

Confidence Intervals for the Mean of Non-Normal Distribution: Transform or Not to Transform

Jolynn Pek¹, Augustine C. M. Wong², Octavia C. Y. Wong¹

¹Department of Psychology, York University, Toronto, Canada

²Department of Mathematics and Statistics, York University, Toronto, Canada

Email: pek@yorku.ca, august@yorku.ca, owong3@my.yorku.ca

How to cite this paper: Pek, J., Wong, A.C.M. and Wong, O.C.Y. (2017) Confidence Intervals for the Mean of Non-Normal Distribution: Transform or Not to Transform. *Open Journal of Statistics*, 7, 405-421. <https://doi.org/10.4236/ojs.2017.73029>

Received: April 17, 2017

Accepted: June 5, 2017

Published: June 8, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In many areas of applied statistics, confidence intervals for the mean of the population are of interest. Confidence intervals are typically constructed assuming normality although non-normally distributed data are a common occurrence in practice. Given a large enough sample size, confidence intervals for the mean can be constructed by applying the Central Limit Theorem or by the bootstrap method. Another commonly used method in practice is the back-transformation method, which takes on the following three steps. First, apply a transformation to the data such that the transformed data are normally distributed. Second, obtain confidence intervals for the transformed mean in the usual manner, which assumes normality. Third, apply the back-transformation to obtain confidence intervals for the mean of the original, non-transformed distribution. The parametric Wald method and a small sample likelihood-based third order method, which can address non-normality, are also reviewed in this paper. Our simulation results suggest that common approaches such as back-transformation give erroneous and misleading results even when the sample size is large. However, the likelihood-based third order method gives extremely accurate results even when the sample size is small.

Keywords

Back-Transformation, Bootstrap, Central Limit Theorem, Delta Method, Maximum Likelihood Estimate, Third Order Method

1. Introduction

In the last two decades, there has been a push in psychological science to im-

prove research reporting with an emphasis on effect size and confidence interval reporting (see American Education Research Association [1]; Cumming [2]; Wilkinson and the Task Force for Statistical Inference [3]). Effect sizes communicate the magnitude and direction of a practically important effect (e.g., treatment decreased depression scores by 13%), and confidence intervals communicate this effect’s estimate precision. The importance of confidence intervals, their basic construction, and interpretation have thus been the focus of several influential pedagogical articles (e.g., see Cumming and Fidler [4]; Cumming and Finch [5]; Greenland *et al.* [6]).

Most, if not all, modern introductory statistics textbooks review and describe the construction of confidence intervals (e.g., see Moore *et al.* [7]). Let (x_1, \dots, x_n) be a sample obtained from a normally distributed population with mean μ and variance σ^2 . Then a $(1 - \alpha)100\%$ confidence interval for μ can be calculated by

$$\left(\bar{x} - t^* \frac{s_x}{\sqrt{n}}, \bar{x} + t^* \frac{s_x}{\sqrt{n}} \right) \tag{1}$$

where

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1},$$

and t^* is the $(1 - \alpha/2)100^{th}$ percentile of the Student t distribution with $(n - 1)$ degrees of freedom. Moreover, when the sample size n is large (usually stated n is larger than 30), then the $(1 - \alpha)100\%$ confidence interval for μ can still be obtained from (1) except that we replaced t^* by z^* , which is the $(1 - \alpha/2)100^{th}$ percentile of the standard normal distribution.

The fundamental assumption underlying the construction of this confidence interval is that the data are normally distributed. However, collected data are usually non-normally distributed in practice (for examples in psychology, see Cain *et al.* [8]; Micceri [9]). In public health research, Bland and Altman [10] reported that serum triglyceride measurements are distributed with positive skewness. In biology, McDonald [11] reported that the number of Eastern mudminnows in Maryland streams are non-normally distributed.

In this paper, we compare various methods for constructing confidence intervals when data are non-normally distributed. Three of the most popular and commonly used methods are the method based on the Central Limit Theorem, the bootstrap method, and the back-transformation method, which are reviewed in Section 2. The parametric based Wald method and likelihood-based third order method are also discussed in Section 2. Note that the popular back-transformation method requires the existence of a transformation such that the transformed data are normally distributed. The selection of such a transformation by the Box-Cox transformation and the Tukey’s ladder of power transformation are briefly discussed in Section 2. Two empirical examples are presented in Section 3 to illustrate that confidence intervals based on the different methods discussed in Section 2 can be vastly different. Simulation results are presented in

Section 4 to compare the accuracy of the methods discussed in this paper and illustrated that the likelihood-based third order method gives extremely accurate coverage probability even when the sample size is small, the Wald method, the Central Limit Theorem method and then bootstrap method all performed poorly when sample size is small but the performance increases when the sample size increases, and the popular back-transformation method should not be used because it does not construct the confidence interval for the correct parameter. Finally, some concluding remarks are given in Section 5.

2. Methodology

This section reviews four commonly used methods, namely the Central Limit Theorem, bootstrap, back-transformation, and Wald for obtaining a confidence interval for the mean of a non-normal distribution. A very accurate likelihood-based method is also introduced in this section.

2.1. Central Limit Theorem Method

Let (x_1, \dots, x_n) be a sample from a non-normal distribution with mean ψ . When the sample size n is large, the Central Limit Theorem gives

$$\frac{\bar{X} - \psi}{\sqrt{\text{var}(\bar{X})}} \rightarrow \mathcal{N}(0,1)$$

where $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ and $\text{var}(\bar{X}) = \frac{\text{var}(X)}{n}$. Since \bar{x} and s_x^2 are the unbiased estimates of ψ and $\text{var}(X)$ respectively; by the Central Limit Theorem, an approximate $(1-\alpha)100\%$ confidence interval for ψ is

$$\left(\bar{x} - z^* \frac{s_x}{\sqrt{n}}, \bar{x} + z^* \frac{s_x}{\sqrt{n}} \right) \quad (2)$$

where z^* is the $(1-\alpha/2)100^{\text{th}}$ percentile of the standard normal distribution.

2.2. Bootstrap Method

The bootstrap method is a popular non-parametric method, which does not require any distributional assumptions. Efron and Tibshirani [12] provide a detailed review of the bootstrap method. The following is an algorithmic approach of obtaining a $(1-\alpha)100\%$ percentile bootstrap confidence interval for the population mean, ψ .

Sample: (x_1, \dots, x_n)

Step 1: Resample the observed sample with replacement and calculate the sample mean for this bootstrap sample.

Step 2: Repeat Step 1 B times, where, typically, $B \geq 200$.

Step 3: Sort the B bootstrapped sample means; the $(\alpha/2)100^{\text{th}}$ and $(1-\alpha/2)100^{\text{th}}$ percentiles give the $(1-\alpha)100\%$ percentile bootstrap confidence interval for the population mean.

Note that as with the Central Limit Theorem method, the bootstrap method

requires the observed sample size to be large so as to be representative of the population.

2.3. Back-Transformation Method

Recall that X is a non-normally distributed random variable with mean ψ . Assume there exists a transformation $g(\cdot)$ such that $Y = g(X)$ is normally distributed with mean μ and variance σ^2 . By the delta method,

$$\psi = E(X) = E[g^{-1}(Y)] \approx g^{-1}(\mu),$$

and an approximate $(1 - \alpha)100\%$ confidence interval for ψ from (1) is

$$\left(g^{-1}\left(\bar{y} - t^* \frac{s_y}{\sqrt{n}}\right), g^{-1}\left(\bar{y} + t^* \frac{s_y}{\sqrt{n}}\right) \right). \tag{3}$$

It is important to note that (3) could be misleading because $g^{-1}(\mu)$ can be very different from ψ . For example, if X follows a lognormal (μ, σ^2) distribution, then $Y = \log(X)$ is distributed as $\mathcal{N}(\mu, \sigma^2)$. It follows that the delta method gives $\psi = E(X) \approx \exp(\mu)$. However, as shown in **Table 1**, $\psi = \exp(\mu + \sigma^2/2)$, which can be quite different from $\exp(\mu)$, especially when σ^2 is large. Consider another example where $Y = \sqrt{X}$ follows a $\mathcal{N}(\mu, \sigma^2)$ distribution. Here, the delta method gives $\psi \approx \mu^2$. However, **Table 1** shows that $\psi = \mu^2 + \sigma^2$, which can be quite different from μ^2 , especially when σ^2 is large.

The rest of this subsection is to provide a systematic way of choosing the transformation $g(\cdot)$. In practice, the most common simple transformations are the logarithmic transformation and square root transformation. Box and Cox [13] proposed a more complicated transformation, which requires the determination of the power parameter. Similarly, Tukey [14] suggested a ladder of power transformation, which also requires the determination of the power parameter. We review Tukey’s method in a later subsection. With an observed sample (x_1, \dots, x_n) , our aim is to obtain confidence intervals for ψ . In this paper, focus is placed on the two most commonly used transformations in practice: the logarithmic transformation and the square root transformation. Note that the transformation methods discussed can be generalized to any known transformation in theory (cf., Box-Cox or Tukey’s transformations).

When observed data are non-normally distributed, a common approach is to first apply a transformation such that the transformed data become somewhat

Table 1. Transformation and the parameter of interest.

λ	Transformation	ψ
-1	reciprocal	does not exist
0	logarithmic	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$
$\frac{1}{2}$	square root	$\mu^2 + \sigma^2$
1	none	μ

normally distributed. In the statistical literature, two very similar families of transformations are frequently discussed: the Box-Cox transformation and Tukey's ladder of power transformation. In particular, Osborne [15] gives a detailed discussion of the application of the Box-Cox transformation. Mathematically, the Box-Cox transformation and Tukey's ladder of power transformation are very similar. Because Tukey's ladder of power transformation is easier to interpret compared to the Box-Cox transformation, we review the ladder of power transformation and suggested criteria to choose an appropriate transformation to address non-normally distributed data below.

Tukey's ladder of power transformation takes the form

$$Y = X^\lambda = \begin{cases} \log(X) & \text{if } \lambda = 0 \\ X^\lambda & \text{if } \lambda \neq 0 \end{cases}$$

where λ is called the power parameter of this transformation, where λ is chosen such that Y is approximately normally distributed. Moreover, λ should be chosen such that the power parameter is easy to interpret. Note that $\lambda = 1$ is equivalent to no transformation. In practice, the popular reciprocal transformation, logarithmic transformation, and square root transformation are equivalent to $\lambda = -1, 0$ and $\frac{1}{2}$, respectively.

Table 1 presents the mean of distribution prior to transformation, ψ , in terms of μ and σ^2 based on the type of transformation used. Since ψ does not exist for the reciprocal transformation, this transformation is not considered in this paper.

With an observed sample, we suggest the choice of λ be based on three criteria:

1. de-trended normal quantile-quantile (Q-Q) plot,
2. p -value of the Shapiro-Wilk test of normality, and
3. skewness.

First, when the de-trended normal Q-Q plot deviates from the horizontal reference line which indicates identical quantiles between the data and a theoretical normal distribution, the plot suggests that the data are likely non-normally distributed. Second, simulation studies by Razali and Wah [16] illustrate that the Shapiro-Wilk test is the most powerful test among all formulated statistical tests for normality. Under the assumption of a normal distribution, the smaller the p -value associated with the Shapiro-Wilk test, the more evidence against the normality assumption. Thus, the transformation which gives the largest p -value of the Shapiro-Wilk test is associated with the least evidence against the transformed data being normally distributed. Finally, with regard to skewness, the normal distribution has skewness 0. In this vein, the transformation which results in a skewness value closest to 0 is most symmetric and would be the preferred transformation.

2.4. Wald Method

As in the previous subsection, we assume that X be a non-normally distributed

random variable with mean ψ and there exists a transformation $g(\cdot)$ such that $Y = g(X)$ is normally distributed with mean μ and variance σ^2 . Moreover, $\psi = \psi(\mu, \sigma^2)$.

The log-likelihood function concerning Y can be written as

$$\ell(\mu, \sigma^2) = a - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \tag{4}$$

where a is an additive constant. Without loss of generality, a is set to zero hereafter. The overall maximum likelihood estimate (MLE), denoted by $(\hat{\mu}, \hat{\sigma}^2)'$, can be obtained by solving

$$\begin{aligned} \left. \frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} \right|_{\hat{\mu}, \hat{\sigma}^2} &= \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (y_i - \hat{\mu}) = 0 \\ \left. \frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} \right|_{\hat{\mu}, \hat{\sigma}^2} &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (y_i - \hat{\mu})^2 = 0. \end{aligned}$$

Hence, we have

$$\hat{\mu} = \bar{y}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 = \frac{(n-1)s_y^2}{n}.$$

The observed information matrix is the negative of the second derivatives of the log-likelihood function with respect to the parameters:

$$j(\mu, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \mu) \\ \frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \mu)^2 \end{pmatrix}.$$

The variance-covariance matrix for $(\hat{\mu}, \hat{\sigma}^2)'$ can be approximated by the inverse of Fisher's expected information matrix, $\{E[j(\mu, \sigma^2)]\}^{-1}$, which, in general, can be difficult to obtain in practice. Nevertheless, the variance-covariance matrix for $(\hat{\mu}, \hat{\sigma}^2)'$ can be approximated by the inverse of the observed information evaluated at the MLE, $j^{-1}(\hat{\mu}, \hat{\sigma}^2)$ where

$$j(\hat{\mu}, \hat{\sigma}^2) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}.$$

It is well-known that $(\hat{\mu}, \hat{\sigma}^2)'$ is asymptotically distributed as normal with mean $(\mu, \sigma^2)'$ and variance $j^{-1}(\hat{\mu}, \hat{\sigma}^2)$.

Recall that the parameter of interest is $\psi = \psi(\mu, \sigma^2)$, and we denote $\hat{\psi} = \psi(\hat{\mu}, \hat{\sigma}^2)$. By the delta method,

$$\widehat{var}(\hat{\psi}) \approx \left(\frac{\partial \ell(\hat{\mu}, \hat{\sigma}^2)}{\partial (\mu, \sigma^2)} \right) j^{-1}(\hat{\mu}, \hat{\sigma}^2) \left(\frac{\partial \ell(\hat{\mu}, \hat{\sigma}^2)}{\partial (\mu, \sigma^2)} \right)'$$

where

$$\frac{\partial \ell(\hat{\mu}, \hat{\sigma}^2)}{\partial(\mu, \sigma^2)} = \left(\begin{array}{c} \frac{\partial \ell(\hat{\mu}, \hat{\sigma}^2)}{\partial \mu} \\ \frac{\partial \ell(\hat{\mu}, \hat{\sigma}^2)}{\partial \sigma^2} \end{array} \right)_{(\hat{\mu}, \hat{\sigma}^2)}.$$

Thus, an approximate $(1 - \alpha)100\%$ confidence interval for ψ is

$$\left(\hat{\psi} - z^* \sqrt{\widehat{\text{var}}(\hat{\psi})}, \hat{\psi} + z^* \sqrt{\widehat{\text{var}}(\hat{\psi})} \right). \tag{5}$$

For the case of the logarithmic transformation (*i.e.*, Tukey’s ladder of power transformation where $\lambda = 0$), the parameter of interest is $\psi = \exp(\gamma)$, where $\gamma = \mu + \sigma^2/2$. Therefore, by the Wald method, a $(1 - \alpha)100\%$ confidence interval for γ is

$$\left(\hat{\gamma} - z^* \sqrt{\widehat{\text{var}}(\hat{\gamma})}, \hat{\gamma} + z^* \sqrt{\widehat{\text{var}}(\hat{\gamma})} \right)$$

where $\hat{\gamma} = \hat{\mu} + \hat{\sigma}^2/2$ and $\widehat{\text{var}}(\hat{\gamma}) \approx \frac{\hat{\sigma}^2}{n} + \frac{\hat{\sigma}^4}{2n}$. Thus, an approximate $(1 - \alpha)100\%$ confidence interval for ψ is

$$\left(\exp\left\{ \hat{\gamma} - z^* \sqrt{\widehat{\text{var}}(\hat{\gamma})} \right\}, \exp\left\{ \hat{\gamma} + z^* \sqrt{\widehat{\text{var}}(\hat{\gamma})} \right\} \right).$$

For the case of the square root transformation (*i.e.*, Tukey’s (1977) ladder of power transformation where $\lambda = \frac{1}{2}$), the parameter of interest is $\psi = \mu^2 + \sigma^2$. Therefore, an approximate $(1 - \alpha)100\%$ confidence interval for ψ is given by (5), where

$$\hat{\psi} = \hat{\mu}^2 + \hat{\sigma}^2 \text{ and } \widehat{\text{var}}(\hat{\psi}) = \frac{2\hat{\sigma}^2(2\hat{\mu}^2 + \hat{\sigma}^2)}{n}.$$

2.5. Likelihood-Based Third Order Method

Both the Central Limit Theorem method and Wald method have a theoretical rate of convergence of $O(n^{-1/2})$, and both the back-transformation method and the bootstrap method have no known rate of convergence. In recent years, many methods have been developed to improve the rate of convergence of existing asymptotic methods. In this subsection, we review the modified signed log-likelihood ratio method by Barndorff-Nielsen [17]. The modified signed log-likelihood ratio statistic is defined as

$$r^* = r^*(\psi) = r(\psi) - \frac{1}{r(\psi)} \log \frac{r(\psi)}{q(\psi)} \tag{6}$$

where

$$r(\psi) = \text{sign}(\hat{\psi} - \psi) \left\{ 2 \left[\ell(\hat{\mu}, \hat{\sigma}^2) - \ell(\hat{\mu}_\psi, \hat{\sigma}_\psi^2) \right] \right\}^{1/2} \tag{7}$$

is the signed log-likelihood ratio statistic, $(\hat{\mu}_\psi, \hat{\sigma}_\psi^2)$ is the constrained MLE obtained by maximizing the log-likelihood function for a given ψ value, and

$q(\psi)$ is a statistic based on the log-likelihood function given in (4). Barndorff-Nielsen [17] showed that r^* is asymptotically distributed as a standard normal distribution with a rate of convergence of $O(n^{-3/2})$. Thus, a $(1-\alpha)100\%$ confidence interval obtained based on r^* is (ψ_L, ψ_U) such that ψ_L and ψ_U satisfies $|r^*(\psi_L)| < z^*$, $|r^*(\psi_U)| < z^*$, and $\psi_L < \psi_U$.

If the model is an exponential family model and the parameter of interest ψ is a component parameter of the canonical parameter, Fraser [18] showed that $q(\psi)$ is the standardized MLE statistic. Given a general model and this idea in mind, Fraser and Reid [19] first approximate the model using an approximate tangent exponential model to obtain the locally defined canonical parameter. Then, they express the parameter of interest in terms of the locally defined canonical parameter and also derived the variance the estimated parameter of interest in this locally defined canonical parameter scale. Thus, $q(\psi)$ is the standardized MLE statistic expressed in the locally defined canonical parameter scale, and the modified signed likelihood ratio statistic can be used to obtain confidence interval for ψ . Details of this algorithmic approach of obtaining r^* is outlined below.

- Notation: $\ell(\theta)$ is the log-likelihood function;
- θ is a k-dimensional vector of parameters;
- $\varphi = \varphi(\theta)$ is a k-dimensional vector of canonical parameters for the exponential family model;
- $\psi = \psi(\theta)$ is a scalar parameter of interest;
- (x_1, \dots, x_n) is the observed data.

Aim: Inference about ψ .

Step 1: From the log-likelihood function, obtain the overall MLE, $\hat{\theta}, \hat{\psi} = \psi(\hat{\theta}), \ell(\hat{\theta})$ and $\hat{j} = j_{\theta\theta}(\hat{\theta})$ can be obtained.

Step 2: Apply the Lagrange multiplier technique to obtain the constrained MLE at $\psi = \psi_0$. More specifically, maximize

$$H(\theta, \lambda) = \ell(\theta) + \lambda(\psi(\theta) - \psi_0)$$

with respect to (θ, λ) , where λ is defined as the Lagrange multiplier. Denote the result of the maximization be $(\tilde{\theta}_{\psi_0}, \tilde{\lambda})$.

Step 3: Define the tilted log-likelihood function as

$$\tilde{\ell}(\theta) = \ell(\theta) + \tilde{\lambda}(\psi(\theta) - \psi)$$

where ψ is a fixed value. Obtain the constrained MLE either from the tilted log-likelihood function or from Step 2, $\tilde{\theta}_\psi, \ell(\tilde{\theta}_\psi) = \tilde{\ell}(\tilde{\theta}_\psi)$ and $\tilde{j}_{\theta\theta}(\tilde{\theta}_\psi)$, which is the matrix of the negative of the second derivatives of the tilted log-likelihood function.

Step 4: The signed log-likelihood ratio statistic is

$$r = \text{sgn}(\hat{\psi} - \psi) \left\{ 2 \left[\ell(\hat{\theta}) - \ell(\tilde{\theta}_\psi) \right] \right\}^{1/2}.$$

Step 5: Define

$$\chi(\theta) = \psi_\theta(\tilde{\theta}_\psi) \varphi_\theta^{-1}(\tilde{\theta}_\psi) \varphi(\theta)$$

where $\psi_\theta(\theta)$ is the first derivative of $\psi(\theta)$ with respect to θ , and $\varphi_\theta(\theta)$ is the first derivative of $\varphi(\theta)$ with respect to θ . This quantity is a recalibration of the parameter of interest ψ in the canonical parameter φ space.

Step 6: The quantity $|\chi(\hat{\theta}) - \chi(\tilde{\theta}_\psi)|$ measures the departure of $\hat{\psi}$ from ψ in φ space.

Step 7: The estimated variance for the departure in φ space is given by

$$\widehat{var}(\chi(\hat{\theta}) - \chi(\tilde{\theta}_\psi)) = \frac{\psi_\theta(\tilde{\theta}_\psi) \tilde{J}_{\theta\theta}^{-1}(\tilde{\theta}_\psi) \psi'_\theta(\tilde{\theta}_\psi) \tilde{J}_{\theta\theta}(\tilde{\theta}_\psi) \|\varphi_\theta(\tilde{\theta}_\psi)\|^{-2}}{\|j_{\theta\theta}(\hat{\theta})\| \|\varphi_\theta(\hat{\theta})\|^{-2}}.$$

Step 8: The standardized MLE departure under the φ scale is given by

$$q = \text{sign}(\hat{\psi} - \psi) \frac{|\chi(\hat{\theta}) - \chi(\tilde{\theta}_\psi)|}{\sqrt{\widehat{var}(\chi(\hat{\theta}) - \chi(\tilde{\theta}_\psi))}}.$$

Step 9: The modified signed log-likelihood ratio statistic is given by

$$r^* = r - \frac{1}{r} \log \frac{r}{q}.$$

Although the algorithm involves many steps, it can easily be implemented into algebraic or statistical software such as MATLAB, Maple and R.

3. Empirical Examples

In this section, the different methods of constructing a confidence interval about the mean of non-normally distributed data are illustrated with two empirical examples. We demonstrate that the results obtained by the methods discussed in this paper can be very different.

3.1. Example 1: Serum Triglyceride Measurements

Bland and Altman [10] considered $n = 278$ serum triglyceride measurements, which had a positively skewed data distribution with an average of 0.51 mmol/l and a standard deviation of 0.22 mmol/l. By applying a base 10 logarithm transformation to the data to obtain a less skewed distribution, the transformed distribution became bell-shaped with an average of -0.33 and a standard deviation of 0.17. By applying the Central Limit Theorem, they report a 95% confidence interval for the mean serum triglyceride measurements to be (0.48, 0.54). Using the back-transformation method, the corresponding interval is (0.45, 0.49). **Table 2** presents the 95% confidence intervals for the mean serum triglyceride

Table 2. 95% confidence interval for the mean serum triglyceride measurements.

Data	Method	95% confidence interval
Original	Central Limit Theorem	(0.48, 0.54)
Log-transformed	Back-transformation	(0.45, 0.49)
Log-transformed	Wald	(0.46, 0.51)
Log-transformed	Third order	(0.46, 0.51)

measurements for the alternative methods reviewed above Note that for this example, the bootstrap method cannot be applied because the original data set is not unavailable.

Bland and Altman [10] noted that the interval obtained by the back-transformation method is actually the 95% confidence interval for the geometric mean of serum triglyceride measurements instead of the mean serum triglyceride measurements, where the latter is the parameter of interest. Stated differently, the back-transformation method does not provide information about the focal parameter of interest (*i.e.*, the mean of the non-normal distribution). From **Table 2**, it can be observed that the results from the Central Limit Theorem method are different from those obtained by the Wald method and third order method. Additionally, the Wald method and third order method give results which agree up to the second decimal place. This observation is not surprising because these two methods theoretically converge to the same answer when the sample size goes to infinity. The only difference is that the third order method will have a faster rate of convergence than the Wald method (*i.e.*, $O(n^{-1/2})$ versus $O(n^{-3/2})$, respectively). The different rates of convergence are more formally illustrated in Section 4.

3.2. Example 2: Abundance of Eastern Mudminnows

McDonald [11] reported on data on the abundance of Eastern mudminnows in Maryland streams, which is reproduced below:

38	1	13	2	13	20	50	9	28	6	4	43
----	---	----	---	----	----	----	---	----	---	---	----

These data are non-normally distributed and McDonald [11] suggested that both the logarithmic and square root transformed data are suitable for analysis because they are more normally distributed compared to the original and other competing transformations. His final analysis makes use of the logarithmic transformed data.

Table 3 presents the 95% confidence intervals for the mean of the non-transformed distribution obtained by applying the Central Limit Theorem method

Table 3. 95% confidence intervals for the mean of the abundance of Eastern mudminnows in Maryland streams.

Data	Method	95% confidence interval
Original	Central Limit Theorem	(9.3, 28.5)
Original	Bootstrap ($B = 5000$)	(10.1, 28.3)
Log-transformed	Back-transformation	(5.0, 24.4)
Log-transformed	Wald	(9.3, 54.4)
Log-transformed	Third order	(11.1, 123.9)
Square root transformed	Back-transformation	(6.7, 26.9)
Square root transformed	Wald	(9.8, 28.0)
Square root transformed	Third order	(11.1, 31.7)

and the bootstrap method with $B = 5000$ to the original data; and the back-transformation method, Wald method, and likelihood-based third order method to both the logarithmic transformed data and square root transformed data.

The results obtained by the methods discussed in this paper are very different for different transformations. In particular, the logarithmic transformation results in a much larger upper bound of the interval compared to the square root transformation. Thus, it is essential to identify which transformation is more appropriate for a given set of data.

The de-trended normal Q-Q plots for the original data, logarithmic transformed data and square root transformed data are shown in **Figure 1**. From

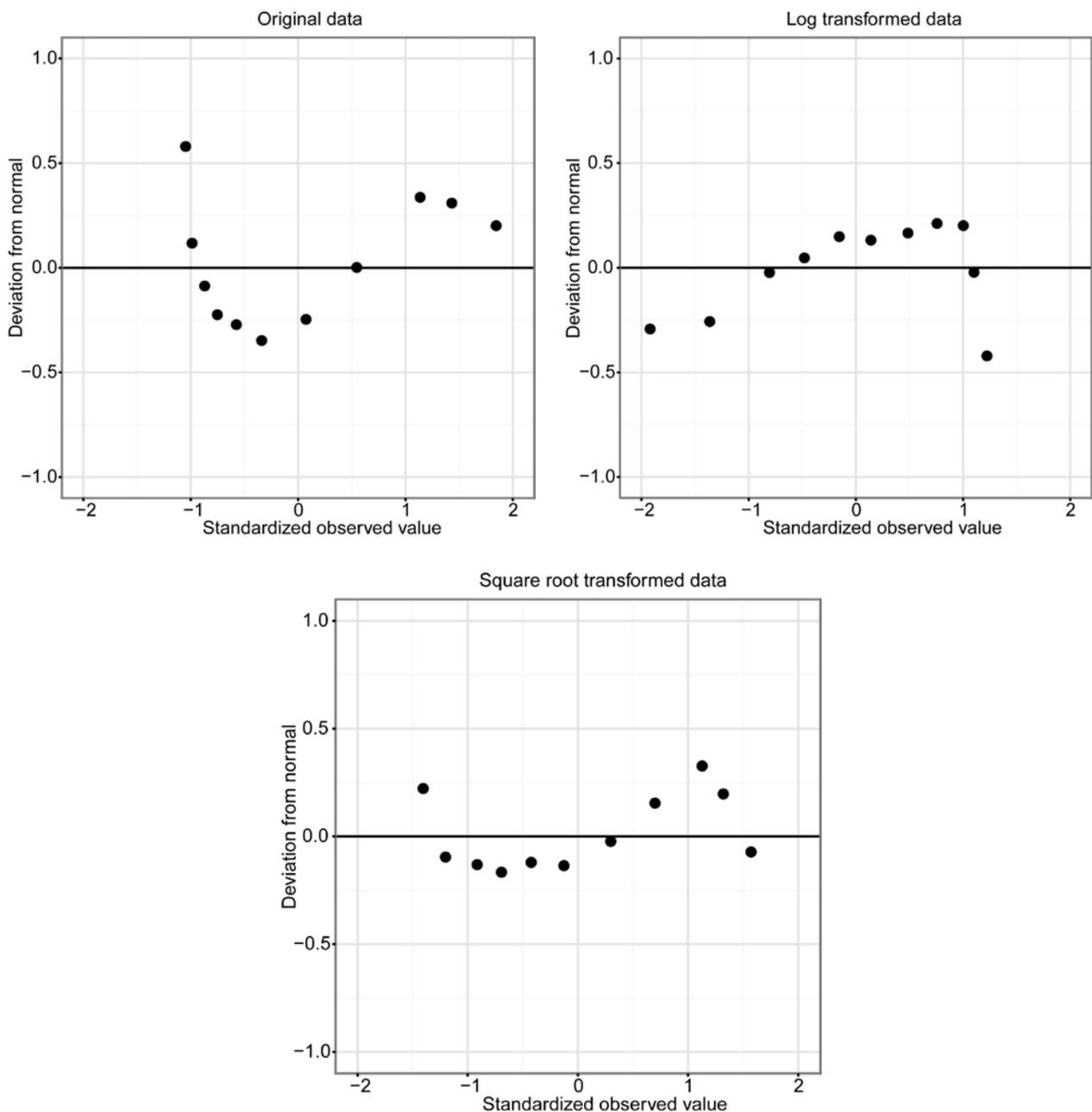


Figure 1. De-trended Normal Q-Q plots for original and transformed data of the abundance of Eastern mudminnows.

these plots, it is obvious that the original data are not normally distributed because the points deviate from the horizontal reference line, which indicates identical quantiles between the data and a theoretical normal distribution. The two sets of transformed data are more closely normally distributed because the points in the de-trended normal Q-Q plots lie more closely to the reference line relative to the original data.

The Shapiro-Wilk test on normality of the original data gives a p -value of 0.1091. The same test on the logarithmic transformed data gives a p -value of 0.5261, and it gives a p -value of 0.6479 on the square root transformed data. Consistent with the de-trended Q-Q plot, the p -values of the Shapiro-Wilk test similarly suggest that the two transformed data sets are more likely to be normally distributed. Additionally, the empirical skewness of the original data, logarithmic transformed data, and square root transformed data are 0.5864, -0.4886 , and 0.1632, respectively. These quantifications of skewness imply that the square root transformed data are more symmetrical than the original data and logarithmic transformed data. Thus, based on the criteria discussed in Section 2.3, the square root transformation is recommended for these data.

4. Simulation Study

A simulation study was carried out to compare the accuracies of the methods discussed in this paper. R code for the simulation is available to the interested reader upon request. For each (n, μ, σ) combination, we generated 10,000 samples from $\mathcal{N}(\mu, \sigma^2)$. These are our simulated transformed samples, and the non-transformed (*i.e.*, original) samples can be obtained by applying the inverse transformation to the simulated data. The transformations examined are the natural logarithm and square root. For each simulated sample, we computed a 95% confidence interval for the mean of the untransformed population from the five reviewed methods: Central Limit Theorem, bootstrap ($B = 5000$), back-transformation, Wald, and likelihood-based third order. The following quantities are recorded: the proportion of true means falling within the 95% confidence interval (coverage proportion), the proportion of true means less than the lower 95% confidence limit (lower error), and the proportion of true means greater than the upper 95% confidence limit (upper error). The nominal values of coverage, lower error, upper error, and bias are: 0.95, 0.025, and 0.025, respectively. We present only a small subset of the simulations we conducted to highlight several key points below, and other simulation results are available upon request.

Table 4 presents results with the natural logarithmic transformed data being generated from $\mathcal{N}(\mu, \sigma^2)$ and the parameter of interest is $\exp\left(\mu + \frac{\sigma^2}{2}\right)$.

It can be observed that the likelihood-based third order method outperforms the other methods especially when the sample size is small; coverage, lower and upper errors associated with the likelihood-based third order method are relatively closer to nominal rates compared to the alternative methods. Among the remaining methods, the Central Limit Theorem method and the bootstrap

Table 4. 95% coverage probability for the logarithmic transformation case.

μ	σ	n	Method	Coverage proportion	Lower error	Upper error
1	2	10	Central Limit Theorem	0.5389	0.0003	0.4608
			Bootstrap ($B = 5000$)	0.5582	0.0007	0.4411
			Back-transformation	0.1965	0.0000	0.8035
			Wald	0.8549	0.0005	0.1446
			Third order	0.9460	0.0262	0.0278
		50	Central Limit Theorem	0.6907	0.0001	0.3092
			Bootstrap ($B = 5000$)	0.7149	0.0013	0.2838
			Back-transformation	0.0000	0.0000	1.0000
			Wald	0.9303	0.0034	0.0063
			Third order	0.9501	0.0238	0.0261
		200	Central Limit Theorem	0.7757	0.0007	0.2236
			Bootstrap ($B = 5000$)	0.7968	0.0021	0.2011
			Back-transformation	0.0000	0.0000	1.0000
			Wald	0.9446	0.0134	0.0420
			Third order	0.9500	0.0261	0.0239
2	0.5	10	Central Limit Theorem	0.8902	0.0129	0.0969
			Bootstrap ($B = 5000$)	0.8793	0.0239	0.078
			Back-transformation	0.8884	0.0047	0.1069
			Wald	0.8997	0.0246	0.0757
			Third order	0.9492	0.0252	0.0256
		50	Central Limit Theorem	0.9340	0.0106	0.0554
			Bootstrap ($B = 5000$)	0.9345	0.0197	0.0458
			Back-transformation	0.5905	0.0003	0.4092
			Wald	0.9415	0.0167	0.0418
			Third order	0.9505	0.0231	0.0264
		200	Central Limit Theorem	0.9430	0.0172	0.0398
			Bootstrap ($B = 5000$)	0.9466	0.0197	0.0337
			Back-transformation	0.0591	0.0000	0.9409
			Wald	0.9452	0.0220	0.0328
			Third order	0.9481	0.0262	0.0257
3	3	10	Central Limit Theorem	0.2739	0.0000	0.7261
			Bootstrap ($B = 5000$)	0.2842	0.0000	0.7158
			Back-transformation	0.0132	0.0000	0.9868
			Wald	0.8335	0.0000	0.1665
			Third order	0.9465	0.0265	0.0270
		50	Central Limit Theorem	0.3983	0.0000	0.6017
			Bootstrap ($B = 5000$)	0.4166	0.0000	0.5834
			Back-transformation	0.0000	0.0000	1.0000
			Wald	0.9256	0.0022	0.0722
			Third order	0.9507	0.0235	0.0258
		200	Central Limit Theorem	0.4936	0.0000	0.5065
			Bootstrap ($B = 5000$)	0.5166	0.0002	0.4834
			Back-transformation	0.0000	0.0000	1.0000
			Wald	0.9418	0.0120	0.0462
			Third order	0.9504	0.0261	0.0235

method give similar results. The Wald method seems to converge faster than the Central Limit Theorem and bootstrap methods. As discussed in Section 2, the back-transformation method gives unacceptable coverage probability because it is constructing confidence intervals about a parameter that is not of interest.

It can be observed that the likelihood-based third order method outperforms the other methods especially when the sample size is small; coverage, lower and upper errors associated with the likelihood-based third order method are relatively closer to nominal rates compared to the alternative methods. Among the remaining methods, the Central Limit Theorem method and the bootstrap method give similar results. The Wald method seems to converge faster than the Central Limit Theorem and bootstrap methods. As discussed in Section 2, the back-transformation method gives unacceptable coverage probability because it is constructing confidence intervals about a parameter that is not of interest.

Table 5 presents results with the square root transformed data being generated from $\mathcal{N}(\mu, \sigma^2)$ and the parameter of interest is $\mu^2 + \sigma^2$.

Similar to results in **Table 4**, we can observe that the likelihood-based third order method outperforms the other methods, especially when sample size is small. In this context, the Central Limit Theorem method and the bootstrap method give similar results and they seem to converge faster than the Wald method. The back-transformation method continues to give unacceptable coverage probability because it constructs confidence intervals about a parameter that is not of interest.

Based on these simulation results, the Central Limit Theorem method, bootstrap method and Wald method converge slowly relative to the likelihood-based third order method. Hence, we recommend using the likelihood-based third order method to obtain confidence intervals for the mean of the non-transformed distribution after applying a normalizing transformation to non-normal data, especially for small sample sizes or large departures from normality. It is important to note that researchers should not use the popular back-transformation method despite its simplicity except for the special case where $\psi = E(X) = g^{-1}(E(Y))$.

More simulations have been performed with the same pattern of results. They are not reported here, but are available upon request.

5. Conclusion

When interest is in constructing a confidence interval about a non-normal distribution, normalizing transformations are typically recommended as a first step. This paper recommends the use of de-trended normal Q-Q plots, the largest p -value of the Shapiro-Wilk test, and quantifications of skewness on the transformed data to determine the power parameter (λ) for Tukey's ladder of power transformation when the exact transformation is unavailable. Our results strongly advise against using the popular back-transformation approach in applied work because it does not construct confidence intervals about the parameter of interest (*i.e.*, the mean of the original distribution). Instead, we recommend the

likelihood-based third order method because of its superior performance in terms of its rate of convergence, coverage, and accuracy relative to the Central Limit Theorem, bootstrap and Wald methods, even when the sample size is small or the distribution is far from being normal.

Acknowledgements

We thank the editor and the referee for their comments. This work was based on O.C.Y. Wong's undergraduate honors thesis. J. Pek was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-04301-2014) and the Early Researcher Award by the Ontario Ministry of Research and Innovation (ER15-11-004). A.C.M. Wong was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-163597-2012).

References

- [1] American Education Research Association (2006) Standards for Reporting on Empirical Social Science Research in AERA Publications. *Educational Researcher*, **35**, 33-40. <https://doi.org/10.3102/0013189X035006033>
- [2] Cumming, G. (2014) The New Statistics: Why and How. *Psychological Science*, **25**, 7-9. <https://doi.org/10.1177/0956797613504966>
- [3] Wilkinson, L. and the Task Force on Statistical Inference (1999) Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, **54**, 594-604. <https://doi.org/10.1037/0003-066X.54.8.594>
- [4] Cumming, G. and Fidler, F. (2009) Confidence Intervals: Better Answers to Better Questions. *Journal of Psychology*, **217**, 15-26. <https://doi.org/10.1027/0044-3409.217.1.15>
- [5] Cumming, G. and Finch, S. (2001) A Primer on the Understanding, Use, and Calculation of Confidence Intervals that Are Based on Central and Noncentral Distributions. *Educational and Psychological Measurement*, **61**, 532-574. <https://doi.org/10.1177/0013164401614002>
- [6] Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C. Goodman, S.N. and Altman, D.G. (2016) Statistical Tests, *P* Values, Confidence Intervals, and Power: A Guide to Misinterpretations. *European Journal of Epidemiology*, **31**, 337-350. <https://doi.org/10.1007/s10654-016-0149-3>
- [7] Moore, D.S., McCabe, G.P. and Craig, B.A. (2014) Introduction to the Practice of Statistics 8th Edition, W.H. Freeman and Company, New York.
- [8] Cain, M.K., Zhang, Z. and Yuan, K.H. (2016) Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation. *Behavior Research Methods*, 1-20. <https://doi.org/10.3758/s13428-016-0814-1>
- [9] Micceri, T. (1989) The Unicorn, the Normal Curve and Other Improbable Creatures. *Psychological Bulletin*, **105**, 156-166. <https://doi.org/10.1037/0033-2909.105.1.156>
- [10] Bland, J.M. and Altman, D.G. (1996) Transformations, means, and confidence intervals. *British Medical Journal*, **312**, 1079. <https://doi.org/10.1136/bmj.312.7038.1079>
- [11] McDonald, J.H. (2014) Handbook of Biological Statistics. Sparky House, Maryland.

- [12] Efron, B. and Tibshirani, R.J. (1994) An Introduction to the Bootstrap. Chapman and Hall, New York.
- [13] Box, G.E. and Cox, D.R. (1964) An Analysis of Transformation (with Discussion). *Journal of the Royal Statistical Society B*, **26**, 211-252.
- [14] Tukey, J.W. (1977) Exploratory Data Analysis. Addison-Wesley, Massachusetts.
- [15] Osborne, J.W. (2010) Improving Your Data Transformation: Applying the Box-Cox Transformation. *Practical Assessment, Research & Evaluation*, **15**, Article 12.
- [16] Razali, N. and Wah, Y. (2011) Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, **2**, 21-33.
- [17] Barndorff-Nielsen, O.E. (1991) Modified Signed Log Likelihood Ratio. *Biometrika*, **78**, 557-561. <https://doi.org/10.1093/biomet/78.3.557>
- [18] Fraser, D.A.S. (1991) Statistical Inference: Likelihood to Significance. *Journal of American Statistical Association*, **86**, 258-265. <https://doi.org/10.1080/01621459.1991.10475029>
- [19] Fraser, D.A.S. and Reid, N. (1995) Ancillaries and Third Order Significance. *Utilitas Mathematica*, **7**, 33-53.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojs@scirp.org

