily responded at by potential sample members. The ideal frame would also list every member of the target population once and only once, contain all of the information necessary to contact any target element, and have data that could be used for stratification or as measures of size. This implies several things about the ideal list or lists used to construct such a frame. Complete coverage of the population of interest is obviously a primary requisite. Nontarget elements on the list would be identifiable as such, allowing them to be removed during frame construction. The list would provide current and accurate name, address, telephone number, size, and stratification data. A final requirement is that the list contain information that permits each element to be categorized as a company headquarters, branch location, franchise, or some other business entity, and identifies any relationships between it's elements, such as ones that are branch offices of the same firm. This would allow the survey statistician to construct a frame free of duplication, and would permit easier resolution of respondent-induced multiplicity problems should they arise.

4.2 Common Problems

Lists which possess all of the ideal characteristics described above rarely, if ever exist. For a comprehensive review of the typical types of imperfections found in list frames, the reader is referred to Hansen, Hurwitz, and

Jabine (1963). In this section we will concentrate on some of the list imperfections which commonly occur in business surveys and impact weighting, and on potential ways in which to circumvent them.

Undercoverage and overcoverage (i.e. inclusion of nontarget elements) are almost always problems in business surveys. Some types of lists commonly used to construct frames for this kind of survey are more prone to these difficulties than others. Lists on which inclusion is strictly voluntary, such as yellow pages directories and trade or professional association membership roles, may have a high degree of undercoverage, and in addition, may not contain a representative cross-section of the industry under study. Some lists specifically exclude elements which do not meet some minimum requirements, thereby providing potentially biased undercoverage. For example, firms with a relatively small net worth may not be included, but in aggregate may account for a significant proportion of the industry's business.

Errors in coverage are also due to the fact that businesses come into and go out of existence at a relatively rapid rate. If frame units are to be locations of firms rather than entire firms, this effect may be increased due to an even quicker rate of change.

In any survey with overcoverage, the eligibility (e.g., membership in the target population) of nonrespondents may be difficult to distinguish. If profile-based nonresponse adjustment is not possible, inability to categorize nonrespondents may require that nonresponse weight adjustments be based on both eligible and ineligible respondents. Such adjustments incorporate an assumption that eligibility is distributed identically in respondents and nonrespondents, an assumption that is difficult to justify in some cases. For example, nontarget sample members may be more likely not to respond simply because they feel that their inclusion in the survey was a mistake, and their response would not count anyway. This type of problem is especially acute in mail surveys, but can be reduced by telephone or personal nonresponse followup.

There are several ways of reducing the degree and impact of undercoverage and overcoverage. Obviously, using the most up-to-date list or lists is an important first step. Oftentimes frames constructed using lists from a variety of sources are more complete than those based on a single list. Screening of some or all of the elements on a list provides identification of nontarget elements and an assessment of the quality of the list. During data collection, well trained interviewers can often compensate for coverage problems by properly classifying ineligibles and reducing the degree of respondent-induced multiplicity.

Extant lists may not be available at the frame unit level chosen for the study, and relatively few have the sort of auxiliary information that allows transformation from the given level to the desired one. In such a case it may be tempting to simply redefine the desired frame unit level. However, as discussed in section 3.1, if the level used is not one the sample members can or will respond at, severe problems with data weighting and quality are likely. A preferable solution, if resources permit, is to contact each element on the list and obtain the necessary information. More realistically, a two (or more) phase sample design can be employed. Data obtained from the first phase sample of list elements is used to construct frame units for the second phase. Weighting may be complicated in multiphase designs because of the introduction at each phase of the potential for nonresponse, multiplicity, and coverage problems. Good instrument design and trained interviewers can usually successfully minimize such difficulties.

Inaccurate auxiliary information on the list or lists used in frame construction can severely impact the weighting task. Inaccuracies may be due to mistakes in the data as it was originally provided or entered on the list, or to changes in business name, address, telephone number, or structure subsequent to inclusion on the list. Whatever it's cause, bad auxiliary data can result in the inability to contact some sample members, which, if overcoverage is also present, leads to the problem of determining the eligibility of nonrespondents. Inaccurate information also increases the difficulty of assuring that each frame unit is unique and so increases the potential for multiplicity. Even if each frame unit is uniquely represented, respondent-induced multiplicity is still possible. Any incorrect auxiliary information on the frame makes the resolution of the analysis unit to frame unit mapping more difficult and less exact, and thus can lead to violation of the assumption of the deterministic nature of the multiplicity factors.

The same methods used to minimize coverage problems are effective in reducing the degree to which incorrect auxiliary data adversely affects the survey. Addresses and telephone numbers can be compared across several lists and inconsistencies noted and resolved. Screening of list elements can also lead to identification and correction of inaccurate data items. Lastly, telephone interviewers and personal enumerators can track down sample members for which there is faulty locator information, reduce the amount of respondent-induced multiplicity, and, should such multiplicity be unavoidable, permit more accurate mapping between analysis unit and frame units.

5. Conclusions

In the ideal case, weighting a list sample of businesses is no more difficult than weighting any list sample of comparable size and design. However, departures from ideal occur in most business surveys, and greatly complicate the calculation of sampling weights. Problems with coverage are routine, and act to increase bias, nonresponse, and the likelihood of multiplicity problems. Because of the variability in the way in which businesses are structured and keep records, there is always potential for frame units not to be appropriate for some sample members. This leads to nonresponse, inaccurate data, and multiplicity. Surveys with multiplicity are especially prone to problems. Resolution of the mapping between analysis units and

frame units can be difficult, assumptions underlying the methods used to correct for multiplicity are often violated, and nonresponse presents problems not yet handled by multiplicity theory.

We have only considered effects on point estimates in this paper, and have not examined concomitant effects on precision. The problems discussed can be readily handled so as not to complicate the calculation of variances. However, multiplicity and frame inefficiency will generally act to increase the sampling error of the survey. While this in itself is cause for concern, it is important to keep in mind that these problems are likely to increase nonsampling error and bias to an even greater extent.

Three general recommendations can be made based on the preceeding discussion. The first is that, except in certain cases, every attempt should be made to avoid having to use multiplicity estimates in business surveys. This means that the business structure and record keeping systems of potential sample members, and the nature of the list or lists used to construct the frame have to be well understood prior to sample selection. Pilot studies, telephone screening, and multiphase sample designs can all be employed to advantage here.

The second recommendation is to use telephone or personal enumeration as the primary mode of data collection. Nonresponse, multiplicity, and respondent misunderstanding will be decreased if there is a trained interviewer involved in the data collection process.

The final recommendation is that a serious nonresponse followup effort, using interviewers rather than mail, be made. Nonresponse interacts with eligibility and multiplicity, and often requires unjustifiable assumptions to be made during data weighting and analysis. Even if the followup effect fails to eliminate the nonresponse, it may permit more informed decision making on the eligibility and multiplicity of the remaining nonrespondents.

All of these recommendations imply a reduced sample size for a fixed amount of resource. The novice survey statistician may therefore balk at first at the associated apparent increase in sampling error. However, if multiplicity, coverage, and nonresponse problems do occur, they will increase both sampling and nonsampling error, and are likely to increase bias. Therefore, in many survey the steps taken to reduce these problems are cost effective ways to reduce total survey error.

6. References

Birnbaum, Z. W. and M. 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, Public Health Service, Washington, D.C.: U.S. Government Printing Office.

Drummond, D. J., Piper, L. L., Pagano, J. R., Waddell, R. D., Woodside, M. B. and F. W. Immerman (in prep) The National Urban Pesticide Applicator Survey, Final Report.

Hansen, M. H., W. N. Hurwitz, and T. 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.

Kish, L. (1965), Survey Sampling, J. Wiley and Sons, Inc., New York.

Lessler, J. T. (1981), "Multiplicity Estimators with Multiple Counting Rules for Multistage Sample Surveys". Proceeding of Social Statistics Section, American Statistical Association, pp. 12-16.

Sirken, M. G. (1970) "Household Surveys with Multiplicity", Journal of the American Statistical Association, Vol. 65, pp. 257-266.

Sirken, M. G. (1972) "Stratified Sampling Survey with Multiplicity", Journal of the American Statistical Association, Vol. 67, pp. 224-227.

Sirken, M. G., B. J. Graubard, and R. W. LaValley (1978), "Evaluation of Census Population Coverage by Network Surveys", Proceedings of Section on Survey Research Methods, American Statistical Association, pp. 239-244.