Int. J. Contemp. Math. Sciences, Vol. 5, 2010, no. 52, 2567 - 2576

Estimation of the Population Variance Using Ranked Set Sampling with Auxiliary Variable

Said Ali Al-Hadhrami

College of Applied Sciences, Nizwa, Oman
abur1972@yahoo.co.uk

Abstract

When the variable of interest Y is correlated to another variable X that has some information, then the use of auxiliary variable may improve the efficiency of the estimation of the variable under the study. In this paper, ratio estimator for the population variance based on Ranked Set Sampling (RSS) was considered. The suggested estimator was compared to the corresponding one based on Simple Random Sampling (SRS). Computer Simulation from some distributions showed that RSS estimator has smaller bias and Mean Square Error (MSE) than that from SRS.

Keywords; Ranked Set Sampling; Ratio Estimator; Population variance; efficiency

1 Introduction

Ranked Set Sampling (RSS) was first used by McIntyre (1952) to estimate the population mean. The method is more efficient when estimating the population mean. However, the gain of RSS over SRS is little when estimating the population variance. Stokes (1980) studied estimation of the population variance using RSS and a bias estimator of σ^2 was defined but it is asymptotically unbiased. Yu et al. (1999) studied some unbiased estimates of σ^2 in the parametric case of a normal population. Tiwari and Kvam (2001) proposed unbiased estimator for σ^2 for location-scale families of symmetric distribution. MacEachern (2002) developed an unbiased estimator of the variance of a population based on RSS. The suggested estimator is better than estimating the variance based on SRS and more efficient than the estimator based on RSS proposed by Stokes (1980). Tiensuwan and Sarikavanij (2003) proposed two unbiased estimators of the variance for multiple cycles and under balanced case. Their proposed estimators are more efficient than the one based on SRS. They showed that there is no unbiased estimate of the population variance based on a single cycle of RSS. Perron and Sinha (2004) showed that for more than one cycle, it is possible to construct a class of quadratic unbiased estimates of σ^2 in both balanced and unbalanced cases. They derived a minimum variance unbiased quadratic nonnegative estimate of σ^2 within a certain class of quadratic estimates. Ahmad (2004) suggested some bootstrap techniques for estimation of variance under RSS. Sengupta and Mukhuti (2006) proposed some unbiased estimators of the variance of the exponential distribution.

In this paper ratio estimator for the population variance is considered and compared to the ratio estimator based on SRS.

2 Ratio Estimator For the Population Variance in SRS

Izaki(1983) suggested the following estimator for the population variance

$$s_1^2 = (s_y^2 / s_x^2) S_x^2 \tag{1}$$

where s_x^2 and s_y^2 are unbiased estimators of population variance S_x^2 and S_y^2, respectively, defined as $s_x^2 = \frac{\sum_{i=`}^{n}(x_i - \bar{x})^2}{n-1}$ and $s_y^2 = \frac{\sum_{i=`}^{n}(y_i - \bar{y})^2}{n-1}$

The MSE of the estimator s_1^2 is

$$MSE(s_{y1}^2) \cong \lambda S_y^4 \left[K_y + K_x - 2\theta\right]$$

where $\lambda = 1/n$, $K_y = \mu_{40}/\mu_{20}^2$, $K_x = \mu_{04}/\mu_{02}^2$, $\theta = \mu_{22}/\mu_{20}\mu_{02}$, and $\mu_{rs} = (1/N)\sum_{i=1}^{N}(y_i - \mu_y)^r (x_i - \mu_x)^s$

3 Some Estimators for the Population Variance Based on RSS

Stokes (1980) suggested for one cycle the following estimator for σ^2

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m}(x_{(i)} - \bar{x}_{RSS})^2}{m-1}$$

where $\bar{X}_{RSS} = (1/m)\sum_{i=1}^{m} x_{(i)}$, with $E(\hat{\sigma}^2) = \sigma^2 + (1/m(m-1))\sum_{i=1}^{m}(\mu_{(i)} - \mu)^2$

The variance of the estimator is

$$Var(\hat{\sigma}^2) = \frac{1}{(m-1)^2}\left[\left(\frac{m-1}{m}\right)^2 \sum_{i=1}^{m}\mu_{4(i)} + 4\sum_{i=1}^{m}\tau_{(i)}^2\sigma_{(i)}^2 + 4\left(\frac{m-1}{m}\right)\sum_{i=1}^{m}\tau_{(i)}\mu_{3(i)}\right]$$
$$+\frac{1}{(m-1)^2}\left[\frac{4}{m^2}\sum_{i<i'=1}^{m}\sigma_{(i)}^2\sigma_{(i')}^2 - \frac{(m-1)^2}{m^2}\sum_{i=1}^{m}\sigma_{(i)}^4\right]$$

Stokes(1980) suggested for r cycles the following estimator for σ^2

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{r}\sum_{i=1}^{m}(x_{(i)}^{(j)} - \bar{x}_{RSS})^2}{rm-1}$$

with $E(\hat{\sigma}^2) = \sigma^2 + (1/m(rm-1))\sum_{i=1}^{m}(\mu_{(i)} - \mu)^2$

The variance of the estimator is

$$Var(\hat{\sigma}^2) = \frac{r}{(m-1)^2}\left[\left(\frac{rm-1}{rm}\right)^2 \sum_{i=1}^{m}\mu_{4(i)} + 4\sum_{i=1}^{m}\tau_{(i)}^2\sigma_{(i)}^2 + 4\left(\frac{rm-1}{rm}\right)\sum_{i=1}^{m}\tau_{(i)}\mu_{3(i)}\right]$$
$$+\frac{r}{(m-1)^2}\left[\frac{4}{rm^2}\sum_{i<i'=1}^{m}\sigma_{(i)}^2\sigma_{(i')}^2 - \frac{2(r-1)-(rm-1)^2}{r^2m^2}\sum_{i=1}^{m}\sigma_{(i)}^4\right]$$

The estimators provided by Stokes are biased. Montip & Sukuman(2003) showed that for one cycle there is no unbiased estimator for σ^2 but for more than one cycle they proposed two unbiased estimators for σ^2 . The first estimator is

$$\hat{\sigma}_{RSS_1}^2 = \frac{W}{mr} + \frac{B}{m(r-1)} \qquad (2)$$

where $W = \sum_{j=1}^{r}\sum_{i=1}^{m}(X_{(i)}^{(j)} - \bar{X}_{RSS}^{(j)})^2$, $B = m\sum_{j=1}^{r}(\bar{X}_{RSS}^{(j)} - \bar{X})^2$ and $\bar{X} = (1/r)\sum_{j=1}^{r}\bar{X}_{RSS}^{(j)}$

They also showed that the variance of the estimator is

$$Var(\hat{\sigma}_{RSS_1}^2) = \frac{T_1 + 4T_2 + 4T_3}{m^2r} - \frac{1}{m^2r}T_5 + \frac{2}{m^2r(r-1)}\left(\sum_{i=1}^{m}\sigma_{(i)}^2\right)^2$$

where $A_1 = \sum_{i=1}^{m}\mu_{4(i)}$, $A_2 = \sum_{i=1}^{m}\tau_{(i)}^2\sigma_{(i)}^2$, $A_3 = \sum_{i=1}^{m}\tau_{(i)}\mu_{3(i)}$, $A_4 = \sum_{i=1}^{m}\sigma_{(i)}^4$, $\mu_{k(i)} = E(X_{(i)} - \mu_{(i)})^k$ and $\tau_{(i)} = \mu_{(i)} - \mu$

The second estimator is

$$\hat{\sigma}^2_{RSS_2} = \frac{rm - m + 1}{m^2 r(r-1)} W^* + \frac{B^*}{mr} \qquad (3)$$

where $W^* = \sum_{j=1}^{r}\sum_{i=1}^{m}(X_{(i)}^{(j)} - \bar{X}_i)^2$, $B^* = r\sum_{j=1}^{m}(\bar{X}_{(i)} - \bar{\bar{X}})^2$

The variance of the estimator is

$$Var(\hat{\sigma}^2_{RSS_2}) = \frac{A_1 + 4A_2 + 4A_3}{m^2 r} - \frac{m^2 r^2 - m^2 r - 2}{m^4 r^2 (r-1)} A_4 + \frac{2}{m^2 r^2}\left(\sum_{i=1}^{m}\sigma^2_{(i)}\right)^2$$

They also computed the variance of the above estimators when the underlying distribution is uniform, exponential and normal. For more details refer to Montip & Sukuman (2003).

4 Ratio Estimator for the Population Variance Based on RSS

In this section we assume that the population variance of X is known, σ^2_X . So we can define a ratio estimator of the variance of Y as follows

$$\sigma^2_{RRSS} = \frac{\hat{\sigma}^2_{RSS_y}}{\hat{\sigma}^2_{RSS_x}} \sigma^2_x \qquad (4)$$

where $\hat{\sigma}^2_{RSS_y}$ and $\hat{\sigma}^2_{RSS_x}$ are unbiased estimators defined in equation (2) .

General form of Taylor series for a bivariate function $h(X,Y)$ is

$$\begin{aligned} h(X,Y) \cong\, & h(\mu_x,\mu_y) + h_y(\mu_x,\mu_y)(Y-\mu_y) + h_x(\mu_x,\mu_y)(X-\mu_x) \\ & + \tfrac{1}{2}(Y-\mu_y)^2 h_{yy}(\mu_x,\mu_y) + (X-\mu_x)(Y-\mu_y)h_{xy}(\mu_x,\mu_y) \\ & + \tfrac{1}{2}(X-\mu_x)^2 h_{xx}(\mu_x,\mu_y) \end{aligned}$$

where $h_{xy}(\mu_x,\mu_y) = \frac{\partial^2 h(X,Y)}{\partial X\,\partial Y}\Big|_{\mu_x,\mu_y}$ and $h_x(\mu_x,\mu_y) = \frac{\partial h(X,Y)}{\partial X}\Big|_{\mu_x,\mu_y}$.

So, applying this expansion to $\hat{\sigma}^2_{RRSS}$ about σ^2_x and σ^2_y gives bias zero when first order expansion is used. However, the second order bivariate Taylor expansion gives

$$\hat{\sigma}^2_{RRSS} \cong \sigma^2_{yRSS} + (\hat{\sigma}^2_{yRSS} - \sigma^2_y) - \frac{\sigma^2_y}{\sigma^2_x}\left(\hat{\sigma}^2_{xRSS} - \sigma^2_x\right)$$
$$+ \frac{\sigma^2_y}{(\sigma^2_x)^2}\left(\hat{\sigma}^2_{xRSS} - \sigma^2_x\right)^2 - \left(\frac{1}{\sigma^2_x}\right)\left(\hat{\sigma}^2_{xRSS} - \sigma^2_x\right)\left(\hat{\sigma}^2_{yRSS} - \sigma^2_y\right)$$

Then the bias of the estimator is

$$Bias(\hat{\sigma}^2_{RRSS}) \cong \frac{\sigma^2_y}{(\sigma^2_x)^2} E\left(\hat{\sigma}^2_{xRSS} - \sigma^2_x\right)^2 - \left(\frac{1}{\sigma^2_x}\right) E\left(\hat{\sigma}^2_{xRSS} - \sigma^2_x\right)\left(\hat{\sigma}^2_{yRSS} - \sigma^2_y\right)$$

This can be expressed as

$$Bias(\hat{\sigma}^2_{RSS}) \cong \frac{\sigma^2_y}{(\sigma^2_x)^2} Var\left(\hat{\sigma}^2_x\right) - \left(\frac{1}{\sigma^2_x}\right) Cov\left(\hat{\sigma}^2_x, \hat{\sigma}^2_y\right)$$

Using first order of the Taylor expansion of σ^2_y about $\hat{\sigma}^2_y$ and $\hat{\sigma}^2_x$, then $\hat{\sigma}^2_{RSS}$ is

$$\hat{\sigma}^2_y \cong \sigma^2_y + (\hat{\sigma}^2_{RSS_y} - \sigma^2_y) - \frac{\sigma^2_y}{\sigma^2_x}(\hat{\sigma}^2_{RSS_x} - \sigma^2_x)$$

Then the variance of the estimator is

$$Var(\hat{\sigma}^2_y) \cong Var(\hat{\sigma}^2_{RSS_y}) - \left(\frac{\sigma^2_y}{\sigma^2_x}\right)^2 Var(\hat{\sigma}^2_{RSS_x}) - 2\left(\frac{\sigma^2_y}{\sigma^2_x}\right) Cov(\hat{\sigma}^2_{RSS_x}, \hat{\sigma}^2_{RSS_y})$$

5 Simulation Study

The behavior of the proposed estimator has been investigated numerically. Let us assume that the variable of interest Y and a concomitant variable X were correlated with a correlation coefficient ρ. The samples were generated from three distributions; bivariate normal, bivariate exponential, and bivariate gamma. For generating bivariate normal, S-Plus codes were used since it is already provided but both exponential and gamma were generated by using the mixture approach method proposed by Minhajuddin et al (2004). The algorithm started with generating from a negative binomial as a prior distribution. Then, conditional on that value, bivariate sample were generated from the posterior distribution.

For each estimator, 10000 samples generated from the distributions and the estimates of σ^2_y were calculated as follows:

For RSS, the estimator $\sigma^2_{RRSS} = (\hat{\sigma}^2_{RSS_y} / \hat{\sigma}^2_{RSS_x})\sigma^2_x$ was used

where

$$\hat{\sigma}^2_{RSS_x} = \frac{W_x}{mr} + \frac{B_x}{m(r-1)} \text{ and } \hat{\sigma}^2_{RSS_y} = \frac{W_y}{mr} + \frac{B_y}{m(r-1)}$$

And for SRS, we used the estimator $\sigma^2_{RSRS} = (\hat{\sigma}^2_{SRS_y} / \hat{\sigma}^2_{SRS_x})\sigma^2_x$. The average of $\hat{\sigma}^2_y$'s and the Mean Square Error (MSE) were calculated respectively using

$$\hat{E} = \tfrac{1}{10000}\sum_{i=1}^{10000}\hat{\sigma}^2_{yi} \text{ and } Va\hat{r} = \tfrac{1}{10000}\sum_{i=1}^{10000}(\hat{\sigma}^2_{yi} - \sigma^2_y)^2 .$$

Different values for ρ and m were used. Table (1) ,(2) and (3) show the bias and the MSE of the estimators from normal distribution, exponential and gamma distribution respectively.

Table (1) Simulation results from normal with $\mu_x = 5,\ \mu_y = 6,\ \sigma^2_x = \sigma^2_y = 1$. Number of cycles=3 with different set size and correlation coefficient ρ

m	ρ	*Bias*		*MSE*		*efficiency*
		SRS	RSS	SRS	RSS	
3		0.0060	0.0037	0.0137	0.0129	1.0608
4		0.0038	0.0009	0.0089	0.0085	1.0439
5	0.99	0.0036	-0.0008	0.0069	0.0064	1.0867
6		0.0031	0.0004	0.0055	0.0049	1.1151
3		0.0606	0.0333	0.1563	0.1416	1.1043
4		0.0399	0.0084	0.0970	0.0855	1.1349
5	0.90	0.0285	-0.0030	0.0693	0.0597	1.1610
6		0.0255	-0.0021	0.0555	0.0462	1.2012
3		0.1636	0.0767	0.6050	0.4639	1.3040
4		0.1109	0.0192	0.3153	0.2325	1.3562
5	0.70	0.0842	0.0049	0.2239	0.1648	1.3583
6		0.0656	-0.0012	0.1667	0.1196	1.3943
3		0.2775	0.1370	1.2397	0.8377	1.4799
4		0.1855	0.0281	0.6556	0.4193	1.5636
5	0.4	0.1348	-0.012	0.4218	0.2650	1.5917
6		0.1148	-0.0111	0.3195	0.1980	1.6133

Table (2) Simulation results from exponential with mean 1 for both X and Y. Number of cycles=3 with different set size and correlation coefficient ρ

m	ρ	*Bias*		*MSE*		*efficiency*
		SRS	RSS	SRS	RSS	
3	0.99	0.0114	0.0136	0.0273	0.0261	1.0494
4		0.0109	0.0076	0.0192	0.0179	1.0713
5		0.0070	0.0056	0.0152	0.0141	1.0813
6		0.0073	0.0058	0.0123	0.0112	1.0981
3	0.90	0.1404	0.1090	0.4077	0.3639	1.1204
4		0.0886	0.0757	0.2370	0.2110	1.1232
5		0.0752	0.0685	0.1835	0.1627	1.1278
6		0.0625	0.0504	0.1490	0.1320	1.1291
3	0.70	0.3652	0.3137	1.9952	1.6093	1.2398
4		0.2662	0.2314	1.1066	0.8841	1.2517
5		0.2069	0.1686	0.7601	0.5998	1.2674
6		0.1729	0.1225	0.5644	0.4398	1.2834
3	0.40	0.7475	0.5890	7.7821	4.7692	1.6317
4		0.5132	0.3865	3.4200	2.0246	1.6892
5		0.4000	0.2999	1.9947	1.1795	1.6911
6		0.3299	0.2323	1.4675	0.8611	1.7042

Table (3) Simulation results from gamma (2,3) Number of cycles=3 with different set size and correlation coefficient ρ .

m	ρ	*Bias*		*MSE*		*efficiency*
		SRS	RSS	SRS	RSS	
3	0.99	0.0149	0.0121	0.0273	0.0252	1.0853
4		0.0090	0.0076	0.0187	0.0167	1.1145
5		0.0069	0.0054	0.0158	0.0140	1.1220
6		0.0072	0.0064	0.0134	0.0118	1.1362
3	0.90	0.1194	0.1178	0.3554	0.3105	1.1447
4		0.0843	0.0654	0.2035	0.1748	1.1647
5		0.0704	0.0581	0.1584	0.1346	1.1774
6		0.0570	0.0478	0.1352	0.1134	1.1926
3	0.70	0.3014	0.2692	1.4113	1.0731	1.3152
4		0.1962	0.1732	0.7348	0.5504	1.3351
5		0.1723	0.1267	0.4828	0.3598	1.3418
6		0.1318	0.1063	0.3741	0.2734	1.3686
3	0.40	0.5902	0.4578	4.1450	3.5744	1.6105
4		0.3909	0.2762	2.3987	1.4469	1.6578
5		0.2704	0.2273	0.9917	0.5967	1.6617
6		0.2493	0.1625	0.8796	0.5143	1.7100

6 Concluding Remarks

From the simulation results given in the tables the following can be concluded:

1- Both the bias and MSE of all the estimators decrease when either the sample size or the correlation coefficient increases.

2- Both the bias and MSE of RSS estimators are smaller than the corresponding ones based on SRS. The relative efficiency of RSS estimators with respect to the corresponding one based on SRS is greater than 1.

3- Both the bias and MSE of the estimators for normal distribution variance are smaller than those from exponential and gamma.

4- The estimators of the variance of exponential distribution have greater bias and MSE than the one for gamma under the given parameters unless the correlation coefficient is near 1.

References

[1] T. Ahmad, Bootstrap Techniques for estimation of variance under ranked set sampling, Atlas conference, institute of engineering and technology, India (2004): 27-29 December.

[2] A.T.M. Minhajuddin , I.R. Harris and W.R. Schucany , Simulating multivariate distributions with specific correlations , Journal of Statistical Computation and Simulation, 74(4)(2004) 599-607.

[3] S.N. MacEachern, O. Ozturk, D.A. Wolf, and G.V. Shark, A new ranked set sample estimator of variance.Journal of the Royal Statistical Society 64(2)(2002), 177-188

[4] G.A. McIntyre, A method of unbiased selective sampling, using ranked sets , Australian J. Agricultural Research, 3(1952), 385-390.

[5] F. Perron, and ,B.K. Sinha , Estimation of Variance based on a ranked set sample, Journal of statistical planning and inference, 120(2004), 21-28

[6] S.L. Stokes, Estimation of variance using judgment order ranked set samples, Biometrics 36(1980), 35-42.

[7] S. Sengupta, and S. Mukhuti, Unbiased Variance estimation in a sample exponential population using ranked set samples. Journal of statistical planning and inference. 136(4) (2006), 1526-1553.

[8] R.C. Tiwari, and P.H. Kvam, Ranked Set Sampling from location-scale families of symmetric distributions, Communications in Statistics: Theoty and Methods 30(2001), 1641-1659.

[9] M. Tiensuwan, and S. Sarikavanij, On estimation of the population variance based on ranked set sample, Journal of Applied Statistical Science, 12(4)(2003), 283-295.

[10] P.L.H. Yau, K. Lam, and B.K. Sinha, Estimation of normal variance based on balanced and unbalanced Ranked Set Samples, Environmental and Ecological Statistics, 6(1999), 23-45

Received: June, 2010