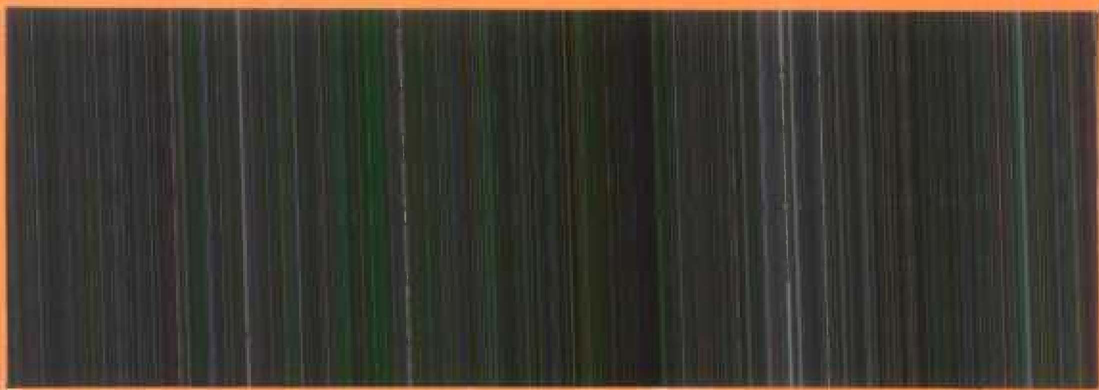
Statistics Canada | Statistique Canada



Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canada

[redacted]

WORKING PAPER NO. BSMD-87-004E
METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. BSMD-87-004E
MÉTHODOLOGIE

VARIANCE ESTIMATION OF MEDIANS IN STRATIFIED SAMPLES

by

J. Kovar

March 1987

* This is a preliminary version.
Do not quote without author's permission. Comments are welcome.

ABSTRACT

This paper addresses the problem of estimating the precision of the estimates of quantiles in stratified samples. Resampling methods, including the balanced half-sample replication and the bootstrap are compared to a method based on Woodruff confidence intervals. The jackknife variance estimator is shown to be inconsistent. Results of empirical studies on bias and stability of these variance estimators are presented and confidence interval coverage probabilities as well as lengths are reported.

SOMMAIRE

Cet article aborde le problème d'estimer la précision des estimés de quantiles dans les échantillons stratifiés. Des méthodes de re-échantillonnage, incluant la méthode des demi-échantillons balancés et le 'bootstrap' sont comparé à une méthode basée sur les intervalles de confiance de Woodruff. On montre que l'estimateur de 'jackknife' de la variance est inconsistent. Les résultats des études empiriques sur le biais et la stabilité de ces estimés de la variance sont présentés. L'article fournit aussi les probabilités de couverture par des intervalles de confiance ainsi que les longueurs.

VARIANCE ESTIMATION OF MEDIANS IN STRATIFIED SAMPLES

John G. Kovar*

1. INTRODUCTION

The estimation of variances of quantiles, medians in particular, has traditionally been somewhat elusive. Linearization methods useful for nonlinear statistics are difficult to implement for functionals such as the quantiles, since density estimation is involved. While Woodruff (1952) has proposed a novel method of estimating confidence intervals for the median, using these intervals to derive variances has been proposed only recently (Rao and Wu, 1986).

Resampling methods have generally been given little consideration ever since Miller (1974) has shown that in the case of the median, the jackknife yields inconsistent variance estimates, even in the iid case. Recent preliminary results due to Rao and Wu (1986) indicate, however, that consistent results can be obtained using the balanced repeated replication (BRR) method.

In this study we compare empirically the variance estimates of the median derived from the traditional Woodruff confidence intervals in stratified samples to various resampling methods, including the BRR, the random group method and the bootstrap. The study concludes that the Woodruff based method yields accurate and inexpensive variance estimates, while the BRR, the random group and a suitably chosen bootstrap method provide good alternatives. The jackknife estimator is confirmed to be inconsistent.

In the following section the estimation methods are described in more detail. Section 3 documents the simulation study and section 4 offers the conclusions and recommendations.

---

* John G. Kovar, Business Survey Methods Division, Statisics Canada, 11th floor, R. H. Coats Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6.

2. ESTIMATION METHODS

Suppose that in a stratified simple random sample, y_{hi} is the value of the characteristic of interest for the i'th sampled observation in stratum h ($h=1,\ldots,L$; $i=1,\ldots,n_h$), that W_h are the corresponding sampling weights, and that n is the total sample size. Let F_n be the sample distribution function, that is,

$$F_n(t) = \Sigma_h W_h F_{nh}(t), \tag{1}$$

where,

$$F_{nh}(t) = \Sigma_i I(y_{hi} \le t) / n_h , \tag{2}$$

and I is the indicator function. The p'th population quantile can then be estimated by

$$q = F_n^{-1}(p) , \tag{3}$$

and in particular, the median by $m = F_n^{-1}(1/2)$. Operationally, m is set equal to the value of the first observation for which the sum of W_h/n_h exceeds one half, when summing over the array of observations sorted by the values of y in ascending order.

Woodruff (1952) proposed a method of setting a confidence interval for the estimated median, m, of size alpha, given by

$$(F_n^{-1}(.5-z_{\alpha/2}s_p) , F_n^{-1}(.5+z_{\alpha/2}s_p)) , \tag{4}$$

where, in the case of stratified sampling, s_p is given by

$$s_p^2 = \Sigma_h W_h^2 F_{nh}(m) (1-F_{nh}(m)) / (n_h-1) . \tag{5}$$

In other words, the confidence interval is obtained by inverting the well known confidence interval for the proportion p (=.5). The variance of m can then be derived from this confidence interval by making the standard normal assumptions, as

$$v_{W(\alpha)} = (L(\alpha) / 2z_{\alpha/2})^2 , \tag{6}$$

where $L(\alpha)$ is the length of the confidence interval of size alpha and $z_{\alpha/2}$ is the upper $\alpha/2$-point of the standard normal distribution. Note that the variance, $v_{W(\alpha)}$, depends on the size, alpha, of the confidence interval. Francisco and Fuller (1986) and Rao and Wu (1986) established the asymptotic consistency of $v_{W(\alpha)}$ for any alpha. For the purposes of this study, eight levels of alpha were chosen, ranging from .01 to .5 .

Secondly, for designs with two units per stratum, the BRR method of variance estimation can be used. To produce the BRR estimates, R orthogonal half replicates were used to recompute the median, say m_r, $r=1,\ldots,R$. In this study , R was chosen to equal 36, the smallest multiple of four greater than 32, the number of strata in the population. The variation between these subsample estimates was then assumed to be the sampling variance of the median (McCarthy, 1966), namely,

$$v_{BRR\text{-}H} = \Sigma_r (m_r - m)^2 / R . \tag{7}$$

An alternate estimator of the variance can be produced by using the complementary half samples in (7) to obtain

$$v_{BRR\text{-}C} = \Sigma_r (m_{-r} - m)^2 / R , \tag{8}$$

where m_{-r} denotes the estimate of the median based on the r'th complementary half sample. Averaging the estimators in (7) and (8), we obtain an estimator customarily referred to as the full BRR estimator, given by

$$v_{BRR\text{-}F} = (v_{BRR\text{-}H} + v_{BRR\text{-}C}) / 2 . \tag{9}$$

A BRR difference estimator can also be defined, and is given by

$$v_{BRR-D} = \Sigma_r \ (\ m_r - m_{-r} \ )^2 \ / \ 4R \ . \tag{10}$$

All four BRR variance estimators were computed and are compared below.

Thirdly, for designs with equal allocation of K units per stratum, a repeated random group method was used. In this case, the sample was randomly split into K subsamples such that exactly one unit from each stratum was assigned to each subsample. An estimate of the median was produced from each such subsample, say m_k. The procedure was repeated 36 times (indexed by r) in order to make the method comparable to the BRR method. Two variance estimators were then defined as follows:

$$v_{RG-F} = \Sigma_r \ (\ \Sigma_k \ (\ m_{rk} - m \ )^2 \ / \ K(K-1) \ ) \ / \ R \tag{11}$$

$$v_{RG-D} = \Sigma_r \ (\ \Sigma_k \ (\ m_{rk} - \bar{m}_r \ )^2 \ / \ K(K-1) \ ) \ / \ R \tag{12}$$

where $\bar{m}_r = \Sigma_k \ m_{rk}/K$. Note that when K=2, the v_{RG-F} and v_{RG-D} estimators have the same form as the v_{BRR-F} and v_{BRR-D} estimators respectively, with the sole exception of how the subsamples are chosen. In other words, for K=2, the random group method provides an approximation of the BRR method, and its generalization in the case of K>2.

Fourthly, the bootstrap method, which relies on recomputing the estimate a large number (B) of times by resampling the original sample, can be used for designs with any allocation scheme. Two bootstrap estimators were considered: the naive bootstrap and the Rao-Wu modified bootstrap (Rao and Wu, 1983). In either case, the original sample is resampled such that n_h^* units in each stratum are selected from the original sample with replacement, and denoted by y_{hi}^*. In the case of the naive bootstrap, a bootstrap distribution function F_n^* is constructed using the resampled y_{hi}^* and inverted to find the median, say m_b, b=1,...,B. The variance estimator is then given by

$$v_{BT} = \Sigma_b \ (\ m_b - m \ )^2 \ / \ B \ . \tag{13}$$

On the other hand, the Rao-Wu bootstrap first constructs the "pseudo-values",

$$\tilde{y}_{hi} = \bar{y}_h + \sqrt{n_h^*/(n_h-1)}\ (y_{hi}^* - \bar{y}_h) , \tag{14}$$

and uses these to construct the distribution function $\tilde{F}_n$. This distribution function can then be inverted to obtain an estimate of the median, m_b, and its variance as in (13). For a justification of the "pseudo-values" in (14), see Rao and Wu (1983). Note that the two methods are equivalent when the bootstrap stratum sample size, n_h^*, is set to n_h-1. In this study the number of bootstrap replicates, B, was chosen to equal 500 in order to produce accurate confidence intervals. It was found that for the purposes of variance estimation alone, B=100 is sufficient.

Finally, six jackknife variance estimators were produced to confirm the suspicion of inconsistency. The notation is that of Rao and Wu (1985). In brief, the jackknife estimator is obtained by recalculating the estimate after deleting one unit from the original sample. The n estimates so obtained are then combined to produce three jackknife estimates. Six variance estimators are obtained by summing the square deviations of the individual delete-one estimates (or their pseudo-values) away from either the jackknife estimates or the original estimate. More precisely, let m^{-hi} be the "delete-one" estimate of the median, where the i'th observation in the h'th stratum is deleted, and let $m^h = \Sigma_i\ m^{-hi}/n_h$. The pseudo values are given by

$$m^{hi} = n_h m - (n_h-1)m^{-hi} , \tag{15}$$

and the jackknife estimates by

$$m_{J1} = (n+1-L)m - \Sigma_h\ (n_h-1)m^h \tag{16}$$

$$m_{J2} = \Sigma_h\ \Sigma_i\ m^{hi} / n \tag{17}$$

$$m_{J3} = \Sigma_h\ (\Sigma_i\ m^{hi} / n_h) / L . \tag{18}$$

The various variance estimators are then constructed as follows:

$$v_{JK1} = \Sigma_h (n_h-1)/n_h \Sigma_i (m^{-hi} - m^h)^2 \quad (19)$$

$$v_{JK2} = \Sigma_h (n_h-1)/n_h \Sigma_i (m^{-hi} - m)^2 \quad (20)$$

$$v_{JK3} = \Sigma_h (n_h-1)/n_h \Sigma_i (m^{-hi} - (\Sigma_h\Sigma_i m^{-hi}/n))^2 \quad (21)$$

$$v_{JK4} = \Sigma_h (n_h-1)/n_h \Sigma_i (m^{-hi} - (\Sigma_h m^h/32))^2 \quad (22)$$

$$v_{JK5} = \Sigma_h 1/((n_h-1)/n_h) \Sigma_i (m^{hi} - m_{J2})^2 \quad (23)$$

$$v_{JK6} = \Sigma_h 1/((n_h-1)/n_h) \Sigma_i (m^{hi} - m_{J3})^2 . \quad (24)$$

With the exception of the Woodruff and bootstrap methods, the confidence intervals were produced using the usual normal assumptions, and are given by

$$(m - z_{\alpha/2}v^{.5} , m + z_{\alpha/2}v^{.5}) , \quad (25)$$

where v is any one of the variance estimators described above. In the case of the Woodruff method, the confidence intervals are derived as indicated in (4). Finally, in the case of the bootstrap, a histogram of the B bootstrap estimates, m_b, was constructed and the confidence intervals were derived therefrom (Rao and Wu, 1983). Both the coverage probabilities as well as the confidence interval half-lengths (the distance from the estimate to the confidence interval end-point) were computed and are compared bellow.

3. THE SIMULATION STUDY

In terms of design, this study resembles the one documented in Kovar (1985), where variance estimates of nonlinear statistics were compared. The reader is referred there for additional details. In summary, the underlying population of the study consists of 32 strata in each of which the variable

Y is assumed to be normally distributed with the parameters E(Y) and S(Y) specified in Table 1. While hypothetical in nature, the population is intended to resemble a real population encountered by Hansen and Tepping (1985) in the National Assessment of Educational Progress study.

The true median of this population was found by solving (3) with $p=.5$ and F_{nh} set to the normal distribution function with the above parameters. The true MSE of the median was approximated by selecting 1000 independent samples according to the criteria set above and averaging the square deviations of the sample medians from the true median. The variance estimators described in the previous section were then compared to this "true" MSE in terms of accuracy and precision.

The simulation study itself consisted of selecting 500 samples from the underlying population and computing the various variance estimators. The accuracy of each of the estimators, v, was measured by the relative average of v, namely

$$\text{rel.ave.} = (\Sigma_s v_s / 500) / MSE , \tag{26}$$

where s indexes the 500 samples. The precision of the estimators was measured by their relative stability, defined by

$$\text{rel.stb.} = (\Sigma_s (v_s\text{-}MSE)^2 / 500)^{.5} / MSE . \tag{27}$$

The study was repeated under two sample designs: one with two units per stratum and the other with five. The results are presented in Table 2. The study was later repeated under different conditions by first doubling the variance of Y in each stratum and secondly by assuming that Y follows a gamma distribution instead of the normal. While, naturally, the results of these eight studies (2 designs x 2 levels of variation x 2 underlying distributions) are somewhat different, they are similar enough so that the overall conclusions that can be derived from Table 2 remain unchanged. As such, for reasons of economy of space, only the initial results are presented here.

The confidence intervals were compared in terms of their coverage probability properties and their lengths. The coverage probabilities are reported in terms of the proportion of intervals which fail to cover the true parameter by falling completely to the right or left of this parameter. Thus for example, a 90% nominal level confidence interval should ideally fail to cover the true median 5% of the time on <u>each</u> side. To compare the lengths of the confidence intervals, we report here the average half lengths, that is, the distance from the estimate to either end point of the interval. These half-lengths are equal in the case of confidence intervals which are symmetric by design, namely those obtained by the BRR, the random group and the jackknife methods. Eight nominal levels of confidence were compared, only one is reported here since the conclusions reached based on any one of them are identical. The results can be found in Table 3.

4. SUMMARY AND CONCLUSIONS

In terms of variance estimation, the Woodruff method yields accurate estimates. Surprisingly, the estimates get better as alpha decreases, not only in terms of accuracy but also in terms of stability. The natural choice of alpha=.05 seems quite reasonable.

For designs that lend themselves to its use, the BRR provides a good alternative to the Woodruff based method. However, we note that the BRR-D method tends to underestimate the true variance, and thus the full estimator, BRR-F, is recommended here. The results in Table 2 also confirm that the random group method is a good approximation of the BRR method as can be seen in the case of the design with two units per stratum. In the case of five units in each stratum, the RG-F estimator performs as well if not better than the Woodruff based estimators, while the RG-D estimator tends to underestimate.

On the other hand, it is quite clear that the naive bootstrap with $n_h^* = n_h$ is inappropriate. In the design with two units per stratum, it underestimates by nearly 40%, but improves when the stratum sample size is increased to 5. As discussed in Rao and Wu (1985), this estimator is expected to underestimate the true variance by a factor of $(n_h-1)/n_h$. The Rao-Wu bootstrap performs satisfactorily, with $n_h^* = n_h-1$ likely the best choice. However, in the case of $n_h^* = n_h-3$, serious overestimation of 54% results.

Finally, we note that the jackknife variance estimators perform poorly, confirming their inconsistency, with v_{JK2} (the most popular estimator in practice) being likely the worst. The estimators overstate the true variance by 30-70% in the design with two units per stratum and perform even worse in the five unit per stratum design. Note also that the standard errors of the jackknife variance estimators are inordinately high. As a point of interest, it can be shown that for designs with equal allocation in all strata, the jackknife variance estimators 3 through 6 are mathematically identical.

Similar conclusions can be drawn regarding the confidence intervals. The jackknife method produces the longest intervals with, paradoxically, substantially lower coverage probabilities than the nominal level. On the other hand, the Woodruff, the BRR and the random group method confidence intervals are all of approximately the same length and all within the sampling error of the true nominal level, even though they all tend to be slightly anticonservative. Both the bootstrap methods that use $n_h^* = n_h$ produce shorter confidence intervals, but at the expense of low coverage probabilities. The best choice of n_h^* is again n_h-1. In this case the confidence intervals are similar to those obtained by the BRR method as can be expected. The Rao-Wu bootstrap with $n_h^* = n_h-3$ produces conservative confidence intervals but at the cost of excessive length.

Overall, we conclude that the Woodruff based method yields accurate variance estimates and good confidence intervals, both in terms of coverage probabilities as well as lengths, at a relatively very low cost. Provided the design is appropriate, the BRR and the random group methods provide good, albeit more expensive, alternatives. For more general designs, the bootstrap method with $n_h^* = n_h-1$ can be used as a relatively expensive alternative but one which yields asymmetrical confidence intervals - a possible advantage in skewed populations. Under no circumstances should the jackknife "delete-one" method be used to estimate variances nor confidence intervals for the median.

ACKNOWLEDGEMENT

Thanks are due to Dr. J.N.K. Rao for proposing the study and for his valuable comments and suggestions.

REFERENCES

Francisco, C.A. and Fuller, W.A. (1986). Estimation of the distribution function with a complex survey. Ph.D. dissertation, Iowa State University.

Hansen, M.H. and Tepping, B.J. (1985). Estimation of variance in NAEP. Unpublished manuscript.

Kovar, J. (1985). Variance estimation of nonlinear statistics in stratified samples. Statistics Canada, Methodology Branch Working paper No. BSMD 85-052E.

McCarthy, P.J. (1966). Replication: An approach to the analysis of data from complex surveys. Vital and Health Statistics, Series 2, Number 14.

Miller, R.G. (1974). The jackknife - a review. Biometrika 61, 1-17.

Rao, J.N.K. and Wu, C.F.J. (1983). Bootstrap inference with stratified samples. Technical Report Series of the Laboratory for Research in Statistics and Probability No.19.

Rao, J.N.K. and Wu, C.F.J. (1985). Inference from stratified samples: Second order analysis of three methods for nonlinear statistics. Journal of the American Statistical Association 80, 620-630.

Rao, J.N.K. and Wu, C.F.J. (1986). Private correspondence.

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. Journal of the American Statistical Association 47, 635-646.

Table 1: The population parameters

Stratum(h)	W_h	$E(Y_h)$	$S(Y_h)$
1	.042	90	12.5
2	.042	75	12.0
3	.042	70	11.0
4	.039	75	11.0
5	.039	70	10.0
6	.037	75	12.0
7	.037	75	11.5
8	.037	75	11.0
9	.037	70	12.0
10	.034	75	11.5
11	.034	70	10.0
12	.034	70	11.0
13	.034	70	11.0
14	.031	75	12.5
15	.031	70	10.0
16	.031	70	9.0
17	.031	70	9.5
18	.031	70	10.0
19	.031	65	10.0
20	.031	60	9.0
21	.031	60	8.0
22	.031	60	10.0
23	.028	70	11.0
24	.028	65	9.0
25	.028	60	10.0
26	.025	70	10.0
27	.025	60	9.0
28	.025	50	7.5
29	.025	50	7.0
30	.020	50	8.0
31	.016	45	7.0
32	.013	45	6.0

Table 2: The relative average variance and its stability under various methods.

Method	$n_h = 2$		$n_h = 5$	
	Rel.Ave	Rel.Stb	Rel.Ave	Rel.Stb
W(.005)	1.046	0.52	1.068	0.39
W(.010)	1.032	0.52	1.050	0.40
W(.025)	1.037	0.57	1.047	0.43
W(.050)	1.052	0.65	1.050	0.48
W(.100)	1.047	0.72	1.082	0.57
W(.150)	1.075	0.82	1.117	0.66
W(.200)	1.076	0.90	1.110	0.74
W(.250)	1.102	1.11	1.149	0.85
BRR-H	1.115	0.70	-	-
BRR-C	1.110	0.69	-	-
BRR-F	1.113	0.68	-	-
BRR-D	0.911	0.56	-	-
RG-F	1.120	0.71	1.053	0.34
RG-D	0.916	0.58	0.991	0.32
NAIVE BT(n_h)*	0.613	0.60	0.934	0.44
R-W BT(n_h)	1.179	0.79	1.122	0.54
R-W BT(n_h-1)	1.109	0.68	1.147	0.53
R-W BT(n_h-3)	-	-	1.544	0.84
JK-1	1.347	2.04	1.681	2.47
JK-2	1.911	3.19	2.807	4.57
JK-3	1.725	2.77	2.234	3.41
JK-4	1.725	2.77	2.234	3.41
JK-5	1.725	2.77	2.234	3.41
JK-6	1.725	2.77	2.234	3.41

* For the bootstrap methods, the bracketed figure indicates the bootstrap stratum sample size.

Table 3: The confidence intervals: Tail coverage probabilities and half-lengths (left and right). (Nominal level = .05 in each tail)

Method	Design							
	$n_h = 2$				$n_h = 5$			
	Tail Prob.		Length		Tail Prob.		Length	
WOODRUFF	.052	.048	3.11	3.18	.066	.054	1.94	1.97
BRR-H	.048	.058	3.24	3.24	-	-	-	-
BRR-C	.058	.060	3.23	3.23	-	-	-	-
BRR-F	.054	.054	3.24	3.24	-	-	-	-
BRR-D	.078	.078	2.93	2.93	-	-	-	-
RG-F	.052	.050	3.24	3.24	.074	.040	1.98	1.98
RG-D	.074	.074	2.93	2.93	.078	.050	1.92	1.92
NAIVE BT(n_h)	.120	.118	2.09	2.26	.088	.086	1.72	1.80
R-W BT(n_h)	.096	.070	3.08	2.71	.088	.060	1.99	1.81
R-W BT(n_h-1)	.056	.064	2.93	3.12	.070	.064	1.92	2.00
R-W BT(n_h-3)	-	-	-	-	.016	.056	1.75	2.54
JK-1	.124	.104	3.07	3.07	.142	.118	2.04	2.04
JK-2	.086	.082	3.63	3.63	.104	.092	2.59	2.59
JK-3	.092	.088	3.47	3.47	.114	.100	2.35	2.35
JK-4	.092	.088	3.47	3.47	.114	.100	2.35	2.35
JK-5	.092	.088	3.47	3.47	.114	.100	2.35	2.35
JK-6	.092	.088	3.47	3.47	.114	.100	2.35	2.35

* In a sample of 500, the probabilities above can be reported within 0.02, 95% of the time.