A MODEL FOR NETWORK (MULTIPLICITY) ESTIMATION OF CENSUS UNDERCOVERAGE

by
Eli S. Marks and Kathleen Ockay
U.S. Bureau of the Census

Introduction

Network or multiplicity estimation involves the use of a sampling frame in which some of the population elements are linked to more than one sampling unit. For example, in a simple household survey, it is usual to sample persons by 'linking' every individual uniquely to some housing unit (most frequently the person's usual residence or the place where he 'stayed' the previous night) and the person would be in the sample if the housing unit to which he is linked is selected for the sample. In a household network survey, one adds to the unique linkage for a simple household survey, linkages to other specified households and the person is in the sample if either his own household or any of the additional households to which he is linked is selected for the sample.

Multiplicity sampling was introduced to increase the frequency with which rare events are reported in the sample and thus, as demonstrated by Sirken (5), reduce the sampling variance. Sirken has also suggested that network sampling could be useful in reducing the bias as well as the variance of some sample estimates. As an example, there is usually a downward bias in reports in a Post-Enumeration Survey of persons living in a housing unit on the census date who have since moved away. Current residents at an address often do not know who the former residents were and even if they do know, they may be unable or unwilling to provide information about them.

Sample surveys can be used to estimate the completeness of census coverage either by (1) trying, for a sample of small areas (segments), to do a very careful job of covering everyone and using this sample to estimate the "true" population; or (2) trying to do an enumeration of a sample of small areas which is completely independent of the census, matching the sample cases to the census to determine which of them were enumerated and taking the percent of the sample cases which were enumerated as an estimate of the completeness of coverage for the entire population. The second method is known as "dual system estimation".

Trying, in a survey, to improve on census coverage has given useful results only where the original census coverage was so poor that any moderately conscientious recanvass in a sample area was bound to turn up a substantial number of households and individuals who were missed in the census. Consequently, since 1950, U.S. Census Bureau use of surveys to estimate census coverage has stressed dual system estimation.

While multiplicity sampling may be useful in partially correcting the biasing effects of undercoverage on estimates from a survey only, it does not necessarily reduce the bias of dual system estimates. Suppose, for example, that a PES multiplicity linkage rule picked up a number of persons enumerated in the Census who would otherwise be missed by the PES but failed to pick up any additional persons not enumerated in the Census. The effect of adding this multiplicity linkage to a linkage to "own household" would be to reduce the bias of an estimate of total population based on the PES alone but to increase the bias of a dual system estimate of total population based on a match of the PES to the Census.

Equation (1) of the Appendix gives a general formula for making survey estimates (both multiplicity and simplex estimates) of the completeness of census coverage. It will be noted that the formulas involve a weight w_{Aij} to be applied to each of the sample observations. These weights should be inversely proportional to the probability of the case being reported in the sample; and this probability should reflect (1) the probability of selecting the reporting household in the sample, (2) the number of households to which the person is "linked" (i.e., the number of households which should report the person) and (3) the probability that a given linked household, if selected for the sample, would actually report the person. The expected value of the estimate is given by equation (2). In this equation, the three factors just mentioned appear.

As indicated by equation (6) of the Appendix, the estimate $\bar{y}_A$ is unbiased (for a self-weighting sample) if the weights are such that the sum of the product of the reporting probability and the weight over all households linked to the person is 1.

Usually, the reporting probabilities are unknown and estimating them is difficult and costly. Thus, most network samples act as if the reporting probabilities are equal for all persons and all linked households and take the weight for each person reported equal to the reciprocal of the number of households linked to the person. This gives the estimate shown in equation (10) and the expected value shown in equation (11) of the Appendix.

The expected value of the estimate of equation (10) can be written, as shown in equation (12) of the Appendix, as the product of the true value $\bar{Y}_A$ and the ratio of the average probability R_{AE} of being reported in the survey for enumerated[1/] persons in class A to the average probability R_A of being reported in the survey for all persons in class A. Thus, if the average probability of being reported in the survey is the same for persons enumerated in the census as it is for persons not enumerated in the census, the estimate $\bar{y}_A$ will be an unbiased estimate of the completeness of coverage of the census.

Alternative Estimates

To study the effect of a particular linkage rule, it is desirable to express the expected value of the multiplicity estimate in terms of estimates based on unique linkages to a single household. To simplify the estimates, we define the "subrules" which make up the "multiplicity counting rule" so that the subrule links any individual to not more than one household. Thus, a "multiplicity counting rule" that involves reports by own household, parents' household(s) and siblings' households would consist of the "subrules" linking people to (1) own household, (2) mother's household, (3) father's household, (4) household of oldest sibling, (5) household of 2nd oldest sibling, ..., (k + 3) household of $_k$th oldest sibling, ..., $(\hat{K} + 3)$ household of $\hat{K}^{th}$ oldest sibling (where $\hat{K}$ = maximum number of siblings for any individual in the population).

To express the multiplicity estimate and its expected value in terms of simplex estimates, we treat the population as divided into "linkage groups" consisting of sets of individuals having the same types of linkages. Thus, if we assume the counting rule involves links to (1) household where person is now living, (2) any other household where person's mother lives and (3) any other household where person's oldest sibling lives, there would be 8 sets linked to: (1) own household only, (2) mother's household only (i.e., person with mother but no own household and no living sibling), (3) oldest sibling's household only, (4) own household and mother's household, (5) own household and oldest sibling's household, (6) mother's household and oldest sibling's household, (7) all three types of household and (8) no household of any of the three types. There will, of course, be no survey reports for Group 8 since these individuals are not linked to any household.

The multiplicity estimate and its expected value can then be expressed as in equation (13) of the Appendix, as weighted averages of estimates ($\bar{y}_{AH}$) of completeness of census enumeration for each of the groups, with the weights r_{AH} (the estimate from the multiplicity sample of the proportion R_{AH} of the total class A population in the group). The group estimates $\bar{y}_{AH}$ are, in turn, averages of estimates $\bar{y}_{AHI}$, weighted by r_{AHI} (the proportion of class A persons in the group reported by a single source).

In equation (13), each of the K_H simplex estimates $\bar{y}_{AHI}$ as well as their weighted average, the multiplicity estimate $\bar{y}_{AH}$, is an estimate of the same value $\bar{Y}_{AH}$, the completeness of enumeration for all class A persons in group H. Similarly, the "weights" r_{AHI} and their average r_{AH} are all estimates of the same proportion R_{AH}. Thus, if we use for each of the M linkage groups, one of the K_H simplex (non-multiplicity) estimates $\bar{y}_{AHI}$ and the corresponding estimate of the proportion r_{AHI}, (instead of the multiplicity estimates $\bar{y}_{AH}$ and r_{AH}) we get the simplex estimate $\bar{y}_{A1}$ shown in equation (14). If we use the estimate $\bar{y}_{AHI}$ based on one "counting subrule" with the proportion r_{AHK} based on another "counting subrule" we get the "composite" simplex estimate $\bar{y}_{A2}$ of equation (15).

If the same counting subrule of linkage to own household is used for all linkage groups, $\bar{y}_{A1}$ is the usual dual system estimate of completeness of coverage. However, one could use different counting subrules for different groups but still have only one linkage for each member of the population. For example, one could use reports from own household where the person has no parents or siblings living in any other household, household of mother if mother lives in another household, and household of the oldest sibling who lives elsewhere if mother is dead. In these cases, sample reports from own household would not be used if the person's mother is living in some other household or if the mother is dead and a sibling is living in some other household.

Note that there usually exists a simplex estimate, $\bar{y}_{A1}$ of equation (14), whose bias is as small as or smaller than the bias of the multiplicity estimate $\bar{y}_A$ of equation (13). For example, a multiplicity counting rule which links persons to own household and to household of mother if mother is living, gives an estimate $\bar{y}_A$ which is the weighted average of estimates from the two simplex counting rules which link persons to (1) own household only and (2) mother's household if mother is living or own household if mother is dead. Thus, if both simplex counting rules give estimates which are biased in the same direction (which is usually the case), the multiplicity estimate must be more biased than the simplex estimate with the smaller bias.

Of course, the least biased simplex estimate $\bar{y}_{A1}$ may have a larger variance than the multiplicity estimate $\bar{y}_A$, so it is possible to get a greater mean square error with the simplex estimate even though it has a smaller bias. Furthermore, it is not always possible to determine with confidence which simplex estimate is least biased, nor to be sure that the direction of bias is the same for all of the estimates. Thus, the network estimate $\bar{y}_A$ can be considered more "robust" than the simplex estimate $\bar{y}_{A1}$ in the sense that, while one can make gains in reducing overall bias if one knows (or guesses correctly) the biases of the $\bar{y}_{AHI}$ values, one can also take substantial losses in increasing overall bias in situations where little is known about the biases of the $\bar{y}_{AHI}$ values.

Against the advantages for the network estimate $\bar{y}_A$ of lower variance and, possible, greater "robustness" must be set the disadvantage of either: (a) obtaining poorer matching information if only the secondary type households[2/] are interviewed for those sample persons reported by a secondary type household; or (b) the increased costs of locating and interviewing the person himself to get better matching data.

These disadvantages can apply, however, to the non-network estimate $\bar{y}_{A1}$ just as much as they do to the network estimate. Furthermore, if $\bar{y}_{A1}$ is used, we may also have trouble getting good matching information for the persons most likely to be missed in the census, so that the increased costs of the multiplicity estimate $\bar{y}_A$ have the offset of being the kind of costs we might elect in any event to reduce the matching bias of our simplex estimate.

Another possibility is to base the estimate r_{AHI} on a different linkage from the estimate $\bar{y}_{AHI}$. That is, we could use the "composite" estimate $\bar{y}_{A2}$ of equation (15). This suggestion derives from the experience with the PES of the Paraguayan Population Census of 1972, reported in Marks (1). This PES was not a multiplicity study since each person was to be linked to only one household. However, the sample was randomly split into two subsamples and different counting rules were used in each subsample. In the A subsample, persons were linked to the household where they were staying on the "census date". In the B subsample, persons were linked to the household where they were living at the time of the PES. Although the expected sample sizes were the same and the obtained sample sizes were fairly close, the number n_{AHI} (and proportion r_{AHI}) of "migrants" was five times as great for counting rule B as for counting rule A. On the other hand, there was little difference between A subsample reports and B subsample reports in the completeness of census coverage estimates, $\bar{y}_{AHI}$ and $\bar{y}_{AHK}$, for either the migrants or the non-migrants. However, the smaller proportion of migrants in subsample A resulted in a significantly higher estimate of completeness of the census from subsample A than from subsample B. On the other hand, matching for counting rule A was simpler, easier (and cheaper), and more reliable than for counting rule B, since, with counting rule A, the census files were searched for "matches" only in the sample enumeration areas, while counting rule B required that one look whereever the person was reported as staying on the census date. As a result, nearly half the rule B "migrants" had to be thrown out in making the estimate $\bar{y}_{AH}$, because they had "insufficient infomation" for doing the matching search. Thus, there may be considerable gain in terms of costs and reliability to using counting rule A reports to estimate the completeness of coverage for "migrants" and considerable gain in terms of overall bias to using counting rule B reports to estimate the proportion of migrants.

Another procedure for using multiplicity to reduce the bias of (dual system) estimates of completeness of census coverage derives from the suggestion by Schmelz, Nathan and Kenvin (4) that a multiplicity survey with an "adjudicated built-in evaluation study" could be used as an alternative to the "usual 'dual system' method for estimating vital events in developing countries".

In the Israeli multiplicity surveys (3 and 4), an "evaluation study", done as a follow-up on the basic survey, involved interviewing another household in the linkage network for a subsample of the persons reported in the basic multiplicity survey. Since this method can be used to give a "3-way match" between the census, the basic multiplicity study and the evaluation study, a "3-system estimate" can be made along the lines discussed in Marks, Seltzer, and Krotki (2).[3/]

While there may be substantial reductions in bias through use of a 3-system estimate, the variance and cost of the 3-system approach must also be considered. In particular, the efficiency of using other households in a multiplicity network as the third system versus using a completely independent system needs investigation.

U.S. Census Bureau Research

The coverage evaluation of the U.S. 1980 Census of Population and Housing will require estimates of undercoverage by State and (at a minimum) major cities and SMSA's.

For this purpose, past experience indicates that no single method of estimating undercoverage will be adequate. Currently, the U.S. Census Bureau is testing out a fairly complex combination of dual system estimates with demographic and statistical analysis. The major source of the dual system estimates would be a post enumeration survey (PES).

Past experiences with dual system estimates of undercoverage based on a PES, point to the likelihood of substantial biases for certain classes of the U.S. population--e.g., for black males, ages 20 to 44. To correct these biases, the U.S. Census Bureau is investigating supplementary techniques. One of these techniques is the use of network (multiplicity) sampling. The goal of current Census Bureau research is to compare alternative estimates from network PES samples with each other and with estimates from non-network PES samples.

In addition to the multiplicity linkages to close relatives used in previous network sample studies, the possibility exists of gain (reduction in the Mean Square Error of the estimate) through network linkages to both the place where a person lives at the time of the PES and the place where he is staying and also to both residence at the time of the PES and residence at the time of the Census. These types of linkages are being studied to try to determine whether they are more efficient for purposes of estimating undercoverage than the "consanguinity" linkages.

The U.S. Census Bureau's initial attempt at studying the use of a network (multiplicity) survey to reduce the bias of estimates of the completeness of census coverage was a "feasibility study" carried out following the pretest for the 1980 Census in Oakland, California, in 1977. In designing the Oakland multiplicity study, the Census Bureau drew heavily on the experience and skill of Dr. Sirken and his colleagues at NCHS and also on the studies done by the Central Bureau of Statistics of Israel. The model presented in this paper was of value in pointing out oversights arising from the nature of the census problem. For example, examination of the model led to revision of the questionnaire originally proposed for use in the Oakland study.

The major difference between the multiplicity studies in the vital events area and the census coverage studies is in the cost-variance structure of the problem. Vital events studies (particularly studies of deaths) are dealing with a "rare event". Adding another linkage will increase the number of events significantly, without any increase in the sample size in terms of number of households and with relatively small increase in costs. However, in measuring completeness of census coverage, adding a type of household link may increase costs almost as much (proportionately) as it reduces variance. Also, while the accuracy and completeness of sample information can be improved by doing interviews with the person's own household, this will increase the field costs--possibly more than it improves the accuracy and reduces the cost of matching.

An important consideration in a PES is the accuracy and completeness of address information needed for matching to the census files. Preliminary results from the Oakland multiplicity study reveal that the ability to provide complete addresses varies by the kind of relative reporting. Parents provided complete addresses for their children 83.6% of the time; children provided complete addresses for parents 80.2% of the time and siblings provided complete addresses for siblings 66.9% of the time. This information suggests that it may be better to ask persons about their adult children than about their siblings. However, it is possible that reports by siblings' households will reduce the bias of the estimates of completeness more than will reports by parents' households (because of selective factors in who is reported).

A more extensive multiplicity study is now being planned for the fall of 1978 in the areas where 1980 Census "dress rehearsals" were taken as of April 4, 1978. It is hoped that further research will provide insights into the questions of bias, variance and cost as they relate to the use of multiplicity surveys in measuring completeness of census enumeration.

Footnotes

1/ The word "enumerated" is used in this paper to mean "enumerated in the census".

2/ By "secondary type household" is meant a linked household of which the person is not a member.

3/ See Chapter VII, Section D.1, pp. 401-408. The number of persons "missed" by all three systems can be estimated using Equation (7.118) on p. 406 of reference (2).

References

1) Marks, Eli S. 1978. The Role of Dual System Estimation in Census Evaluation. In Developments in Dual System Estimation of Population Size and Growth, edited by Karol Krotki, University of Alberta Press, Edmonton, Alberta, Canada.

2) ____________, William Seltzer, and Karol J. Krotki. 1974. Population Growth Estimation: A Handbook of Vital Statistics Measurement. New York, N. Y.: The Population Council.

3) Nathan, Gad, U. O. Schmelz, and J. Kenvin. 1977. Multiplicity Study of Marriages and Births in Israel. National Center for Health Statistics, Vital and Health Statistics, Series 2, Number 70. Washington, D.C.: U. S. Government Printing Office.

4) Schmelz, U. O.; G. Nathan; and J. Kenvin. 1977. Multiplicity Study of Births and Deaths in Judea-Samaria and Gaza Strip-North Sinai. Technical Publication Series No. 44. Central Bureau of Statistics, Jerusalem.

5) Sirken, Monroe J. 1970. Household Surveys with Multiplicity. Journal of the American Statistical Association, 65: 257-266.

APPENDIX

General formula for estimating completeness of coverage of the census from survey results:

$$(1)\quad \bar{y}_A = \frac{y_A}{n_A} = \frac{\sum_i^k \sum_j^{n_{Ai}} w_{Aij}\, y_{Aij}}{\sum_i^k \sum_j^{n_{Ai}} w_{Aij}} = \frac{\sum_j^{n_A} \sum_i^{k_{Aj}} w_{Aij}\, y_{Aij}}{\sum_j^{n_A} \sum_i^{k_{Aj}} w_{Aij}}$$

where $\bar{y}_A$ = estimated completeness of census enumeration of persons in class A;

y_A = number of persons in class A reported in the sample who were enumerated in the census;

n_A = number of persons in class A reported in the sample;

k = number of sample households;

n_{Ai} = number of persons in class A reported by i^{th} sample household 1/;

w_{Aij} = weight to be applied to j^{th} person in class A reported by the i^{th} sample household;

y_{Aij} = 1 if j^{th} person in class A reported by i^{th} sample household was enumerated in the census; = 0, if otherwise;

k_{Aj} = number of sample households reporting j^{th} sample person in class A (usually k_{Aj} = 1 for all j -- i.e., while the j^{th} person may be linked to several households, only one of these is in the sample).

The expected value of $\bar{y}_A$ is 2/:

$$(2)\ E\bar{y}_A = \bar{Y}'_A = \frac{Y'_A}{N'_A} = \frac{\sum_I^K \frac{k}{K} \sum_J^{N_{AI}} R_{AIJ}\, w_{AIJ}\, y_{AJ}}{\sum_I^K \frac{k}{K} \sum_J^{N_{AI}} R_{AIJ}\, w_{AIJ}}$$

$$= \frac{\sum_J^{N_A} y_{AJ} \sum_I^{K_{AJ}} R_{AIJ}\, w_{AIJ}}{\sum_J^{N_A} \sum_I^{K_{AJ}} R_{AIJ}\, w_{AIJ}}$$

where K = number of households in population;

K_{AJ} = number of households that should report the J^{th} person in class A (i.e., the number of households to which the person is linked);

N_{AI} = number of persons in class A who should be reported by $_I$th (population) household;

N_A = total number of persons in class A;

R_{AIJ} = probability (for fixed I,J) that I^{th} household will report the $_J$th person in class A;

w_{AIJ} = weight to be applied to the J^{th} person in class A who should if reported by the $_I$th (population) household;

y_{AJ} = 1 if the J^{th} person in class A was enumerated in the census; = 0 if otherwise[3/];

and:

$$(3)\ Y'_A = \sum_I^K \sum_J^{N_{AI}} R_{AIJ}\, w_{AIJ}\, y_{AJ} = \sum_J^{N_A} y_{AJ} \sum_I^{K_{AJ}} R_{AIJ}\, w_{AIJ}$$

$$(4)\ N'_A = \sum_I^K \sum_J^{N_{AI}} R_{AIJ}\, w_{AIJ} = \sum_J^{N_A} \sum_I^{K_{AJ}} R_{AIJ}\, w_{AIJ}$$

The completeness of census coverage for class A:

$$(5)\ \bar{Y}_A = \frac{Y_A}{N_A} = \frac{\sum_J^{N_A} y_{AJ}}{N_A}$$

Conditions under which $\bar{Y}_A$ will be unbiased:

$$(6)\ \sum_I^{K_{AJ}} R_{AIJ}\, w_{AIJ} = 1$$

$$(7)\ w_{AIJ} = \frac{1}{\sum_I^{K_{AJ}} R_{AIJ}} = \frac{1}{R_{AJ}} \text{ for all I}$$

$$(8)\ w_{AIJ} = \frac{1}{K_{AJ}\, R_{AIJ}}$$

In the case where there are no reporting errors (i.e., $R_{AIJ} = 1$ for all I,J):

$$(9)\ w_{AIJ} = \frac{1}{K_{AJ}}$$

Substituting (9) in equations (1) and (2) results in:

$$(10)\ \bar{y}_A = \frac{\sum_i^k \sum_j^{n_{Ai}} \frac{y_{Aij}}{K_{Aj}}}{\sum_i^k \sum_j^{n_{Ai}} \frac{1}{K_{Aj}}}$$

and:

$$(11)\ \bar{Y}'_A = \frac{Y'_A}{N'_A} = \frac{\sum_J^{N_A} y_{AJ}\, \bar{R}_{AJ}}{\sum_J^{N_A} \bar{R}_{AJ}}$$

where

$$\bar{R}_{AJ} = \frac{\sum_I^{K_{AJ}} R_{AIJ}}{K_{AJ}}$$ = average probatility of the J^{th} person in class A being reported survey.

Alternative way of expressing $\bar{Y}'_A$ since $y_{AJ} = 0$ for persons not enumerated in the census;

$$(12)\ E\bar{y}_A = \bar{Y}_A = \frac{\sum_J^{Y_A} \bar{R}_{AJ}}{\sum_J^{N_A} \bar{R}_{AJ}} = \frac{Y_A}{N_A} \frac{\bar{R}_{AE}}{\bar{R}_A} = \bar{Y}_A \frac{\bar{R}_{AE}}{\bar{R}_A}$$

where

$$\bar{R}_{AE} = \frac{\sum_J^{Y_A} \bar{R}_{AJ}}{Y_A}$$ = average probability of enumerated person in class A being reported in the survey

$$\bar{P}_A = \frac{\sum_J^{N_A} \bar{R}_{AJ}}{N_A}$$ = average probability of all persons in class A being reported in the survey

Multiplicity estimate expressed in terms of weighted averages of estimates of completeness of census enumeration for linkage groups:

$$(13)\ \bar{y}_A = \sum_H^M r_{AH}\, \bar{y}_{AH} = \sum_H^M \frac{1}{K_H} \sum_I^{K_H} r_{AHI}\, \bar{y}_{AHI}$$

$$= \frac{\sum_H^M \frac{1}{K_H} \sum_I^{K_H} \sum_j^{n_{AHI}} y_{AHIj}}{\sum_H^M \frac{1}{K_H} \sum_I^{K_H} n_{AHI}}$$

where M = number of linkage type groups

K_H = number of types of household to which persons in H^{th} group are linked

n_{AHI} = number of sample reports of group H persons in class A reported under the I^{th} counting rule.

$n_{AH} = \sum_I^{K_H} n_{AHI}$ = number of group H persons of class A reported in the sample

$$r_{AHI} = \frac{n_{AHI}}{\sum_{H}^{M} \frac{1}{K_H} \sum_{I}^{K_H} n_{AHI}}$$ = estimate of proportion of class A persons in H^{th} group based on reports under I^{th} counting rule

$$r_{AH} = \frac{n_{AH}}{K_H \sum_{H}^{M} \frac{1}{K_H} \sum_{I}^{K_H} n_{AHI}}$$ = estimate of the proportion of class A persons in H^{th} group based on all reports.

y_{AHIj} = enumeration state (=1 if enumerated, 0 if missed) for j^{th} group H person in class A reported under the I^{th} counting rule.

$$\bar{y}_{AHI} = \frac{\sum_{j}^{n_{AHI}} y_{AHIj}}{n_{AHI}}$$ = estimate of completeness of enumeration for group H persons in class A based on reports under I^{th} counting rule only.

$$\bar{y}_{AH} = \frac{\sum_{I}^{K_H} n_{AHI}\, \bar{y}_{AHI}}{n_{AH}} = \frac{\sum_{I}^{K_H} \sum_{j}^{n_{AHI}} y_{AHIj}}{\sum_{I}^{K_H} n_{AHI}}$$

= estimate of completeness of enumeration of group H persons in class A based on all sample reports of such persons

Simplex estimate obtained by using any of the K_H counting rules to give estimates of $\bar{Y}_{AH}$ and R_{AH}:

$$(14)\quad \bar{y}_{A1} = \sum_{H}^{M} r_{AHI}\, \bar{y}_{AHI} = \frac{\sum_{H}^{M} \sum_{j}^{n_{AHI}} y_{AHIj}}{\sum_{H}^{M} n_{AHI}}$$

Basing r_{AHI} on a different linkage from the estimate $\bar{y}_{AHI}$:

$$(15)\quad \bar{y}_{A2} = \sum_{H}^{M} r_{AHK}\, \bar{y}_{AHI} = \frac{\sum_{H}^{M} \frac{n_{AHK}}{n_{AHI}} \sum_{j}^{n_{AHI}} y_{AHIj}}{\sum_{H}^{M} n_{AHK}}$$

where $r_{AHK} = \frac{n_{AHK}}{\sum_{H}^{M} n_{AKH}}$ = estimate of proportion of all class A persons in group H based on reports under K^{th} counting rule.

Footnotes

1/ Note: If $n_{Ai} = 0$ the sums $\sum_{j}^{n_{Ai}} w_{Aij}$ and $\sum_{j}^{n_{Ai}} w_{Aij}\, y_{Aij}$ are zero--i.e., any sum over zero terms is equal to zero.

2/ Formulas shown are for a self-weighting sample. For nonself-weighting samples, k/K in equation (2) would have to be replaced by the factors appropriate to the sample design.

3/ y_{AJ} is treated here as a constant. Actually, it could be considered as a variable and we could define $Y_{AJ} = \bar{y}_{AJ}$ = probability that $_J$th person in class A who would be reported in the survey would be enumerated in a census taken under the conditions that prevailed for the census actually taken. This change would have only minimal effect on the model outlined here.