INADMISSIBILITY OF CUSTOMARY ESTIMATORS IN SAMPLING OVER TWO OCCASIONS

By P. K. PATHAK* and T. J. RAO

Indian Statistical Institute

SUMMARY. Inadmissibility of customary estimators of population total in sampling over two occasions is demonstrated by providing more efficient estimators under some well-known schemes of sampling over two occasions.

1. Improved estimators for SRS scheme

In this paper, we consider two schemes of sampling over two occasions: (1) when simple random sampling (srs) (without replacement) is used on both the occasions and (2) when probability proportional to size (pps) sampling is used on both the occasions.

Under scheme 1, a sample of size n is selected by simple random sampling (without replacement) from the population on the first occasion. Then on the second occasion m units of the first sample are retained and $(n-m)$ units are selected independently by simple random sampling (without replacement) from the whole population. For estimating the total of the population on the second (current) occasion, the customary estimator (Cochran, 1963) of the population total is given by

$$t_2 = \phi(N\bar{y}_{2u})+(1-\phi)N\bar{y}'_{2m}$$

where

N = population size,

$\bar{y}_{hu}$ = mean of the unmatched portion on occasion $h (h = 1, 2)$,

$\bar{y}_{hm}$ = mean of matched portion on occasion $h (h = 1, 2)$,

$\bar{y}_h$ = mean of the whole sample on occasion $h (h = 1, 2)$,

$\bar{y}'_{2m} = \bar{y}_{2m}+b(\bar{y}_1-\bar{y}_{1m})$, is the regression estimate based on the matched portion

and ϕ is determined such that the variance of t_2 is minimised.

Now, let the $(n-m)$ units selected independently on the second occasion be represented by $s = (s_1, s_2)$, where $s_1 = (u_{11}, u_{12}, \ldots, u_{1m_2})$ denotes the sample units of s which come from the matched portion and $s_2 = (u_{21}, u_{22}, \ldots, u_{2l_2})$ denotes the remaining units.

The following theorem now provides an estimator more precise than t_2 and thus proves the inadmissibility of t_2.

Theorem 1 : *Let*

$$t_2^* = N\phi\left\{\frac{m_2\,\bar{y}_{2m}+l_2\,\bar{y}_{2l_2}}{m_2+l_2}\right\}+N(1-\phi)\bar{y}'_{2m},$$

*Now with University of Illinois, U.S.A.

7

where $\bar{y}_{2l_2}$ = mean of the units based on the sample s_2. Then $E(t_2^*) = E(t_2)$ and for any convex loss function t_2^* does not have greater expected loss than t_2.

Proof: We have

$$E[t_2 | s_2, s_m] = \frac{N\phi}{l_2+m_2}[m_2 E(\bar{y}_{2m_2} | s_m) + l_2 \bar{y}_{2l_2}] + N(1-\phi)\bar{y}'_{2m} = t_2^*, \qquad \ldots (1)$$

where $\bar{y}_{2m_2}$ is the mean of units based on the sample s_1.

Therefore,

$$E(t_2^*) = E(t_2) \text{ by (1)}.$$

Now as a consequence of Jensen's inequality, it follows from (1) that t_2^* does not have greater expected loss than t_2.

This completes the proof of the theorem.

Although the main object of this paper is to demonstrate the inadmissibility of the customary estimators of the population total in sampling over two occasions, we, however, provide below for completeness (Corollaries 1 and 2) the gain in efficiency of t_2^* over t_2 and an upper bound to it.

Corollary 1: *If squared error is the loss function, the gain in efficiency on using t_2^* is given by*

$$E(t_2^*-t_2)^2 = N^2\phi^2S^2\frac{m-1}{N-1}\left(\frac{1}{n-m}-\frac{1}{N}\right).$$

Proof:

$$E(t_2^*-t_2)^2 = N^2\phi^2 E\left[\frac{m_2^2}{(l_2+m_2)^2}(\bar{y}_{2m_2}-\bar{y}_{2m})^2\right]$$

$$= \frac{N^2\phi^2}{(n-m)^2} E[Em_2^2(\bar{y}_{2m_2}-\bar{y}_{2m})^2 | m_2; s_m]$$

$$= \frac{N^2\phi^2}{(n-m)^2} E\left[m_2\left(1-\frac{m_2}{m}\right)S_m^2\right]$$

$$= \frac{N^2\phi^2S^2}{(n-m)^2}\left[(n-m)\frac{m}{N}-\frac{1}{m}\{E^2(m_2)+V(m_2)\}\right]$$

$$= \frac{N^2\phi^2S^2}{(n-m)^2}\left[(n-m)^2\frac{m}{N}\left\{\frac{1}{n-m}-\frac{1}{N}-\frac{1}{m}\left(\frac{1}{n-m}-\frac{1}{N}\right)\frac{N-m}{N-1}\right\}\right]$$

$$= N^2\phi^2S^2\frac{m}{N}\left(\frac{1}{n-m}-\frac{1}{N}\right)\frac{m-1}{N-1}\cdot\frac{N}{m}$$

$$= N^2\phi^2S^2\frac{m-1}{N-1}\left(\frac{1}{n-m}-\frac{1}{N}\right)$$

Corollary 2 : *An upper bound to the relative gain in efficiency of* t_2^* *over* t_2 *is equal to* $\frac{u}{n}\cdot\frac{m-1}{N-1}\Big/1-\frac{u}{n}\cdot\frac{m-1}{N-1}$ *when* $\phi < \frac{u}{n}$.

Proof : We have from Cochran (1963)

$$V(t_2) = N^2S^2\left(\frac{n-u\rho^2}{mn}-\frac{1}{N}\right)\left(\frac{1}{u}-\frac{1}{N}\right)\Big/\left(\frac{n^2-u^2\rho^2}{mnu}-\frac{2}{N}\right)$$

where ρ is the correlation between units over two occasions and $u = n-m$. The relative gain in efficiency of t_2^* over t_2 is defined to be

$$G = \frac{E(t_2^*-t_2)^2}{V(t_2^*)}.$$

Since $V(t_2^*) = V(t_2)-E(t_2-t_2^*)^2$, we get by virtue of Corollary 1

$$G = \frac{\phi[(m-1)/(N-1)]}{1-\phi[(m-1)/(N-1)]} \text{ where } \phi = \left(\frac{n-u\rho^2}{mn}-\frac{1}{N}\right)\Big/\left(\frac{n^2-u^2\rho^2}{mun}-\frac{2}{N}\right).$$

In practice $\frac{n}{N} < \frac{1}{4}$ and $\rho > \frac{1}{2}$ so that $\phi < \frac{u}{n}$. Therefore

$$G < \frac{u}{n}\cdot\frac{m-1}{N-1}\Big/1-\frac{u}{n}\cdot\frac{m-1}{N-1}.$$

Further in practice we have

$$\frac{m}{n} < \frac{1}{2} \text{ and so } \frac{m-1}{N-1} = \frac{m-1}{n-1}\cdot\frac{n-1}{N-1} < \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}$$

so that
$$G < \frac{u}{8n}\Big/\left(1-\frac{u}{8n}\right) = \frac{1}{8}\left(1-\frac{m}{n}\right)\Big/\left(1-\frac{1}{8}\left(1-\frac{m}{n}\right)\right) < 7\% \text{ approximately.}$$

When one is dealing with finite populations and the sampling fraction is large, a modified sampling scheme has been suggested by the referee in which $(n-m)$ units are selected on the second occasion from the $(N-n)$ units in the population not sampled on the first occasion. It seems that the estimator similar to t_2 under the new scheme would fare better than t_2^*. Since our aim has been to demonstrate the inadmissibility of the customary estimators, we do not go into detailed comparisons of these estimators.

2. Improved estimators for pps scheme

We now consider the other sampling scheme over two occasions where pps sampling is used on both the occasions (Des Raj, 1965). In this scheme, a sample s_f of size n is selected by pps with replacement on the first occasion and on the second occasion a simple random sample of size m is selected without replacement from s_f and an independent sample of $(n-m)$ units is selected by pps with replacement from the whole population.

As an estimator of the population total on the second occasion Des Raj considers the following estimator :

$$z_2 = \phi z_{2u} + (1-\phi) z'_{2m}$$

where

$$z_{hu} = \frac{1}{n-m} \Sigma \frac{y_{hi}}{p_i}$$

is the estimate of the population total based on the unmatched sample, on occasion $h (h = 1, 2)$,

$$z_{hm} = \frac{1}{m} \Sigma \frac{y_{hi}}{p_i}$$

is the estimate of the population total based on the matched sample, on occasion $h (h = 1, 2)$,

$$z_h = \frac{1}{n} \Sigma \frac{y_{hi}}{p_i}$$

is the estimate of the population total based on the whole sample, on occasion $h (h = 1, 2)$ and

$$z'_{2m} = (z_{2m} - z_{1m}) + z_{1n}$$

is the difference estimate of the population total on the second occasion, based on the matched portion.

Now, let s_m denote the matched sample and represent the unmatched sample on the second occasion by $s = (s_1, s_2)$, where $s_1 = (u_{11}, u_{12}, \ldots, u_{1m_2})$ denotes the sample units which come from the matched portion and $s_2 = (u_{21}, u_{22}, \ldots, u_{2l_2})$ denotes the remaining units. We then have the following theorem.

Theorem 2 : *Let*

$$z_2^* = \frac{\phi}{n-m} \Big[\sum_2 (y_{2i}/p_i) + \Big(m_2 \Big(\sum_{i=1}^{m} y_{2i} \Big) \Big) \Big/ \sum_{i=1}^{m} p_i \Big] + (1-\phi) z'_{2m},$$

where Σ_2 denotes the summation over the units in s_2 and the summation in the second term within braces is taken over the matched portion. Then $E(z_2^*) = E(z_2)$ and for any convex loss function z_2^* does not have greater expected loss than z_2.

Proof : We have

$$E[z_2 \big| s_2, s_m] = (1-\phi) z'_{2m} + \frac{\phi}{n-m} \Big[\Sigma_2 \frac{y_{2i}}{p_i} + E \sum_{s_1} \frac{y_{2i}}{p_i} \Big| s_m \Big]$$

$$= z_2^*.$$

$$\therefore \quad E(z_2^*) = E(z_2).$$

That z_2^* does not have greater expected loss than z_2 for any convex loss function follows from Jensen's inequality.

Corollary 3 : *If squared error is the loss function then the gain in efficiency in using z_2^* is given by*

$$E(z_2-z_2^*)^2 = \frac{\phi_m^2}{(n-m)^2}\Big\{\Sigma Y_i^2 - \underset{i\neq j}{\Sigma\Sigma} Y_iY_j(P_i+P_j) + (m-1)\underset{i\neq j}{\Sigma\Sigma}\frac{Y_i^2P_j^2}{P_i}$$

$$-(m-2)\underset{i\neq j\neq k}{\Sigma\Sigma\Sigma} Y_iY_jP_k^2\Big\}$$

Proof : $E(z_2-z_2^*)^2 = c^2E\Big[\sum_1^{m_2}\Big((y_{2i}/p_i)-\Big(\sum_1^m y_{2i}\Big/\sum_1^m p_i\Big)\Big)\Big]^2$, where $c = \dfrac{\phi}{n-m}$

$$= c^2E\left[E\left\{\frac{1}{(\Sigma p_i)^2}\left(\sum_1^{m_2}\left(\frac{y_{2i}}{p_i/\Sigma p_i}-\Sigma y_{2i}\right)\right)^2\Big/ m_2, s_m\right\}\right]$$

$$= c^2E\left[\frac{m_2}{(\Sigma p_i)^2}\,V\left(\frac{y_{2i}}{p_i/\Sigma p_i}\right)\right]$$

$$= c^2E\left[\frac{m_2}{(\Sigma p_i)^2}\cdot\frac{1}{2m(m-1)}\sum^m\underset{i\neq j}{\sum^m}\left(\frac{y_{2i}}{p_i/\Sigma p_i}-\frac{y_{2j}}{p_j/\Sigma p_j}\right)^2\right]$$

$$= c^2E\left[\frac{\left(\sum_{k=1}^m p_k\right)}{2(m-1)}\sum^m\underset{i\neq j}{\sum^m}\left(\frac{y_{2i}}{p_i}-\frac{y_{2j}}{p_j}\right)^2\right]$$

$$= \frac{c^2}{2(m-1)}\left[2(m-1)E\left\{(\Sigma p_i)\,\Sigma\frac{y_i^2}{p_i^2}\right\}-2E\left\{(\Sigma p_i)\underset{i\neq j}{\Sigma\Sigma}\frac{y_iy_j}{p_ip_j}\right\}\right]$$

$$= \frac{c^2}{2(m-1)}\Big[2(m-1)\Big\{m\,\Sigma y_i^2 + m(m-1)\underset{i\neq j}{\Sigma\Sigma}\frac{Y_i^2P_j^2}{P_i}\Big\}$$

$$-2\Big\{m(m-1)\underset{i\neq j}{\Sigma\Sigma} Y_iY_jP_i + m(m-1)\underset{i\neq j}{\Sigma\Sigma} Y_iY_jP_j$$

$$+m(m-1)(m-2)\underset{i\neq j\neq k}{\Sigma\Sigma\Sigma} Y_iY_jP_k^2\Big\}\Big]$$

$$= \frac{\phi_m^2}{(n-m)^2}\Big[\Sigma Y_i^2 - \underset{i\neq j}{\Sigma\Sigma} Y_iY_j(P_i+P_j) + (m-1)\underset{i\neq j}{\Sigma\Sigma}\frac{Y_i^2P_j^2}{P_i}$$

$$-(m-2)\underset{i\neq j\neq k}{\Sigma\Sigma\Sigma} Y_iY_jP_k^2\Big] \text{ on simplification.}$$

A sampling scheme which would probably fare better than the above described scheme of sampling over two occasions can be suggested in a manner similar to that given in the equal probability case.

ACKNOWLEDGEMENT

The authors are grateful to the referee for his valuable suggestions.

REFERENCES

COCHRAN, W. G. (1963) : *Sampling Techniques*, (2nd Edition), Wiley, New York.

DES RAJ (1965) : On sampling over two occasions with probability proportionate to size. *Ann. Math. Stat.*, 36, 327–30.

Paper received : August, 1966.