ON THE USE OF HORVITZ-THOMPSON ESTIMATOR IN MIXED SAMPLE SURVEY DESIGNS

M. Ahmad, Uni. of Petroleum & Minerals, Dhahran, S. Arabia
M. Hanif, El-Fateh University, Tripoli, Libya

Abstract

A general theory of sampling with unequal probabilities with replacement was developed by Hansen and Hurwitz (1943) and without replacement by Horvitz and Thompson (1952). Hanif and Brewer (1979) introduced a single generalized Horvitz-Thompson (GHT) estimator of total with or without replacement where the population units may appear more than once in the sample with the condition that the total number of appearances is fixed. In this paper, a more general single estimator of population total is discussed where Hanif and Brewer (1979) and Brewer et al (1979) Poisson sampling estimators are particular cases. A ratio estimator is also developed. Neyman type optimum allocation is derived for stratified random sample.

1. Introduction

In certain sample designs it may be difficult to categorize a selection procedure as either 'with replacement' or 'without replacement'. It is of some interest to devise a procedure for such a mixed sample design. Hanif and Brewer (1979) developed an estimator

$$y'_{GHT} = \sum_{I=1}^{N} \frac{\delta_I\, y_I}{\mu_I} \qquad (1)$$

where μ_I is the expected number of population units appearing in the sample, δ_I is defined such that $E(\delta_I) = \mu_I$. The variance of y'_{GHT} is

$$V(y'_{GHT}) = \sum_{I=1}^{N} \frac{Y_I^2}{\mu_I} + \sum_{I,J=1}^{N}\sum \mu_{IJ} \frac{Y_I Y_J}{\mu_I \mu_J} - Y^2 \qquad (2)$$

where $Y = \sum_{I=1}^{N} Y_I$

2. A Generalization of Hanif-Brewer Estimator

Suppose δ_I is the number of times the Ith population unit and δ_{IJ} is the number of times the ordered pair (I,J) appear in the sample. Let

$$\delta_{IJ} = \begin{cases} \delta_I \delta_J & , \quad I \neq J \\ \delta_{II} & , \quad \text{otherwise} \end{cases} \qquad (3)$$

Consider an estimator

$$Y'_G = \sum_{I=1}^{N} \frac{\delta_I Y_I}{\lambda_I} \qquad (4)$$

which is an unbiased estimator of population total provided $E\delta_I = \lambda_I$ and λ_I is a fixed quantity. If λ_I is a random variable, then y'_G is not an unbiased estimator of the population total. For example, if λ_I equals $m\Pi_I/n$, where m is a random number such that E(m)=n and n and Π_I are constants, then $E(\lambda_I) = E(m)\Pi_I/n = \Pi_I$. Now $E(\delta_I \lambda_I^{-1}) = E_I(\delta_I | \lambda_I)$ $E_2(\lambda_I^{-1})$, where E_I is the expectation over δ_I given λ_I and E_2 is the expectation over the number of repeated units in the sample.

The asymptotic variance of (4) is

$$V(y'_G) = \sum_I \frac{Y_I^2}{\lambda_I^4} E^2(\lambda_I)V(\delta_I) + E^2(\delta_I)V(\lambda_I) - E(\lambda_I) \times$$
$$E(\delta_I)\mathrm{Cov}(\lambda_I,\delta_I) + \sum_{I\neq J}\sum [E(\lambda_I)E(\lambda_J)\mathrm{Cov}(\delta_I,\delta_J) - \{E(\lambda_I)$$
$$\times E(\delta_J) + E(\lambda_J)E(\delta_I)\}\mathrm{Cov}(\lambda_I,\delta_J) + E(\delta_I)E(\delta_J)$$
$$\mathrm{Cov}(\lambda_I,\lambda_J)]$$

where $E(\delta_{IJ}) = \lambda_{IJ}$, $I \neq J$.

If $\delta_{II} = \delta_I(\delta_I - 1)$ and λ_I is fixed, then $E(\delta_I^2) = \lambda_{II} + \lambda_I$, and $V(y'_G)$ reduces to Hanif and Brewer (1979) variace expression (2).

3. A Ratio Estimator

Suppose $\lambda_I = \frac{m}{n}\Pi_I$, where m is a random variable, then the expression (4) becomes a ratio estimator as defined by Brewer et al (1979). The asymptotic variance of y'_{RG} is

$$V(y'_{RG}) = \frac{1}{n^2}[n^2 \mathrm{Var}(y'_G) + Y_G^2 \mathrm{Var}(m) - 2nY_G \mathrm{Cov}(y'_G, m)] \qquad (6)$$

where $y'_{RG} = \frac{n}{m}\sum_{I=1}^{N}\frac{\delta_I Y_I}{\Pi_I}$, and for any selection procedure

$$\mathrm{Var}(y'_G) = \sum_{I=1}^{N}(1-\Pi_I)\frac{Y_I^2}{\Pi_I} + \sum_{I\neq J=1}^{N}\sum(\Pi_{IJ} - \Pi_I\Pi_J)\frac{Y_I Y_J}{\Pi_I \Pi_J}$$

$$\mathrm{Var}(m) = \sum(1-\Pi_I)\Pi_I + \sum_{I\neq J=1}^{N}\sum(\Pi_{IJ} - \Pi_I\Pi_J) ,$$

$$\mathrm{Var}(y'_G, m) = \sum_{I=1}^{N}(1-\Pi_I)Y_I + \sum_{I\neq J=1}^{N}\sum(\Pi_{IJ} - \Pi_I\Pi_J)\frac{Y_I}{\Pi_I}$$

and

$E(m) = n.$.

4. Optimum Allocation in Stratified Random Random Sampling

Suppose a random sample of size n_h is drawn from an hth stratum of population of size N_h. Suppose the population of size N is divided into k strata. Suppose the estimator of hth population total and its variance are given by equations (4) and (5) respectively. Then the estimator of stratified population total is

$$y'_{SG} = \sum_{h=1}^{k}\sum_{I=1}^{N_h}\frac{\delta_{Ih} Y_{Ih}}{\lambda_{Ih}} \qquad (7)$$

and its variance is

$$\mathrm{Var}(y'_{SG}) = \sum_{h=1}^{k}\mathrm{Var}(y'_{RGh}).$$

If $\delta_{Ih} = a_h P_h$ and $\lambda_{Ih} = E(a_h)$, then the expression (7) reduces to the ratio estimator given by Doss et al (1978).

If $\lambda_{Ih} = n_h P_h$, $\lambda_{IJh} = n_h(n_h - 1)P_{IJh}$, as defined by Hanif and Brewer (1979), then $\mathrm{Var}(y'_{SG})$ $= \sum_{h=1}^{k}\frac{1}{n_h}V_h + C$ where $V_h = \sum_{I=1}^{N_h}\frac{Y_{Ih}^2}{P_{Ih}} - \sum_{I,J=1}^{N_h}\sum P_{IJh} \times \frac{Y_{Ih}Y_{Jh}}{P_{Ih}P_{Jh}}$ and $C = \sum_{h=1}^{k}\sum_{I,J=1}^{N_h}\sum P_{IJh}\frac{Y_{Ih}Y_{Jh}}{P_{Ih}P_{Jh}} - \sum_{h=1}^{k} Y_h^2$.

V_h and C are independent of n_h.

For fixed $n = \sum_{h=1}^{k} n_h$, n_h is obtained by minimizing $Var(y'_{SG})$ as $n_h = n\sqrt{V_R}/\sum_{h=1}^{k}\sqrt{V_h}$.

Similarly, if the total cost is fixed as $C = \sum_{h=1}^{k} c_h n_h$, then n_h is obtained by minimizing $Var(y'_{SG})$ as $n_h = C(\sqrt{V_h/c_h})/\sum_{h=1}^{k}\sqrt{V_h c_h}$.

REFERENCES

1. Brewer, K.R.W., Early, L.J., and Hanif,M.(1979) Poisson, Modified Poisson and Collocated Sampling. J. Statist. Planning and Inference, Vol.7.
2. Brewer, K.W.R., and Hanif, M. (1979). Generalization of the Horvitz and Thompson Estimator. J. Math and Statist., University of Panjab, Vol. XII.
3. Doss, D.C., Hartley, H.O. and Somayajulu, G.R. (1979). An exact small theory for post-stratification. J. Statist. Planning and Inf. 3, 235-247.
4. Hansen, M.H., and Hurwitz, M.N. (1943). On the Theory of Sampling from a Finite Population. Ann. Math. Statist. Vol. 14, 333-362.
5. Horvitz, D.G., and Thompson, D.J. (1952). A Generalization of Sampling wtthout replacement from a Finite Universe. J. Amer.Statist. Assoc. Vol. 47, 663-685.