

A simulation study on some confidence intervals for the population standard deviation

Moustafa Omar Ahmed Abu-Shawiesh¹, Shipra Banik²
and B. M. Golam Kibria³

Abstract

In this paper a robust estimator against outliers along with some other existing interval estimators are considered for estimating the population standard deviation. An extensive simulation study has been conducted to compare and evaluate the performance of the interval estimators. The exact and the proposed robust method are easy to calculate and are not overly computer-intensive. It appears that the proposed robust method is performing better than other confidence intervals for estimating the population standard deviation, specifically in the presence of outliers and/or data are from a skewed distribution. Some real-life examples are considered to illustrate the application of the proposed confidence intervals, which also supported the simulation study to some extent.

MSC: 62F10, 62F40

Keywords: Breakdown point, bootstrapping, confidence interval, coverage probability, standard deviation, Q_n estimator, robust estimator, outlier, scale estimator.

1. Introduction

Point estimates are of limited value, since we cannot attach to them statements regarding the amount of confidence that they have estimated the unknown parameter. Of great value is an interval estimate, an estimate about which we can make statements of confidence (Daniel, 1990). The confidence interval is defined as an estimated range of values that is likely to include an unknown population parameter. If independent samples

¹ Department of Mathematics, Faculty of Science, Hashemite University, Zarqa 13115, Jordan, mabushawiesh@hu.edu.jo

² School of Engineering and Computer Science, Independent University, Bangladesh, Dhaka 1229, Bangladesh, banik@secs.iub.edu.bd

³ Department of Mathematics and Statistics, Florida International University, Modesto A. Maidique Campus, Miami, FL 33199, USA, kibriag@fiu.edu

Received: February 2011

Accepted: May 2011

are taken repeatedly from the same population, and the confidence interval is calculated for each sample, then a certain percentage, called the confidence level of the interval, will include the unknown population parameter.

Scale estimators are very important in many statistical applications. The sample standard deviation is the most common scale estimator that provides a logical point estimate of the population standard deviation, σ . Unfortunately, the sample standard deviation, S , is very sensitive to the presence of outliers in the data. Furthermore, S is not necessarily the most efficient or meaningful estimator of scale in skewed and leptokurtic distributions and it is notable that it is not robust to slight deviations from normality (Tukey, 1960). 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 (Bonett, 2006). Also, according to Gorard (2004), S has no obvious intuitive meaning because squaring before summing and then taking the square root makes the resulting figure difficult to understand, which restricts any subsequent intuitive interpretation. Nevertheless, S is the most efficient scale estimator for the normal distribution often used to construct the $100(1 - \alpha)\%$ confidence interval for σ . The standard error of S is a scale multiple of the actual parameter being estimated. In this paper, we are looking for a scale estimator which is robust, has a closed form, and easy to compute as an alternative to S . The Rousseeuw-Croux estimator, Q_n might be a more meaningful measure of variation and may be preferred to S . It is the most efficient scale estimator for the normal distribution often used to construct the $100(1 - \alpha)\%$ confidence interval for σ .

The exact $100(1 - \alpha)\%$ confidence interval for the population standard deviation, σ , is based on the assumption that the underlying distribution of the data is normal with no outliers, but what would happen if the data are not from a normal distribution instead of heavier tails or from a skewed distribution. The statistical literature shows that robust methods might give more meaningful measures of scale and are indeed more resistant to departures from normality and presence of outliers than S . Therefore, the need for alternatives to the exact $100(1 - \alpha)\%$ confidence interval for σ comes to play. The statistical literature is full of robust confidence intervals for the mean, for example Tukey and McLaughlin (1963); Huber (1964); Dixon and Tukey (1968); De Wet and van Wyk (1979); Gross (1973, 1976); Bickel and Doksum (1977, p. 375); Kim (1992) and Clark (1994). For small-sample inference about variance and its transformations we refer to Longford (2010) among others. The problem of constructing robust confidence interval for the population standard deviation, σ , has received much less attention. Here, by a robust confidence interval, we mean that its actual coverage probability is close to the specified confidence level $(1 - \alpha)$ with a short length of the confidence interval.

In this paper, an approximate confidence interval for the population standard deviation, σ , for one sample problems that is much less sensitive to the presence of outliers and/or to departure from normality is proposed. The proposed method provides an alternative to the exact $100(1 - \alpha)\%$ confidence interval for σ based on Q_n . The performance of the proposed method is investigated through a Monte Carlo simulation study based

on various evaluation criteria such as coverage probability, average width and standard deviation of the width. The coverage probability naturally varies from distribution to distribution for a given procedure, but a good procedure should keep this variation small. Furthermore, we want a confidence interval whose endpoints are generally close together, thus a small average width is good (Gross, 1976). A set of real data is employed to illustrate the results given in the paper.

The organization of the paper is as follows. In Section 2, we presented Rousseeuw-Croux estimator, Q_n , for estimating σ , and discussed outliers. The proposed confidence intervals for σ are presented in Section 3. A Monte Carlo simulation study is conducted in Section 4. Some real life data are analyzed in Section 5. Finally, some concluding remarks based on simulation and numerical examples are given in Section 6.

2. Rousseeuw-Croux estimator, Q_n and outliers

2.1. The Rousseeuw-Croux estimator

Rousseeuw and Croux (1993) proposed two robust estimators for scale, the S_n and Q_n estimators. They can be used as initial or ancillary scale estimators in the same way as the median absolute deviation (MAD) but they are more efficient and not biased towards symmetric distributions. The breakdown point of the S_n estimator is 50% and its efficiency is 58%, while Q_n has the same breakdown point but its efficiency at normal distributions is very high, about 82%. Due to its high efficiency and other good properties, the Q_n estimator is considered in this paper. Mosteller and Tukey (1977) define two types of robustness as follows:

1. Resistance: This means that changing a small part even by a large amount of the data does not cause a large change in the estimate.
2. Robustness of efficiency: This means that the statistic has high efficiency in a variety of situations rather than in any one situation. Efficiency means that the estimate is close to the optimal estimate given that the distribution of the data is known.

Many statistics have one of these properties. However, it can be difficult to find statistics that have both resistance and robustness of efficiency. The most common estimate of scale, S is the most efficient estimate of scale if the data come from a normal distribution. However, S is not robust in the sense that changing even one value can dramatically change the computed value of S ; that is, it has poor resistance. In addition, it does not have robustness or efficiency for non-normal data. MAD and the inter-quartile range (IQR) are the two most commonly used robust alternatives to S . MAD in particular is a very robust scale estimator. However, MAD does not have

particularly high efficiency for data (37% for normal data) and also MAD has an implicit assumption of symmetry, that is it measures the distance from a measure of central location (the median). Rousseeuw and Croux (1993) proposed the Q_n estimate of scale as an alternative to MAD. It shares desirable robustness properties with MAD (50% breakdown point, bounded influence function). In addition, it has significantly better normal efficiency (82%) and it does not depend on symmetry.

2.1.1. Definition of Q_n

The estimator Q_n for a random sample X_1, X_2, \dots, X_n with model distribution F is defined as:

$$Q_n = 2.2219 \{ |X_i - X_j| ; i < j ; i = 1, 2, 3, \dots, n ; j = 1, 2, 3, \dots, n \}_{\{g\}} \quad (1)$$

where $g = \binom{h}{2} \binom{n}{2} / 4$ and $h = \lfloor \frac{n}{2} \rfloor + 1$ (i.e., roughly half the number of observations). Here the symbol $\binom{\cdot}{\cdot}$ represents the combination and the symbol $\lfloor \cdot \rfloor$ is used to take only the integer part of a fraction. The Q_n estimator is the g -th order statistic of the $\binom{n}{2}$ inter-point distances. The value 2.2219 is chosen to make Q_n a consistent estimator of scale for normal data. Rousseeuw and Croux (1993) have derived the unbiasing factor dn so that $dn \times Q_n$ becomes an unbiased estimator of σ for the case of normal distribution. These values of dn are provided here in Table 2.1 as a function of n . The scatterplot between n and dn is presented in Figure 2.1. We can observe from both Table 2.1 and Figure 2.1 that dn is sensitive to the sample sizes.

Table 2.1: The values of the unbiasing factor dn .

n	dn	n	dn
2	0.399	14	0.787
3	0.994	15	0.915
4	0.512	16	0.808
5	0.844	17	0.924
6	0.611	18	0.826
7	0.857	19	0.931
8	0.669	20	0.840
9	0.872	21	0.938
10	0.725	22	0.853
11	0.887	23	0.943
12	0.759	24	0.863
13	0.903	25	0.647

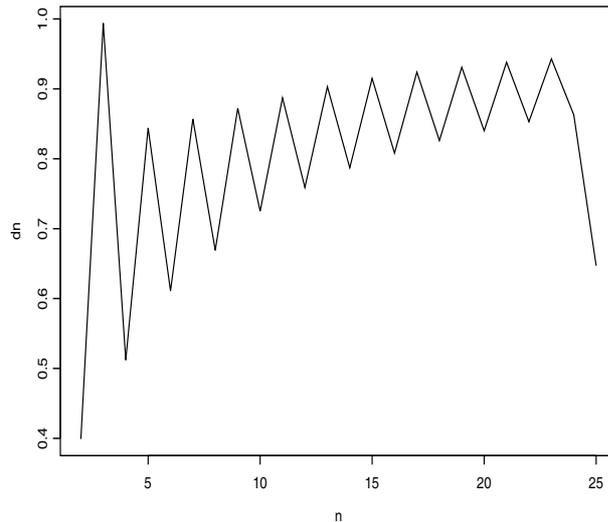


Figure 2.1: Scatterplot between n and dn .

An approximation result of dn for larger values of n is given by Croux and Rousseeuw (1992) as follows:

$$d_n = \begin{cases} \frac{n}{n+1.4} & \text{for odd values of } n \\ \frac{n}{n+3.8} & \text{for even values of } n \end{cases} \quad (2)$$

2.1.2. Properties of Q_n

The Q_n estimator has a simple and explicit formula, which is equally suitable for asymmetric distributions. The main properties of the Q_n estimator investigated by Rousseeuw and Croux (1993) are given below:

1. For any sample $X = \{X_1, X_2, \dots, X_n\}$ in which no two points coincide, the break-down point of the scale estimator Q_n is given by $\varepsilon^*(Q_n, X) = \frac{\lfloor \frac{n}{2} \rfloor}{n}$.
2. For $F = \Phi$, where $\Phi(x)$ is the standard normal distribution function, the value of d is given by $d = \frac{1}{\sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right)} = 2.2219$. With this constant d , Q_n has bias in small samples.
3. The influence function of Q_n estimator is smooth and unbalanced (see Rousseeuw and Croux, 1993). For a model distribution F which has a density f , the influence function of Q_n is given by

$$IF(x; Q, F) = d \frac{\frac{1}{4} - F(x+d^{-1}) + F(x-d^{-1})}{\int f(y+d^{-1})f(y)dy}$$

4. The gross-error sensitivity of the Q_n estimator is larger than those of MAD and S_n estimators and its value is $\gamma^*(Q, \Phi) = \sup_x |IF(x; Q, \Phi)| = 2.069$.
5. The asymptotic variance of Q_n in the case of normal distribution is given by $V(Q, \Phi) = 0.6077$ and this yields an efficiency of 82.27%. This is very high relative to the MAD estimator whose efficiency at normal distribution is only 36.74% and S_n whose efficiency is 58.23%. Using a simulation study, Rousseeuw (1991) concluded that the estimator Q_n is more efficient than MAD and S_n estimators. However, Q_n loses some of its efficiency for small sample sizes.
6. The square of the Q_n estimator, that is, $(Q_n)^2$, can be used as an estimate of σ^2 . Even though both Q_n and $(Q_n)^2$ are biased estimators of σ and σ^2 respectively, they are efficient estimators of their respective targets (Rousseeuw, 1991).

2.2. Estimators and outliers

The presence of outliers in the data set is one of the most important topics in statistical inference. An outlier can be defined as observations which appear to be inconsistent with the remaining set of data. Outliers can be contaminants, i.e. arising from other distributions or can be typical observations generated from the assumed model (Barnett, 1988). Therefore, outliers need very special attention because a small departure from the assumed model can have strong negative effects on the efficiency of classical estimators for location and scale (Tukey, 1960). In this section, a simple numerical example taken from Rousseeuw (1991) is given to show the effect of outliers on S and Q_n estimators. Suppose we have five measurements of a concentration without outliers given as follows:

$$5.59, 5.66, 5.63, 5.57, 5.60$$

Let us now suppose that one of these concentrations has been wrongly recorded so that the data have an outlier value and become as follows:

$$5.59, 5.66, 5.63, 55.7, 5.60$$

Based on these two data sets, the values of the two estimators are calculated and given in Table 2.2.

Table 2.2: Values of the estimators for the example.

	Scale Estimator	Data Set
	Without Outlier	With Outlier
S	0.0354	22.4
Q_n	0.066657	0.066657

From Table 2.2, we notice that the value of the single outlier has changed the value of S , which becomes very large. The robustness of the Q_n estimator is clear where the value of it is the same for the two data sets.

3. Proposed robust confidence interval for σ

3.1. Exact confidence interval for σ

Let X_1, X_2, \dots, X_n be a random sample of size n from the normal distribution, i.e., $X_i \sim N(\mu, \sigma^2)$ for all i , then $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2$ where $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is the sample variance. The exact $100(1 - \alpha)\%$ confidence interval for a population variance σ^2 is given as follows:

$$P\left(\frac{(n-1)S^2}{\chi_{\left(\frac{\alpha}{2}, n-1\right)}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\left(1-\frac{\alpha}{2}, n-1\right)}^2}\right) = 1 - \alpha \quad (3)$$

where $\chi_{\frac{\alpha}{2}}^2$ and $\chi_{1-\frac{\alpha}{2}}^2$ are the $\left(\frac{\alpha}{2}\right)^{th}$ and $\left(1 - \frac{\alpha}{2}\right)^{th}$ percentile points of the χ^2 distribution with $(n-1)$ degrees of freedom. Taking the square root of the endpoints of equation (3) gives a $100(1-\alpha)\%$ confidence interval for σ as follows:

$$\left(\sqrt{\frac{(n-1)S^2}{\chi_{\left(\frac{\alpha}{2}, n-1\right)}^2}}, \sqrt{\frac{(n-1)S^2}{\chi_{\left(1-\frac{\alpha}{2}, n-1\right)}^2}} \right) \quad (4)$$

The exact confidence interval for σ^2 in (3) is hypersensitive to minor violations of the normality assumption. Scheffe (1959, p. 336) show that (3) has an asymptotic coverage probability of about 76, 63, 60 and 51 for the Logistic, the Student t(7), the Laplace and Student t(5) distributions respectively. The result is disturbing because these symmetric distributions are not easily distinguished from a normal distribution unless the sample size is large. Also, the exact confidence interval for σ^2 in (3) as demonstrated by Lehman (1986, p. 206) is highly sensitive to the presence of outliers and / or to departure from normality. However, as pointed out by Lehman, the sample size n may have to be rather large for the asymptotic result to give a good approximation.

3.2. Robust confidence intervals

In this section, we will propose the new robust confidence interval for estimating the population standard deviation σ . Instead of assuming $X_i \sim N(\mu, \sigma^2)$, let X_1, X_2, \dots, X_n are the random samples of size n from a continuous, independent and identically distributed random variable. The random variable T is defined as the ratio,

$$T = \frac{d_n Q_n}{\sigma} \quad (5)$$

where the expression $d_n Q_n$ acts as an unbiased estimator of σ so that $E(T)=1$ for normal distribution. Based on Rousseeuw and Croux(1993), for larger values of n , the following asymptotic result can be used:

$$T = \frac{d_n Q_n}{\sigma} \sim N\left(1, \frac{1}{1.65n}\right) \quad (6)$$

The following approximation result can be obtained:

$$d_n Q_n \sim N\left(\sigma, \frac{1}{1.65n}\sigma^2\right) \quad (7)$$

Therefore from (7), we can get the following pivotal quantity:

$$\frac{d_n Qn - \sigma}{\frac{1}{1.28\sqrt{n}}\sigma} \sim N(0, 1) \quad (8)$$

Now, using the above pivotal quantity, we can derive the $100(1 - \alpha)\%$ robust confidence interval for σ as follows:

$$\begin{aligned} P\left(q_1 < \frac{d_n Qn - \sigma}{\frac{1}{1.28\sqrt{n}}\sigma} < q_2\right) &= 1 - \alpha \\ \Rightarrow P\left(\frac{q_1}{1.28\sqrt{n}} + 1 < \frac{d_n Qn}{\sigma} < \frac{q_2}{1.28\sqrt{n}} + 1\right) &= 1 - \alpha \\ \Rightarrow P\left(\frac{1.28\sqrt{n} * d_n Qn}{q_1 + 1.28\sqrt{n}} < \sigma < \frac{1.28\sqrt{n} * d_n Qn}{q_2 + 1.28\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

where $q_1 = Z_{\frac{\alpha}{2}}$ and $q_2 = Z_{1-\frac{\alpha}{2}}$ are the $(\frac{\alpha}{2})^{th}$ and $(1 - \frac{\alpha}{2})^{th}$ percentile points of the standard normal distribution so that the length is minimum. Therefore, the $100(1 - \alpha)\%$ robust confidence interval for σ , is as follows:

$$\left(\frac{1.28\sqrt{n} * d_n Qn}{Z_{\frac{\alpha}{2}} + 1.28\sqrt{n}}, \frac{1.28\sqrt{n} * d_n Qn}{Z_{1-\frac{\alpha}{2}} + 1.28\sqrt{n}} \right) = \left(\frac{D Qn}{Z_{\frac{\alpha}{2}} + D_1}, \frac{D Qn}{Z_{1-\frac{\alpha}{2}} + D_1} \right) \quad (9)$$

where the values of the factors $D = 1.28\sqrt{n} * d_n$ and $D_1 = 1.28\sqrt{n}$.

An approximation result of D for larger values of n can be calculated as follows:

$$D = \begin{cases} (1.28\sqrt{n}) \left(\frac{n}{n+1.4} \right), & \text{for odd values of } n \\ (1.28\sqrt{n}) \left(\frac{n}{n+3.8} \right), & \text{for even values of } n \end{cases} \quad (10)$$

The squaring of the endpoints of equation (9) gives a $100(1 - \alpha)\%$ confidence interval for σ^2 .

3.3. Bonett confidence interval

Let X_1, X_2, \dots, X_n be a random sample of size n from the normal distribution, that is, $X_i \sim N(\mu, \sigma^2)$ for all i . Scheffe (1959) found in his simulation study, the exact CI for σ

does not have an asymptotic coverage probability for non-normal distributions. Bonett (2006) proposed the following $(1-\alpha)100\%$ confidence interval (CI) for σ as

$$LCL = \exp \{ \ln(c\hat{\sigma}^2) - Z_{\alpha/2} se \} \quad \text{and} \quad UCL = \exp \{ \ln(c\hat{\sigma}^2) + Z_{\alpha/2} se \}$$

where $Z_{\alpha/2}$ is two-sided critical z-value, $se = c[\{\hat{\gamma}_4(n-3)/n\}/(n-1)]^{1/2}$, $c = n/(n - Z_{\alpha/2})$ and $\hat{\gamma}_4 = n \sum_i (Y_i - \hat{\mu})^4 / (\sum_i (Y_i - \hat{\mu})^2)^2$.

3.4. Cojbasic and Tomovic (CT) CI

Based on t -statistic, Cojbasic and Tomovic (2007) proposed the following nonparametric bootstrap t CI:

$$I_{boot} = S^2 - \hat{t}^{(\alpha)} \sqrt{\widehat{var}(S^2)}$$

where $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$ is the sample variance, $\hat{t}^{(\alpha)}$ is a α percentile of T^* defined as $T^* = \frac{S^{2*} - S^2}{\sqrt{\widehat{var}(S^{2*})}}$, S^{2*} is a bootstrap replication of statistic S^2 and $\widehat{var}(S^2)$ is a consistent estimator of the variance, defined by $2\sigma^4/(n-1)$.

3.5. Some bootstrap CIs

Let $X^{(*)} = X_1^{(*)}, X_2^{(*)}, \dots, X_n^{(*)}$, where the i -th sample is denoted by $X(i)$ for $i = 1, 2, \dots, B$ and B is the number of bootstrap samples. We proposed the following bootstrap CIs for the sample σ :

Non-parametric bootstrap CI

Compute σ for all bootstrap samples and then order the sample SDs of each bootstrap samples as follows:

$$S_{(1)}^* \leq S_{(2)}^* \leq S_{(3)}^* \cdots \leq S_{(B)}^*$$

CI for population σ :

$$LCL = S_{[(\alpha/2)B]}^* \quad \text{and} \quad UCL = S_{[(1-\alpha/2)B]}^*$$

CI for population σ :

$$LCL = S \sqrt{(n-1)/\chi_{\alpha/2, (n-1)}^{*2}} \quad \text{and} \quad UCL = S \sqrt{(n-1)/\chi_{1-\alpha/2, (n-1)}^{*2}}$$

where $\chi_{\alpha/2}^{*2}$ and $\chi_{1-\alpha/2}^{*2}$ are the $(\alpha/2)$ -th and $(1 - \alpha/2)$ -th sample quantiles of $\chi^2 = \frac{(n-1)S^2}{\hat{\sigma}_B^2}$ and $\hat{\sigma}_B = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\bar{x}_i^* - \bar{\bar{x}})^2}$ and \bar{x}_i^* is the i -th bootstrap sample mean, $\bar{\bar{x}}$ is the bootstrap mean and $\hat{\sigma}_B$ is the bootstrap standard deviation.

Bootstrap robust CI

CI for population σ :

$$LCL = \frac{DQn}{Z_{\alpha/2}^* + D_1} \quad \text{and} \quad UCL = \frac{DQn}{Z_{1-\alpha/2}^* + D_1}$$

where $Z_{\alpha/2}^*$ and $Z_{1-\alpha/2}^*$ are the $(\alpha/2)$ th and $(1 - \alpha/2)$ th sample quantiles of the bootstrap test statistic, $Z_i^* = \frac{(\bar{x}_i^* - \bar{\bar{x}})}{\hat{\sigma}_B}$.

We note that all proposed confidence intervals except exact and Bonett do not require any distributional assumptions. However, bootstrap methods are computer intensive, where as others are very easy to compute. The exact method works better for any sample size when the data are from the normal distribution.

4. Simulation study

Our basic objective is to investigate some efficient estimators of σ by a simulation study. Since a theoretical comparison among the intervals is not possible, a simulation study has been made to compare the performance of the estimators.

4.1. Simulation technique

The flowchart of our simulation is as follows:

1. We use sample sizes $n = 5, 10, 20, 30, 50, 70$ and 100.
2. Random samples are generated from symmetric and skewed distributions:
 - (a) Normal distribution with mean 3 and SD 1.
 - (b) Chi-square distribution with df 1.
 - (c) Lognormal distribution with mean 1 and SD 0.80.

We used 5000 simulation replications and 1500 bootstrap samples for each n . 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), the coverage probability will be exact or close to $(1 - \alpha)$. So the coverage probability is a useful criterion for evaluating the confidence interval. Another criterion is the width of

the confidence interval. A shorter length width gives a better confidence interval. It is obvious that when coverage probability is the same, a smaller width indicates that the method is appropriate for the specific sample. In order to compare the performance of the various intervals, the following criteria are considered: coverage probabilities (below, cover and above), mean and SD of the widths of the resulting confidence intervals. The

Table 4.1: Coverage properties for $N(3,1)$ distribution with skewness 0.

Approaches	Measuring Criteria	Sample Sizes						
		5	10	20	30	50	70	100
Exact	Below rate	0.0230	0.0242	0.0230	0.0214	0.0250	0.0234	0.0276
	Cover rate	0.9494	0.9492	0.9524	0.9554	0.9512	0.9544	0.9484
	Over rate	0.0276	0.0266	0.0246	0.0232	0.0238	0.0222	0.0240
	Mean width	2.1400	1.1113	0.6934	0.5427	0.4075	0.3413	0.2826
	SD width	0.7807	0.2631	0.1129	0.0711	0.0412	0.0289	0.0202
Robust	Below rate	0.1142	0.0634	0.0498	0.0376	0.0318	0.0306	0.0320
	Cover rate	0.8098	0.9041	0.9160	0.9276	0.9388	0.9438	0.9402
	Over rate	0.0760	0.0352	0.0342	0.0348	0.0294	0.0256	0.0278
	Mean width	2.6107	1.2849	0.7810	0.6058	0.4528	0.3786	0.3134
	SD width	1.4262	0.3828	0.1545	0.0940	0.0528	0.0367	0.0252
Bonett	Below rate	0.4352	0.1918	0.0890	0.0594	0.0372	0.0290	0.0274
	Cover rate	0.5292	0.7972	0.9060	0.9362	0.9582	0.9674	0.9678
	Over rate	0.0356	0.0110	0.0050	0.0044	0.0046	0.0036	0.0048
	Mean width	0.5073	0.6494	0.5901	0.5238	0.4327	0.3776	0.3222
	SD width	0.1992	0.1835	0.1264	0.0954	0.0621	0.0464	0.0339
Non-para Bootstrap	Below rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.7728	0.7930	0.9998	1.0000	1.0000	1.0000	1.0000
	Over rate	0.2272	0.2070	0.0002	0.0000	0.0000	0.0000	0.0000
	Mean width	0.9662	0.7504	0.5589	0.4699	0.3727	0.3197	0.2695
	SD width	0.3723	0.2312	0.1399	0.1035	0.0666	0.0497	0.0365
Parametric Bootstrap	Below rate	0.0266	0.0000	0.0006	0.0024	0.0124	0.0036	0.0092
	Cover rate	0.8472	0.9624	0.5398	0.8606	0.9730	0.5292	0.9758
	Over rate	0.1262	0.0376	0.4596	0.1370	0.0146	0.4672	0.0150
	Mean width	1.9244	3.4834	0.8099	0.6397	0.4608	0.2805	0.3289
	SD width	0.7021	0.8246	0.1318	0.0838	0.0466	0.0238	0.0235
Robust Bootstrap	Below rate	0.1104	0.0614	0.0470	0.0324	0.0290	0.0338	0.0328
	Cover rate	0.7560	0.9002	0.9154	0.9290	0.9490	0.9372	0.9394
	Over rate	0.1336	0.0384	0.0376	0.0386	0.0220	0.0290	0.0278
	Mean width	2.7531	1.3338	0.7980	0.6210	0.4679	0.3667	0.3109
	SD width	1.5039	0.3974	0.1578	0.0964	0.0546	0.0355	0.0250
CT Bootstrap	Below rate	0.0654	0.0368	0.0010	0.0012	0.0002	0.0000	0.0000
	Cover rate	0.9346	0.9632	0.9990	0.9988	0.9998	1.0000	1.0000
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	0.9073	0.5657	0.3581	0.2655	0.1679	0.1343	0.1032
	SD width	0.3310	0.1339	0.0583	0.0348	0.0170	0.0114	0.0074

below and above rates of a confidence interval is the fraction out of 5000 samples that resulted in an interval that lies entirely above and below the true value of the population mean. The coverage probability is found as the sum of the lower rate and upper rate and then subtracted from total probability 1. Simulation results are tabulated in Tables 4.1, 4.2 and 4.3 for normal, chi-square and log-normal distributions respectively. For

Table 4.2: Coverage properties for χ_1^2 distribution with skewness 2.83.

Approaches	Measuring Criteria	Sample Sizes						
		5	10	20	30	50	70	100
Exact	Below rate	0.2062	0.2494	0.2694	0.2762	0.2550	0.2562	0.2532
	Cover rate	0.7084	0.6374	0.5922	0.5722	0.5714	0.5766	0.5640
	Over rate	0.0854	0.1132	0.1384	0.1516	0.1736	0.1672	0.1828
	Mean width	2.6007	1.4303	0.9242	0.7355	0.5671	0.4733	0.3952
	SD width	1.9051	0.7663	0.3623	0.2407	0.1451	0.1028	0.0719
Robust	Below rate	0.0260	0.0074	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.5410	0.6018	0.8746	0.9672	0.9972	0.9994	1.0000
	Over rate	0.4330	0.3908	0.1252	0.0328	0.0028	0.0006	0.0000
	Mean width	1.8389	0.8986	0.4863	0.3592	0.2632	0.2132	0.1747
	SD width	1.6597	0.5313	0.2027	0.1243	0.0693	0.0484	0.0335
Bonett	Below rate	0.2888	0.4280	0.2906	0.2358	0.1634	0.1406	0.1160
	Cover rate	0.6266	0.5186	0.6796	0.7388	0.8158	0.8436	0.8680
	Over rate	0.0846	0.0534	0.0298	0.0254	0.0208	0.0158	0.0160
	Mean width	0.6715	1.0754	1.1966	1.1636	1.0610	0.9628	0.8628
	SD width	0.5506	0.7581	0.7435	0.6528	0.5057	0.4333	0.3485
Non-para Bootstrap	Below rate	0.0110	0.0052	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.9876	0.9948	0.9972	0.9966	0.9924	0.9454	0.6380
	Over rate	0.0014	0.0000	0.0026	0.0034	0.0076	0.0546	0.3620
	Mean width	1.2724	1.3299	1.2379	1.1433	1.0177	0.9143	0.8157
	SD width	1.0062	0.9391	0.7437	0.6133	0.4682	0.3920	0.3163
Parametric Bootstrap	Below rate	0.0472	0.0600	0.0030	0.0052	0.0002	0.0002	0.0400
	Cover rate	0.6048	0.6202	0.6954	0.6602	0.9482	0.5050	0.7466
	Over rate	0.3480	0.3198	0.3016	0.3346	0.0516	0.4948	0.2134
	Mean width	5.9921	2.0968	3.1846	1.7225	1.5354	1.9256	0.7131
	SD width	4.3895	1.1233	1.2484	0.5638	0.3927	0.4184	0.1298
Robust Bootstrap	Below rate	0.0280	0.0082	0.0006	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.5300	0.5798	0.8400	0.9666	0.9964	0.9992	1.0000
	Over rate	0.4420	0.4120	0.1594	0.0334	0.0036	0.0008	0.0000
	Mean width	1.5955	0.8422	0.4538	0.3538	0.2615	0.2043	0.1739
	SD width	1.4400	0.4979	0.1892	0.1224	0.0689	0.0464	0.0333
CT Bootstrap	Below rate	0.0968	0.1190	0.1186	0.1674	0.1218	0.2750	0.1414
	Cover rate	0.9032	0.8810	0.8814	0.8326	0.8782	0.7250	0.8586
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	1.3284	0.7912	0.4827	0.3365	0.2444	0.1814	0.1402
	SD width	0.9731	0.4239	0.1892	0.1102	0.0625	0.0394	0.0255

all simulated distributions, we also provided the coverage probabilities in Table 4.4. For more on the simulation techniques, we refer Baklizi and Kibria (2009) and Banik and Kibria (2010a,b) and references therein.

Table 4.3: Coverage properties for the Lognormal (1.0,0.80) distribution with skewness 3.69.

Approaches	Measuring Criteria	Sample Sizes						
		5	10	20	30	50	70	100
Exact	Below rate	0.1710	0.2600	0.3038	0.3280	0.3266	0.3494	0.3498
	Cover rate	0.7568	0.6460	0.5640	0.5266	0.5124	0.4784	0.4720
	Over rate	0.0722	0.0940	0.1322	0.1454	0.1610	0.1722	0.1782
	Mean width	6.4430	3.4512	2.2911	1.8185	1.3942	1.1700	0.9744
	SD width	4.8451	1.9012	1.0438	0.6992	0.4350	0.3197	0.2311
Robust	Below rate	0.0268	0.0040	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.7252	0.5714	0.7442	0.9086	0.9900	0.9986	1.0000
	Over rate	0.2480	0.4246	0.2556	0.0914	0.0100	0.0014	0.0000
	Mean width	5.6323	2.6675	1.5611	1.1971	0.8873	0.7353	0.6064
	SD width	3.9645	1.0909	0.4237	0.2594	0.1436	0.1011	0.0684
Bonett	Below rate	0.2620	0.4720	0.3510	0.3030	0.2406	0.2166	0.1816
	Cover rate	0.6678	0.4852	0.6118	0.6688	0.7382	0.7660	0.8046
	Over rate	0.0702	0.0428	0.0372	0.0282	0.0212	0.0174	0.0138
	Mean width	1.6464	2.5439	2.9784	2.9380	2.7665	2.5728	2.3409
	SD width	1.4319	1.9865	2.3489	2.1990	1.9159	1.6832	1.5210
Non-para Bootstrap	Below rate	0.0312	0.1258	0.0200	0.0086	0.0078	0.0000	0.0002
	Cover rate	0.9688	0.8742	0.9800	0.9914	0.9922	1.0000	0.9998
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	3.0850	3.0532	3.0013	2.8159	2.5739	2.3781	2.1623
	SD width	2.5675	2.3907	2.2831	1.9829	1.6510	1.4309	1.2690
Parametric Bootstrap	Below rate	0.0082	0.0052	0.0038	0.2056	0.0014	0.0674	0.0056
	Cover rate	0.8972	0.7640	0.7832	0.6352	0.6152	0.8756	0.9782
	Over rate	0.0946	0.2308	0.2130	0.1592	0.3834	0.0570	0.0162
	Mean width	20.3072	10.9107	6.9319	2.2693	4.2257	2.7495	3.6072
	SD width	15.2710	6.0106	3.1581	0.8726	1.3183	0.7512	0.8555
Robust Bootstrap	Below rate	0.0428	0.0062	0.0004	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.8148	0.6230	0.7108	0.8904	0.9870	0.9984	1.0000
	Over rate	0.1424	0.3708	0.2888	0.1096	0.0130	0.0016	0.0000
	Mean width	3.5183	2.2211	1.4099	1.1484	0.8495	0.7064	0.5978
	SD width	2.4765	0.9083	0.3827	0.2488	0.1375	0.0971	0.0674
CT Bootstrap	Below rate	0.3916	0.4424	0.2970	0.3686	0.1390	0.1400	0.0182
	Cover rate	0.6084	0.5576	0.7030	0.6314	0.8610	0.8600	0.9818
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	2.4976	1.6585	1.0762	0.8497	0.5779	0.4639	0.3291
	SD width	1.8782	0.9137	0.4903	0.3267	0.1803	0.1267	0.0781

Table 4.4: Coverage Probabilities for All Approaches and Distributions.

Distribution	Approaches	Sample Size (n)						
		5	10	20	30	50	70	100
Normal	Exact	0.9494	0.9492	0.9525	0.9554	0.9512	0.9544	0.9484
	Robust	0.8098	0.9041	0.9160	0.9276	0.9388	0.9438	0.9402
	Bonett	0.5292	0.7972	0.9060	0.9362	0.9582	0.9674	0.9678
	NP-Boot	0.7728	0.7930	0.9998	1.0000	1.0000	1.0000	1.0000
	Par-Boot	0.8472	0.9624	0.5398	0.8606	0.9730	0.5292	0.9758
	Robust Boot	0.7560	0.9002	0.9154	0.9290	0.9490	0.9372	0.9394
	CT Boot	0.9346	0.9632	0.9990	0.9988	0.9998	1.0000	1.0000
χ_1^2	Exact	0.7084	0.6374	0.5922	0.5722	0.5714	0.5766	0.5640
	Robust	0.5410	0.6018	0.8746	0.9672	0.9972	0.9994	1.0000
	Bonett	0.6266	0.5186	0.6796	0.7388	0.8158	0.8436	0.8680
	NP-Boot	0.9876	0.9948	0.9972	0.9966	0.9924	0.9454	0.6380
	Par-Boot	0.6048	0.6202	0.6954	0.6602	0.9482	0.5050	0.7466
	Robust Boot	0.5300	0.5798	0.8400	0.9666	0.9964	0.9992	1.0000
	CT Boot	0.9032	0.8810	0.8814	0.8326	0.8782	0.7250	0.8586
Log-normal	Exact	0.7568	0.6460	0.5640	0.5266	0.5124	0.4784	0.4720
	Robust	0.7252	0.5714	0.7442	0.9086	0.9900	0.9986	1.0000
	Bonett	0.6678	0.4852	0.6118	0.6688	0.7382	0.7660	0.8046
	NP-Boot	0.9688	0.8742	0.9800	0.9914	0.9922	1.0000	0.9998
	Par-Boot	0.8972	0.7640	0.7832	0.6352	0.6152	0.8756	0.9782
	Robust Boot	0.8148	0.6230	0.7108	0.8904	0.9870	0.9984	1.0000
	CT Boot	0.6084	0.5576	0.7030	0.6314	0.8610	0.8600	0.9818

4.2. Results discussion

The MATLAB programming language was used to run the simulation and to make the necessary tables. The performance of the selected techniques in Section 3 for normal distribution is examined first and the simulated results are tabulated in Table 4.1.

The results in Table 4.1 suggested that when sampling from a normal distribution, the performance of the estimators do not differ greatly. However, for small sample sizes, the exact method has coverage probability close to 0.95, followed by CT Bootstrap, the proposed robust method, the robust bootstrap method and Bonnet performed the worse. The parametric bootstrap method performed better than the non-parametric bootstrap method for all sample sizes. When measuring criterion is average width, it is observed that the CT bootstrap interval performed well as compare to others, followed by Bonett and the non-parametric interval. The average width of the exact method is observed closed to the proposed robust method.

The next simulation compares the performance of the proposed intervals for a variety of non-normal distributions. Results are depicted for chi-square and log-normal in Tables 4.2 and 4.3 respectively. The results in Table 4.2 and Table 4.3 suggested that when sampling from a skewed distribution, the proposed robust method, Non-

parametric bootstrap, robust bootstrap performed better compared to others in the sense of coverage probability and average width. It is clear that the proposed robust method is superior to the exact method when the data are from a non-normal population.

When the data are from normal distributions or sample sizes are large, the exact method would be considered as it is easy to compute and has coverage probability close to the nominal size compared to the rest. Since in real life the distributions of the data are unknown or do not follow the normality assumption for most of the cases, our proposed robust confidence interval would be recommended, as it does not required any distributional assumption and is easy to compute compared to the bootstrap methods. Even though some of the bootstrap methods are as good as our proposed robust method, it is not advisable to use them as they are very computer intensive. However, for a computer expert researcher, the non-parametric bootstrap method can be recommended.

5. Applications to real data

In this section, we will present some real life examples to illustrate the application and the performance of the selected intervals.

5.1. Example 1

This example is taken from Hogg and Tanis (2001, page 359). 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 observations:

481, 537, 513, 583, 453, 510, 570, 500, 457, 555

618, 327, 350, 643, 499, 421, 505, 637, 599, 392

The sample mean, standard deviation and skewness of data are 507.5, 89.75 and -0.3804 and respectively. Shapiro-Wilk Normality Test ($W = 0.9667$, p -value = 0.6834) suggested that the data follow a normal distribution. The resulting 95% confidence intervals for different methods and the corresponding confidence widths are given in Table 5.1.

From Table 5.1, we observed that when the data under consideration has a normal distribution, the confidence intervals widths for both exact and robust methods are approximately the same, but as expected, the exact method provided the shortest intervals widths among the two methods. From the above Table, we observed that the non-parametric bootstrap interval has the narrowest width followed by the Bonett interval. It is noted that the CT bootstrap has the widest width.

Table 5.1: The 95% Confidence Intervals for the Butterfat Data.

Method	95% Confidence Interval	Width
Exact	(68.255,131.087)	62.832
Robust	(63.261,129.137)	65.876
Bonett	(63.910,114.789)	50.879
Non-para Bootstrap	(62.515,106.951)	44.435
Parametric Bootstrap	(74.656,130.790)	56.133
Robust Bootstrap	(74.970,155.960)	80.989
CT Bootstrap	(76.851,169.375)	92.523

5.2. Example 2

This example is taken from Weiss (2002, page 291). 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 observations:

47, 39, 62, 49, 50, 70, 59, 53, 55, 0, 65, 63, 53, 51, 50, 72, 45

The sample mean, standard deviation and skewness of these data are 51.94, 16.08 and -2.11 respectively. Shapiro-Wilk Normality Test ($W = 0.8013$, p -value = 0.0021) suggested that the data do not follow a normal distribution. The zero value may be a recording error or due to a person in the random sample who does not eat chicken for some reason (e.g., a vegetarian) and may be considered as an outlier. Now, if we remove the outlier 0 pound from the sample data, the sample mean, standard deviation and skewness of data are 55.19, 9.21 and 0.33 respectively. Shapiro-Wilk Normality Test ($W = 0.9651$, p -value = 0.7539) suggested that the data follow a normal distribution. The resulting 95% confidence intervals for different methods and the corresponding confidence widths based on the two types of data are given in Table 5.2.

From Table 5.2, we observed that the non-parametric bootstrap interval has the narrowest width followed by the Bonett interval. It is also noted that the CT bootstrap has the widest width than others. From Table 5.2 it is also observed that when the outlier is removed from the data, then the confidence interval for exact and robust methods are very similar and approximately have the same interval width, although the robust interval width is slightly shorter. The value of the outlier does not affect so much the proposed robust confidence interval and therefore the exact confidence interval for the population standard deviation, σ , should be avoided in the presence of outliers. In general, an outlier should not be removed without careful consideration. Simply removing an outlier because it is an outlier is unacceptable statistical practice.

Table 5.2

Method	95% Confidence Interval		Width	
	Original Data	Abridged Data	Original Data	Abridged Data
Exact	(11.978 , 24.478)	(7.485 , 16.330)	12.50	8.845
Robust	(6.804, 14.255)	(5.193, 11.635)	7.451	6.442
Bonett	(8.798,26.362)	(6.396,11.813)	17.564	5.416
Non-para Bootstrap	(6.789,23.977)	(6.158,11.470)	17.188	5.311
Parametric Bootstrap	(10.676,37.332)	(7.506,13.971)	26.655	6.464
Robust Bootstrap	(7.800,17.840)	(6.398,13.991)	10.040	7.593
CT Bootstrap	(16.374,40.284)	(9.737, 18.698)	23.910	8.961

Also, our result if we had blindly finding a confidence interval without first examining the data would have been invalid and misleading. In this case we can use the proposed robust confidence interval which is resistant to outliers.

6. Concluding remarks

This paper proposes an approximate confidence interval for estimating the population standard deviation, σ , based on a robust estimator and compares its performance with other proposed intervals. A simulation study has been conducted to compare the performance of the estimators, and shows that the proposed robust confidence interval for all distributions considered performs well and had a good coverage probability compared to the exact method especially for non-normal distributions. It appears that the sample size (n) has significant effect on the proposed confidence interval. We observed that if the population is really normal, the exact confidence interval for the population standard deviation, σ , performs slightly better than the proposed robust method. If the distribution is highly skewed, the coverage probability of the proposed robust method becomes close to $1 - \alpha$ and improves as the sample size increases. Actually, if the population is really non-normal, the exact confidence interval for the population standard deviation, σ , can be arbitrarily bad. A single outlier makes it worse than useless. To illustrate the findings of the paper we considered some real life examples which also supported the simulation study to some extent. Finally, among all proposed intervals, the robust, exact, non-parametric boot strap and CT bootstrap intervals are promising and can be recommended for the practitioners. However, both exact and proposed robust intervals are easy to compute and are not computer intensive.

Acknowledgements

The authors are thankful to the referees and editor of the journal for their valuable and constructive comments which certainly improved the quality and presentation of the paper. This paper was completed while Dr. B. M. Golam Kibia was on sabbatical leave (2010-2011). He is grateful to Florida International University for awarding him the sabbatical leave, which gave him excellent research facilities.

References

- Baklizi, A. and Kibria, B. M. G. (2009). One and two sample confidence intervals for estimating the mean of skewed populations: an empirical comparative study. *Journal of Applied Statistics*, 36, 601–609.
- Banik, S. and Kibria, B. M. G. (2010a). Comparison of some parametric and nonparametric type one sample confidence intervals for estimating the mean of a positively skewed distribution. *Communications in Statistics-Simulation and Computation*, 39, 361–389.
- Banik, S. and Kibria, B. M. G. (2010b). Comparison of some test statistics for testing the mean of a right skewed distribution. *Journal of Statistical Theory and Applications*, 9, 77–90.
- Barnett, V. (1988). Outliers and order statistics. *Communication in Statistics: Theory and Methods*, 17, 2109–2118.
- Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden Day, San Francisco, USA.
- Bonett, D. G. (2006). Approximate confidence interval for standard deviation of nonnormal distributions. *Computational Statistics and Data Analysis*, 50, 775–782.
- Clark, B. R. (1994). Empirical evidence for adaptive confidence interval and identification of outliers using methods of trimming. *Australian Journal of Statistics*, 36(1): 45–58.
- Cojbasic, V. and Tomovic, A. (2007). Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples. *Computational Statistics & Data Analysis*, 51, 5562–5578.
- Croux, C. and P. J. Rousseeuw (1992). Time efficient algorithms for two highly robust estimators of scale. *Computational Statistics*, Volume 1, eds. Y. Dodge and J. Whittaker, Heidelberg: Physika-Verlag, 411–428.
- Daniel, W. W. (1990). *Applied Nonparametric Statistics*, second edition, Duxbury, USA.
- De Wet, T. and van Wyk, W.J. (1979). Efficiency and robustness of Hogg's adaptive trimmed means. *Communications in Statistics, Theory and Methods*, 8, 117–128.
- Dixon, W. J. and Tukey, J. W. (1968). Approximate behavior of the distribution of winsorized t (trimming/winsorizing 2). *Technometrics*, 10, 83–93.
- Gorard, S. (2004). Revisiting a 90-year-old debate: the advantages of the mean deviation. *Paper presented at the British Educational Research Association Annual Conference*. University of Manchester, UK.
- Gross, A. M. (1973). A Monte Carlo swindle for estimators of locations. *Applied Statistics*, 22, 347–353.
- Gross, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *American Statistical Associations*, 71, 409–416.
- Hogg, R. V. and E. A. Tanis (2001). *Probability and Statistical Inference*. Sixth edition, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.
- Huber, P. J. (1964). Estimation of a location parameter. *The Annals of Mathematical Statistics*, 35, 73–101.
- Kim, Seong-Ju (1992). The metrically trimmed mean as a robust estimator of location. *The Annals of Statistics*, 20, 1534–1547.

- Lehman, E. L. (1986). *Testing Statistical Hypothesis*, John Wiley, New York.
- Longford, N. T. (2010). Small-sample inference about variance and its transformations, *SORT*, 34, 1, 3–20.
- Mosteller and Tukey (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, USA.
- Rousseuw, P. J. (1991). Tutorial to robust statistics. *Journal of Chemometrics*, 5, 1–20.
- Rousseuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273–1283.
- Scheffe, H. (1959). *The Analysis of Variance*. Wiley, New York.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In Olkin, I., *et al.* (Eds.) *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*. Stanford: Stanford University Press. 448–485.
- Tukey, J. W. and McLaughlin, D. H. (1963). Less confidence and significance procedures for location based on a single sample: trimming / winsorizing 1, *Sankhya A*, 25, 331–352.
- Weiss, A. N. (2002). *Introductory Statistics*, Sixth edition, Addison-Wesley, USA.

