RANKED SET SAMPLING FOR THE PRODUCT ESTIMATOR

Carlos N. Bouza[1]
Facultad de Matemática y Computación
Universidad de La Habana.
San Lázaro y L. CP 10400. Habana, Cuba

ABSTRACT
This paper is devoted to the study of the estimation of the population mean using of product estimator under ranked set . Strategies for different alternative estimators are developed and compared with theirs simple random sampling with replacement counterparts. The sample errors of them are compared analytically.

KEY WORDS: product estimators, gain in accuracy, order statistics

MSC 62D05

RESUMEN
En este trabajo se estudia la estimación de la media poblacional usando el estimador de producto bajo muestreo por conjunto ordenados. Estrategias para diferentes estimadores alternativos son desarrolladas y comparadas con ls contrapartidas basadas en muestreo simple aleatorio con reemplazo. Los errores de muestreo son comprados analíticamente

1. INTRODUCTION

McIntire (1952) proposed the method of ranked set sampling (rss). He proposed that the units may be ranked visually and noticed the existence of a gain in accuracy with respect to the use of the sample mean when simple random sampling (srs) with replacement (srswr) was used. Dell and Clutter (1972) and Takahasi and Wakimoto (1968) provided mathematical support to his claims. This method has proved to be better than the corresponding simple random sampling with replacement (srswr) counterpart for different estimators. See for example Bouza (2001), Samawi et al.(1996), El-Neweihi and Sinha (2000) for example.

This paper is motivated by the recent study developed by Zai-zai-and Jun-ling (2001) on the product estimator. It presents srs strategies for the product estimator under stratified sampling. We compare the classic product based estimators with their rss alternatives which are developed in section 3. Comparisons are made using their variances. Zai-zai and Jung-ling (2001) results on stratified sampling are extended to rss.

2.THE PRODUCT ESTIMATOR

The product estimator is close related to the ratio estimator The behavior of the ratio estimator under rss and srswr has been studied by Bouza (2001) using the ratio of rss means when the unit is ranked, Samawi et al.(1996) developed a similar analysis considering that the auxiliary variable *X* is ranked.

Consider a population *U* with *N* individuals and the set of values of variables $\{(X_i, Y_i), i=1,..,N\}$. A srswr sample of size *n* is selected and we obtain $\{(x_j, y_j, j=1,..,N\}$. The values of *X* are known for all the individuals in *U* and we compute

[1]. bouza@matcom.uh.cu

$$\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}, \qquad \bar{y} = \frac{\sum_{j=1}^{n} y_j}{n} \qquad \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

We want to estimate the population mean of the interest variable *Y*

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$$

The product estimator is given by

$$\bar{y}_0 = \frac{\bar{x}\bar{y}}{\bar{X}} \tag{2.1}$$

and its variance is

$$V(\bar{y}_0) = \frac{1}{n}\left[\sigma_Y^2 + 2R\sigma_{XY} + R^2\sigma_X^2\right] \tag{2.2}$$

where

$$V(Z) = \sigma_Z^2, \qquad Z = X, Y; \quad Cov(X,Y) = \sigma_{XY} = \rho\sigma_X\sigma_Y; \quad R = \frac{\bar{Y}}{\bar{X}}$$

See Chadhuri-Voos (1991). They analyzed the behavior of the product estimator using a model approach. The model considered fixed that

$Y=\alpha+\beta X+\varepsilon$

and

$\sum\varepsilon_i = \sum X_i\ \varepsilon_i, =0.$

The parameters are given by $\beta = \sigma_{XY} / \sigma_X^2$, $\qquad \alpha = \bar{Y} - \beta\bar{X}$.

The auxiliary information can be used also for computing the common ratio estimator

$$\bar{y}_r = \bar{X}\frac{\bar{y}}{\bar{x}} = \bar{X}r$$

Its variance is given by

$$V(\bar{y}_r) = \frac{1}{n}\left[\sigma_Y^2 + R^2\sigma_X^2 - 2R\sigma_{XY}\right]$$

The product estimator (2.1) is more accurate than it when the correlation between X and Y is negative. Under the model it is preferred whenever $\alpha > \frac{\overline{X}}{2}$, see for a discussion of these facts Chaudhuri-Voos (1991).

The following algorithm provides a description of ranked set sampling .

Rss implementation

```
Input r, m
i=0 and t=0
  While t<r+1 do
    While i<m+1 do
      Select a sample s_i of size |s_i|=m using srswr
      Rank the sampled units with respect to the variable of interest ξ=Y
      Measure Y in the unit with rank i (ξ_(i:m)t)
      i=i+1
    End
  t=t+1
End
```

Hence the rss sample is the sequence of order statistics (os) $\xi_{(1:1)t},\ldots, \xi_{(m:m)t}$, where the sub index $(j:h)t$ identifies the statistic of order j $(os(j))$in the h-th sample $(s_{(h)})$ in the cycle $t=1,..,r$. We have $n=mr$ observations and r of them are of $os(j)$, $j=1,..., m$. Take μ_ξ as the mean of a variable of interest ξ, its rss estimator is

$$\mu_{(rss)\xi}=\Sigma^{r}{}_{t=1}\, \Sigma^{m}{}_{i=1}\xi_{(i:m)t}/rm \tag{2.3}$$

It is unbiased and its variance is given by

$$V(\mu_{(rss)\xi})= \Sigma^{m}{}_{i=1}\, \sigma^2{}_{\xi(i:m)}/rm^2=[\sigma^2{}_{\xi}- \Delta_{\xi(:m)}]/n \tag{2.4}$$

where

$$\sigma^2{}_{\xi(i:m)} =E[\xi_{(i:m)}-E(\xi_{(i:m)})]^2$$

$$\Delta_{(i:m)}= E[\xi_{(i:m)}]-\mu_{\xi}.$$

and

$$\Delta_{\xi(m)} =m^{-1}\, \Sigma^{m}{}_{i=1}\, \Delta^2{}_{\xi(i:m)}$$

See Takahashi-Wakimoto (1968) for a detailed discussion of these results.

The second term of (2.4) is the gain in accuracy due to the use of rss instead of srswr.

Let us study the behavior of a rss strategy in the estimation of the mean of Y when the product estimator is used. Two variables are involved $\xi=X$, ***Y*.** The auxiliary variable X is known and the rss counterpart of (2.1) is

$$\overline{y}_{0(rss)} = \frac{\mu_{(rss)X}\,\mu_{(rss)Y}}{\overline{X}} \tag{2.5}$$

From (2.2) and (2.5) is obtained, by direct, calculation that

$$V(\bar{y}_{0(rss)}) = \frac{1}{n}\left[\sigma_Y^2 + R^2\sigma_X^2\right] + 2R\rho\lambda - \frac{1}{n}\left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]$$
$$\lambda = \frac{1}{\text{n}}\sqrt{\left[\sigma_Y^2 - \Delta_{(Y|m)}\right]\left[\sigma_X^2 - \Delta_{(X|m)}\right]}; \qquad \rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{2.6}$$

The last term of (2.6) is the gain in accuracy of (2.5) with respect to (2.1). Then the rss estimator of μ_X is preferred to the arithmetic means computed from a samples selected by srswr if

$2nR\rho\lambda\text{-}\Delta_{(Y|m)}\ \text{-}\Delta_{(X|m)} < 0$

Note that $n\lambda < \sigma_Y\sigma_x$. Therefore

$2nR\rho\lambda\text{-}\Delta_{(Y|m)}\ \text{-}\Delta_{(X|m)} < 2nR\rho\sigma_{XY}\text{-}\Delta_{(Y|m)}\ \text{-}\Delta_{(X|m)}$.

The preference for (2.5) is determined by the sign of

$$G(b,a) = V(\bar{y}_{0(rss)}) - V(\bar{y}_a) \qquad a, b = 0(rss), \qquad r, \qquad 0$$

The results of its comparison with the ratio and product estimator under srswr are given in the following theorem:

Theorem 2.1. The rss- product estimator (2.5) is more accurate than the srswr ratio estimator if

$$R \le \sqrt{\frac{\Delta_{(Y|m)} + \Delta_{(X|m)}}{2n\sigma_X^2}}$$

It is more accurate than its srswr counterpart always.

Proof:

Calculating the expression $G(0(rss),r)$ we obtain that

$$G(r,0(rss)) < 2R^2\sigma_X^2 - \frac{1}{n}\left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]$$

As

$$G(0(rss),0) = -\frac{2R(\sigma_{X,Y} - n\lambda) + \left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n} < -\frac{\left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n}$$

The second hypothesis holds. □

Following the results of Zai-zai and Jun-ling (2001) the separate product estimator for stratified sampling is

$$\bar{y}_{0(s)} = \sum_{h-1}^{H} W_h \frac{\bar{y}_h \bar{x}_h}{\bar{X}_h} \tag{2.7}$$

where the means are computed for each strata, $W_h = N_h/N$, N_h is the size of strata h, n_h is the size of its sub sample and N is the number of individuals in the whole population. Its approximate variance is

$$V(\bar{y}_{o(s)}) = \sum_{h=1}^{H} W_h^2 \frac{\sigma_{Yh}^2 + R^2{}_h \sigma_{Xh}^2 + 2R_h \sigma_{XY}}{n_h}, \qquad \sigma_{Zh}^2 = \sum_{j=1}^{N_h} \frac{(Z_{hj} - \overline{Z}_h)^2}{N_h}, Z = X, Y \tag{2.8}$$

Due to the structure of (2.7) its rss extension is easily derived by replacing the srswr estimator by (2.5) in each strata resulting in that

$$\bar{y}_{o(rss-s)} = \overline{\sum_{h=1}^{H} W_h} \frac{\bar{y}_{(rss)h} \bar{x}_{(rss)h}}{\overline{X}_h}, \qquad \bar{z}_{(ss)h} = \sum_{t=1}^{t_h} \overline{\sum_{j=1}^{m_h}} \frac{z_{(j|m_h)t_h}}{n_h}, z = x, y \qquad n_h = m_h t_h \tag{2.9}$$

is its expression. The comparisson of (2.7) and (2.9) yields that

$$G(0(rss-s), 0(s)) = \sum_{h=1}^{H} W_h^2 \left[\frac{2R_h(\sigma_{X,Y} - n\lambda) - \left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n} \right] < -\sum_{h=1}^{H} W_h^2 \frac{\left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n}$$

The combined product estimator

$$\bar{y}_{0(c)} = \frac{\left(\sum_{h=1}^{H} W_h \bar{y}_h\right)\left(\sum_{h=1}^{H} W_h \bar{x}_h\right)}{\overline{X}} \tag{2.10}$$

has as variance

$$V(\bar{y}_{o(c)}) \approx \sum_{h=1}^{H} W_h^2 \frac{\sigma_{Yh}^2 + R^2 \sigma_{Xh}^2 + 2R\sigma_{XY}}{n_h}, \tag{2.11}$$

Using a plug-in-rule we obtain the combined rss estimator

$$\bar{y}_{0(rss-c)} = \frac{\left(\sum_{h=1}^{H} W_h \bar{y}_{(rss)h}\right)\left(\sum_{h=1}^{H} W_h \bar{x}_{(rss)h}\right)}{\overline{X}} \tag{2.12}$$

The gain in accuracy is derived and we will prefer it if is satisfied the inequality:

$$G(0(rss-c), 0(c)) = \sum_{h=1}^{H} W_h^2 \left[\frac{2R(\sigma_{X,Y} - n\lambda) - \left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n} \right] < -\sum_{h=1}^{H} W_h^2 \frac{\left[\Delta_{(Y|m)} + \Delta_{(X|m)}\right]}{n}$$

Hence the rss alternatives to the estimators studied by Zai-zai and Jung-li (2001) are more accurate. These results support the proof of the following theorem:

Theorem 2. Take a population divided into H strata of size N_h, and assume that independent sub samples of size n_h are selected from each of them with srswr. The rss product estimators (2.9) and (2.12) are more accurate than (2.7) and (2.10).

Acknowledgements: This paper was developed within a project with the Institute of Mathematics of China supported by the Academy of Sciences of the Developing World (TWAS).

Received November 2006
Revised March 2007

REFERENCES

[1] BOUZA, C.N. : (2001): Model Assisted : Ranked set sampling . **Biometrical J**., Berlin. 43, 154-163

[2] CHAUDHURI A. and VOS. J.W.E. (1991): Unified theory and strategies of survey sampling. **North Holland,** Amsterdam.

[3] EL-NEWEIHI E. and SINHA B.K. (2000): Reliability Estimation based on ranked set sampling. **Commun Statisti-Theory and Methods.** 29, 1583-1595.

[4] DELL, T. R. and CLUTTER J.L. (1972): Ranked set sampling theory with order statistics background.. **Biometrics,**, 28, 545-555.

[5] MCINTIRE, G. A. (1952): A method for unbiased sampling using ranked sets. **Australian J. of Agricultural Research, Melbourne**, 3, 385-390.

[6] SAMAWI, H.M., M.S. AHMED and W. ABU DAYYEH (1996:): Estimation of a ratio using Ranked sert sampling**. Biometrica**l **J**. 36, 753-764

[7] PATIL G.P., SINHA A.K. and TAILLIE, C.(1995): Ranked set sampling coherent ranking and size biased permutations. **J. Statistical Planning and Inference**. 63, 311-324.

[8] TAKAHASHI K. and WAKIMOTO, M. (1968): On unbiased estimates of the population mean based on the sample stratified by means ordering. **Annals of the Institute of Mathematical Statistics**. 20, 1031-1047.

[9] ZAI-ZAI, Y. and JUNG-LING M. (2001): Separate product estimator and combined product estimator under stratified sampling. **J. Engineering Mathematics**, 18, 8-12.