Jour. Ind. Soc. Ag. Statistics
56(3), 2003 : 230-236

Estimation of Variance in Normal Population

Sharad Bhatnagar
CCS Haryana Agricultural University, Hisar–125004 (Haryana)
(Received : July, 2000)

SUMMARY

A general class of estimators for the variance in normal population is proposed and their large sample properties are studied. The efficiencies of the estimators of variance have been discussed according to mean squared error criterion and through simulation results.

Key words : Bias, Absolute bias, Mean squared error, Coefficient of variation, Normal population, Asymptotic, Simulation.

1. Introduction

Singh and Pandy [6] considered the problem of estimating the variance in normal population by utilizing the knowledge of the coefficient of variation. In the absence of any prior knowledge of coefficient of variation, operational estimators of variance were proposed and analyzed by Pandey [3], Pandy and Singh [4] and Singh [5] keeping in view the biases of these estimators. Pandey [3] proposed two estimators which have the same mean squared error, but one of his estimators has smaller bias than the other. Pandy and Singh [4] presented an estimator which has smaller bias and mean squared error than Pandey's [3] estimator under certain condition. Singh [5] proposed an estimator which has the same mean squared error as Pandey [3], but it has smaller bias than Pandey's [3] estimator under certain conditions on the fixed characterizing scalar.

In the present paper, a class of estimators has been proposed and its bias and mean squared error have been worked out up to order $0(n^{-2})$ and $0(n^{-3})$ respectively. The efficiencies of the estimator have also been compared according to mean squared error criterion in Section 2 and through simulation technique in Section 3.

2. Estimators and their Properties

Suppose a random sample of size n is drawn from a normal distribution

$N(\mu, \sigma^2)$ with $C = \frac{\sigma^2}{\mu^2}$ as the relative variance and σ^2 as the variance.

An unbiased estimator of σ^2 is

$$P_{00} = s^2 \tag{2.1}$$

whose relative variance is

$$RV(P_{00}) = E\left(\frac{s^2 - \sigma^2}{\sigma^2}\right)^2 \tag{2.2}$$

$$= \frac{2}{n}$$

Writing the variance $\sigma^2 = C\mu^2$, another simple estimator of σ^2

$$t = \hat{C}\hat{\mu}^2 \tag{2.3}$$

$$= \frac{s^2}{\bar{y}^2}\left(\bar{y}^2 - \frac{s^2}{n}\right)$$

For small value of n in normal population, the unbiased estimator of $\hat{\mu}^2 = \left(\bar{y}^2 - \frac{s^2}{n}\right)$ may assume negative values with positive probability (c.f. Das [2]). For this reason, we have not considered the estimator t.

Bhatnagar [1] developed a class of estimator of $\hat{\mu}^2$ as

$$t_{kg} = \bar{y}^2\left(1 + g\frac{s^2}{n\bar{y}^2} + k\frac{s^4}{n^2\bar{y}^4}\right)^{-1} \tag{2.4}$$

where g and k are characterizing scalars.

Keeping in view the estimators t and t_{kg}, we may define the following general class of estimators of σ^2

$$P_{kg} = s^2\left(1 + g\frac{s^2}{n\bar{y}^2} + k\frac{s^4}{n^2\bar{y}^4}\right)^{-1} \tag{2.5}$$

Pandey [3] studied two particular cases viz., P_{01} and P_{11} while Pandy and Singh [4] proposed the following estimator

$$p_s = s^2\left(1+\frac{s^2}{n\bar{y}^2}\right)^{-1}\left(1+\frac{4s^2}{n\bar{y}^2}+\frac{2s^4}{n(n-1)\bar{y}^4}\right)^{-1} \tag{2.6}$$

$$= s^2\left[1+\frac{5s^2}{n\bar{y}^2}+\frac{2(3n-2)}{(n-1)}\frac{s^4}{n^2\bar{y}^4}+\frac{2s^4}{n^2(n-1)\bar{y}^6}\right]^{-1}$$

$$\approx s^2\left(1+\frac{5s^2}{n\bar{y}^2}+\frac{6s^4}{n^2\bar{y}^4}\right)^{-1} = P_{65} \quad \text{(say)}$$

Further Singh [5] proposed the estimator of variance as

$$T_\alpha = s^2\left[1-\frac{s^2}{n\bar{y}^2}\left(1+\frac{s^2}{n\bar{y}^2}\right)^{-\alpha}\right] \tag{2.7}$$

$$\approx s^2\left(1+\frac{s^2}{n\bar{y}^2}-\alpha\frac{s^4}{n^2\bar{y}^4}\right)^{-1} = P_{-\alpha 1} \quad \text{(say)} \qquad [4]$$

where α is a scalar.

Setting $g = 5$, $k = 6$ we obtain the estimator P_{65} of Pandy and Singh [4] while putting $g = 1$, $k = -\alpha$ yields the estimator $P_{-\alpha 1}$ proposed by Singh [5].

Supposing g and k to be fixed scalars, the relative bias to order $0(n^{-2})$ and relative mean squared error to order $0(n^{-3})$ of the estimator P_{kg} are

$$RB(P_{kg}) = -g\frac{C}{n}+\left(g^2-k-3g-\frac{2g}{C}\right)\frac{C^2}{n^2} \tag{2.8}$$

and

$$RM(P_{kg}) = \frac{2}{n}+g(gC-8)\frac{C}{n^2}$$

$$+2g\left[12(g-1)-g(g-5)C+\frac{k}{g}(gC-6)\right]\frac{C^2}{n^3} \tag{2.9}$$

The term of order $0(n^{-2})$ in the expression (2.8) for relative bias has negative value for the positive values of scalars.

From (2.9), we find that the estimator P_{kg} dominates over conventional estimator P_{00} of σ^2 for any value of sample size n if

$$g(gC-8) < 0 \tag{2.10}$$

and

$$k < g\left[\frac{g(g-5)C - 12(g-1)}{(gC-6)}\right] \tag{2.11}$$

Thus we observe that with respect to the mean squared error criterion to order $0(n^{-2})$, the estimator P_{kg} with $0 < g < \frac{8}{C}$ dominates over S^2 for value of k according to inequality (2.10). Larger gain is found when (2.11) holds.

It may be pointed out that the second term in the relative mean square error of the estimator P_{kg} attains its minimum at $g = \frac{4}{C}$. Substituting $g = \frac{4}{C}$ in (2.9) yields the following expression

$$\frac{2}{n} - \frac{16}{n^2} + 8\left[\frac{8(C+4)}{C} - k\frac{C}{2}\right]\frac{C}{n^3} \tag{2.12}$$

implying the larger reduction in relative mean squared error in obtained when

$$k > \frac{16(C+4)}{C^2} \tag{2.13}$$

Further, it is observed from (2.9) that all the estimators of this general class P_{kg} have identical relative mean squared error upto $0(n^{-2})$ for any value of k, they differ with respect to the terms of order $0(n^{-3})$.

Thus it follows that for the positive values of the scalars the estimator P_{kg} when $g < \frac{6}{C}$ and P_{-kg} when $g > \frac{6}{C}$ are more useful for reducing the relative mean squared error.

Comparing the estimator P_{01} with P_{11}, we find

$$RB(P_{11}) < RB(P_{01}) \tag{2.14}$$

and

$$RM(P_{11}) < RM(P_{01}) \qquad \text{if } C < 6$$
$$RM(P_{11}) > RM(P_{01}) \qquad \text{if } C > 6 \tag{2.15}$$

Further, when we compare the estimators P_{01} and p_s, we obtain

$$RB(p_s) < RB(P_{01}) \qquad \text{if } C < 1.33 \tag{2.16}$$

$$RM(p_s) > RM(P_{01}) \qquad \text{if } C > 1.33 \tag{2.17}$$

Comparing the estimators P_{01} and T_α, we find

$$RB(P_{01}) < RB(T_\alpha) \tag{2.18}$$

$$RM(T_\alpha) < RM(P_{01}) \qquad \text{if } C > 6, \alpha > 0 \tag{2.19}$$

Taking $g = 1$, $k = -1$ in (2.5), we find the estimator P_{-11} of variance as

$$P_{kg} = s^2\left(1 + g\frac{s^2}{n\bar{y}^2} + k\frac{s^4}{n^2\bar{y}^4}\right)^{-1} \tag{2.20}$$

Comparing the relative biases and relative mean squared errors of the estimator P_{-11} with P_{01} and P_{11}, we observe that

$$RB(P_{11}) < RB(P_{01}) < RB(P_{-11}) \tag{2.21}$$

and

$$RM(P_{11}) < RM(P_{01}) < RM(P_{-11}) \quad \text{if } C < 6$$

$$RM(P_{-11}) < RM(P_{01}) < RM(P_{11}) \quad \text{if } C > 6 \tag{2.22}$$

Thus the estimator P_{11} has smaller bias in comparison to the estimators P_{01} and P_{-11}. According to mean squared error criterion, the estimator P_{11} when $C < 6$ and P_{-11} when $C > 6$ should be chosen in preference to the estimator P_{01}.

3. Simulation Result

The nature of the expressions for relative bias and relative mean squared error of the estimators P_{11}, P_{01} and P_{-11} are asymptotic so that the deduction of any inference about their efficiencies is difficult. With this in view the efficiencies of the estimators have been compared through simulation technique. 1500 random samples (each sample is one run) of sizes 10, 20 and 50 have been generated from normal populations $N(5, 5^2)$, $N(6, 10^2)$, $N(7, 15^2)$, $N(8, 20^2)$ and $N(9, 25^2)$. Relative biases and relative mean squared errors of the estimators have been calculated (Table 1). Assuming C is known for the same population, the relative biases and relative mean squared errors of the estimators have also worked out (Table 2).

In both the cases, absolute relative bias of all the estimators decreases with an increase in sample size whereas it increases as coefficient of variation increases. The estimator P_{-11} has minimum absolute bias than the estimators P_{01} and P_{11}.

Similarly, the relative mean squared error of all estimators decreases with an increase in sample size whereas it increases as coefficient of variation increases. Further the estimator P_{11} dominates over P_{01} and P_{-11} for $C \leq 6.25$ and the estimator P_{-11} is found to be more efficient than P_{01} and P_{11} for $C > 6.25$ which is satisfying the condition (2.22).

Table 1. Absolute relative biases and relative mean squared error of the estimators when C is unknown

Population sampled	Sample size (n)	Relative biases P_{11}	P_{01}	P_{-11}	Rel. mean squared errors P_{11}	P_{01}	P_{-11}
$N(5, 5^2)$	10	0.1978	0.1820	0.1663	0.11623	0.13117	0.14611
	20	0.0746	0.0712	0.0678	0.07997	0.08163	0.08328
	50	0.0237	0.0233	0.0228	0.03701	0.03710	0.03719
$N(6, 10^2)$	10	0.7127	0.6059	0.4991	0.26633	0.32469	0.38305
	20	0.2705	0.2421	0.2137	0.09179	0.09926	0.10673
	50	0.0817	0.0774	0.0730	0.03563	0.03610	0.03657
$N(7, 15^2)$	10	1.4681	1.1898	0.9114	1.19074	1.23105	1.27137
	20	0.5376	0.4604	0.3831	0.23455	0.23795	0.24135
	50	0.1538	0.1413	0.1288	0.04560	0.04580	0.04601
$N(8, 20^2)$	10	2.0997	1.6639	1.2280	2.46188	2.40943	2.35699
	20	0.7969	0.6646	0.5324	0.48854	0.47169	0.45484
	50	0.2206	0.1985	0.1765	0.06574	0.06449	0.06323
$N(9, 25^2)$	10	2.6809	2.0937	1.5066	3.96836	3.77315	3.57795
	20	1.1583	0.9425	0.7267	1.00274	0.93175	0.86077
	50	0.3165	0.2786	0.2406	0.11165	0.10597	0.10028

Table 2. Absolute relative biases and relative mean squared error of the estimators when C is known

Population sampled	Sample size (n)	Relative biases P_{11}	P_{01}	P_{-11}	Rel. mean squared errors P_{11}	P_{01}	P_{-11}
$N(5, 5^2)$	10	0.1500	0.1400	0.1300	0.12800	0.13800	0.14800
	20	0.0625	0.0600	0.0575	0.08225	0.08350	0.08475
	50	0.0220	0.0216	0.0212	0.03718	0.03726	0.03734
$N(6, 10^2)$	10	0.5648	0.4877	0.4105	0.17668	0.22641	0.27613
	20	0.2107	0.1914	0.1721	0.07895	0.08517	0.09138
	50	0.0670	0.0640	0.0609	0.03517	0.03557	0.03597
$N(7, 15^2)$	10	1.1836	0.9727	0.7619	0.75867	0.81805	0.87743
	20	0.4107	0.3580	0.3053	0.15027	0.15769	0.16512
	50	0.1208	0.1124	0.1039	0.03946	0.03994	0.04041
$N(8, 20^2)$	10	1.9219	1.5313	1.1406	2.06328	2.04375	2.02422
	20	0.6367	0.5391	0.4414	0.31924	0.31680	0.31436
	50	0.1769	0.1613	0.1456	0.05141	0.05125	0.05109
$N(9, 25^2)$	10	2.7121	2.1167	1.5213	4.05758	3.85324	3.64890
	20	0.8709	0.7221	0.5732	0.57946	0.55392	0.52837
	50	0.2319	0.2081	0.1843	0.07016	0.06852	0.6689

When C is unknown or known, the relative mean squared error of the estimators is compared to examine the effect of approximations (Table 3). It is observed that there is a deviation from – 0.00015 to –0.08922 and from 0.00046 to 0.39860 in the relative mean squared error of the estimators.

Table 3. Difference in the value of relative mean squared error of the estimators when C is unknown and known

Poplulation	Sample size (n)	Difference in relative mean squared error		
		P_{11}	P_{01}	P_{-11}
$N(5, 5^2)$	10	–0.01177	–0.00683	–0.00189
	20	–0.00228	–0.00187	–0.00147
	50	–0.00017	–0.00016	–0.00015
$N(6, 10^2)$	10	0.08965	0.09828	0.10692
	20	0.01284	0.01409	0.01535
	50	0.00046	0.00053	0.00060
$N(7, 15^2)$	10	0.43207	0.41300	0.39394
	20	0.08423	0.08026	0.07623
	50	0.00614	0.00586	0.00056
$N(8, 20^2)$	10	0.39860	0.36568	0.33277
	20	0.16930	0.15489	0.14048
	50	0.01433	0.01324	0.01214
$N(9, 25^2)$	10	–0.08922	–0.08009	–0.07095
	20	0.42328	0.37783	0.33240
	50	0.04149	0.03745	0.03339

ACKNOWLEDGEMENT

The author wishes to thank the referee for his critical and constructive comments which led to improve the paper. The author is also thankful to Dr. S.D. Batra and Dr. B.K. Hooda, Deptt. of Math. & Statistics for their valuable help in computation work.

REFERENCES

[1] Bhatnagar, S. (1986). Estimation of the square of mean in normal population. *Statistica, anno*, **XLVI(4)**, 521-525.

[2] Das, B. (1975). Estimation of μ^2 in normal population. *Cal. Stat. Assn. Bull.*, **24,** 135-140.

[3] Pandey, B.N. (1981). On estimation of variance in normal distribution. *Jour. Ind. Soc. Agril. Stat.*, **33(3)**, 1-6.

[4] Pandy, M. and Singh, J. (1981). Estimation of variance of normal population. *Comm. Stat. – Theory Methods,* **A10(18)**, 1879-1884.

[5] Singh, H.P. (1990). Estimation of parameters in normal parents. *Jour. Ind. Soc. Agril. Stat.,* **42(1)**, 98-107.

[6] Singh, J. and Pandy, M. (1975). On utilization of known coefficient of variation in estimation of variance. *J. Scientific Res.,* B.H.U., **25(1-2),** 165-172.