Metrika (1998) 47:203–213

Metrika

Quantile Interval Estimation in Finite Population Using a Multivariate Ratio Estimator

M. RUEDA GARCÍA and A. ARCOS CEBRIÁN

Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Granada, E-18071 Granada, Spain
e-mail: mrueda@goliat.ugr.es

E. ARTÉS RODRÍGUEZ

Departamento de Estadística e Investigación Opertiva, Universidad de Almería, Spain
e-mail: eartes@ualm.es

Abstract: *A new method to derive confidence intervals for quantiles in a finite population is presented. This method uses multi-auxiliary information through a multi-variate ratio type estimator of the population distribution function.*

Key Words and Phrases: ratio type estimator, finite population quantile, confidence interval.

1 Introduction

In survey practice it is often of interest to estimate the median and other finite population quantiles of a skewed variable such as income. While there is an extensive literature on the estimation of means and totals in sample surveys, relatively less effort has been devoted to the development of efficient methods for estimating finite population quantiles. Furthermore, emphasis has been laid on the use of auxiliary information for improving the estimates through indirect methods of estimation as ratio-type and regression estimators. These have been extensively studied for the population mean and have been introduced for the variance parameter.

In this paper we present results pertaining to similar investigations concerning quantile estimation. Sample medians have long been recognized as simple robust alternatives to sample means, for estimating location of heavy-tailed or markedly skewed populations from simple random samples. Its simplicity relative to other robust estimates, indicated its choice for investigation in designs other than simple random sampling.

Recently, the point estimation for finite population median and other finite population quantiles in presence of only one auxiliary variable has received considerable attention. Relevant references are *Chambers* and *Dunstan* (1986), *Rao*, *Kovar* and *Mantel* (1990), *Kuk* and *Mak* (1989), *Mak* and *Kuk* (1993).

Section 2 revises the confidence interval method for constructing confidence intervals proposed by *Woodruff*, (1952).

In Section 3 we present an alternative confidence interval for a finite population quantile when the quantile of an auxiliary variable is known.

Section 3 is the extension to multivariate case of ratio method in section 2.

Section 4 shows a simulation study.

2 Woodruff's Method

As usual, let $y_1, \dots, y_N$ be the values of the population elements $U_1, \dots, U_N$, for the variable of interest y. For any y $(-\infty < y < \infty)$, the population distribution $F_Y(y)$ is defined as the proportion of elements in the population that are less than or equal to y. The population β quantile is

$$Q_Y(\beta) = \inf\{y|F_Y(y) \geq \beta\} = F_Y^{-1}(\beta) .$$

The problem is to estimate the population quantile $Q_Y(\beta)$, using data y_k for $k \in s$, where s is a simple random sample. The general procedure is based on obtaining an estimated distribution function, $\hat{F}_Y(y)$, and then we estimate $Q_Y(\beta)$ as $\hat{Q}_Y(\beta) = \hat{F}_Y^{-1}(\beta)$, when the inverse $\hat{F}_Y^{-1}$ is to be understood in the same way as F_Y^{-1} above.

Woodruff describes a large sample procedure for determining confidence intervals for a finite population median: For any two constants d_1 and d_2, and for any values of $Q_Y(\beta)$,

$$P\{d_1 \leq \hat{F}_Y(Q_Y(\beta)) \leq d_2\} \simeq P\{\hat{F}_Y^{-1}(d_1) \leq Q_Y(\beta) \leq \hat{F}_Y^{-1}(d_2)\} .$$

Hence, it follows that for any d_1 and d_2 constants such that $P\{d_1 \leq \hat{F}_Y(Q_Y(\beta)) \leq d_2\} = 1 - \alpha$, the interval $[\hat{F}_Y^{-1}(d_1), \hat{F}_Y^{-1}(d_2)]$ is a $100(1-\alpha)\%$ approximate confidence interval for $Q_Y(\beta)$.

In simple random sampling, as $n\hat{F}_Y(Q_Y(\beta))$ is hypergeometrically distributed variable, then $E(\hat{F}_Y(Q_Y(\beta))) = F_Y(Q_Y(\beta)) = \beta$ and $V(\hat{F}_Y(Q_Y(\beta))) = \frac{1-f}{n}\beta(1-\beta)$. If the sample size n is sufficiently large then $\hat{F}_Y(Q_Y(\beta))$ is approximately normal, and we would choose the smallest confidence interval as

$$d_1 = \beta - z_{\alpha/2}\left\{\frac{1-f}{n}\beta(1-\beta)\right\}^{1/2} \quad \text{and} \quad d_2 = \beta + z_{\alpha/2}\left\{\frac{1-f}{n}\beta(1-\beta)\right\}^{1/2} ,$$

where $f = n/N$ is the sampling fraction and $z_{\alpha/2}$ denotes the upper $1-\alpha/2$ percentage point of the standard normal distribution.

3 Ratio Method to Derive Confidence Intervals for a Finite Population Quantile

Let y and x respectively be the survey variable and the auxiliary variable related to y. We assume that the population quantile $Q_X(\beta)$ of x is known. From the sample of n units from a population of size N we observe (x_k, y_k) where $k \in s$.

Considering the ratio type estimator

$$\hat{F}_R(Q_Y(\beta)) = \frac{\hat{F}_Y(Q_Y(\beta))}{\hat{F}_X(Q_X(\beta))} F_X(Q_X(\beta))$$

and assuming that $\hat{F}_R(Q_Y(\beta))$ is approximately normally distributed (see *Kuk* and *Mak*, 1989) we would chosse

$$c_1 = \beta - z_{\alpha/2}\{V(\hat{F}_R(Q_Y(\beta)))\}^{1/2} \quad \text{and} \quad c_2 = \beta + z_{\alpha/2}\{V(\hat{F}_R(Q_Y(\beta)))\}^{1/2}$$

and thus

$$P\left(\hat{F}_Y^{-1}\left(c_1 \frac{\hat{F}_X(Q_X(\beta))}{\beta}\right) \le Q_Y(\beta) \le \hat{F}_Y^{-1}\left(c_2 \frac{\hat{F}_X(Q_X(\beta))}{\beta}\right)\right) = 1-\alpha \, .$$

To evaluate the variance $V(\hat{F}_R(Q_Y(\beta)))$ we make the variables

$$e_0 = \frac{\hat{F}_Y(Q_Y(\beta)) - F_Y(Q_Y(\beta))}{F_Y(Q_Y(\beta))} \, , \quad e_1 = \frac{\hat{F}_X(Q_X(\beta)) - F_X(Q_X(\beta))}{F_X(Q_X(\beta))} \, .$$

Then

$$\hat{F}_R(Q_Y(\beta)) = F_Y(Q_Y(\beta)) \frac{1+e_0}{1+e_1} \, .$$

Taylor's series expansion yields

$$(\hat{F}_R(Q_Y(\beta)) - F_Y(Q_Y(\beta)))^2 \simeq F_Y(Q_Y(\beta))^2(e_0^2 + e_1^2 - 2e_0e_1) ,$$

and thus

$$\begin{aligned} V(\hat{F}_R(Q_Y(\beta))) &\simeq F_Y(Q_Y(\beta))^2(E(e_0^2) + E(e_1^2) - 2E(e_1e_0)) \\ &= \beta^2\left(\frac{1}{\beta^2} V(\hat{F}_X(Q_X(\beta))) + \frac{1}{\beta^2} V(\hat{F}_Y(Q_Y(\beta))) \right. \\ &\quad \left. - \frac{2}{\beta^2} \operatorname{Cov}(\hat{F}_X(Q_X(\beta)), \hat{F}_Y(Q_Y(\beta)))\right) \\ &= 2\,\frac{1-f}{n}\,\beta(1-\beta) - 2\operatorname{Cov}(\hat{F}_X(Q_X(\beta)), \hat{F}_Y(Q_Y(\beta))) . \end{aligned} \tag{3.1}$$

To evaluate $\operatorname{Cov}(\hat{F}_X(Q_X(\beta)), \hat{F}_Y(Q_Y(\beta)))$, consider now the two-way classification

(3.2)

	$x \leq Q_X(\beta)$	$x > Q_X(\beta)$	
$y \leq Q_Y(\beta)$	$n_{11} \backslash N_{11}$	$n_{12} \backslash N_{12}$	$n_{1\cdot} \backslash N_{1\cdot}$
$y > Q_Y(\beta)$	$n_{21} \backslash N_{21}$	$n_{22} \backslash N_{22}$	$n_{2\cdot} \backslash N_{2\cdot}$
	$n_{\cdot 1} \backslash N_{\cdot 1}$	$n_{\cdot 2} \backslash N_{\cdot 2}$	

where n_{11} denotes the number of units in the sample with $x \leq Q_X(\beta)$ and $y \leq Q_Y(\beta)$. Similarly N_{11} is the number of units in the population with $x \leq Q_X(\beta)$ and $y \leq Q_Y(\beta)$. Hence, $(n_{11}, n_{12}, n_{21}, n_{22})$ is a hypergeometrically distributed random variable

$$(n_{11}, n_{12}, n_{21}, n_{22}) \simeq HG(N, n, N_{11}, N_{12}, N_{21}) ,$$

and $n\hat{F}_Y(Q_Y(\beta)) = n_{11} + n_{12}$ and $n\hat{F}_X(Q_X(\beta)) = n_{11} + n_{21}$ we obtain

$$\begin{aligned} \operatorname{Cov}(n\hat{F}(Q_Y(\beta)), n\hat{F}(Q_X(\beta))) &= \operatorname{Cov}(n_{11} + n_{12}, n_{11} + n_{21}) \\ &= V(n_{11}) + \operatorname{Cov}(n_{11}, n_{12}) + \operatorname{Cov}(n_{11}, n_{21}) + \operatorname{Cov}(n_{12}, n_{21}) . \end{aligned}$$

Furthermore, it can be seen that

$$\mathrm{Cov}(n_i, n_j) = -\frac{N-n}{N-1}\, n\, \frac{N_i N_j}{N^2}\,, \quad V(n_{ij}) = \frac{N-n}{N-1}\, n\, \frac{N_{ij}}{N}\left(1 - \frac{N_{ij}}{N}\right)$$

and thus

$$\mathrm{Cov}(n\hat{F}_Y(Q_Y(\beta)), n\hat{F}_X(Q_X(\beta))) = \frac{N-n}{N-1}\frac{n}{N^2}(N_{11}N_{22} - N_{12}N_{21})$$

Substituting the last expression in (3.1) we have

$$V(\hat{F}_R(Q_Y(\beta))) \simeq \frac{1-f}{n}\, 2\left(\beta(1-\beta) - \frac{N_{11}N_{22} - N_{12}N_{21}}{N^2}\right) \tag{3.3}$$

and using Cramer's V coefficient

$$\phi(\beta) = \frac{N_{11}N_{22} - N_{12}N_{21}}{\sqrt{N_{1\cdot}N_{2\cdot}N_{\cdot 1}N_{\cdot 2}}}$$

(3.3) can be writen as follows

$$V(\hat{F}_R(Q_Y(\beta))) \simeq 2\,\frac{1-f}{n}\,\beta(1-\beta)(1-\phi(\beta))\,.$$

A similar cross-classification as before

(3.4)

	$x \le Q_X(\beta)$	$x > Q_X(\beta)$	
$y \le \hat{Q}_Y(\beta)$	$\tilde{n}_{11}\backslash\tilde{N}_{11}$	$\tilde{n}_{12}\backslash\tilde{N}_{12}$	$\tilde{n}_{1\cdot}\backslash\tilde{N}_{1\cdot}$
$y > \hat{Q}_Y(\beta)$	$\tilde{n}_{21}\backslash\tilde{N}_{21}$	$\tilde{n}_{22}\backslash\tilde{N}_{22}$	$\tilde{n}_{2\cdot}\backslash\tilde{N}_{2\cdot}$
	$\tilde{n}_{\cdot 1}\backslash\tilde{N}_{\cdot 1}$	$\tilde{n}_{\cdot 2}\backslash\tilde{N}_{\cdot 2}$	

yields,

$$\tilde{\phi}(\beta) = \frac{\tilde{n}_{11}\tilde{n}_{22} - \tilde{n}_{12}\tilde{n}_{21}}{\sqrt{\tilde{n}_{1\cdot}\tilde{n}_{2\cdot}\tilde{n}_{\cdot 1}\tilde{n}_{\cdot 2}}}$$

as an estimator for $\phi(\beta)$ and thus

$$\hat{V}(\hat{F}_R(Q_Y(\beta))) = \frac{1-f}{n} 2\beta(1-\beta)(1-\tilde{\phi}(\beta)) \tag{3.5}$$

is an estimator for $V(\hat{F}_R(Q_Y(\beta)))$.

Finally, if we denote $\tilde{r}_i(\beta) = \frac{\hat{F}_X(Q_X(\beta))}{\beta} \tilde{c}_i$ and

$$\tilde{c}_i(\beta) = \beta + (-1)^i z_{\alpha/2}\left\{\frac{1-f}{n} 2\beta(1-\beta)(1-\tilde{\phi}(\beta))\right\}^{1/2}, \quad i = 1,2,$$

the $100(1-\alpha)\%$ confidence interval for $Q_Y(\beta)$ is $[\hat{F}_Y^{-1}(\tilde{r}_1(\beta)), \hat{F}_Y^{-1}(\tilde{r}_2(\beta))]$.

Note that, for example if $\beta = 0.5$, $Q_Y(\beta) = M_Y$, the population median, from $V(\hat{F}(M_Y)) = \frac{1-f}{n} 0.25$, we obtain $V(\hat{F}_R(M_Y)) < V(\hat{F}_Y(M_Y)) \Leftrightarrow \phi > 1/2$, that is, the ratio type estimator $\hat{F}_R$ for the population distribution function is more precise than the usual estimator $\hat{F}_Y$ if and only if an association between $y \leq M_Y$ and $x \leq M_X$ exists, where $M_X = Q_X(0.5)$.

4 The Multivariate Case

Let y and x_i $(i = 1, \ldots, l)$ respectively be the survey variable and the auxiliary variables related to y. We assume that the population quantiles $Q_{x_i}(\beta)$ of x_i $(i = 1, \ldots, l)$ are known.

From the sample of n units from a population of size N we observe (x_{i_k}, y_k) where $k \in s, i = 1, \ldots, l$.

We are going to use this reasoning to obtain an alternative confidence interval which is based on the construction of a multiple ratio estimator of the population distribution function.

Consider the ratio-type estimators:

$$\hat{F}_{R_i}(Q_Y(\beta)) = \frac{\hat{F}_Y(Q_Y(\beta))}{\hat{F}_{X_i}(Q_{X_i}(\beta))} \beta, \quad i = 1, \ldots, l.$$

These estimators constructed for each auxiliary variable, can combine to increase precision, in the following way:

$$\hat{F}_R^M(Q_Y(\beta)) = \sum_{i=1}^{l} \omega_i \hat{F}_{R_i}(Q_Y(\beta)),$$

where $\omega = (\omega_1, \ldots, \omega_l), \sum_{i=1}^{l} \omega_i = 1$, is a weighting function.

Let c_1 and c_2 be constants such that

$$P\{c_1 \leq \hat{F}_R^M(Q_Y(\beta)) \leq c_2\} = 1 - \alpha$$

Then

$$P\left\{\frac{c_1}{\beta}\left(\sum_{i=1}^{l}\frac{\omega_i}{\hat{F}_{X_i}(Q_{X_i}(\beta))}\right)^{-1} \leq \hat{F}_Y(Q_Y(\beta)) \leq \frac{c_2}{\beta}\left(\sum_{i=1}^{l}\frac{\omega_i}{\hat{F}_{X_i}(Q_{X_i}(\beta))}\right)^{-1}\right\}$$
$$= 1 - \alpha$$

and

$$\left[\hat{F}_Y^{-1}\left(\frac{c_1}{\beta}\left(\sum_{i=1}^{l}\frac{\omega_i}{\hat{F}_{X_i}(Q_{X_i}(\beta))}\right)^{-1}\right), \hat{F}_Y^{-1}\left(\frac{c_2}{\beta}\left(\sum_{i=1}^{l}\frac{\omega_i}{\hat{F}_{X_i}(Q_{X_i}(\beta))}\right)^{-1}\right)\right],$$

(if $\hat{F}_Y$ is a continuous and strictly increasing function) is approximately a $100(1-\alpha)\%$ confidence interval for the population quantile $Q_Y(\beta)$.

As $\hat{F}_{R_i}(Q_Y(\beta))(i = 1, \ldots, l)$ is approximately normally distributed, (see *Kuk* and *Mak*, 1989), then the multiple estimator $\hat{F}_R^M(Q_Y(\beta))$ is also approximately normally distributed with expected value $F_Y(Q_Y(\beta)) = \beta$. Therefore we would choose

$$c_1 = \beta - z_{\alpha/2}\{V(\hat{F}_R^M(Q_Y(\beta)))\}^{1/2}, \quad c_2 = \beta + z_{\alpha/2}\{V(\hat{F}_R^M(Q_Y(\beta)))\}^{1/2}.$$

To evaluate this variance, we take into account that

$$V(\hat{F}_R^M(Q_Y(\beta))) = \sum_{i=1}^{l}\omega_i^2 V(\hat{F}_{R_i}(Q_Y(\beta)))$$
$$+ 2\sum_{i<j}\omega_i\omega_j \operatorname{Cov}(\hat{F}_{R_i}(Q_Y(\beta)), \hat{F}_{R_j}(Q_Y(\beta))).$$

From (3.5)

$$\hat{V}(\hat{F}_{R_i}(Q_Y(\beta))) = \frac{1-f}{n} 2\beta(1-\beta)(1-\tilde{\phi}_i(\beta)), \quad i = 1, \ldots, l$$

and using a similar two-way classification with the variables X_i and X_j as (3.2) (note that the coefficients $\phi_{ij}(\beta)$ are known $(i,j = 1,\dots,l; i \neq j)$, we obtain

$$\mathrm{Cov}(\hat{F}_{X_i}(Q_Y(\beta)), \hat{F}_{X_j}(Q_Y(\beta))) = \frac{1-f}{n}\beta(1-\beta)\phi_{ij}(\beta)\,,$$

$$\mathrm{Cov}(\hat{F}_{R_i}(Q_Y(\beta)), \hat{F}_{R_j}(Q_Y(\beta)),) \simeq \frac{1-f}{n}\beta(1-\beta)[1-\tilde{\phi}_i(\beta)-\tilde{\phi}_j(\beta)+\phi_{ij}(\beta)]\,,$$

$$\hat{V}(\hat{F}_R^M(Q_Y(\beta)))$$

$$\simeq 2\frac{1-f}{n}\beta(1-\beta)\left[\sum_{i=1}^{l}\omega_i^2(1-\tilde{\phi}_i(\beta))+\sum_{i<j}\omega_i\omega_j(1-\tilde{\phi}_i(\beta)-\tilde{\phi}_j(\beta)+\phi_{ij}(\beta))\right]$$

and therefore the interval $[\hat{F}_Y^{-1}(\tilde{r}_1), \hat{F}_Y^{-1}(\tilde{r}_2)]$ is a $100(1-\alpha)\%$ confidence interval for $Q_Y(\beta)$, where

$$\tilde{r}_k = \frac{\beta + (-1)^k z_{\alpha/2}[\hat{V}(\hat{F}_R^M(Q_Y(\beta)))]^{1/2}}{\beta \sum_{i=1}^{l} \frac{\omega_i}{\hat{F}_{X_i}(Q_{X_i}(\beta))}}\,.$$

Now we derive a confidence interval from the optimum multiple ratio estimator. We have seen above how to obtain an "optimum" multiple ratio-type estimator, $\hat{F}_R^M(Q_Y(\beta))$, from a weighting function $\omega = (\omega_1,\dots,\omega_l)$ such that $\sum \omega_i = 1$. Further, we have also seen how to derive a confidence interval for the population median from this multiple ratio estimator.

Consider now the problem of choosing the most appropriate weights. The solution is simple: we will select the weights ω_i $(i = 1,\dots,l)$ which increase the accuracy of the multiple ratio estimator.

If we define $\omega = (\omega_1,\dots,\omega_l)'$ and $\hat{F}_R = (\hat{F}_{R_1}(Q_Y(\beta)),\dots,\hat{F}_{R_l}(Q_Y(\beta)))'$, then, we have

$$\hat{F}_R^M(Q_Y(\beta)) = \omega'\hat{F}_R \quad \text{and} \quad V(\hat{F}_R^M(Q_Y(\beta))) = \omega' A \omega\,,$$

where $A = (a_{ij}), a_{ij} = \mathrm{Cov}(\hat{F}_{R_i}(Q_Y(\beta)), \hat{F}_{R_j}(Q_Y(\beta))), a_{ii} = V(\hat{F}_{R_i}(Q_Y(\beta)))$.

The criterion for optimality of the weight vector $\omega = (\omega_1,\dots,\omega_l)$ with $\sum \omega_i = 1$ is to minimize the variance. To obtain the extremum we make use of the generalized Cauchy-Schwarz inequality, and, since A is positive semidefinite, it follows that the optimum ω is given by

$$\hat{\omega} = \frac{A^{-1}e}{e'A^{-1}e},$$

where $e = (1, \ldots, 1)'$.

Insertion of $\hat{\omega}$ in the variance yields

$$V_{min}(\hat{F}_R^M(Q_Y(\beta))) = \frac{1}{\hat{\omega}'A^{-1}\hat{\omega}}.$$

5 Conclusions: Simulation Studies

In this section we look into the properties of the method proposed above by means of simulation studies. Samples of different sizes from a population have been generated. For each sample we compute the length of confidence intervals using two methods: the classical method, followed by the method based on the ratio estimator.

We have observed that the real coverage is approximately the same and the average length of the confidence intervals using the proposed method is lower than those found in the other method.

In conclusion, we have been able to achieve a substantial improvement in the accuracy of the estimates using the intervals obtained from the proposed method.

The population considered consists of 1500 families living in an Andalucian province (*Fernández* and *Mayor*, 1994). In this we have considered three variables: the main variable y denotes the cost of food, and two auxiliary variables x_1 and x_2 which denote family income and other cost, respectively.

We have taken 500 samples of the following sizes 30, 35, 40, 45, 50 and 100, and from these samples we have constructed the confidence intervals: firstly, without any auxiliary information; secondly, using only x_1 as the auxiliary variable; thirdly, using only x_2 as the auxiliary variable; and lastly, using both of the above variables. The β-quantile is the population median and $100(1-\alpha)\% = 95\%$.

In this especial bivariate case, the weights that yield the optimum multivariate ratio estimator are given by

$$\hat{\omega} = \frac{A^{-1}e}{e'A^{-1}e} = \left(\frac{\tilde{\phi}_1 - \phi_{12}}{\tilde{\phi}_1 + \tilde{\phi}_2 - 2\phi_{12}}, \frac{\tilde{\phi}_2 - \phi_{12}}{\tilde{\phi}_1 + \tilde{\phi}_2 - 2\phi_{12}}\right).$$

Table 1. Empirical comparisons of confidence intervals for household population median.

		$100(1-\alpha)\% = 95\%$		
n	method	$\bar{l}$	Cove	σ_l^2
30	classical:	1120.17	0.960	87294.19
	ratio x_1:	938.80	0.922	145253.39
	ratio x_2:	1235.39	0.912	217188.25
	ratio x_1 and x_2:	840.05	0.900	103313.34
35	classical:	1045.72	0.980	60084.93
	ratio x_1:	845.46	0.924	110498.32
	ratio x_2:	1103.08	0.924	132910.30
	ratio x_1 and x_2:	775.64	0.918	88970.29
40	classical:	987.63	0.960	52752.10
	ratio x_1:	794.27	0.918	86697.00
	ratio x_2:	1034.04	0.930	103167.12
	ratio x_1 and x_2:	723.11	0.928	67534.16
45	classical:	803.04	0.952	39796.46
	ratio x_1:	718.94	0.944	55193.70
	ratio x_2:	979.87	0.938	87757.36
	ratio x_1 and x_2:	670.09	0.938	48405.64
50	classical:	783.71	0.938	38776.56
	ratio x_1:	694.02	0.916	57016.51
	ratio x_2:	934.30	0.926	67488.43
	ratio x_1 and x_2:	642.77	0.928	48087.54
100	classical:	563.52	0.952	13216.27
	ratio x_1:	465.28	0.954	14197.85
	ratio x_2:	623.32	0.952	20003.33
	ratio x_1 and x_2:	431.55	0.940	11797.69

In Table 1 we can see the average lengths of the intervals, the variance of these lengths and the proportion of times that these intervals relate to the true median, for a 95% confidence level.

As we can observe for all the sample sizes, the multiple method that we proposed produces intervals with an average length inferior to the other three methods. The use of the variable x_1 only, produces intervals whose average length is lower than the direct methods, which does not occur when only the variable x_2 is considered. However, when we consider both auxiliary variables using the multiple method proposed, it produces much smaller intervals. We can also see that for the smaller sizes there is a real coverage less than the nominal coverage, and as the sample size increases, this problem disappears.

References

Chambers RL, Dunstan R (1986) Estimating distribution functions from survey data. Biometrika 73:597–604

Kuk AYC, Mak TK (1989) Median estimation in presence of auxiliary information. J. R. Statist. Soc. B 51:261–269

Fernández García FR, Mayor Gallego JA (1994) Muestreo en poblaciones finitas: Curso básico. P.P.U., Barcelona

Mak TK, Kuk AYC (1993) A new method for estimating finite-population quantiles using auxiliary information. The Canadian Journal of Statistics 21:29–38

Rao JNK, Kovar JG, Mantel HJ (1990) On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika 77:365–375

Woodruff RS (1952) Confidence intervals for medians and other position measures. J. Amer. Statist. Assoc. 47:635–646

Received: 01.09.97