AN ESTIMATOR OF CHANGE FROM AN UPDATED FRAME

Chapman P. Gleason, U.S. Department of Agriculture

Introduction

Estimation of totals and change is a basic procedure from periodic surveys. Current text books (Cochran [1], Raj [6], Hansen, Hurwitz, Madow [3]) treat only the case of successive sampling from finite populations when the sampling frame remains fixed between occasions. Lowerre [5] treats the birth/death estimation problem by using regression estimators for estimates of totals on the second occasion. Population sizes on both occasions are unknown. The case where population sizes are known (in all strata) on both occasions and where the number of births and deaths to the "old" sampling frame to construct the "new" sampling frame are also known will be treated here. A combined ratio estimate will be used to estimate the total and the percent change on the current occasion.

As an example consider a frame of business, units can:

1) merge with another unit,
2) spin off a subsidary,
3) go out of business, or
4) a new unit can be created and go into business.

In general only two types of changes can happen to a sampling frame; a frame can:

1) add units (called birth or adds), or
2) loose units (called deaths or deletes).

Any type of merger or subsidary spinoff of a new company or business enterprise can be treated as a death to the "old" sampling frame and an add to the "new" sampling frame. A unit which remains the same on each sampling frame is said to be "static." Changes in stratum membership are treated as deletes to one stratum and adds to a stratum on the new sampling frame. The mathematical and statistical formulation of the problem will be given in the next section.

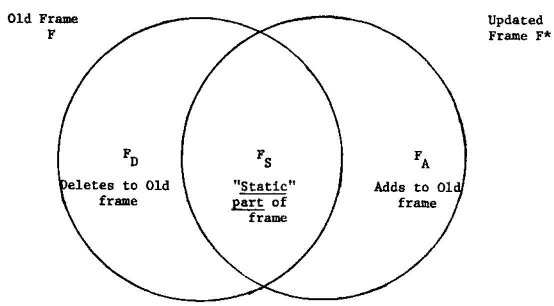
Sampling Design and Estimators

Let F be a stratified sampling frame at a previous point in time (call it the "old" frame), and let F* be the sampling frame F updated with adds F_A, and deletes F_D. Suppose the number of strata and the stratum definitions do not change between updates. However, the i^{th} unit can change stratum membership in the update process. So in set notation,

$$F^* = (F - F_D) \cup F_A.$$

Conceptually the following figure depicts the situation, $F_S = (F \cap F^*)$ is the "static" part of the frame

Figure 1



Let S be a sample of size n selected from frame F on the previous occasion and S* be an independently drawn sample of size n* from frame F* on the current occssion, further let S' be a subsample of size n' selected from S. Let S_A be a sample of size n_A for F_A.

Enumerate the samples S', S* and S_A on the current occasion. The previous sample S has already been collected. The subscript h will denote stratum membership. Let y^*_{hi} be the value of the i^{th} unit in the h^{th} stratum on the current occasion selected from the S* sample, let y'_{him} and y_{him} be the value of a variable on the i^{th} unit in the S' subsample on the current occasion and the previous occasion respectively. The m subscript denotes matched.

Further let the number of units in the h^{th} stratum on the sampling frame F be N_h, similarily define the number on the frame F* as N^*_h, and let N_{hA} be the population size of the add frame F_A for the h^{th} stratum. The design can be summarized as:

	Frame F	Frame F*	Frame F_A
Population size h^{th} stratum	N_h	N^*_h	N_{hA}
Samples selected from frame	previous, current S,S'	S*	S_A
Sample size in h^{th} stratum	n_h, n'_h	n^*_h	n_{hA}
Measurement for i^{th} unit in the h^{th} stratum	y_{him}, y'_{him}	y^*_{hi}	y_{hAi}

We will consider two estimators of the population total for the frame F*. The first: a simple unbiased estimator of the total is based on the sample S* from frame F*. The second estimator is a combined ratio estimator; the S' subsample of S estimates for the population F_S and the S_A sample is used to estimate for the population F_A.

Mathematically, a more complex description of the estimators is given below. The simple unbiased estimator of the total for the F* frame on the current occasion using only the S* sample is

$$(1)\quad \hat{Y}_{F^*} = \sum_{h=1}^{L} N^*_h \bar{y}_{h^*},$$

where,

$$y^*_h = \sum_{i=1}^{n^*_h} y^*_{hi}/n^*_h \text{ i.e. the } h^{th} \text{ stratum mean.}$$

Another estimator of the frame F* total based on the sample of adds (S_A) and the subsample S' is the combined ratio estimate:

$$(2)\quad \hat{Y}_{R,F^*} = \hat{Y}_{F_A} + \hat{Y}_{F_S}$$

where,

$$\hat{Y}_{F_A} = \sum_{h=1}^{L} N_{hA} \bar{y}_{hA},$$

$$\bar{y}_{hA} = \sum_{i=1}^{n_{hA}} y_{hAi}/n_{hA},$$ (add sample mean for the h^{th} stratum) and

$$(3)\quad \hat{Y}_{F_S} = \hat{R}\,\hat{Y}_F = \left(\frac{\sum_{h=1}^{L} N_h \bar{y}'_{hm}}{\sum_{h=1}^{L} N_h \bar{y}_{hm}}\right)\left(\sum_{h=1}^{L} N_h \bar{y}_h\right).$$

So the estimator $\hat{Y}_{R,F*}$ is the sum of a combined ratio estimate of the portion of the frame which is static from the previous occasion to the current occasion (F_S) and a simple unbiased estimator of the add population total.

Other estimators could also be used for the situation just described. Consider the frames F_D, F_S, and F_A as the domains a, ab (ba) and b in the multiple frame situation described by Hartley [2]. Hartley's estimator could also be used to estimate for the frame F* total.

Variances

In this section we show the variances of the estimators, (1) and (2). Now the variance of (1) is the usual variance of a simple linear unbiased estimator of the total, namely:

$$(4)\quad \mathrm{Var}(Y_{F*}) = \Sigma\, N_h^{*2}(1 - f_h^*)\,\frac{\mathrm{var}(y_h^*)}{n_h^*}$$

where $f_h^* = n_h^*/N^*$ = the finite population correction factor for the updated frame F*. We now derive the variance of the estimator (2).

$$\mathrm{Var}(\hat{Y}_{R,F*}) = \mathrm{Var}(\hat{Y}_{F_A} + \hat{Y}_{F_S})$$

$$= \mathrm{Var}(\hat{Y}_{F_A}) + \mathrm{Var}(\hat{Y}_{F_S}),$$

since the add sample S_A was selected independent of S'.

Now,

$$\mathrm{Var}(\hat{Y}_{F_S}) = \Sigma\, N_{hA}^{\;2}\,(1 - f_{hA})\,\frac{\mathrm{Var}(\bar{y}_{hA})}{n_{hA}}$$

where $f_{hA} = n_{hA}/N_{hA}$, the sampling fraction for the add frame. Now consider

$$\mathrm{Var}(\hat{Y}_{F_S}) = \mathrm{Var}(\hat{R}\,\hat{Y}_F),$$

where, R is defined in (3) and, $\hat{Y}_F = \Sigma N_h \bar{y}_h$.

$\hat{R}\,\hat{Y}_F$ is a double sampling combined ratio estimator of the total for frame F on the 2nd occasion. But noting that units that were deleted from the frame F are 0 on the current occasion, and since $F = F_D \cup F_S$, any estimator of F_D on the current occasion is 0, therefore $\hat{R}\,\hat{Y}_F$ estimates F_S for the current occasion.

Since S' was a subsample of S and estimates the total Y for the frame F on the 2nd occasion, generalizing the variance formula which appears in Knoijn [4, page 126] to a combined ratio estimator for a stratified random sample we get

$$(5)\quad \mathrm{Var}(\hat{R}\,\hat{Y}_F) = \sum_{h=1}^{L} N_h^2 \left\{ \frac{N_h - n_h}{N_h}\cdot\frac{\mathrm{Var}(y'_h)}{n_h} \cdot \frac{(n_h - n'_h)}{n'_h}\,\frac{1}{} \cdot (\mathrm{Var}(y'_h) - 2\,\hat{R}\,\mathrm{Cov}(y'_h, y_h) + \hat{R}^2\,\mathrm{Var}(y_h)) \right\}$$

We note that if $n_h = n'_h$ (i.e. reinterview the entire sample S), then (3) becomes

$$\hat{Y}_{F_S} = \frac{\Sigma\, N_h \bar{y}'_m}{(\Sigma\, N_h\, y_{hm})}\,(\Sigma\, N_h \bar{y}_h) = \Sigma\, N_h \bar{y}'_h)$$

and (5) becomes the variance of the appropriate linear unbiased estimator. Now if $n_h = N_h$ (i.e. the previous survey was a census) the estimator (3) is the ordinary combined ratio estimator with a known total, (5) reduces to the appropriate formula given in Cochran [1, page 176 formula (6.25)].

Application

The Economics, Statistics, and Cooperative Service conducts a quarterly survey in December, March, June and September to estimate hog inventories. A sample is selected from a stratified list frame, this sample is used four quarters with simple linear unbiased estimates ($\sum_{h=1}^{L} N_h \bar{y}_h$) made of totals for each item each quarter. The list frame is updated with births, deaths, and stratum changes after the September survey and a completely new sample is drawn from the updated frame for the next sequence of quarterly surveys. In the theory section developed earlier the old frame is F and the updated frame is F*.

A State Statistical office agreed to collect data from a subsample of their September 1979 sample (S' in the theory developed earlier) in addition to the required sample for December 1979 (S*). Data was already available for the entire September 1979 sample (S). Further, adds to the sampling frame F were identified and a Sample S_A selected from this frame (F_A in our theory section). We now have the necessary data to compute the estimators (1) and (2) and compare their precision using (4) and (5). Table 1 gives the estimates and their precision.

Table 1. Estimates of Total Hogs

Occasion	Estimator (000)	Estimate (000)	S.E. (000)	C.V. (%)	Sample Size
Previous	$\hat{Y}_F$	3,265	149	4.6	1801
Current	$\hat{Y}_{F*}$	3,650	113.9	3.1	1800
Current	$\hat{Y}_{R,F*}$	3,549	493	13.9	507

Note that the sample size on the estimator $\hat{Y}_{R,F*}$ is about 3.5 times smaller than the sample size for estimator $\hat{Y}_{F*}$. For a fixed size sample, estimator $\hat{Y}_{R,F*}$ is more precise than $\hat{Y}_{F*}$, since

$$\rho > 1/2 \frac{CV(\text{Previous occasion})}{CV(\text{Current occasion})}, \text{ in all strata}$$

[1, p. 165]. This is shown in Table 2.

Table 2. Correlation Coefficients between September and December 'total hogs and pigs' based on subsample S' of September sample S.

Stratum Definition (No. hogs)	Correlation Coefficient	$\frac{1}{2}\left[\frac{CV(\text{September})}{CV(\text{December})}\right]$
No livestock	1.00	.75
No hogs	.83	.54
1-99	.79	.50
100-199	.63	.58
200-299	.92	.56
300-399	.69	.49
600-1249	.97	.53

Conclusion

This paper attempts to apply existing theory to an important problem in estimation from periodic surveys when the sampling frame changes between survey periods. More research is needed in the following areas:

1) More efficient estimation of both change and totals on the current occasion.
2) Optimum allocation of the sample (given the estimators of course).

Bibliography

[1] Cochran, W.G., "Sampling Techniques," John Wiley and Sons, New York, 2nd Edition.

[2] Hartley, H.D., "Multiple Frame Surveys," Proceedings of the Social Statistics Section, ASA, 1962.

[3] Hansen, M.H., Hurwitz, W.N., and Madow, W.G., "Sampling Survey Methods and Theory," Vol I, John Wiley and Son Inc., New York.

[4] Konign, H.S., "Statistical Theory of Sample Survey Design and Analysis" North-Holland, New York, New York.

[5] Lowerre, James M. "Sampling For Change," Proceeding on the Social Statistics Section, Joint Statistical Meetings, Washington, D.C., 1979.

[6] Raj. D., Sampling Theory. McGraw Hill, 1968.