QUALITY CONTROL SAMPLING IN FOOD STAMP AND AFDC PROGRAMS

Robert Vogel and Donna Brogan, Emory University

I. Introduction

Georgia, like all states, is required by federal regulation to administer and monitor AFDC and Food Stamp Programs. In addition, they are required to report every six months on the estimated proportion of cases in error and the estimated average over-payment for each program. This paper describes some of the problems and proposed solutions involved in the construction of sampling plans for estimating these attributes in each program.

II. Definition of Sampling Frames and Estimation Procedures

In Georgia there are approximately 230,000 Food Stamp cases and 90,000 AFDC cases. Of those receiving public assistance from these programs, approximately 69,000 cases receive both benefits. Currently, there are two lists comprised of the names of the payees for each type of case, i.e. one list for food stamps and one for AFDC. Cross references between the two lists are unavailable, and there is no common list providing the names of cases receiving both benefits. In fact, the true status of a case in some instances cannot be determined until the field interview.

If we define the food stamp list and the AFDC list to be two distinct sampling frames, the easiest approach from an administrative point of view is to take two simple or systematic random samples, one from each list. Although this approach facilitates the organization and reporting of results, since each program has its own QC (quality control) workers who prefer to interview cases only in their programs, it is costly. The average cost in collecting data is approximately 438 minutes for a single AFDC case and 371 minutes for a single Food Stamp case, while only 685 minutes is needed to collect data on a food stamp and AFDC case simultaneously.

Therefore, an integrated sampling plan was proposed in order to take advantage of simultaneous interviews. Hartley's (1962, 1974) sampling plan was considered appropriate with some modification. The AFDC and Food Stamp lists constitute the two sampling frames, and they are partitioned with their 'overlap' into three mutually exclusive domains as specified below:

domain f: N_f cases with food stamps only
domain a: N_a cases with AFDC only
domain fa: N_{fa} cases with both AFDC and food stamps

Let N_F and N_A denote the frame size of the food stamp and AFDC list, respectively. Note that $N_F = N_f + N_{fa}$ and $N_A = N_{af} + N_{fa}$. For the Georgia data base,

$N_F \doteq 230{,}000$ $\quad N_A \doteq 90{,}000$

$N_f \doteq 161{,}000$ $\quad N_{fa} \doteq 69{,}000$ $\quad N_a \doteq 21{,}000$

Assume now that a simple or systematic random sample is selected independently from each of the two frames, with sample size n_A and n_F from the AFDC and food stamp frames, respectively. If a sample AFDC case is found to have food stamps, then the QC interview on food stamps is done also. Similarly, if a sample food stamp case is found to have AFDC, then the QC interview on AFDC is done also. The sample then is poststratified into four mutually exclusive domains:

domain f: case has food stamps only
domain a: case has AFDC only
domain fa': case selected from food stamp frame but also has AFDC
domain fa'': case selected from AFDC frame but also has food stamps

Let the sample means in the four post-strata be $\bar{y}_f$, $\bar{y}_a$, $\bar{y}'_{fa}$ and $\bar{y}''_{fa}$, where the variable y can be dollar error or can be a (1, 0) variable indicating error or no error. Note that both $\bar{y}'_{fa}$ and $\bar{y}''_{fa}$ estimate the population parameter $\bar{Y}_{fa}$ in the domain of size N_{fa}. They will be weighted with weights p and q (where $p + q = 1$) to estimate $\bar{Y}_{fa}$. Hartley's (1962) second case is used to estimate the population total Y, i.e.

$$(1)\quad \hat{Y}_H = N_f\bar{y}_f + N_{fa}(p\bar{y}'_{fa} + q\bar{y}''_{fa}) + N_a\bar{y}_a$$

and, ignoring fpc,

$$(2)\quad V(\hat{Y}_H) \doteq N_F^2\left\{\sigma_f^2(1-\alpha) + p^2\sigma_{fa}^2\alpha\right\}/n_f + N_A^2\left\{\sigma_a^2(1-\beta) + q^2\sigma_{fa}^2\beta\right\}/n_A$$

where $\alpha = N_{fa}/N_F$, $\beta = N_{fa}/N_A$ and σ_f^2, σ_a^2 and σ_{fa}^2 are the population variances of the y variable in the three population domains f, a, and fa.

Hartley's intent was to use two or more frames to get better coverage when sampling for a single attribute. In our case, though, each list gives 100% coverage for each benefit, and our intent is to use the 'overlap' as a means of cost reduction. Thus, when the variable y refers to a food stamp variable, $N_a = 0$. Likewise, when the variable y refers to an AFDC variable, $N_f = 0$. Thus, equations (1) and (2) can be particularized for food stamps (F) and AFDC (A) as follows:

$$(3)\quad \hat{Y}_{HF} = N_f\bar{y}_f + N_{fa}(p\bar{y}'_{fa} + q\bar{y}''_{fa})$$

$$(4)\quad V(\hat{Y}_{HF}) \doteq \frac{N_F^2}{n_F}\left\{\sigma_f^2(1-\alpha) + p^2\sigma_{Ffa}^2\alpha\right\} + \frac{N_A^2q^2\beta\sigma_{Ffa}^2}{n_A}$$

and

$$(5)\quad \hat{Y}_{HA} = N_a\bar{y}_a + N_{fa}(p\bar{y}'_{fa} + q\bar{y}''_{fa})$$

$$(6)\quad V(\hat{Y}_{HA}) \doteq \frac{N_A^2}{n_A}\left\{\sigma_a^2(1-\beta) + q^2\beta\sigma_{Afa}^2\right\} + \frac{N_F^2p^2\alpha\sigma_{Afa}^2}{n_F}$$

where σ_{Ffa}^2 is the variance of the 'overlap' when sampling for food stamps and σ_{Afa}^2 is the variance of the 'overlap' when sampling for AFDC. Note, that $\sigma_{Afa}^2 = \sigma_{Ffa}^2 = \sigma_{fa}^2$.

Solving equations (4) and (6) to find the optimal values of p and q yield the same results, namely:

(7) $p = n_F N_A (n_F N_A + n_A N_F)^{-1}$

(8) $q = n_A N_F (n_F N_A + n_A N_F)^{-1}$

III. Definition of Precision (Sample Size)

The mimimum sample size for a simple or systematic random sample required by each sponsoring federal agency is 1200 whenever there are more than 60,000 cases in the frame. Since cost is a function of sample size and since it is desired to minimize cost, the minimum required sample size was always used, i.e. a systematic random sample of size 1200 from each frame. Thus, the precision of the proposed integrated plan is compared to the precision obtained from a systematic random sample of size 1200 from each frame. If the half width of the 95% confidence interval obtained by the proposed plan is less than or equal to that obtained by a systematic random sample of size 1200, then the precision of the plan is considered acceptable. Now, in order also to obtain the minimum required sample sizes needed from each frame, i.e. 1200, the following system of equations is solved for N_A and N_F:

$$1200 = n_A + \alpha_{nF}$$

$$1200 = n_F + \beta_{nA}$$

Thus, if a simple or systematic random sample of size n_A is selected from the AFDC frame, and a simple or systematic random sample of size n_F is selected from the food stamp frame, and if sample cases in the overlap frame are interviewed by the QC worker on both AFDC and food stamps, then about 1200 interviews will be obtained on food stamp cases and 1200, likewise, on AFDC cases.

IV. A Problem with Overlap

An additional problem needs to be addressed in regard to the 'overlap'. As mentioned earlier, the true status of the case cannot always be determined until the field worker interviews the recipient. Occasionally there are some problems in defining the overlap because a Food Stamp case is not always congruent with the AFDC case with which it overlaps.

The simple situation is one in which the AFDC case and the Food Stamp case comprise the same set of persons. Thus, this case is in the overlap domain and contributes one case to N_{fa}. A typical non-congruent situation is one in which the AFDC case is included in a larger group of persons which comprise the Food Stamp case. Here, the AFDC case and the Food Stamp case belong to the 'overlap' domain and contribute one case to N_{fa}.

A more unusual but possible situation is one in which two or more AFDC cases comprise a single Food Stamp case or are included in a larger group of persons that comprise a single Food Stamp case. For example, consider a Food Stamp case which includes n ($n \geq 2$) separate AFDC cases, plus possibly other persons who are not included in the n AFDC cases. This type of case contributes n AFDC cases to N_{fa}. If a Food Stamp case is selected from the food stamp frame and the field worker finds that it contains within it n AFDC cases, then the field worker adds to the sample all n AFDC cases. However, if an AFDC case is selected from the AFDC frame, and the field worker finds that the selected AFDC case belongs to a Food Stamp case which also includes (n-1) other AFDC cases ($n \geq 2$), then two alternatives are possible to adjust for the fact that this Food Stamp case has n chances of being selected via the sample from the AFDC frame. One alternative is to always include the Food Stamp case in the sample but then weight it by 1/n to adjust for its higher probability of selection. The second alternative is for the field worker to select the Food Stamp case into the sample with probability 1/n via a random number table. Of course, the other (1-1) AFDC cases would not be included unless they were also selected from the AFDC frame or the actual Food Stamp case was selected from the food stamp frame.

V. Numerical Example

A comparison of the proposed sampling plan with the current sampling plan, based on actual data collected over a six-month audit period in Georgia, follows. First, let the variable y for food stamps be 1 if an error in payment is present and 0 if no payment error is present. The sample data are:

$$\bar{y}_f = .21, \quad \bar{y}'_{fa} = .26, \quad \bar{y}''_{fa} = .32$$

$$s^2_f = .166, \quad s^{2'}_{fa} = .192, \quad s^{2''}_{fa} = .218$$

Solving equation (9) yields

$$n_F = 364, \quad n_A = 1091$$

By equation (4), using the sample values,

$$v(\hat{Y}_{HF})/(230,000)^2 = .000339$$

and the half width of the 95% confidence interval is:

$$(1.96)\sqrt{.000339} = .03608$$

The overall sample proportion of Food Stamp cases in error is .241. The half width of the 95% confidence interval based on a systematic random sample of size 1200 is:

$$(1.96)\sqrt{(.241)(.759)/1199} = .02421$$

Letting now the variable y be a (1, 0) variable for an error in the AFDC payment, the sample data are:

$$\bar{y}_a = .108, \quad \bar{y}'_{fa} = .128, \quad \bar{y}''_{fa} = .116$$

$$s^2_a = .096, \quad s^{2'}_{fa} = .112, \quad s^{2''}_{fa} = .103$$

By equation (6), using the sample values, we obtain

$$v(\hat{Y}_{HA})/(90,000)^2 = .0000852$$

and the half width of the 95% confidence interval is:

$$(1.96)\sqrt{.0000852} = .01809$$

The overall sample proportion of cases in error for AFDC is .1154. The half width of the 95% confidence interval based on a systematic random sample of size 1200 is:

$$1.96\sqrt{(.1154)(.8846)/1199} = .01809$$

In comparing the cost of the two methods, the expected cost of two systematic random samples of size 1200 is: 1200 (371 + 438) = 971414 minutes or 16190 hours. The expected cost of the proposed integrated plan is: 255(371 + 438) + 945 (685) = 854225 minutes or 14237 hours. Hence, the total savings by using the proposed method is approximately 1953 hours, a 13% reduction in hours.

In summary, the proposed integrated plan yields equal precision for AFDC, poorer precision for food stamps, and costs approximately 13% less per six months period to administer. Since the major concern is to reduce cost, this plan meets that objective. The question is whether or not this savings is worth the reduced precision in the food stamp estimate. Note that the federal agencies, in essence, required two precision criteria to be met for the point estimate of error rate: minimum sample size of 1200 and half-width of confidence interval not to exceed that obtained with a systematic random sample of size 1200. Logically, only one of these should be required, and, most likely, the latter. It can be shown that the precision of the error rate estimate for food stamps can be increased by decreasing n_A below 1091 and increasing n_F beyond 364. This will result in fewer than 1200 AFDC cases in the resulting integrated sample, although the precision of the error rate estimate for AFDC does not suffer very much and the precision is increased for food stamps. The cost of this integrated plan is considerably less than two independent systematic random samples of size 1200.

Additional activity currently underway will reduce the joint interview time below 685 minutes, thus giving the integrated sampling plan even more of an economic advantage.