P-ISSN 2586-9000
E-ISSN 2586-9027
Homepage : https://tci-thaijo.org/index.php/SciTechAsia
Science & Technology Asia
Vol.27 No.2 April - June 2022
Page: [42-57]

Original research article

Regression-in-Ratio Estimator and Confidence Interval for the Population Mean for Data with Outliers

Patarawan Sangnawakij*

Faculty of Science and Technology, Department of Mathematics and Statistics, Thammasat University, Pathum Thani 12120, Thailand

Received 10 April 2021; Received in revised form 12 June 2021
Accepted 21 June 2021; Available online 29 June 2022

ABSTRACT

Parameter estimation is a method in statistical inference widely used in several areas of research. However, in a sampling survey, many types of data are obtained. Suitable statistical tools are then necessary in consideration. In this paper, we focus on the data included outliers in the variable of interest. The ratio estimators used information on an auxiliary variable and two robust regression estimators based on the least quantile of squares and bisquare methods are then interested to estimate the population mean in simple random sampling. Novel variance estimation of these estimators derived based on the Fieller method is proposed to construct the confidence intervals. Moreover, simulations in many situations when outliers are available are studied. The results show that the proposed confidence interval using variance estimator derived from the Fieller method provides the coverage probability greater than the confidence interval using the mean square error derived based on the Taylor series expansion in all cases in the study. A real-world dataset on apple products in Turkey is analyzed to confirm the practical application of our estimators.

Keywords: Fieller method; Interval estimation; Ratio estimator; Robust regression; Simple random sampling

1. Introduction

Information on the nature in defined populations of a variety of characteristics is a necessity for social services, marketing, public health, politicians, and others. For reasons related to timeless cost, and size of population, this information is obtained by use of sample surveys [1]. Sampling surveys are therefore important in the procedure of any research. Simple random sampling (SRS) is a type of sample survey widely applied in applications. A simple random sample obtained from this technique is chosen by following two steps: first, assign a number from one to each element in population of size N, and then choose

***Corresponding author:** patarawan.s@gmail.com
doi: 10.14456/scitechasia.2022.23

a sample of size n, where $n \le N$, from these numbers by use of a random process, for example, a table of random numbers and a calculator with a random number generator. So, when we take a sample using SRS with replacement, each of the N^n possible of size n has the same probability of selection of $1/N^n$. In another type, when we take a sample using SRS without replacement, each of the $C_{N,n}$ possible of size n has the same probability of selection equal to $1/C_{N,n}$, where $C_{N,n}$ denotes the combination of N taken n at a time without repetition [2]. An important process is then after the data are collected. We need to analyze the data using statistical inference to conclude the findings of the sample survey and refer to the character of the target population. Parameter estimation, both point and interval estimation, is often used in this process. In general, parameters of interest in quantity outcomes or related to the continuous variable are the mean, total, standard deviation, ratio or difference of two population means. For the quality outcomes, they are the proportion, difference of proportions, and percentage. We devote this paper to discuss the estimation of the population mean in SRS without replacement (SRS/WOR), which is widely used in actual sample surveys.

Let us give the idea for estimating the population mean. Consider a simple random sample of size n drawn from the population of size N using SRS/WOR. The population mean of the study variable is denoted as μ_Y. The sample mean which is an unbiased estimator of μ_Y is given by $\bar{y} = \sum_{i=1}^{n} y_i / n$. Suppose that we have an auxiliary variable X, where X has high correlation with the variable Y. As noted in literatures, the use of information of auxiliary variable can reduce the error in estimation and provide a more accurate estimate of the true parameter [3-4]. Now, we wish to estimate the population mean of Y using information of X. In this special situation, Cochran [5] proposed the traditional ratio estimator with a smaller variance than the sample mean $\bar{y}$. It is given as $\bar{y}_{tr} = (\bar{y}/\bar{x})\mu_X = r\mu_X$, where $\bar{x}$ and μ_X are the sample mean and population mean of the auxiliary variable, respectively. The ratio of two sample means $r = \bar{y}/\bar{x}$ is the estimator for $R = \mu_Y / \mu_X$. Here, $\bar{y}_{tr}$ is an asymptotically unbiased estimator of μ_Y. The estimated variance of $\bar{y}_{tr}$ is

$$\mathrm{var}(\bar{y}_{tr}) = \frac{f}{n(n-1)}\left(\sum_{i=1}^{n} y_i^2 - 2r\sum_{i=1}^{n} x_i y_i + r^2 \sum_{i=1}^{n} x_i^2\right),$$

where $f = (N-n)/N$.

A further estimator has been developed. When the relationship of the study and auxiliary variables are significantly different from the origin, the regression estimator is used to obtain greater accuracy of μ_Y. It is given as

$$\bar{y}_{reg} = \bar{y} + b_{ls}(\mu_X - \bar{x}),$$

where b_{ls} is the simple linear regression coefficient estimator for parameter β, obtained from the ordinary least square (LS) method. $\bar{y}_{reg}$ is an asymptotic unbiased estimator for μ_Y. The variance of $\bar{y}_{reg}$ is approximated by

$$\mathrm{var}(\bar{y}_{reg}) = \frac{f}{n}\left(\frac{n-1}{n-2}\right)\left(s_y^2 - b_{ls}^2 s_x^2\right),$$

where s_y^2 and s_x^2 are the sample variances [1]. Note that when the variable of interest and auxiliary variable have a strong correlation, $\bar{y}_{reg}$ will have lower variance than $\bar{y}$ [2]. If there has no relationship between the two variables or the value of regression coefficient is significantly close to zero, $\bar{y}_{reg}$ is equivalent to the simple estimate $\bar{y}$.

Modifying the two estimators above, Ray and Singh [6] proposed the ratio (regression-in-ratio) estimator for the mean. It is given by

$$\bar{y}_{RS,ls} = \frac{\bar{y} + b_{ls}(\bar{x}^{\alpha} - \mu_X^{\alpha})}{\bar{x}^{\gamma}} \mu_X^{\gamma},$$

where α and γ are constants. Kadilar and Cingi [7] suggested $\alpha = \gamma = 1$ and introduced the ratio estimator

$$\bar{y}_{ls} = \frac{\bar{y} + b_{ls}(\mu_X - \bar{x})}{\bar{x}} \mu_X. \quad (1.1)$$

which is widely used in applications. The estimated mean square error (MSE) is obtained from the method based on the second-order Taylor series expansion [8]

$$\text{mse}(\bar{y}_{ls}) = \frac{f}{n}\Big(r^2 s_x^2 - 2r s_{xy} + s_y^2 + 2b_{ls} r s_x^2 + b_{ls}^2 s_x^2 - 2b_{ls} s_{xy}\Big), \quad (1.2)$$

where s_{xy} is the sample covariance of the two variables and $f = (N-n)/N$. Papers related to the regression-in-ratio estimator include Al-Jararha and Al-Haj [9], Banerjie and Tiwari [10], Bhushan et al. [11], Koyuncu and Kadilar [12], Lawson [13], Misra et al. [14], Raza et al. [15], Singh [16], Singh and Tailor [17], Singh et al. [18], Subramani [19], Yan and Tian [20], and Zaman [21].

As noted in the beginning of this section, mean estimation is involved in many researches. However, a limitation of this method is that it must be applied for the data without outliers. Outliers are a small number of observed values that are significantly different from the main group of data [22]. When they present, the estimated mean will have low accuracy. This is because outliers increase the variability in the data [23]. In such a case, $\bar{y}$ and $\bar{y}_{ls}$ may face the problem. They can be an unbiased estimator and provide a larger MSE [24]. Unfortunately, removing outliers for some or all values from the data should be avoided, as it may be a cause of large bias in estimation, if in fact, they represent the nature of the data. To address this problem, Kadilar et al. [25] constructed the ratio estimator using a robust regression. The estimator and its estimated MSE formulas are given by

$$\bar{y}_M = \frac{\bar{y} + b_M(\mu_X - \bar{x})}{\bar{x}} \mu_X \quad (1.3)$$

and

$$\text{mse}(\bar{y}_M) = \frac{f}{n}\Big(r^2 s_x^2 - 2r s_{xy} + s_y^2 + 2b_M r s_x^2 + b_M^2 s_x^2 - 2b_M s_{xy}\Big), \quad (1.4)$$

respectively, where b_M is the regression coefficient for β obtained from the Huber-M or M-estimator [26]. Similarly, ratio estimators have been introduced using further robust regression methods which are less sensitive to outliers, for example, least absolute deviations, least median of squares, and least trimmed squares [21]. A $(1-\alpha)100\%$ confidence interval based on the large-sample method for μ_Y can be constructed using $\bar{y}_{pr} \pm z_{\alpha/2}\sqrt{\text{mse}(\bar{y}_{pr})}$, where $\bar{y}_{pr}$ is a general estimator and $z_{\alpha/2}$ is the $(1-\alpha/2)100$th percentile of the standard normal distribution.

In this paper, two robust regression estimators based on the least quantile of squares and biweight, or bisquare approaches [27-28] are applied. A highlight of this paper is that the novel confidence interval is proposed. The Fieller method [29] is used to derive the variance of the ratio estimator. These are thoroughly explained in Section 2. The performance of the proposed method is investigated using simulations. Coverage probabilities of the confidence intervals for the mean based on the new variance and existing MSE are compared. The results are presented in Section 3. In Section 4, a real-world dataset on apple production in Turkey is used to illustrate the methods. Finally, Section 5 reports our conclusions.

2. Statistical Methodology

In this section, we provide literatures of ratio estimators based on the least quantile of squares (LQS) and bisquare methods. The novel variance of the ratio estimator is derived using the Fieller approach. Then, we construct the confidence interval to improve the efficiency in estimating μ_Y.

2.1 Ratio estimator

As noted in Kadilar et al. [25], the ordinary least square method in a simple linear regression is sensitive to outliers which decrease the efficiency in estimating μ_Y. So, they used the regression coefficient based on the Huber-M estimator. However, Zaman and Bulut [24] showed that the ratio-type estimator using a robust regression, namely the bisquare method, less sensitive to outliers was more efficient. Additionally, the LQS is a robust statistic for outliers widely used in regression analysis. We therefore consider these two approaches in this paper.

Firstly, the ratio estimator using the LQS is applied. It is of the form

$$\bar{y}_{lqs} = \frac{\bar{y} + b_{lqs}(\mu_X - \bar{x})}{\bar{x}}\mu_X, \qquad (2.1)$$

where b_{lqs} is the regression coefficient for β from the LQS method. Note that LQS is obtained from minimizing the residual related to the $q-\text{th}$ ordered absolute residual. This can then be written as $b_{lqs} \in \min_\beta |e_{(q)}|$ [30]. Following the method in Kadilar et al. [25], the estimated MSE of $\bar{y}_{lqs}$ based on Taylor series expansion is given as

$$\text{mse}(\bar{y}_{lqs}) = \frac{f}{n}\Big(r^2 s_x^2 - 2rs_{xy} + s_y^2 + 2b_{lqs} r s_x^2 + b_{lqs}^2 s_x^2 - 2b_{lqs} s_{xy}\Big). \qquad (2.2)$$

The LQS estimator has no closed-form solution, but it can be calculated using statistical packages, for example, the R programming language (https://www.r-project.org/) via the MASS package with lqs function.

Secondly, the bisquare regression is considered. The idea of this robust statistic is to find the coefficient for β that minimizes a loss function of the residuals, denoted as $b_{Bisq} \in \min_\beta \sum_{i=1}^{n} \rho(e_i)_{Bisq}$, where the function is given by

$$\rho(e_i)_{Bisq} = \frac{a^2}{2}\left(1 - \left[1 - \left(\frac{e_i}{k}\right)^2\right]^3\right);$$

$|e_i| \le a$ and $\rho(e_i)_{Bisq} = a^2/6$; otherwise, for $i = 1, 2, ..., n$. The bisquare's weight function is

$$w(e_i)_{Bisq} = \left(1 - \left(\frac{e_i}{a}\right)^2\right)^2;$$

$|e_i| \le a$ and $w(e_i)_{Bisq} = 0$; otherwise, where a is a turning constant and k is the number of independent variable or auxiliary variable. The principal idea of the bisquare and M-estimation methods are similar, but the latter uses the function

$$\rho_M(e_i) = \frac{e_i^2}{2};$$

$|e_i| \le a$ and $\rho_M(e_i) = a|e_i| - a^2/2$; otherwise. For more details see John and Weisherg [31]. The MASS package with rlm and lqs functions in the R language is helpful in computation. Therefore, the ratio estimator noted in Zaman and Bulut [24] is given by

$$\bar{y}_{Bisq} = \frac{\bar{y} + b_{Bisq}(\mu_X - \bar{x})}{\bar{x}}\mu_X \qquad (2.3)$$

with the estimated MSE

$$mse(\bar{y}_{Bisq}) = \frac{f}{n}\Big(r^2 s_x^2 - 2rs_{xy} + s_y^2 + 2b_{Bisq} r s_x^2 + b_{Bisq}^2 s_x^2 - 2b_{Bisq} s_{xy}\Big). \qquad (2.4)$$

Note that the MSEs presented in (2.2) and (2.4) are derived from the Taylor series expansion, and can be used to construct the large sample-based confidence interval for μ_Y. However, we suggest a method to find the variance of ratio estimator in the next section.

2.2 New variance of ratio estimator and confidence interval

In this section, we highlight the Fieller method to obtain the variance of the ratio estimator. The process of this method is a way of expressing ratio as a linear combination of two random variables which made computation of the confidence interval of the ratio relatively simple. It is often used in the finite population, where the data are typically not assumed to come from a specific distribution [32]. The details for constructing the variance of estimator proposed in this paper are given in the following.

Theorem 1. Let (y_i, x_i) be two random samples of size n drawn from the population (Y_i, X_i) of size N using SRS/WOR. Suppose that μ_X is a known population mean of the auxiliary variable. Then, the estimated variance of

$$\bar{y}_{pr} = \frac{\bar{y} + b(\mu_X - \bar{x})}{\bar{x}}\mu_X$$

using the Fieller method is

$$var(\bar{y}_{pr}) = \frac{\mu_X^2 \hat{\sigma}_V^2 + \bar{y}_{pr}^2 \hat{\sigma}_W^2}{\bar{x}^2}, \qquad (2.5)$$

where $\hat{\sigma}_V^2 = \left(\frac{N-n}{N}\right)\frac{(n-1)}{n(n-2)}\left(s_y^2 - b^2 s_x^2\right)$

and $\hat{\sigma}_W^2 = \left(\frac{N-n}{N}\right)\frac{s_x^2}{n}$.

Proof. Let (y_i, x_i), for $i = 1, 2, ..., n$ be two simple random samples drawn from the population (Y_i, X_i) of size N. $\bar{y}_{pr}$ is a regression-in-ratio estimator for the population mean μ_Y and b is a generic regression coefficient estimator. Now, we assume that

$$E(\bar{y}_{pr}) = E\left(\frac{V\mu_X}{W}\right) = \mu_Y,$$

where $V = \bar{y} + b(\mu_X - \bar{x})$, $W = \bar{x}$, and $E(V\mu_X - W\mu_Y) = 0$. The ratio $V\mu_X / W$ is called a *linear combination* of V and W. According to the statistical theory, if the two random variables are normal distribution, the distribution of a linear combination of two normal variables is also a normal [33]. Applied to our case, since V and W are assumed to be normal distributions for large n, $V\mu_X - W\mu_Y$ is also a normal distribution. Therefore, we have the function $V\mu_X - W\mu_Y \sim N(0, \sigma_{pr}^2)$, where the variance $\sigma_{pr}^2 = \mu_X^2\sigma_V^2 + \mu_Y^2\sigma_W^2$, and the pivotal statistic is given by

$$\frac{V\mu_X - W\mu_Y}{\sqrt{\sigma_{pr}^2}} \sim N(0,1). \qquad (2.6)$$

It can be seen that V is the regression estimator and W is the sample mean in SRS/WOR. The variances of V and W, denoted as σ_V^2 and σ_W^2, are estimated by $\hat{\sigma}_V^2$ and $\hat{\sigma}_W^2$, respectively. Thus, σ_{pr}^2 is approximated as $\hat{\sigma}_{pr}^2 = \mu_X^2\hat{\sigma}_V^2 + \bar{y}_{pr}^2\hat{\sigma}_W^2$. Substituting $\hat{\sigma}_{pr}^2$ into (2.6), we have the pivot

$$Z = \frac{V\mu_X - W\mu_Y}{\sqrt{\mu_X^2\hat{\sigma}_V^2 + \mu_Y^2\hat{\sigma}_W^2}} \sim N(0,1)$$

as $n \to \infty$. The distribution of Z does not depend on the unknown parameter. Based on the normal approximation, the $(1-\alpha)100\%$ confidence interval for μ_Y is therefore derived by

$$\begin{aligned} 1-\alpha &= P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) \\ &= P(\bar{y}_{pr} - z_{\alpha/2} \times SE \le \mu_Y \\ &\qquad \le \bar{y}_{pr} + z_{\alpha/2} \times SE) \end{aligned} \quad (2.7)$$

where $SE = \sqrt{\mu_X^2\hat{\sigma}_V^2 + \bar{y}_{pr}^2\hat{\sigma}_W^2} / W$, $z_{\alpha/2}$ is the $(1-\alpha/2)100$th percentile of the standard normal distribution, and α is the significant level. The advantage of the Fieller method is that we obtain the estimated variance of the ratio estimator $\bar{y}_{pr}$ from the probability statement (2.7). It is given by

$$\text{var}(\bar{y}_{pr}) = \frac{\mu_X^2\hat{\sigma}_V^2 + \bar{y}_{pr}^2\hat{\sigma}_W^2}{\bar{x}^2},$$

establishing (2.5). □

We note that the proposed variance estimator corresponds to a given ratio estimator. In this paper, we will study the performance of the confidence intervals for μ_Y using $\text{var}(\bar{y}_{pr})$ with the four ratio estimators, namely $\bar{y}_{ls}$, $\bar{y}_M$, $\bar{y}_{Bisq}$, and $\bar{y}_{lqs}$. Thus, the confidence interval for μ_Y is given in the form $\bar{y}_{pr} \pm z_{\alpha/2}\sqrt{\text{var}(\bar{y}_{pr})}$.

3. Simulation Study

In this section, the regression-in-ratio estimators, $\bar{y}_{ls}$, $\bar{y}_M$, $\bar{y}_{lqs}$, and $\bar{y}_{Bisq}$, were evaluated through simulations in various situations, using the R statistical package [34]. The performance of these point estimators was investigated in terms of bias, MSE, and variance. The coverage probability was used to conduct the confidence intervals using the MSE derived from Taylor series expansion and variance obtained from the Fieller method. Simulation settings and results were given as follows.

3.1 Simulation setting

We considered a population of size $N = 100{,}000$. The data for the auxiliary variable X_i were generated from the normal distribution with $\mu_X = 20$ and $\sigma_X^2 = 10$. The study variable Y_i was obtained from the simple linear relationship: $Y_i = \beta_0 + \beta X_i + \varepsilon_i$, where $\beta_0 = 5$, $\beta = 1$, and ε_i was generated from a standard normal distribution. This provided that X_i and Y_i had a strong correlation with the true parameter mean $\mu_Y = 25$. From (Y_i, X_i), the sample (y_i, x_i) of size n was selected using SRS without replacement, where $n = 40$, 60, 120, and 200. Following the boxplot criteria for detecting outliers, we considered three degrees of outliers. The outliers can be constructed by $y_i^* = y_{[n-c+1:n]} + d \times IQR(y_{[1:n]})$, where the degrees or level of outliers were $d = 1.5$, 2, and 3, referring to mild to extreme outliers. The number of outliers was given as $c = 1$, 3, 4, and 5, and the inter quantile range of y or $IQR(y)$ was estimated by the difference between third and first quartiles.

We performed $M = 10{,}000$ simulation runs on the generated dataset. On average, the bias, MSE, and variance of estimator were computed by

$$\text{ABias}(\bar{y}_{pr}) = \frac{1}{M}\sum_{m=1}^{M}(\bar{y}_{pr})_m - \mu_Y,$$

$$\text{AMSE}(\bar{y}_{pr}) = \frac{1}{M}\sum_{m=1}^{M}\text{mse}(\bar{y}_{pr})_m,$$

and

$$\text{AVar}(\bar{y}_{pr}) = \frac{1}{M}\sum_{m=1}^{M} \text{var}(\bar{y}_{pr})_m,$$

respectively. In our case, $\bar{y}_{pr}$ will be denoted as $\bar{y}_{ls}$, $\bar{y}_{M}$, $\bar{y}_{lqs}$, or $\bar{y}_{Bisq}$. The ratio estimator that has a small error, while its bias is closer to zero, is a good performance.

For interval estimation, the coverage probability was used to compare the performance of the confidence interval obtained from the two approaches. It was estimated by

$$\text{ACP}(\bar{y}_{pr}) = \frac{c(L \le \mu_Y \le U)}{M},$$

where $c(L \le \mu_Y \le U)$ is the number of simulation runs for the parameter of interest μ_Y that lies within the lower limit L and upper limit U of the confidence interval. A preferred 95% confidence interval of μ_Y would have the coverage probability greater than or equal to the nominal coverage criteria of 0.9464 with a short interval length. This suggests the confidence interval covers the true parameter and outperforms the compared estimator. Simulation results are given in the next section.

3.2 Simulation results

In this simulation, the correlation between X and Y was estimated by 0.9312, which showed the strong relationship between the two variables. The ratio estimator is then suitable to use in this case. According to Table 1 given in the appendix, $\bar{y}_{lqs}$ and $\bar{y}_{Bisq}$ had bias closer to zero than $\bar{y}_{ls}$ and $\bar{y}_{M}$ in general cases, especially when sample size $n < 120$. The values of bias were decreased when n increased, but they were increased when the degree of outliers (d) or number of outliers (c) was large. From Table 2, it can be seen that $\bar{y}_{Bisq}$ provided smaller MSEs than the compared estimators in all cases in the study. Especially, it was much smaller than the MSE of $\bar{y}_{ls}$. When ordered by the best performance in terms of MSE, the estimators ranked as $\bar{y}_{Bisq}$, $\bar{y}_{lqs}$, $\bar{y}_{M}$, and $\bar{y}_{ls}$. Moreover, the MSEs decreased as n increased, but they increased if d or c was large. From these results, we conclude that the point estimators $\bar{y}_{lqs}$ and $\bar{y}_{Bisq}$ have a great efficiency to estimate the population mean for data with outliers. In particular, the latter estimator provides the smallest MSE referred to have more accuracy in estimation.

From Table 3, the variances of the ratio estimators obtained from the Fieller method proposed in this paper are presented. It was found that $\bar{y}_{ls}$ provided the variance smaller than $\bar{y}_{M}$, $\bar{y}_{lqs}$, and $\bar{y}_{Bisq}$. In detail, the variances of these estimators were decreased when n increased, but they were increased for large d or c. In fact, the MSE and variance can be used to build the confidence interval using the large-sample method. However, it can be seen in this study that the MSE and variance of the estimators were different. Thus, one may lead to the low efficiency in interval estimation. The question arises now as to which approach is superior to use in interval estimation in SRS/WOR when outliers are available. Since there has been no paper that compared the performance of the confidence intervals using the MSEs in (1.2), (1.4), (2.2), and (2.4) it is therefore addressed in this section.

The performance of the 95% confidence intervals using MSE and variance given in Section 2 was investigated and given in Table 4. The results showed that the coverage probabilities of the confidence intervals using the MSEs from Taylor series expansion were much lower than the nominal coverage level at 0.9464 in all cases in the study. Particularly, when d or c was large, those coverage probabilities were very poor. They were not entirely suggested to use for estimating the mean when outliers occurred, not only extreme ($d \geq 2$) but also mild level of outliers ($d = 1.5$). In contrast, the novel confidence intervals using variance estimation based on the Fieller method performed well in terms of coverage probability in many cases. The confidence interval based on the bisquare method had the coverage probability greater than the compared estimators. It satisfied the nominal coverage level in almost all situations. Except for $c = 5$ and $d \geq 2$, the coverage probability was slightly lower than 0.9464. The confidence interval based on the LQS was greater than the target probability when $c \leq 4$ and $d \leq 2$. The confidence interval based on the M-estimator was satisfied only when $c \leq 3$. However, the confidence interval based on the least square method was unsatisfied, as its coverage probability was much lower than 0.9464. Table 5 shows that the expected lengths of the confidence intervals using MSE were slightly smaller than those of the confidence intervals using the proposed variance. However, it is very important to note that the confidence interval using MSE and the confidence interval related to the LS regression with the Fieller's variance cannot cover the parameter mean that we need to estimate (see Table 4).

In summary, the simulation results suggest that $\bar{y}_{Bisq}$ provides a good performance in terms of bias and mean square error in estimation. It is suitable to be the point estimator to estimate the population mean when outliers are available. This is similar to the results given in Zaman and Bulut [24], but they did not consider many situations in simulations. So, our current study is more intensive. The variance of the ratio estimators using the Fieller method proposed in this paper is generally superior to the mean square error based on Taylor series expansion method for constructing the confidence interval. This is because the coverage probability of the confidence interval based on the novel variance estimation provides a desirable nominal coverage probability and short interval length. We therefore conclude that our interval estimation is appropriate for estimating the population mean for data with outliers. In such cases, it is clear that the regression-in-ratio estimator using the coefficient estimate from the LS method and the confidence interval based on the mean square error derived by Taylor series expansion must be avoided.

4. Application to Real Data

We illustrated the use of point and confidence intervals for the population mean using the real-world example. The data were consisted of the apple production amount (study variable, Y in tons) and number of apple trees (auxiliary variable, X), where 1 unit = 100 trees in 106 villages of a region in Turkey. The data were also mentioned in Kadilar et al. [25] that they came from Institute of Statistics, Republic of Turkey. We point out here that the dataset used in this section was different from the previous work, as it came from different villages. Here, the mean of the apple trees from all villages was found to be $\mu_X = 27{,}422$ trees. The objective is that we need to estimate the weight of the apple production. Using SRS/WOR, 30 villages were selected. Apple trees and product of apples in the sample villages were counted and weighted. It can be seen that there was a highly positive correlation between these two variables with a sample correlation of 0.97 (see Fig. 1 (a)). Furthermore, there were four outliers from 30 observations (13.33%) in the dataset (see also Fig. 1 (b)). The sample means were given by $\bar{x} = 27{,}680.67$ and $\bar{y} = 15{,}94.87$ and sample standard deviations were $s_x = 47{,}739.13$ and $s_y = 32{,}31.21$.

(a)

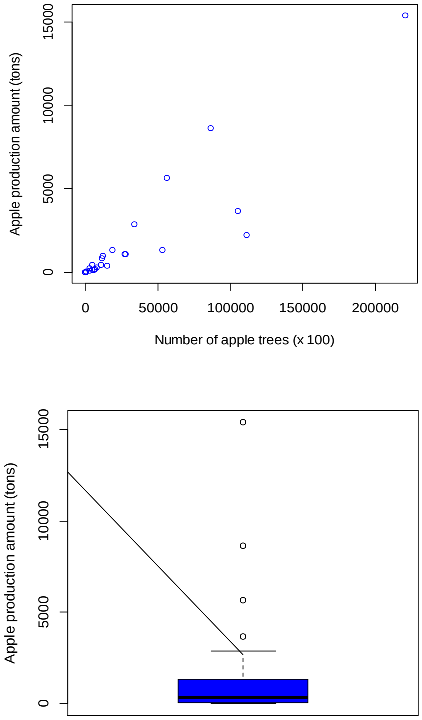

Fig. 1. Scatter plot of the two variables of interest (a) and boxplot of the data on apple production amount (b) from villages in Tukey

Using the ratio estimators, we had $\bar{y}_{ls}$ =1,564.23, $\bar{y}_{M}$ =1,562.18, $\bar{y}_{lqs}$ =1,570, and $\bar{y}_{Bisq}$ =1,569.99, with the standard errors obtained from the Fieller method 467.79, 400.54, 587.49, and 587.33, respectively. The confidence intervals were CI_{ls} = (647.38, 2,481.08), CI_{M} = (777.13, 2,347.23), CI_{lqs} = (418.53, 2,721.46), and CI_{Bisq} = (418.85, 2,721.13) with interval lengths 1,833.70, 1,570.10, 2,302.93, and 2,302.29, respectively. From the results, the estimated mean using ratio estimators had little difference from each other. This was similar to the simulation results, where they showed small and similar bias in estimation. The estimated variances related to the LQS and bisquare regression were similar and greater than those of LS and M estimators as presented in the simulation study. However, simulations have shown that the confidence intervals using LS and M estimators hardly covered the population mean for data with outliers. We then concluded by following $\bar{y}_{Bisq}$ that the mean of apple production amount was 1,569.99 tons with the 95% confidence interval of 418.85 to 2,721.13 tons.

5. Conclusions

In this paper, the main objective was to introduce the new confidence interval for the population mean in simple random sampling without replacement. The Fieller method widely used in interval estimation is applied. From this process, we obtain the variance of the ratio estimator. In our simulations, we show that the regression-in-ratio estimator using the bisquare estimator is a good point estimator for the population mean. The performance of the confidence intervals using the MSE from Taylor series expansion and proposed variance are compared. Unfortunately, the results show that the confidence interval based on the MSE has low performance in terms of coverage probability in all cases, as its coverage probability is lower than the target level. We observe that the MSE is too small so that the lower and upper limits from this method is not sufficient to cover the true mean, leading to bias in interval estimation. Meanwhile, the coverage probability of the proposed confidence interval using the Fieller variance with the robust bisquare regression hits the target probability in general cases. Except when the data have extreme outliers, its coverage probability is slightly lower than the nominal level. Based on our findings, we therefore recommend to use the ratio estimator $\bar{y}_{Bisq}$ with the Fieller confidence interval for estimating the population mean when outliers occur in simple random sampling.

Acknowledgements

The author would like to thank the editor and the reviewers for their valuable comments and suggestions. We are also very grateful to Professor Dr. Cem Kadilar for supporting the real data used in this paper. This study was supported by Thammasat University Research Fund, Contact No. TUFT 12/2564.

References

[1] Thompson SK. Sampling. New Jersey: John Wiley and Sons; 2012.

[2] Levy PS, Lemeshow S. Sampling of populations: methods and applications. New Jersey: John Wiley and Sons; 2008.

[3] Jitthavech J, Lorchirachoonkul V. Estimators in simple random sampling: Searls approach. Songklanakarin J Sci Technol 2013;35(6):749-60.

[4] Jiang Z, Walker K, Shi D, Cao J. Improving generalizability coefficient estimate accuracy: a way to incorporate auxiliary information. Metho Innov 2018;11(2):1-14.

[5] Cochran WG. Sampling techniques. New York: John Wiley and Sons; 1977.

[6] Ray SK, Singh KR. Difference-cum-ratio type estimator. J Indian Stat Assoc 1981;19:147-51.

[7] Kadilar C, Cingi H. Ratio estimators in simple random sampling. Appl Math Compu 2004;151(3):893-902.

[8] Oehlert GW. A note on the delta method. Am Stat 1992;46(1):27-9.

[9] Al-Jlararha J, Al-Haj EM. A ratio estimator under general sampling design. Austrian J Stat 2012;41(2):105-15.

[10] Banerjie J, Tiwari N. Improved ratio type estimator using jackknife method of estimation. J Reli Stat Stu 2011;4(1):53-63.

[11] Bhushan S, Pandey A, Katara S. A class of estimators in double sampling using two auxiliary variables. J Reli Stat Stu 2008;1(1):67-73.

[12] Koyuncu N, Kadilar C. Efficient estimators for the population mean. Hacet J Math Stat 2009;38(2):217-25.

[13] Lawson N. New ratio estimators for population mean in simple random sampling using robust regression. J Appl Sci 2020;19(1):51-8.

[14] Misra GC, Shukla AK, Yadav SK. A comparison of regression methods for improved estimation in sampling. J Reli Stat Stu 2009;2(2):85-90.

[15] Raza A, Noor-ul-Amin M, Hanif M. Regression-in-ratio estimators in the presence of outliers based on redescendingm-estimator. J Reli Stat Stu 2019;12(2):1-10.

[16] Singh GN. On the improvement of product method of estimation in sample surveys. J Indian Soc Agric Stat 2003;56(3):267-5.

[17] Singh HP, Tailo R. Use of known correlation coefficient in estimating the finite population means. Stat Tran 2005;6(4):555-60.

[18] Singh R, Malik S, Chaudhary MK, Verma HK, Adewara AA. A general family of ratio-type estimators in systematic sampling. J Reli Stat Stu 2012;5(1):73-82.

[19] Subramani J. A modified approach in linear regression estimator in simple random sampling. J Adv Res Appl Math Stat 2016;1(2):1-8.

[20] Yan Z, Tian B. Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In: Zhu R, Zhang Y, Liu B, Liu C, editors. Communications in Computer and Information Science. Berlin, Heidelberg: Springer; 2010. p. 103-10.

[21] Zaman T. (2019). Improvement of modified ratio estimators using robust regression methods. Appl Math Compu 2019;348:627-31.

[22] Barnett V, Lewis T. Outliers in statistical data. Biom J 1995;37(2):256.

[23] Wada K. Outliers in official statistics. Jpn J Stat Data Sci 2020;3:669-91.

[24] Zaman T, Bulut H. Modified ratio estimators using robust regression methods. Comm Stat Theor Meth 2019;48(8):2039-48.

[25] Kadilar C, Candan M, Cingi H. Ratio estimators using robust regression. Hacet J Math Stat 2007;36(2):181-8.

[26] Huber PJ. Robust estimation of a location parameter. Ann Math Stat 1964;35(1):73-101.

[27] Beaton AE, Tukey JW. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 1974;16(2):147-85.

[28] Rousseeuw PJ. Least median of squares regression. J Am Stat Assoc 1984;79(388):871-80.

[29] Fieller EC. The biological standardization of Insulin. J R Stat Soc Series B 1940;7(1):1-64.

[30] Bertsimas D, Mazumder R. Least quantile of squares regression via modern optimization. Ann Stat 2014;42(6):2494-525.

[31] John F, Weisberg S. An R companion to applied regression. California: SAGE Publication; 2011.

[32] Sherman M, Maity A, Wang S. Inferences for the ratio: Fieller's interval, log ratio, and large sample-based confidence intervals. Adv Stat Anal 2011;95(3):313-23.

[33] Beyene J, Moineddin R. Methods for confidence interval estimation of a ratio parameter with application to location quotients. BMC Med Res Methodol 2005;32(5):1-7.

[34] R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020.

Appendix

Simulation results noted in Section 3 are given in Tables 1-5.

Table 1. Bias of the four regression-in-ratio estimators from simulations.

c	d	n	$\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$
1	1.5	40	0.1715	0.1714	0.1707	0.1703
		60	0.1462	0.1443	0.1436	0.1437
		120	0.0767	0.0761	0.0760	0.0759
		200	0.0328	0.0328	0.0328	0.0327
	2	40	0.2273	0.2262	0.2263	0.2258
		60	0.1625	0.1614	0.1607	0.1611
		120	0.0775	0.0775	0.0772	0.0776
		200	0.0639	0.0633	0.0631	0.0632
	3	40	0.3596	0.3549	0.3543	0.3541
		60	0.2366	0.2349	0.2355	0.2346
		120	0.1115	0.1114	0.1113	0.1114
		200	0.0729	0.0724	0.0723	0.0723
3	1.5	40	0.5147	0.5120	0.5106	0.5107
		60	0.3395	0.3382	0.3373	0.3378
		120	0.1658	0.1653	0.1651	0.1651
		200	0.1150	0.1145	0.1144	0.1143
	2	40	0.6518	0.6501	0.6488	0.6492
		60	0.4420	0.4414	0.4410	0.4411
		120	0.2515	0.2489	0.2483	0.2482
		200	0.1361	0.1359	0.1360	0.1359
	3	40	1.0125	1.0033	1.0007	1.0008
		60	0.6444	0.6488	0.6419	0.6419
		120	0.3578	0.3540	0.3533	0.3534
		200	0.2026	0.2022	0.2021	0.2021

c	d	n	$\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$
4	1.5	40	0.6420	0.6430	0.6405	0.6404
		60	0.4444	0.4436	0.4435	0.4430
		120	0.2063	0.2074	0.2028	0.2018
		200	0.1318	0.1317	0.1319	0.1317
	2	40	0.9002	0.8932	0.8886	0.8893
		60	0.6206	0.6147	0.6134	0.6128
		120	0.2893	0.2897	0.2809	0.2810
		200	0.1773	0.1772	0.1771	0.1772
	3	40	1.2924	1.2940	1.2905	1.2906
		60	0.8715	0.8654	0.8648	0.8649
		120	0.4819	0.4744	0.4730	0.4732
		200	0.2626	0.2633	0.2534	0.2534
5	1.5	40	0.8352	0.8330	0.8295	0.8305
		60	0.5369	0.5379	0.5139	0.5138
		120	0.2632	0.2645	0.2625	0.2625
		200	0.1721	0.1719	0.1719	0.1718
	2	40	1.0852	1.0844	1.0841	1.0850
		60	0.7413	0.7384	0.7367	0.7373
		120	0.3855	0.3835	0.3828	0.3828
		200	0.2256	0.2253	0.2254	0.2252
	3	40	1.6279	1.6249	1.6240	1.6232
		60	1.0797	1.0832	1.0644	1.0642
		120	0.5429	0.5435	0.5235	0.5244
		200	0.3064	0.3101	0.3011	0.3011

Table 2. Mean square error of the four regression-in-ratio estimators based on Taylor series expansion from simulations.

c	d	n	MSE of estimator $\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$	c	d	n	MSE of estimator $\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$
1	1.5	40	0.4481	0.3961	0.3821	0.3782	4	1.5	40	0.5231	0.4031	0.3237	0.3187
		60	0.2901	0.2644	0.2580	0.2560			60	0.3324	0.2612	0.2292	0.2276
		120	0.1426	0.1351	0.1333	0.1329			120	0.1533	0.1302	0.1229	0.1218
		200	0.0845	0.0816	0.0809	0.0807			200	0.0885	0.0790	0.0758	0.0758
	2	40	0.4747	0.4005	0.3870	0.3826		2	40	0.5896	0.4115	0.3384	0.3372
		60	0.3032	0.2671	0.2608	0.2589			60	0.3633	0.2625	0.2340	0.2310
		120	0.1462	0.1357	0.1336	0.1335			120	0.1632	0.1307	0.1232	0.1225
		200	0.0856	0.0815	0.0809	0.0806			200	0.0922	0.0789	0.0761	0.0758
	3	40	0.5290	0.4158	0.4030	0.3991		3	40	0.7791	0.4926	0.4381	0.4350
		60	0.3322	0.2760	0.2693	0.2681			60	0.4529	0.2983	0.2741	0.2721
		120	0.1532	0.1368	0.1345	0.1321			120	0.1888	0.1391	0.1326	0.1318
		200	0.0879	0.0816	0.0807	0.0807			200	0.1022	0.0819	0.0794	0.0790
3	1.5	40	0.5011	0.3918	0.3396	0.3380	5	1.5	40	0.5439	0.4299	0.3106	0.3039
		60	0.3198	0.2601	0.2374	0.2354			60	0.3369	0.2599	0.2172	0.2164
		120	0.1495	0.1309	0.1257	0.1246			120	0.1565	0.1296	0.1200	0.1191
		200	0.0869	0.0793	0.0772	0.0769			200	0.0896	0.0783	0.0750	0.0745
	2	40	0.5520	0.3948	0.3454	0.3421		2	40	0.6229	0.4414	0.3318	0.3306
		60	0.3454	0.2612	0.2402	0.2378			60	0.3792	0.2657	0.2274	0.2254
		120	0.1579	0.1318	0.1266	0.1256			120	0.1673	0.1294	0.1196	0.1194
		200	0.0901	0.0796	0.0777	0.0772			200	0.0943	0.0784	0.0750	0.0747
	3	40	0.7056	0.4603	0.4211	0.4167		3	40	0.8238	0.5271	0.4471	0.4429
		60	0.4197	0.2895	0.2710	0.2689			60	0.4882	0.3118	0.2815	0.2797
		120	0.1784	0.1379	0.1321	0.1302			120	0.1980	0.1402	0.1322	0.1314
		200	0.0977	0.0815	0.0795	0.0792			200	0.1067	0.0823	0.0791	0.0788

Table 3. Variance of the four regression-in-ratio estimators based on the Fieller method from simulations.

			Variance of estimator							Variance of estimator			
c	d	n	$\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$	c	d	n	$\bar{y}_{ls}$	$\bar{y}_{M}$	$\bar{y}_{lqs}$	$\bar{y}_{Bisq}$
1	1.5	40	0.4537	0.5010	0.5137	0.5169	4	1.5	40	0.5311	0.6607	0.7500	0.7432
		60	0.2966	0.3192	0.3236	0.3264			60	0.3325	0.4037	0.4356	0.4369
		120	0.1437	0.1501	0.1515	0.1520			120	0.1515	0.1723	0.1792	0.1798
		200	0.0842	0.0866	0.0870	0.0874			200	0.0892	0.0976	0.0997	0.1003
	2	40	0.4798	0.5485	0.5612	0.5651		2	40	0.5934	0.8030	0.8908	0.8921
		60	0.3011	0.3329	0.3391	0.3403			60	0.3656	0.4739	0.5059	0.5076
		120	0.1457	0.1546	0.1556	0.1565			120	0.1643	0.1953	0.2022	0.2030
		200	0.0853	0.0886	0.0890	0.0894			200	0.0917	0.1034	0.1058	0.1061
	3	40	0.5393	0.6530	0.6652	0.6695		3	40	0.7895	1.1883	1.2777	1.2818
		60	0.3339	0.3862	0.3927	0.3936			60	0.4568	0.6468	0.6783	0.6807
		120	0.1532	0.1674	0.1691	0.1693			120	0.1867	0.2371	0.2443	0.2446
		200	0.0889	0.0943	0.0949	0.0950			200	0.1021	0.1213	0.1236	0.1241
3	1.5	40	0.5079	0.6223	0.6770	0.6786	5	1.5	40	0.5508	0.6795	0.8158	0.7822
		60	0.3212	0.3790	0.4009	0.4025			60	0.3396	0.4207	0.4645	0.4647
		120	0.1503	0.1669	0.1718	0.1725			120	0.1553	0.1804	0.1897	0.1901
		200	0.0878	0.0943	0.0961	0.0963			200	0.0901	0.1002	0.1032	0.1036
	2	40	0.5671	0.7440	0.7987	0.8024		2	40	0.6253	0.8474	0.9912	0.9794
		60	0.3469	0.4321	0.4540	0.4561			60	0.3864	0.5153	0.5606	0.5619
		120	0.1567	0.1805	0.1856	0.1861			120	0.1669	0.2037	0.2131	0.2134
		200	0.0911	0.1003	0.1019	0.1024			200	0.0938	0.1082	0.1114	0.1115
	3	40	0.7145	1.0298	1.0872	1.0905		3	40	0.8472	1.3010	1.4480	1.4463
		60	0.4265	0.5738	0.5949	0.5976			60	0.4877	0.7169	0.7614	0.7630
		120	0.1791	0.2184	0.2232	0.2240			120	0.2003	0.2632	0.2721	0.2729
		200	0.0989	0.1139	0.1155	0.1160			200	0.1073	0.1312	0.1343	0.1346

Table 4. Coverage probability of the 95% confidence interval for the population mean.

			Coverage probability							**Coverage probability**			
c	d	n	CI_{ls}	CI_M	CI_{lqs}	CI_{Bisq}	c	d	n	CI_{ls}	CI_M	CI_{lqs}	CI_{Bisq}
Using mean square error based on Taylor series expansion													
1	1.5	40	0.9235	0.9220	0.9216	0.9213	4	1.5	40	0.8172	0.7907	0.7650	0.7667
		60	0.9297	0.9289	0.9281	0.9287			60	0.8483	0.8286	0.8169	0.8170
		120	0.9393	0.9391	0.9389	0.9389			120	0.8957	0.8884	0.8846	0.8851
		200	0.9427	0.9424	0.9424	0.9426			200	0.9141	0.9104	0.9094	0.9098
	2	40	0.9206	0.9203	0.9199	0.9200		2	40	0.7501	0.6909	0.6475	0.6508
		60	0.9279	0.9268	0.9269	0.9271			60	0.7911	0.7468	0.7287	0.7268
		120	0.9368	0.9369	0.9367	0.9367			120	0.8697	0.8570	0.8518	0.8516
		200	0.9373	0.9378	0.9378	0.9379			200	0.8959	0.8895	0.8886	0.8881
	3	40	0.9053	0.9032	0.9027	0.9026		3	40	0.6769	0.5527	0.5068	0.5068
		60	0.9189	0.9167	0.9179	0.9165			60	0.7309	0.6487	0.6265	0.6257
		120	0.9339	0.9330	0.9331	0.9329			120	0.7863	0.7461	0.7391	0.7387
		200	0.9365	0.9360	0.9364	0.9360			200	0.8589	0.8391	0.8365	0.8362
3	1.5	40	0.8481	0.8319	0.8170	0.8196	5	1.5	40	0.7613	0.7242	0.6594	0.6748
		60	0.8868	0.8761	0.8712	0.8712			60	0.8163	0.7861	0.7588	0.7602
		120	0.9121	0.9082	0.9063	0.9067			120	0.8759	0.8644	0.8599	0.8594
		200	0.9253	0.9230	0.9223	0.9226			200	0.8975	0.8915	0.8898	0.8904
	2	40	0.8251	0.7922	0.7741	0.7746		2	40	0.6907	0.6113	0.5334	0.5423
		60	0.8589	0.8367	0.8307	0.8315			60	0.7457	0.6813	0.6431	0.6443
		120	0.8853	0.8792	0.8766	0.8774			120	0.8257	0.7965	0.7853	0.7864
		200	0.9170	0.9148	0.9130	0.9133			200	0.8682	0.8546	0.8518	0.8508
	3	40	0.7500	0.6731	0.6533	0.6515		3	40	0.5619	0.4063	0.3194	0.3246
		60	0.8141	0.7653	0.7542	0.7539			60	0.6445	0.5129	0.4719	0.4717
		120	0.8503	0.8321	0.8289	0.8291			120	0.7567	0.6876	0.6711	0.6709
		200	0.8896	0.8821	0.8799	0.8803			200	0.8324	0.8062	0.7992	0.8008
Using variance based on the Fieller method													
1	1.5	40	0.9416	0.9665	0.9679	0.9733	4	1.5	40	0.8370	0.9167	0.9540	0.9531
		60	0.9436	0.9610	0.9603	0.9654			60	0.8774	0.9377	0.9582	0.9607
		120	0.9456	0.9581	0.9567	0.9607			120	0.9024	0.9431	0.9527	0.9559
		200	0.9471	0.9545	0.9536	0.9564			200	0.9229	0.9466	0.9506	0.9539
	2	40	0.9380	0.9692	0.9700	0.9757		2	40	0.7964	0.9105	0.9519	0.9524
		60	0.9387	0.9634	0.9636	0.9686			60	0.8291	0.9263	0.9479	0.9511
		120	0.9409	0.9563	0.9556	0.9602			120	0.8637	0.9291	0.9400	0.9427
		200	0.9489	0.9583	0.9563	0.9597			200	0.9030	0.9400	0.9471	0.9479
	3	40	0.9304	0.9754	0.9753	0.9793		3	40	0.6715	0.8869	0.9411	0.9442
		60	0.9298	0.9682	0.9689	0.9722			60	0.7466	0.9081	0.9426	0.9468
		120	0.9389	0.9632	0.9642	0.9647			120	0.8410	0.9271	0.9482	0.9488
		200	0.9455	0.9600	0.9607	0.9625			200	0.8674	0.9302	0.9465	0.9478
3	1.5	40	0.8844	0.9458	0.9641	0.9693	5	1.5	40	0.7824	0.8640	0.9362	0.9427
		60	0.9012	0.9489	0.9611	0.9645			60	0.8336	0.9133	0.9477	0.9473
		120	0.9250	0.9555	0.9599	0.9636			120	0.8904	0.9370	0.9496	0.9531
		200	0.9311	0.9488	0.9532	0.9546			200	0.9068	0.9406	0.9478	0.9498
	2	40	0.8452	0.9368	0.9601	0.9645		2	40	0.7170	0.8520	0.9303	0.9225
		60	0.8764	0.9487	0.9597	0.9652			60	0.7643	0.8859	0.9219	0.9247
		120	0.9062	0.9493	0.9538	0.9582			120	0.8491	0.9164	0.9429	0.9438
		200	0.9181	0.9442	0.9477	0.9506			200	0.8920	0.9347	0.9437	0.9447
	3	40	0.7944	0.9442	0.9617	0.9641		3	40	0.5636	0.8009	0.8878	0.8909
		60	0.8187	0.9354	0.9496	0.9529			60	0.6708	0.8658	0.9149	0.9047
		120	0.8722	0.9391	0.9465	0.9468			120	0.7760	0.8961	0.9421	0.9428
		200	0.9062	0.9479	0.9517	0.9529			200	0.8177	0.9021	0.9435	0.9445

Table 5. Expected length of the 95% confidence interval for the population mean.

c	d	n	Expected length CI_{ls}	CI_M	CI_{lqs}	CI_{Bisq}	c	d	n	Expected length CI_{ls}	CI_M	CI_{lqs}	CI_{Bisq}
Using mean square error based on Taylor series expansion													
1	1.5	40	2.6101	2.4524	2.3972	2.3954	4	1.5	40	2.8113	2.4718	2.2011	2.2276
		60	2.1148	2.0185	1.9856	1.9864			60	2.2468	1.9906	1.8602	1.8578
		120	1.4698	1.4305	1.4186	1.4181			120	1.5294	1.4084	1.3640	1.3621
		200	1.1393	1.1194	1.1145	1.1135			200	1.1641	1.1000	1.0780	1.0778
	2	40	2.6805	2.4616	2.4111	2.4056		2	40	3.0014	2.5077	2.2644	2.2666
		60	2.1456	2.0131	1.9837	1.9813			60	2.3491	1.9988	1.8807	1.8756
		120	1.4880	1.4338	1.4231	1.4216			120	1.5758	1.4113	1.3673	1.3665
		200	1.1479	1.1202	1.1172	1.1144			200	1.1900	1.1009	1.0827	1.0793
	3	40	2.8372	2.5157	2.4722	2.4636		3	40	3.4134	2.7216	2.5540	2.5462
		60	2.2438	2.0443	2.0160	2.0140			60	2.6225	2.1266	2.0332	2.0287
		120	1.5264	1.4426	1.4301	1.4308			120	1.6991	1.4574	1.4224	1.4179
		200	1.1638	1.1218	1.1154	1.1160			200	1.2549	1.1225	1.1016	1.1024
3	1.5	40	2.7614	2.4430	2.2612	2.2642	5	1.5	40	2.8707	2.5584	2.1657	2.2660
		60	2.2056	1.9889	1.8936	1.8921			60	2.2730	1.9957	1.8273	1.8258
		120	1.5139	1.4165	1.3866	1.3813			120	1.5481	1.4063	1.3523	1.3486
		200	1.1520	1.1009	1.0885	1.0842			200	1.1735	1.0973	1.0727	1.0699
	2	40	2.9137	2.4668	2.3024	2.2948		2	40	3.0771	2.5930	2.2394	2.2754
		60	2.2943	1.9955	1.9062	1.9032			60	2.4042	2.0141	1.8593	1.8546
		120	1.5496	1.4153	1.3836	1.3813			120	1.6035	1.4119	1.3587	1.3564
		200	1.1774	1.1061	1.0915	1.0895			200	1.2056	1.1003	1.0760	1.0736
	3	40	3.2822	2.6511	2.5254	2.5174		3	40	3.5332	2.8251	2.5897	2.5932
		60	2.5303	2.1033	2.0297	2.0256			60	2.7195	2.1742	2.0605	2.0555
		120	1.6517	1.4532	1.4264	1.4222			120	1.7479	1.4701	1.4256	1.4228
		200	1.2219	1.1162	1.1016	1.1005			200	1.2746	1.1194	1.0954	1.0949
Using variance based on the Fieller method													
1	1.5	40	2.7037	2.8406	2.8740	2.8861	4	1.5	40	2.9171	3.2542	3.4624	3.4507
		60	2.1569	2.2389	2.2563	2.2648			60	2.2913	2.5237	2.6223	2.6268
		120	1.4944	1.5271	1.5340	1.5367			120	1.5492	1.6527	1.6845	1.6888
		200	1.1425	1.1587	1.1590	1.1635			200	1.1728	1.2272	1.2429	1.2442
	2	40	2.7752	2.9687	3.0047	3.0135		2	40	3.1132	3.6166	3.8126	3.8194
		60	2.2022	2.3157	2.3328	2.3413			60	2.4056	2.7380	2.8267	2.8362
		120	1.5097	1.5545	1.5603	1.5641			120	1.6032	1.7474	1.7774	1.7812
		200	1.1537	1.1764	1.1793	1.1811			200	1.1979	1.2731	1.2878	1.2895
	3	40	2.9421	3.2337	3.2639	3.2752		3	40	3.5457	4.3466	4.5134	4.5195
		60	2.3120	2.4873	2.5004	2.5111			60	2.6999	3.2133	3.2914	3.2981
		120	1.5542	1.6230	1.6297	1.6321			120	1.7192	1.9380	1.9647	1.9687
		200	1.1708	1.2058	1.2079	1.2103			200	1.2627	1.3762	1.3908	1.3915
3	1.5	40	2.8667	3.1682	3.3037	3.3085	5	1.5	40	2.9767	3.3006	3.6180	3.5424
		60	2.2560	2.4488	2.5168	2.5242			60	2.3359	2.5982	2.7286	2.7324
		120	1.5375	1.6199	1.6436	1.6472			120	1.5687	1.6908	1.7325	1.7348
		200	1.1667	1.2088	1.2182	1.2218			200	1.1789	1.2427	1.2614	1.2638
	2	40	3.0077	3.4440	3.5708	3.5822		2	40	3.1838	3.7055	4.0082	3.9820
		60	2.3520	2.6250	2.6901	2.6976			60	2.4565	2.8387	2.9601	2.9669
		120	1.5683	1.6828	1.7038	1.7089			120	1.6258	1.7958	1.8347	1.8380
		200	1.1814	1.2403	1.2505	1.2531			200	1.2133	1.3026	1.3189	1.3230
	3	40	3.3884	4.0608	4.1721	4.1811		3	40	3.6849	4.5588	4.8133	4.8103
		60	2.5904	3.0112	3.0683	3.0747			60	2.7879	3.3817	3.4854	3.4900
		120	1.6718	1.8485	1.8713	1.8727			120	1.7639	2.0212	2.0559	2.0584
		200	1.2280	1.3176	1.3259	1.3296			200	1.2884	1.4248	1.4403	1.4434