Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/130501

Approximate Asymptotic Confidence Interval for the Population Standard Deviation based on the Sample Gini's Mean Difference

Moustafa O. A. Abu-Shawiesh[1,*], *Aamir Saghir* [2] *and B. M. Golam Kibria*[3]

[1] Dept. of Mathematics, Faculty of Science, The Hashemite University, Al-Zarqa, 13115, Jordan
[2] Dept. of Mathematics, Mirpur University of Science and Technology, Mirpur-10250, Pakistan
[3] Dept. of Mathematics and Statistics, Florida International University, Miami, Florida, USA

Received: 27 Feb. 2019, Revised: 24 Mar. 2019, Accepted: 19 Apr. 2019
Published online: 1 Sep. 2019

Abstract: In this paper, an approximate asymptotic confidence interval for the population standard deviation (σ) is constructed based on the sample Gini's Mean Difference (GMD). The estimated Coverage Probability (CP) and the Average Width (AW) of the proposed approximate asymptotic confidence interval were studied by means of a Monte-Carlo simulation under different settings and compared with two-widely used methods, namely the exact method and the Bonnet (2006) method. It appears that the proposed approximate asymptotic confidence interval method based on GMD performing well comparing to the exact method for some selected distributions. Two real-life data examples are analyzed to illustrate the implementation of the several methods which also supported the results of the simulation study to some extent.

Keywords: Robust estimation, standard deviation, Gini's mean difference, confidence interval, non-normal distribution, pivotal quantity, coverage probability, average width, Monte-Carlo simulation.

1 Introduction

The confidence intervals for variances are known to be hypersensitive to minor violations of the normality assumption and are sensitive to amounts of kurtosis that may pass unnoticed in the handling of the data [1]. The classical estimation of the confidence intervals for the population standard deviation (σ), which is considered very important in many statistical applications, is based on the sample standard deviation (S), which is the most common scale estimator that provides a logical point estimate of the population standard deviation (σ). Unfortunately, S is very sensitive to the deviations from the assumed distribution. It is also not robust in the sense that changing even one value can dramatically change the computed value of it. Furthermore, S is not necessarily the most efficient or meaningful estimator of σ in skewed distributions and it is notable that it is not robust to slight deviations from normality [2]. Even that S has a good efficiency in platykurtic and moderately leptokurtic distributions but the classic inferential methods for it may perform poorly in realistically non-normal distributions [3]. Nevertheless, S is known as the most efficient scale estimator for the normal distribution often used to construct the $(1-\alpha)100\%$ confidence interval (CI) for σ.

The exact $(1-\alpha)100\%$ confidence interval (CI) for σ is generally based on the assumption that the underlying distribution of the data is normal. However this assumption is seldom fulfilled in applications as discussed by many authors including [4,5]. So, one main reason for considering alternatives to S for estimating σ, is its low efficiency at the non-normal distributions. The statistical literature shows that robust methods might give more meaningful measures of σ and are indeed more resistant to departures from normality than S, see for example [6] - [11].

Therefore, we are looking for an estimator, as an alternative to S, which is robust, has a closed form, and is easy to compute. The sample Gini's Mean Difference (GMD) might be a more meaningful measure of variation and may be preferred to the sample standard deviation (S). We think that the reconstruction of the utility of Gini's mean difference as a variability measure out of the

* Corresponding author e-mail: mabushawiesh@hu.edu.jo

normal framework is good. Moreover, we think that the idea of using the sample mean difference to build confidence intervals for variability is worth to be investigated. Many authors studied the performance of the Gini's mean difference to estimate variability in normal as well as in non-normal samples. They concluded that the GMD performed better for non-normal scenario, possibly for heavy-tailed distributions and it is also a very close competitor to S, while other estimators are not close competitor to S when samples are drawn from non-normal distributions, see for example [4, 12, 13, 14, 15].

In this paper, an approximate asymptotic confidence interval for the population standard deviation (σ) for one sample problem is proposed when samples are drawn from non-normal distributions. The proposed confidence interval is based on GMD. Also, the normal distribution of our pivotal quantity used to construct this confidence interval holds only asymptotically. Therefore, the proposed method provides an approximate asymptotic alternative confidence interval to the exact $(1-\alpha)100\%$ confidence interval for σ. The performance of the proposed approximate asymptotic method is investigated through a Monte-Carlo simulation study based on various evaluation criteria such as Coverage Probability (CP) and Average Width (AW). Furthermore among other methods developed to construct a confidence interval for σ, as Bonnet method [3] is best, so we have considered it and the exact methods in our simulation study comparison.

The rest of the paper is organized as follows: In Section 2, we present the definition of the sample Gini's mean difference and also state clearly its main properties. The pivotal quantity is derived in Section 3. In Section 4, the exact, the Bonett method [3] and the proposed approximate asymptotic confidence interval method for σ are introduced. A Monte-Carlo simulation study is conducted in Section 5. Two real-life data are analyzed in Section 6. Finally, some concluding remarks based on the simulation study and the real-life examples results are given in Section 7.

2 The Sample Gini's Mean Difference

The sample Gini's Mean Difference (GMD) was proposed as a measure of income inequality by the Italian mathematician Corrado Gini [16]. Gini (1912) introduced his statistic as he did not agree with the variance as a measure of variation. He thought that variability should not depend on location and variance, so he suggested using the mean difference which defined as the average of the absolute differences of all pairs of values in a population, even that, the Gini's mean difference, is not as widely employed as σ is. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n, then GMD is defined as follows:

$$GMD = \frac{\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} |X_i - X_j|}{\binom{n}{2}}$$

$$= \frac{2}{n(n-1)}\left[\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} |X_i - X_j|\right] \qquad (1)$$

The Gini's mean difference combines the advantages of standard deviation and mean deviation. The GMD may be more appropriate in case of small departure from normality. It is known that the GMD has asymptotic relative efficiency of 98% at normal distribution and more efficient than S if the normal distribution is contaminated by a small fraction [17] - [19]. For a random sample $X_1, X_2, \ldots, X_n$ from $N(\mu, \sigma^2)$ we have $E((\sqrt{\pi}/2)GMD) = \sigma$ [18], that is, it is an unbiased estimator of σ. The statistic $(\sqrt{\pi}/2)GMD$ is also known as the Downton's estimator [20]. In fact, David in [18] had shown that the GMD is highly efficient as compared to S. Unlike other estimators designed for measuring the variability of a random variable, the GMD is independent of any central measure of localization. Gini in [16] concludes that Gini's mean difference places less weight on the extremes and provides some modest protection against outliers with little loss in efficiency. It performs well over a wide range of distributions, including much heavier than normal distribution tails. However, the main advantage of Gini's mean difference is its finite-sample performance. First of all, it is unbiased at all distributions with finite first moments. Second, its finite-sample variance is known, which allows for instance better approximate confidence intervals. Neither of that is true for the standard deviation [21].

3 The Pivotal Quantity Derivation

In this section, we derive, based on [22], the pivotal quantity that will be used in this paper to construct the proposed approximate asymptotic confidence interval for σ. For the case of the normal distribution, $N(\mu, \sigma^2)$, the following relation between the population standard deviation (σ) and the Gini's mean difference (Δ) is given in [22]:

$$\Delta = \frac{2\sigma}{\sqrt{\pi}}$$

and

$$Var(\Delta) = \frac{4\sigma^2}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right] \qquad (2)$$

A first conclusion from [22] is about the estimator $\hat{\Delta}$ of Δ which is based on the following two relations:

(i) $E(\hat{\Delta}) = \Delta$

(ii) $Var(\hat{\Delta}) \to 0, \quad as \quad n \to \infty$

Hence $\hat{\Delta}$ is a mean-squared-error consistent estimator of Δ, and we have to notice that for the existence of $Var(\hat{\Delta})$, it is sufficient that $\sigma^2 < \infty$. Now, we consider the standardized sample statistic (Z) given in [23,24] which according to them as $\hat{\Delta}$ is asymptotically normally distributed, then the asymptotic distribution of Z is a standard normal, that is:

$$Z = \frac{\hat{\Delta} - \Delta}{\sqrt{Var(\hat{\Delta})}} \sim N(0,1) \quad (3)$$

The standardized sample statistic (Z) is used as our pivotal quantity to construct the proposed approximate asymptotic confidence interval for σ using the estimates $\hat{\Delta}$ and $Var(\hat{\Delta})$:

(i) $\Delta = \frac{2\sigma}{\sqrt{\pi}} \Rightarrow \hat{\Delta} = \frac{2\hat{\sigma}}{\sqrt{\pi}}$

(ii) $\hat{\sigma} = GMD \Rightarrow \hat{\Delta} = \frac{2\hat{\sigma}}{\sqrt{\pi}} = \frac{2GMD}{\sqrt{\pi}}$

$$Var(\hat{\Delta}) = \frac{4\hat{\sigma}^2}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right]$$

$$\Rightarrow Var(\hat{\Delta}) =$$

$$\frac{4GMD^2}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right]$$

(iii)

$$\sqrt{Var(\hat{\Delta})} =$$

$$\sqrt{\frac{4GMD^2}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right]}$$

$$= 2GMD\sqrt{\frac{1}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right]}$$

$$= 2GMD * B$$

where $B = \sqrt{\frac{1}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2)-2(2n-3)}{\pi}\right]}$

The values of the factor B are calculated for sample sizes $2,3,4,\ldots,49$ and given in Table 1 below.

(iv) The pivotal quantity (Z) is defined after simplifications as follows:

$$Z = \frac{\hat{\Delta} - \Delta}{\sqrt{Var(\hat{\Delta})}} = \frac{\frac{2GMD}{\sqrt{\pi}} - \frac{2\sigma}{\sqrt{\pi}}}{2GMD * B}$$

$$= \frac{GMD - \sigma}{\sqrt{\pi} * GMD * B} \sim N(0,1) \quad (4)$$

Table 1: The values of factor B

Sample size (n)	B	Sample size (n)	B
2	0.426062	26	0.081022
3	0.296007	27	0.079436
4	0.239511	28	0.077939
5	0.206283	29	0.076524
6	0.183822	30	0.075183
7	0.167359	31	0.073910
8	0.154634	32	0.072700
9	0.144420	33	0.071547
10	0.135989	34	0.070448
11	0.128878	35	0.069397
12	0.122774	36	0.068392
13	0.117462	37	0.067430
14	0.112783	38	0.066507
15	0.108621	39	0.065621
16	0.104888	40	0.064769
17	0.101515	41	0.063950
18	0.098447	42	0.063161
19	0.095641	43	0.062401
20	0.093062	44	0.061667
21	0.090681	45	0.060959
22	0.088474	46	0.060274
23	0.086420	47	0.059612
24	0.084503	48	0.058972
25	0.082708	49	0.058351

4 The Confidence Interval for the Population Standard Deviation

In this section, we introduce the two widely-used confidence interval methods for the population standard deviation (σ), namely the exact method and the Bonett method [3], with the proposed approximate asymptotic confidence interval method.

4.1 The Exact Confidence Interval for the Population Standard Deviation

Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ where S^2 is the sample variance, then the exact $(1-\alpha)100\%$ confidence interval for σ is given as follows:

$$CI = \left[\sqrt{\frac{(n-1)S^2}{\chi^2_{(\frac{\alpha}{2},n-1)}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}}\right] \quad (5)$$

where $\chi^2_{(\frac{\alpha}{2},n-1)}$ and $\chi^2_{(1-\frac{\alpha}{2},n-1)}$ be the $(\frac{\alpha}{2}) * 100 - th$ and $(1-\frac{\alpha}{2}) * 100 - th$ percentiles points of the chi-square (χ^2) distribution with $(n-1)$ degrees of freedom.

4.2 The Bonett Confidence Interval for the Population Standard Deviation

Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, then Bonett in [3] proposed the following $(1-\alpha)100\%$ confidence interval for σ:

$$CI = \left(\sqrt{exp\left\{ln(cS^2) - Z_{\frac{\alpha}{2}} Se\right\}}, \sqrt{exp\left\{ln(cS^2) + Z_{\frac{\alpha}{2}} Se\right\}} \right) \quad (6)$$

where $Z_{\frac{\alpha}{2}}$ is the two-sided critical Z-value, Se is a small sample adjustment given as $Se = c\sqrt{\{\hat{\gamma}_4^* - (n-3)/n\}/(n-1)}$, $c = \frac{n}{\left(n - Z_{\frac{\alpha}{2}}\right)}$, $\hat{\gamma}_4^* = \frac{(n_0\tilde{\gamma}_4 + n\tilde{\gamma}_4)}{(n_0+n)}$ where $\tilde{\gamma}_4$ is a prior estimate of γ_4 with trim-proportion equal to $\frac{1}{\{2(n-4)^{1/2}\}}$ and $\tilde{\gamma}_4 = \frac{n\sum_{i=1}^{n}(Y_i-\hat{\mu})^4}{\left(\sum_{i=1}^{n}(Y_i-\hat{\mu})^2\right)^2}$ where $\hat{\mu}$ is the trimmed mean of the data.

4.3 The Approximate Asymptotic Confidence Interval for the Population Standard Deviation

In this section, the proposed approximate asymptotic confidence interval for σ based on the pivotal quantity (Z) given in equation (4) is constructed. Let $X_1, X_2, \ldots, X_n$ be an independently and identically distributed (iid) random sample of size n from a population with finite mean μ and finite variance σ^2. Let $Z_{\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$ be the $(\frac{\alpha}{2}) * 100 - th$ and $(1 - \frac{\alpha}{2}) * 100 - th$ percentiles, then the proposed $(1-\alpha)100\%$ approximate asymptotic confidence interval for σ will be derived and constructed as follows:

$$P\left(Z_{\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Rightarrow P(Z_{\frac{\alpha}{2}} < \frac{GMD - \sigma}{\sqrt{\pi} * GMD * B} < Z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

$$\Rightarrow P(\sqrt{\pi} * GMD * B * Z_{\frac{\alpha}{2}} - GMD < -\sigma < \sqrt{\pi} * GMD * B * Z_{1-\frac{\alpha}{2}} - GMD) = 1 - \alpha$$

$$P(-\sqrt{\pi} * GMD * B * Z_{\frac{\alpha}{2}} + GMD > \sigma > -\sqrt{\pi} * GMD * B * Z_{1-\frac{\alpha}{2}} + GMD) = 1 - \alpha$$

$$P(-\sqrt{\pi} * GMD * B * Z_{1-\frac{\alpha}{2}} + GMD < \sigma < -\sqrt{\pi} * GMD * B * Z_{\frac{\alpha}{2}} + GMD) = 1 - \alpha$$

$$P(GMD(1 - \sqrt{\pi} * B * Z_{1-\frac{\alpha}{2}}) < \sigma < GMD(1 - \sqrt{\pi} * B * Z_{\frac{\alpha}{2}}) = 1 - \alpha$$

Thence, the proposed $(1-\alpha)100\%$ approximate asymptotic confidence interval for σ is obtained, which is:

$$CI = \left(GMD(1 - \sqrt{\pi} * B * Z_{1-\frac{\alpha}{2}}), GMD(1 - \sqrt{\pi} * B * Z_{\frac{\alpha}{2}}\right) \quad (7)$$

5 The Simulation Study and Results

In this section, the performance of the three confidence intervals is illustrated and compared by a Monte-Carlo simulation study. All simulations are performed using programs written in the R statistical software for windows. The aim of this simulation is to study the effect of non-normality on the three confidence intervals based on several non-normal distributions. In order to make the comparisons among various confidence intervals, the following criteria are considered: CP and AW of the resulting confidence intervals. It is acknowledged that the CP and AW are useful criterion for evaluating the confidence intervals. A shorter length width gives a better confidence interval and it is obvious that when coverage probability for all method is same, then a smaller width indicates that the method is appropriate for the specific sample.

The most common 95% confidence interval ($\alpha = 0.05$) for the confidence coefficient is used. It is well known that if the data are from a symmetric distribution (or n is large), then CP will be exact or close to $(1-\alpha)$. Another criterion is the width of the confidence interval. A shorter length width gives a better confidence interval. We have used 50,000 simulation replications for each one of the following sample sizes: 10, 25, 50, 75 and 100. We obtain the $(1-\alpha)100\%$ confidence interval $CI = (L, U)$ and then estimate CP and AW, respectively, by using the following two formulas:

$$CP = \frac{\#(L \leq \sigma \leq U)}{50000}$$

and

$$AW = \frac{\Sigma_{i=1}^{50000}(U_i - L_i)}{50000}$$

The simulated data are generated from: Standard Normal distribution, $t(5)$ distribution, Exponential distribution, Gamma and Beta distributions respectively. The simulation results are shown in Table 2 to Table 6.

The performance of the three confidence intervals for the normal distribution is examined first. The results in Table 2 suggested that when sampling from a normal distribution, the performance of the estimators for the three methods do not differ greatly. The CP is 95%, the same as the nominal value. The AW for the exact and the

Table 2: CP and AW for N(0, 1)

CP	AW	CP	AW	CP	AW	
10	0.9486	2.39	0.9495	1.57	0.9482	2.42
25	0.9497	2.20	0.9513	0.97	0.9483	2.23
50	0.9449	2.17	0.9511	0.78	0.9510	2.21
75	0.9497	2.09	0.9500	0.56	0.9500	2.19
100	0.9485	1.98	0.9512	0.05	0.9510	2.01

Table 3: CP and AW for t(5)

Exact CI	Bonett CI	Proposed GMD CI				
CP	AW	CP	AW	CP	AW	
10	0.8207	1.24	0.9380	1.18	0.8726	1.22
25	0.7367	1.08	0.9213	1.05	0.8563	1.07
50	0.8070	0.97	0.9155	0.93	0.8345	0.94
75	0.7886	0.85	0.8990	0.79	0.8085	0.80
100	0.7895	0.78	0.8523	0.73	0.7980	0.76

Table 4: CP and AW for Exponential (2)

CP	AW	CP	AW	CP	AW	
10	0.7045	2.10	0.8780	2.89	0.7298	1.85
25	0.6858	2.00	0.8834	2.84	0.7144	1.78
50	0.6870	1.93	0.8935	1.98	0.7234	1.75
75	0.6545	1.89	0.8996	1.95	0.6978	1.71
100	0.6450	1.77	0.9179	1.87	0.6920	1.64

Table 5: CP and AW for Gamma (1, 6)

Exact CI	Bonett CI	Proposed GMD CI				
CP	AW	CP	AW	CP	AW	
10	0.7022	0.17	0.8779	0.22	0.7666	0.12
25	0.6893	0.27	0.8872	0.29	0.7402	0.24
50	0.6545	0.56	0.8434	0.59	0.7234	0.45
75	0.5890	0.62	0.7456	0.67	0.6765	0.54
100	0.5042	0.71	0.6987	0.72	0.6246	0.60

Table 6: CP and AW for Beta (3, 3)

	Exact CI	Bonett CI	Proposed GMD CI			
	CP	AW	CP	AW	CP	AW
10	0.9750	1.14	0.9554	0.94	0.9598	0.91
25	0.9801	1.07	0.9530	0.85	0.9579	0.81
50	0.9790	0.98	0.9519	0.82	0.9560	0.79
75	0.9813	0.93	0.9511	0.79	0.9520	0.74
100	0.9820	0.79	0.9504	0.71	0.9502	0.68

Table 7: The 95% confidence intervals for butterfat data

Confidence Interval Method	95% CI for σ	Width
Exact CI	(68.257 , 131.088)	62.831
Bonett CI	(63.910 , 114.789)	50.879
Proposed GMD CI	(62.449 , 122.144)	59.695

Table 8: The 95% confidence intervals for chicken consumption data

Confidence Interval Method	95% CI for σ	Width
Exact CI	(11.980 , 24.492)	12.512
Bonett CI	(08.798 , 26.362)	17.564
Proposed GMD CI	(09.145 , 19.113)	9.968

proposed approximate asymptotic method based on GMD are about equal whereas it is shorter for the Bonett method.

The next simulation compares the performance of the three confidence intervals for a variety of non-normal distributions.

The results of Table 3 show that the Bonett method is more robust than both the exact and the proposed approximate asymptotic method. The CP ranges from 85% to 93%. The proposed approximate asymptotic method shows that the CP for it is better than that for the exact method especially for small samples. The CP in all methods decreases with increasing sample sizes. Concerning AW, the exact and the proposed approximate asymptotic method have approximately the same AW whereas the Bonett method AW is slightly shorter.

The results of Table 4 shows obviously that the CP for the three methods diverts away from the nominal value. The best CP is for the Bonett method and the shortest AW is for the proposed approximate asymptotic method. The CP ranges from 64% to 70% in exact method, from 87% to 91% in Bonett method and from 69% to 72% in the proposed approximate asymptotic method. Its shows that the CP for the proposed approximate asymptotic method is better than that for the exact method especially for small samples. The CP in all methods decreases with increasing sample sizes. The results of Table 5 shows obviously that the CP for the three methods diverts away from the nominal value. The best CP is for the Bonett method and the shortest AW is for the proposed approximate asymptotic method. The CP ranges from 50% to 70% in exact method, from 69% to 87% in Bonett method and from 62% to 76% in the proposed approximate asymptotic method. Its shows that the CP for the proposed approximate asymptotic method is better than that for the exact method especially for small samples. The CP in all methods decreases with increasing sample sizes.

As it can be seen from the results of Table 6 the CP ranges from 97% to 98% for the exact method. It is 95% in Bonett and proposed approximate asymptotic methods. The CP in Bonett and proposed approximate asymptotic methods decreases with increasing sample sizes, but for the exact method it is fluctuates with sample size changes. We note that the CP for the exact method is larger than the nominal value and then than that for the Bonett and proposed approximate asymptotic methods. It shows obviously that the CP for the Bonett and proposed approximate asymptotic methods is very good and approximately the same and it is better than that for the exact method.

In conclusion, all confidence interval methods considered in this paper are sensitive to moderate deviations from the normality. Their CP is going close to each other's when the sample size n is sufficiently large. The Bonett method proves the best coverage among the three estimation methods for sampled distributions, but the proposed approximate asymptotic method proves better CP than the exact methods in all sampled distributions.

6 Applications to Real Data

In this section, we present two real-life examples to illustrate the implementation and performances of the three confidence intervals given in the present paper.

6.1 Example I

This example is taken from [25]. The data set represents the amount of butterfat in pounds produced by a typical cow during a 305-day milk production period between her first and second calves. The butterfat production for a random sample of size $n = 20$ cows measured by a farmer yielding the following data: 481, 537, 513, 583, 453, 510, 570, 500, 457, 555, 618, 327, 350, 643, 499, 421, 505, 637, 599, 392. The sample mean, the sample standard deviation and the skewness for the data are 507.5, 89.75 and -0.3804 respectively. The value of the Shapiro-Wilk Normality test (W= 0.9667, p-value = 0.6834) suggested that the data follows a normal distribution. The resulting 95% confidence intervals for the three methods and the corresponding confidence widths are given below in Table 7. From the results in Table 7, we have observed that when the data has a normal distribution, the confidence intervals widths for the exact and proposed approximate asymptotic methods are closed to each other and about equal whereas it is shorter for the Bonett method. These results support the results of the Monte-Carlo simulation study to some extent.

6.2 Example II

This example is taken from [26]. The data set represents the last year's chicken consumption in pounds for people on USA published by the USA Department of Agriculture in Food Consumption, Prices, and Expenditures. The last year's chicken consumption, in pounds, for a random sample of size $n = 17$ people yielded the following data: 47, 39, 62, 49, 50, 70, 59, 53, 55, 0, 65, 63, 53, 51, 50, 72, 45.

The sample mean, the sample standard deviation and the skewness for these data are 51.94, 16.08 and -2.11 respectively. The Shapiro-Wilk Normality test value (W = 0.8013, p-value = 0.0021) suggested that the data does not follow a normal distribution. The resulting 95% confidence intervals for the three different methods are given below in Table 8.

From the results in Table 8, we have observed that the proposed approximate asymptotic method has the narrowest width followed by the exact method. It is also noted that the Bonett method has the widest width than the other two methods. This results supported the results of the Monte-Carlo simulation study to some extent.

7 Perspective

The paper proposes an asymptotic confidence interval for estimating the population standard deviation (α) based on the sample Gini's Mean Difference (GMD). A Monte-Carlo simulation study has been conducted to compare the performance of the proposed method with the exact method and the Bonett method in [3]. It appears that the sample size n has significant effect on the proposed approximate asymptotic method. Also, all confidence interval methods considered in this paper are sensitive to moderate deviations from the normality. Their CP goes close to each other's when the sample size n is sufficiently large. To illustrate the findings of the paper we considered two real-life examples which also supported the simulation study results to some extent. As a result, when the population is non-normal and for small to moderate sample sizes, the proposed approximate asymptotic confidence interval method is recommended to be used. For the further research, we can derive an Asymptotically Distribution Free (ADF) standard error of the Gini mean difference, and we then compare the Coverage Probability (CP) of the Gini CI using the ADF standard error with the coverage probability (CP) of the ADF, MAD, CI as proposed by [27]. We believe that an ADF Gini CI might have narrower relative width than the ADF MAD CI.

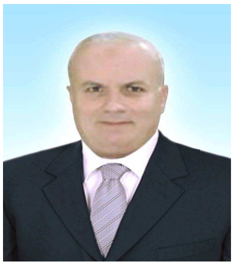
Acknowledgement

We thank the anonymous referees for thorough review and highly appreciate the helpful comments and useful suggestions, which significantly contributed to improving the quality of the manuscript.

References

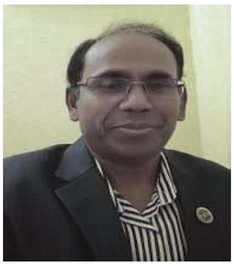
[1] G. W. Snedecor, and W. G. Cochran, *Statistical Methods*, 7th Edition, Iowa State University Press Iowa USA, (1980).
[2] J. W. Tukey, *A survey of sampling from contaminated distributions*, in contributions to probability and statistics, essays in honor of Harold Hotelling, Olkin, I., et al. Ed. Stanford University Press, Stanford USA, 448-485, (1960).
[3] D. G. Bonett, Approximate confidence interval for standard deviation of nonnormal distributions, *Computational Statistics and Data Analysis*, **50**, 775-782 (2006).
[4] A. Saghir, and Z. Lin, Designing of Gini-chart for exponential, t, logistic and laplace distributions, *Communications in Statistics-Simulation and Computation*, **44**, 2387-2409 (2015).
[5] S. Banik, A. N. Albatineh, M. O. Abu-Shawiesh, and B. M. G. Kibria, Estimating the population standard deviation with confidence interval: a simulation study under skewed and symmetric conditions, *International Journal of Statistics in Medical Research*, **3**, 356-367 (2014).
[6] H. A. David, Early sample measures of variability, *Statistical Science*, **13**, 368-377 (1998).
[7] A. M. Barham, and S. Jeyaratnam, Robust confidence interval for the variance, *Journal of Statistical Computation and Simulation*, **62**, 189-205 (1999).
[8] S. Niwitpong, and P. Kirdwichai, Adjusted Bonett confidence interval for standard deviation of non-normal distributions, *Thailand Statistician*, **6**, 1-6 (2008).
[9] B. D. Burch, Estimating kurtosis and confidence intervals for the variance under nonnormality, *Journal of Statistical Computation and Simulation*, **84**, 2710-2720 (2014).
[10] H. E. Akyüz, and H. Gamgam, Interval estimation for nonnormal population variance with kurtosis coefficient based on trimmed mean, *Türkiye Klinikleri Biyoistatistik Dergisi*, **9**, 213-221 (2017).
[11] H. E. Akyüz, H. Gamgam, and A. YalASçnkaya, Interval estimation for the difference of two independent nonnormal population variances, *Gazi University Journal of Science*, **30**, 117-129 (2017).
[12] M. Schoonhoven, and J. M. M. Does, The X control chart under non-normality, *Quality and Reliability Engineering International*, **26**, 167-176 (2010).
[13] M. Schoonhoven, M. Riaz, and J. M. M. Does, Design and analysis of control charts for standard deviation with estimated parameters, *Journal of Quality Technology*, **43**, 307-333 (2011).
[14] S. A. Abbasi, and A. Miller, On proper choice of variability control chart for normal and non-normal processes, *Quality and Reliability Engineering International*, **28**, 279-296 (2012).
[15] A. Saghir, and Z. Lin, A Study on the robustness of G-chart to non-normality, *Communications in Statistics Simulation and Computation*, **43**, 2241-2251 (2014).
[16] C. Gini, Variabilitiae mutabilitia, *contribututoallo studiodelle distribuzioni e dellerelazoine statistiche*, Studion Economicoiuredice dell'Universitiadie Cagliari **3 (Part 2) i-iii**, 3-159 (1912).
[17] S. M. Stigler, Studies in the history of probability and statistics, XXXII Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika*, **60**, 439-445 (1973).

[18] H. A. David, Gini's mean difference rediscovered, *Biometrika*, **55**, 573-575 (1968).
[19] C. Gerstenberger, and D. Vogel, *On the efficiency of Gini's mean difference*, Cornel University Library, arxiv.org/abs/1405.5027 (2014).
[20] F. Downton, Linear estimates with polynomial coefficients, *Biometrika*, **53**, 129-141 (1966).
[21] S. Yitzhaki, Gini's mean difference: a superior measure of variability for non-normal distributions, *International Journal of Statistics*, **LXI, 2**, 285-316 (2003).
[22] M. Zenga, M. Polisicchio, and F. Greselin, The variance of Gini's mean difference and its estimators, *STATISTICA*, **anno LXIV, n.3**, 455-475 (2004).
[23] W. Hoeffding, A class of statistics with asymptotically normal distribution, *The Annals of Mathematical Statistics*, **19**, 293-325 (1948).
[24] H. A. David , and H. N. Nagraja, *Order Statistics*, 3rd Edition, John Wiley and Sons, New York USA, (2003).
[25] R. V. Hogg, E.A. and Tanis, *Probability and Statistical Inference*, 6th Edition, Prentice-Hall Inc., New Jersey USA, (2001).
[26] Weiss A. N., *Introductory Statistics*, 6th Edition, Addison-Wesley, New York USA, (2002).
[27] D. G. Bonett, and E. Seier, Confidence intervals for mean absolute deviations, *American Statistician*, **57**, 233-236 (2003).



Moustafa Abu-Shawiesh is an Associate Professor of statistics in the department of mathematics at the Hashemite University (HU). He also taught at the King Faisal University, Saudi Arabia and Nizwa University, Oman. Dr. Abu-Shawiesh has diverse research interests, mainly, Applied Statistics, Estimation Theory, Statistical Process Control, artificial neural networks, Statistical Inference and Simulation Study. Since 1997, he has more than 30 full research articles either published or accepted in different internationally well-reputed statistical journals. Dr. Abu-Shawiesh is a member of editorial board of many international statistical journals, including Annals of Management Science and Journal of Mathematics and Statistics. He also served as a peer reviewer to many international journals in Statistics, Mathematics and Management.

Aamir Saghir is currently working as Associate Professor of Statistics in Mirpur University of Science and Technology (MUST), Mirpur-10250 (AJK), Pakistan. He has obtained his PhD degree in Probability theory and Mathematical Statistics from Zhejiang University P.R. China in 2014. He has obtained MS degree in Statistics from Quaid-e-Azam University Islamabad Pakistan in 2008. His current research interests include Statistical Process Monitoring and Probability Modelling. He published more than 34 articles in internationally well reputed ISI journals.. He also served as a peer reviewer to many international journals in Statistics and Mathematics.



B. M. Golam Kibria is a Professor in the department of mathematics and statistics at Florida International University (FIU). He also taught at The University of British Columbia and The University of Western Ontario, Canada and Jahangirnagar University, Bangladesh. Dr. Kibria has diverse research interests, mainly, Applied Statistics, Distribution Theory, Quality Control, Linear Model, Ridge Regression, Statistical Inference and Simulation Study. Since 1993, he has about 180 full research articles either published or accepted in different internationally well-reputed statistical journals. He is the recipient of several awards: Certificate of Excellence in Reviewing Award -2018, awarded by the Journal of Advances in Mathematics and Computer Sciences. FIU Top Scholar Award2016, FIU College of Arts, Science and Education Research Award 2016, Asadul Kabir Gold Medal from Jahangirnagar University among others. Dr. Kibria is the Editor-in-Chief of the Journal of Probability and Statistical Science and member of editorial board of many international statistical journals, including Communications in Statistics -A, B and C. He is the elected Fellow of Royal Statistical Society.