THE COUNTING RULE STRATEGY IN SAMPLE SURVEYS

Monroe G. Sirken, National Center for Health Statistics

1. INTRODUCTION

Frequently, the enumeration units in sample surveys are not identical to the population elements whose parameters are being estimated. For example, Table 1 on the following page lists four illustrative surveys in which the enumeration units are different from the population elements. Clearly, in surveys of this type, it is necessary to specify rules for linking elements to the enumeration units where the elements are eligible and required to be counted in the survey. The counting rule adopted in the survey should be viewed as a design factor for improving the efficiency of the survey. We have investigated [4], [5], [6] some of the statistical properties of counting rules. In this expository report, we summarize some of those findings, and we describe the strategy for selecting counting rules in sample surveys.

2. COUNTING RULES

Counting rules specify the conditions for linking enumeration units to population elements. It is understood that elements are eligible and required to be reported by the enumeration units to which they are linked by the rule and that the elements are ineligible to be reported by any other units. Conventional counting rules have the property of uniquely linking every population element to one and only one enumeration unit. For instance, the de jure residence rule is a conventional counting rule in surveys in which households are the enumeration units and persons are the population elements. The rule links every person to one and only one household, namely his de jure residence.

Conventional rules are appealing because they assure that every element is eligible to be counted once and only once. However, as we shall see, survey estimates based on conventional rules are sometimes subject to intolerably large sampling and measurement errors. Consequently, we have been investigating the statistical properties of estimators based on multiplicity rules that would make them subject to smaller errors than conventional rules. Multiplicity rules specify conditions that permit more than one enumeration unit to be linked to the same elements.

Of particular interest is a class of multiplicity rules which have the property of supplementing the condition of a conventional rule with other conditions for linking elements to enumeration units. Multiplicity rules of this type have two desirable properties. First, they assure that every element is linked to at least one enumeration unit. Second, they permit the survey to produce several sets of estimates--one set based on the conventional rule and each of the other sets based on a multiplicity rule that incorporates the condition of the conventional rule. In this manner, the estimate based on the conventional rule is preserved. For instance, the rule that links persons to their own residence as well as the residences of their siblings and children would produce four sets of estimates based on the following rules:

(1) persons are linked to their own residence;

(2) persons are linked to their own residence and to the residences of their siblings;

(3) persons are linked to their own residence and to the residences of their children;

(4) persons are linked to their own residence and to the residences of their siblings and children

The first is a conventional rule. The others are multiplicity rules.

Table 2 presents conventional and multiplicity rules which have been reported in the literature [7], [2], [3], [9] for each of the illustrative surveys listed in Table 1. It will be noted that in each example, the multiplicity rule supplements the condition specified by the conventional rule. Table 2 also compares the population elements that are eligible to be counted at an enumeration unit in compliance with the conventional and the multiplicity rule.

3. COUNTING RULE ESTIMATORS

In surveys based on conventional rules, all elements have the same probability of being selected in the sample but in surveys based on multiplicity rules they do not. Unbiased estimators for multiplicity rules adjust for the different probabilities by appropriately weighting the elements. Several kinds of unbiased estimators have been proposed [1], [8], each estimator being based on a different system of weights. Some estimators require matching the elements enumerated in the survey to eliminate duplicate reports. One unbiased estimator [11] that does not eliminate duplicate reports, we shall refer below to as the multiplicity estimator.

The multiplicity estimator assigns a weight to every element everytime it is enumerated. It assigns a weight of unity to elements that are linked to a single enumeration unit, and to elements linked to multiple units, it assigns a weight, greater than zero and less than unity, everytime they are enumerated. The multiplicity estimator is unbiased if the sum of the weights assigned to the multiple linked elements is equal to unity. The multiple linked element may be assigned the same or a different weight each time it is enumerated. If it is assigned the same weight each time, its weight is the inverse of the number of units it is linked to.

Table 1. Illustrative Surveys in Which Enumeration Units are not the Same as Population Elements

Survey	Enumeration Unit	Population Element	Variate
1	Housing unit	Decedent	Hospital utilization in last year of life
2	Areal segment	Farm	Number of farms, and acres of farmland
3	Medical source	Patient with a disease	Medical costs of treating disease
4	Line of text of a statistical report	Statistical statement	Statements failing statistical standards

Table 2. Conventional and Multiplicity Rules in Illustrative Surveys

Survey	Rule		Elements Counted at an Enumeration Unit	
	Conventional	Multiplicity	Conventional Rule	Multiplicity Rule
1	Decedents are linked to their former de jure residence.	Decedents are linked to their former de jure residence and to the de jure residences of surviving relatives.	Deaths of persons who resided in the housing unit	Deaths of persons who resided in the housing unit or whose relatives reside in the unit
2	Farms are linked to the segment in which the farm headquarters are located.	Farms are linked to all segments overlapped by the farms.	Farms whose headquarters are located within the segment's boundaries	Farms entirely or partially overlapping the segment's boundaries
3	Patients are linked to the medical source with primary responsibility for treating their disease.	Patients are linked to all medical sources who treated them for the disease.	Patients for whom the medical source had primary responsibility for treatment	Patients treated by the medical source for the disease
4	Statements are linked to the line of text which contains the first word of the statement.	Statements are linked to all lines of text overlapped by the statement.	Statements starting on the line of text	Statements overlapping the line of text

In a formal sense, the estimator based on the conventional rule is a special case of the multiplicity estimator in which each of the elements is assigned a weight of unity. However, there is an important difference between the conventional and multiplicity estimators. The conventional estimator presumes that the weight assigned each element is equal to unity because by definition every element is uniquely linked by the conventional rule to one enumeration unit. On the other hand, the weights assigned elements by the multiplicity estimator are rarely known beforehand and usually require collecting ancillary information in the survey.

The procedure for determining the weights that are assigned to elements and the source of the ancillary information needed to calculate the weights varies for surveys based on multiplicity rules. For instance, Table 3 below presents this information for the surveys that are based on the multiplicity rules described in Table 2.

4. COUNTING RULE STRATEGY

The objective of the counting rule strategy is to select the optimum rule for linking elements to enumeration units and to assign the optimum sets of weights to enumeration units that are linked to the same elements. The rule and weights are considered to be optimum in the sense that they minimize the combined effect of sampling and measurement errors on the survey estimates subject to the cost constraints of conducting the survey. Although we have focused on surveys in which the population elements are different from the enumeration units, it is noteworthy that the counting rule strategy is also applicable to surveys in which the population elements are the enumeration units.

In this section, we describe some conditions which may make multiplicity rules preferable to conventional rules. We consider in turn, the effect of counting rules on sampling errors, on measurement errors, and on the costs of conducting the survey.

Sampling Errors

Since the counting rule specifies the conditions for linking population elements to enumeration units, the distribution of the population elements among the enumeration units is a function of the rule. Changing the counting rule modifies the population distribution and thereby alters the sampling distribution and the sampling variance of the survey estimate. It would be a mistake to conclude, however, that the estimators based on multiplicity rules are necessarily subject to smaller sampling variance than estimators based on conventional rules. They are not, except under special conditions such as when the multiplicity rule links at most one element to an enumeration unit. Nevertheless, when the population distribution based on the conventional rule is highly skewed, multiplicity rules offer a strategy for redistributing the elements among the enumeration units and thereby decreasing the variance of the population distribution. Two conditions contributing to the skewness of the population distribution based on the conventional rule are: (1) few enumeration units are linked to population elements and (2) the distribution of the variate among the population elements is highly skewed. When either or both of these conditions exist, multiplicity rules may also exist which would have the effect of redistributing the weighted elements among the enumeration units so that the resulting distribution would be less skewed.

Table 3. Weights Used in Illustrative Surveys Based on Multiplicity Rules

Survey	Weight	Source of Information about Weight
1	Inverse of the number of housing units linked to the death.	Unit where the death is enumerated reports the number of different housing units which are either residences of surviving relatives or the former residence of the decedent.
2	Fraction of the farm area overlapped by the boundaries of the segment	Areal maps
3	Inverse of the number of medical sources treating the patient	Source where patient is enumerated identifies the referral and referring sources which are followed up in the survey
4	Inverse of the number of lines overlapped by the statistical statement	Clerk counts the number of lines of text that are overlapped by the statement

A multiplicity rule that links every enumeration unit to every population element would not be subject to any variance. Unfortunately, this rule would be subject to unacceptably large measurement errors in virtually all surveys. Although the rule is impractical, it serves to illustrate the point that consideration must be given to the joint effect of sampling and measurement errors in selecting a counting rule.

Two kinds of measurement errors associated with counting rules are coverage errors and response errors.

Coverage Errors

Coverage errors occur when enumeration units either fail to report population elements to which they are linked or erroneously report elements to which they are not linked by the counting rule. In most surveys underreporting appears to be a far more serious problem than overreporting.

Multiplicity rules are often apropos when it is difficult to implement conventional rules.

If some characteristic of the population leads to multiple links of the population elements with enumeration units, it is natural to think in terms of multiplicity rules. Thus, if patients have encounters with several physicians, it may be easier to implement a multiplicity rule that links the patients to all their medical sources than to implement a conventional rule that defines a condition that would uniquely link every patient to a single medical source.

If the attribute being measured in a household survey defines a population of persons that does not reside in the area covered by the survey, it is extremely difficult to implement conventional residence rules. Thus, it would be virtually impossible to collect emigration statistics in cross sectional surveys using de jure residence rules. Multiplicity rules based on consanguine relations are often more apropos than conventional rules in these surveys. For instance, the conventional rule which links the deceased person to his former residence is subject to gross underreporting of deaths in single retrospective surveys of population change because the death of an individual is often followed by the dissolution or emigration of the household. It appears [10] that a multiplicity rule that links decedents to the residences of surviving relatives misses fewer deaths than a conventional rule that links decedents to their former residence.

Multiplicity rules may sometimes be subject to smaller coverage errors than conventional rules in population surveys which collect sensitive information. For instance, a person with a stigmatized attribute, such as an alcoholic or a drug addict might be less likely to report himself than he would be to report his close friends or relatives living elsewhere who were drug addicts or alcoholics since he could report them anonymously without identifying them either by name or by their place of residence.

Response Errors

Response errors represent the invalid or incomplete reports by the enumeration units of the characteristics of elements to which they are linked by the counting rule. Sometimes one of the enumeration units is "best" in the sense of having more complete and accurate information about the variate being measured than any other unit that is linked to the same element. The "best" unit may or may not be the unit linked to the element by the conventional rule. The surviving spouse, for example, might be a best informant for the decedent, whether or not she lived with the decedent when he died. If every unit linked to an element by the multiplicity rule knows the identity and location of the "best" unit, it might be cost effective to introduce a supplementary strategy to determine the value of the variate for the elements which are enumerated at units which are not the "best" units. The strategy would involve reenumerating these elements at their "best" enumeration units.

Survey Costs

A multiplicity rule is not necessarily more efficient than the conventional rule it incorporates even though for a fixed size sample of enumeration units, it is subject to less combined sampling error and measurement error. This is so because the survey based on a multiplicity rule is more costly. There are two reasons why the survey costs are greater for a multiplicity than for a conventional rule: First, the enumeration unit cost is greater because the average number of elements linked to an enumeration unit is greater for the multiplicity rule than for the conventional rule. Second, the estimator based on the multiplicity rule, and its variance require ancillary information to determine the weights assigned to elements which is not needed by the conventional estimator. Since this ancillary information is frequently collected in the survey, it adds to the survey cost.

5. SUMMARY

The counting rule strategy treats the counting rule as a design factor for improving the efficiency of the survey. The strategy selects the optimum counting rule for linking population elements to enumeration units and assigns the optimum sets of weights to the elements that are linked to more than one enumeration unit. We have indicated the types of survey conditions in which the sampling errors and measurement errors of the estimate based on a conventional rule may be reduced by selecting a multiplicity rule. However, the cost of conducting a survey for a fixed size sample of enumeration units is invariably larger for a multiplicity rule than for a conventional rule. To apply the counting rule strategy, therefore, requires information that compares the cost coefficients as well as the error components associated with different counting rules.

References

[1] Birnbaum, Z. W. and Sirken, M. G., "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, P.H.S. Publication No. 1000-Series 2-No. 11, Public Health Service, Washington, D.C.: U.S. Government Printing Office, 1965, pp. 1-8.

[2] Hendricks, W. A., Searls, D. T., and Horvitz, D. G., "A Comparison of Three Rules for Associating Farms and Farmland with Sample Area Segments in Agricultural Surveys," Estimation of Areas in Agricultural Statistics, Food and Agriculture Organization of the United Nations, Rome, 1965, pp. 191-198.

[3] Kramm, E. R., Crane, M. M., Sirken, M. G. and Brown, M. L., "A Cystic Fibrosis Pilot Survey in Three New England States," American Journal of Public Health, December 1962, pp. 2041-2057.

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

[5] _____, "Stratified Sample Surveys with Multiplicity," Journal of the American Statistical Association, 67 (March 1972), pp. 224-227.

[6] _____, "Variance Components of Multiplicity Estimators," Biometrics, 28 (September 1972), pp. 869-873.

[7] _____, "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, 10 (August 1973), pp. 469-478.

[8] _____, "Effect of Counting Rules on Sampling Errors and Survey Costs," unpublished manuscript.

[9] _____, and Levy, P. S., "Multiplicity Estimation of Proportions Based on Ratios of Random Variables," Journal of the American Statistical Association, 69 (March 1974), pp. 68-73.

[10] _____, and Royston, P. N., "Underreporting of Births and Deaths in Household Surveys of Population Change," Proceedings of the Social Statistics Section, American Statistical Association, (1973), pp. 412-415.