# SMALL SAMPLE CONFIDENCE INTERVALS IN LOG SPACE BACK-TRANSFORMED FROM NORMAL SPACE

THESIS

Jason E. Tisdel

Second Lieutenant, USAF

AFIT/GAM/ENC/06-02

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

WRIGHT-PATTERSON AIR FORCE BASE, OHIO

| **Report Documentation Page** | | | | | | *Form Approved* *OMB No. 0704-0188* |
|---|---|---|---|---|---|---|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. | | | | | | |
| 1. REPORT DATE<br>**13 JUN 2006** | | 2. REPORT TYPE | | | | 3. DATES COVERED |
| 4. TITLE AND SUBTITLE<br>**Small Sample Confidence Intervals in Log Space Back-Transformed from Normal Space** | | | | | | 5a. CONTRACT NUMBER<br>5b. GRANT NUMBER<br>5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S)<br>**Jason Tisdel** | | | | | | 5d. PROJECT NUMBER<br>5e. TASK NUMBER<br>5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Air Force Institute of Technology,Graduate School of Engineering and Management,2950 Hobson Way,Wright Patterson AFB,OH,45433-7765** | | | | | | 8. PERFORMING ORGANIZATION REPORT NUMBER<br>**AFIT/GAM/ENC/06-02** |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | | | | | 10. SPONSOR/MONITOR’S ACRONYM(S)<br>11. SPONSOR/MONITOR’S REPORT NUMBER(S) |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT<br>**Approved for public release; distribution unlimited.** | | | | | | |
| 13. SUPPLEMENTARY NOTES | | | | | | |
| 14. ABSTRACT<br>**The logarithmic transformation is commonly applied to a lognormal data set to improve symmetry, homoscedasticity, and liearity. Simple to implement and easy to understand, the logarithm function transforms the original data to closely resemble a normal distribution. Analysis in the normal space provides point estimates and confidence intervals, but transformation back to the original space using the naive approach yields confidence intervals of impractical width. The naive approach offers results that are often inadequate for practical purpose. We present an alternative approach that provides improved results in the form of decreased interval width, increased confidence level, or both. Our alternative approach yields dramatically improved results at small sample sizes drawn from the right tail of the lognormal distribution.** | | | | | | |
| 15. SUBJECT TERMS | | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON | |
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | | **60** | | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

AFIT/GAM/ENC/06-02

# SMALL SAMPLE CONFIDENCE INTERVALS IN LOG SPACE BACK-TRANSFORMED FROM NORMAL SPACE

THESIS

Presented to the Faculty of the

School of Engineering and Management of the

Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Jason E. Tisdel, B.S.

Second Lieutenant, USAF

June 2006

AFIT/GAM/ENC/06-02

# SMALL SAMPLE CONFIDENCE INTERVALS IN LOG SPACE BACK-TRANSFORMED FROM NORMAL SPACE

Jason E. Tisdel, B.S.

Second Lieutenant, USAF

Approved:

| | |
|---|---|
| Edward D. White III<br>Thesis Advisor | Date |
| Mark A. Abramson<br>Committee Member | Date |
| Robert E. Neher Jr<br>Committee Member | Date |

*Abstract*

The logarithmic transformation is commonly applied to a lognormal data set to improve symmetry, homoscedasticity, and linearity. Simple to implement and easy to understand, the logarithm function transforms the original data to closely resemble a normal distribution. Analysis in the normal space provides point estimates and confidence intervals, but transformation back to the original space using the naive approach yields confidence intervals of impractical width. The naive approach applies the exponential function $e$ to the parameter of interest in normal space to obtain the corresponding parameter of interest in the original space. The naive approach offers results that are often inadequate for practical purposes. We present an alternative approach that provides improved results in the form of decreased interval width, increased confidence level, or both. Our alternative approach yields dramatically improved results at small sample sizes drawn from the right tail of the lognormal distribution.

*To my wife and son*

*Acknowledgements*

I wish to express my sincere gratitude and appreciation to Dr. White, my thesis advisor. For guidance, assistance, patience, and advice, I am grateful. I am thankful to Lt. Col Abramson and Maj. Neher, both who served as committee members, for assistance in working out the details and keeping me on my game. Without the aforementioned assistance, this thesis would never have come to fruition.

Jason E. Tisdel

*Table of Contents*


| | | Page |
|---|---|---|
| Abstract | | iv |
| Acknowledgements | | vi |
| List of Figures | | ix |
| List of Tables | | x |
| I. | Introduction | 1 |
| | 1.1 General Issue | 1 |
| | 1.2 Specific Issue | 4 |
| | 1.3 Research Objectives | 5 |
| | 1.4 Synopsis of Research | 5 |
| II. | Literature Review | 6 |
| | 2.1 Distribution of a Variate whose Logarithm is Normally Distributed | 6 |
| | 2.2 The Lognormal Distribution | 8 |
| | 2.3 Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance | 9 |
| | 2.4 Calculating Confidence Intervals for the Mean of a Lognormally Distributed Variable | 10 |
| | 2.5 Confidence Intervals for the Log-normal Mean | 13 |
| | 2.6 Confidence Intervals for the Mean of a Log-Normal Distribution | 15 |
| | 2.7 Comparison Between Prior Research and Current Research | 17 |
| III. | Methodology & Results | 18 |
| | 3.1 Background & Scope | 18 |
| | 3.2 Simulation Approach | 18 |
| | 3.3 Naive Approach | 19 |

Page


3.4 Alternative Approaches . . . . . . . . . . . . . . . . . . . . . . . 21

3.4.1 Approach 1: $\hat{y} \pm \sqrt[m]{\hat{y}}$ . . . . . . . . . . . . . . . . . . 21

3.4.2 Approach 2: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . 22

3.4.3 Approach 3: $\hat{y} \pm \frac{1}{\sqrt{\frac{n}{a}}}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . 22

3.4.4 Approach 4: $\hat{y} \pm \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . 22

3.5 Naive vs. Proposed Approach . . . . . . . . . . . . . . . . . . 25

IV. Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

Appendix A. Functions $g(t)$ and $g_1(t)$ . . . . . . . . . . . . . . . . . . . . 31

Appendix B. C factors: Land's Exact Confidence Limits . . . . . . . . . . 33

Appendix C. Results of Zhou and Gao Simulation . . . . . . . . . . . . . . 35

Appendix D. Results of Olsson Simulation . . . . . . . . . . . . . . . . . . 38

Appendix E. Simulation Code . . . . . . . . . . . . . . . . . . . . . . . . 39

Appendix F. Simulation Results . . . . . . . . . . . . . . . . . . . . . . . 44

Appendix G. Percent to Standard z-score Conversion . . . . . . . . . . . . 56

Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

Vita . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

*List of Figures*


| Figure | | Page |
|---|---|---|
| 1 | Compression of Values with the Natural Logarithm | 2 |
| 2 | Stretching of Values with the Natural Logarithm | 3 |
| 3 | Equation trendlines | 23 |
| 4 | Percentile Relative to Standard z-score | 29 |

*List of Tables*


| Table | | Page |
|---|---|---|
| 1 | Summary statistics for sample data | 15 |
| 2 | Percentiles and associated values in log space and normal space | 19 |
| 3 | Values of $t_{n-1,\alpha/2}$ where $n$ = sample size and $\alpha = 0.95$ | 20 |
| 4 | Naive Approach Simulation Results – $z$-statistic | 20 |
| 5 | Naive Approach Simulation Results – $t$-statistic | 21 |
| 6 | Values of $c$ providing superior results | 22 |
| 7 | Equation comparison at percentile $p$ | 23 |
| 8 | Percentile Dependent Equation Simulation Results | 24 |
| 9 | Small Sample and High Percentile Confidence Level and Interval Width | 26 |
| 10 | Comparison of Naive Approaches with $p$ equation | 27 |
| 11 | The function $g(t)$ | 31 |
| 12 | The function $g(t)$ | 32 |
| 13 | The function $g_1(t)$ | 32 |
| 14 | One-sided (upper) confidence limits - 15 degrees of freedom | 33 |
| 15 | One-sided (upper) confidence limits - 15 degrees of freedom | 34 |
| 16 | Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 11$ | 35 |
| 17 | Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 101$ | 36 |
| 18 | Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 400$ | 37 |
| 19 | Percent of all intervals that cover the true parameter value | 38 |
| 20 | 10th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$ | 44 |
| 21 | 25th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$ | 45 |
| 22 | 50th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$ | 46 |
| 23 | 75th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$ | 47 |
| 24 | 90th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$ | 48 |

Table Page


25 10th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . . . . . . . . 49

26 25th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . . . . . . . . 50

27 50th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . . . . . . . . 51

28 75th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . . . . . . . . 52

29 90th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ . . . . . . . . . . . . . . . . . . . . . . . . 53

30 10th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$ . . . . . . . . . . . . . . . . . . . . . . . . 54

31 25th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$ . . . . . . . . . . . . . . . . . . . . . . . . 54

32 50th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$ . . . . . . . . . . . . . . . . . . . . . . . . 54

33 75th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$ . . . . . . . . . . . . . . . . . . . . . . . . 55

34 90th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$ . . . . . . . . . . . . . . . . . . . . . . . . 55

35 Percent to Standard z-score Conversion . . . . . . . . . . . . . . . . . 56

# SMALL SAMPLE CONFIDENCE INTERVALS IN LOG SPACE BACK-TRANSFORMED FROM NORMAL SPACE

## *I. Introduction*

### *1.1 General Issue*

In regression modeling, transformations are often applied to satisfy the homogeneity of variance assumption and to linearize the fit as much as possible (Transformations to Improve Fit, 2005). In practice, the logarithmic transformation often works well to satisfy these requirements. Thus a data set may be transformed to normal space by taking the logarithm of each data point. The two most commonly applied logarithms are the natural logarithm (base $e$) and the common logarithm (base 10). It is necessary to specify which logarithm is being applied and the same logarithm must be used throughout the regression. However, switching between different kinds of logarithms involves multiplying by the proper constant.

According to Dallal, "there are three reasons why logarithms should interest us" (Dallal, 2005).

- First, many statistical techniques work best with data that are single-peaked and symmetric (symmetry).
- Second, when comparing different groups of subjects, many techniques work best when the variability is roughly the same within each group (homoscedasticity).
- Third, it is easier to describe the relationship between variables when it's approximately linear (linearity). (Dallal, 2005).

When these conditions are not true in the original data, they can often be achieved by applying a logarithmic transformation. However, since the logarithm of a non-positive number does not exist, a positive constant must be added to a data set not bounded below by zero. This allows for the use of a logarithmic transformation but shifts the central tendency of the data. Once again, the specific logarithm used is not as important as maintaining the same logarithm throughout the regression, as it is easy to change between logarithms.

A logarithmic transformation produces from a skewed right distribution, a distribution in which “the left tail (the smaller values) is tightly packed together and the right tail (the larger values) is widely spread apart,...a data set that is closer to symmetric” (Simon, 2005). Symmetry is accomplished by compressing the upper tail of the distribution while stretching out the lower end (Dallal, 2005). The logarithm function compresses together large data values (values greater than 1) and stretches small values apart (values less than 1). The further the data points are from 1, the greater the effect of the logarithm function. The compression and stretching of the logarithm only have a significant impact with data having a wide range, i.e. the maximum value is at least three times larger than the minimum value (Simon, 2005). The compression and stretching of values are demonstrated with the natural logarithm in Figures 1 and 2.

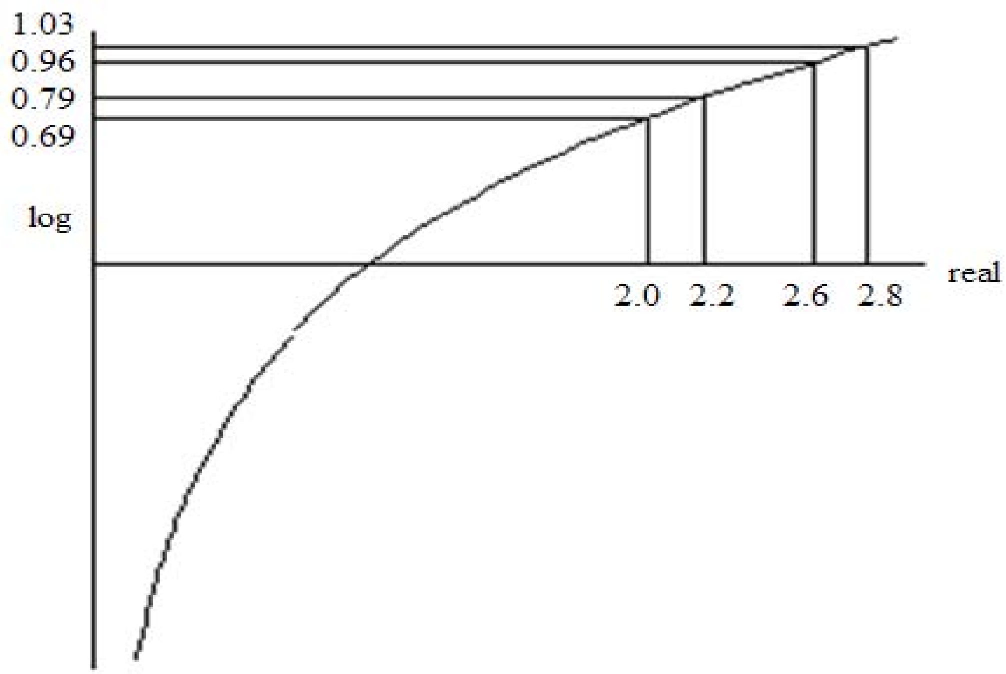


Figure 1 Compression of Values with the Natural Logarithm

In Figure 1, the first two values of 2.0 and 2.2 have respective logarithms of 0.69 and 0.79 which are much closer together than the original data points. The second two values of 2.6 and 2.8 have respective logarithms of 0.96 and 1.03 which are compressed even more. In Figure 2, the first two values of 0.4 and 0.45 have respective logarithms of -0.92 and -0.80 which are further apart, or stretched, in relation to the original data. The second two values of 0.2 and 0.25 have respective logarithms of -1.61 and -1.39 which are stretched even

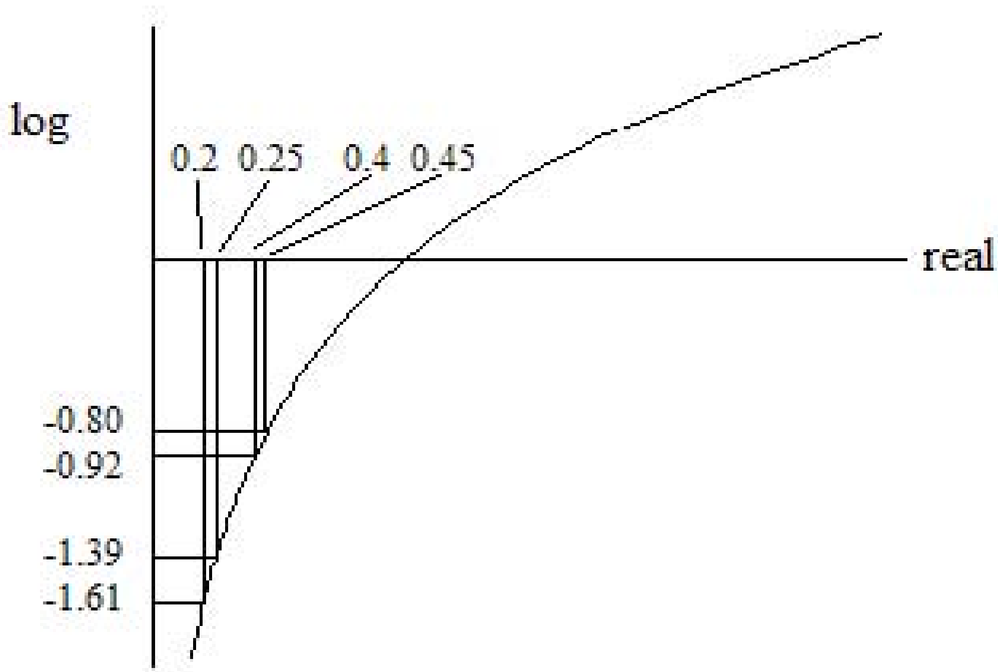


Figure 2 Stretching of Values with the Natural Logarithm

further. Thus, applying a logarithmic transformation to a right skewed distribution often results in a more symmetric distribution. As a note of caution, a logarithmic transformation could actually make things worse in a symmetric or left skewed distribution (Simon, 2005).

In addition to producing a more symmetric distribution, a logarithmic transformation can improve homoscedasticity. Homoscedasticity implies constant variability over the range of the dependent variable or similarity of within-group variability. When data is partitioned into groups, it is common for groups with larger values to have greater within-group variability (Dallal, 2005). "A logarithmic transformation will often make the within-group variability more similar across groups" (Dallal, 2005). This is due to the compression property of the logarithmic transformation demonstrated in Figure 1. The logarithmic transformation compresses groups with larger standard deviations more than it compresses groups with smaller standard deviations (Simon, 2005). We next look at the third of Dallal's reasons as to why the logarithm should interest us: linearity.

When a statistical model is used to describe the relationship between two measurements, there is no guarantee that the association between the measurements will be linear. When logarithmic transformations are applied to both variables, the association often be-comes linear. As Dallal stated earlier, an approximately linear relationship is much easier to describe.

Another factor to consider is the data points classified as outliers on the high end, or to the far right, of a distribution. The compression of large values under the logarithmic transformation can pull the outlier back in closer to the data (Simon, 2005). However, if a value is close to the low end of the distribution, or to the far left and less than one, a logarithmic transformation can force a non-outlier to become an outlier, due to the stretching property of the logarithm.

We have now seen that a logarithmic transformation may improve symmetry, homoscedasticity, and linearity. This transformation is used "because sometimes it's easier to analyze or describe something in terms of log-transformed data than in terms of the original values" (Dallal, 2005).

### *1.2 Specific Issue*

Once data has been transformed, using a logarithm function, and analysis performed, it is often necessary to back-transform and report results on the original scale. This back-transformation is cause for both concern and interest. Certain parameters in the original scale are easily obtained, while other parameters are not so trivial.

The mean, median, and confidence interval for the median of the original scale can be obtained from the log scale. The median (or geometric mean) on the original scale is found by the back-transform of the mean on the log-scale. Similarly, the confidence interval for the median in the original space is found by back-transforming the confidence interval for the mean on the log scale. The median and the confidence interval are merely simple back-transforms from the log scale to the original scale. However, the simplicity ends at this point. It is possible to obtain the mean on the original scale from the mean on the log scale. Schwarz provides the following equation for log-normal data,

$$\overline{Y}_{\text{original}} = \exp\left(\overline{Y}_{\text{transformed}} + \frac{s^2}{2}\right).$$

Distributions transformed using a function other than the logarithm will result in a different formula for the mean on the original scale (Schwarz, 2005). The back-transformation of the standard deviation does not work. A close approximation, given by Schwarz, is

$$s_{\text{untransformed}} = (s_{\text{transformed}}) \exp\left(\overline{Y}_{\text{transformed}}\right).$$

Now that the specific issue concerning the back-transformation from normal space to the original (log) space has been identified, we turn our focus to the specific research objectives.

### *1.3 Research Objectives*

The lack of a back-transformation for the standard deviation motivates the primary focus of this paper. This problem arose after reviewing *Estimate at Completion: A Regression Approach to Earned Value* (Tracy, 2005). In this thesis, a data set was transformed using the natural logarithm to obtain normalization and symmetry. Upon back-transforming the data to the original scale, confidence intervals in the original space appeared quite large. In particular, the variance and associated confidence intervals seemed large. This research looks to reduce the width of the confidence interval while maintaining or increasing the confidence level when back-transforming from normal space to the original space.

### *1.4 Synopsis of Research*

The logarithmic transformation can be highly beneficial in regression models. The logarithm function can be applied to a data set to facilitate the description and the computations of a particular distribution. Back-transforming from the log space to the original space is of special interest. In the following chapter, this thesis examines previous studies dealing with this back-transformation. We introduce a series of alternative approaches for improving and/or reducing the width of the confidence intervals after back-transformation in Chapter III. Through this series we propose an approach that yields improved results over the naive approach. The results of the proposed approach are discussed in detail in Chapter IV.

## *II. Literature Review*

A logarithmic transformation is often applied to a data set to facilitate the ease of obtaining parameter estimates and confidence intervals about these estimates. However, in many applications it is desirable to have these estimates and confidence intervals in terms of the untransformed data, rather than in terms of the transformed data. Thus, we first explore previous research focusing on this topic.

### *2.1 Distribution of a Variate whose Logarithm is Normally Distributed*

According to Finney, "the object ... is to derive, from the sufficient statistics for the normal distribution obtained from the transformed data, efficient estimates both of the mean and of the variance of the population" (Finney, 1941). Let X be a variate of the original data and let $\mathrm{Y} = \log \mathrm{X}$ be normally distributed with mean $\mu$ and variance $\sigma^2$. Then it can be shown that

$$\mathrm{E}[\mathrm{X}^r] = e^{r\mu + \frac{1}{2}r^2\sigma^2}. \tag{1}$$

From (1), the moments of the distribution may be obtained; in particular, the mean, $\theta$, and variance, $\delta^2$, are given by

$$\theta = e^{\mu + \frac{1}{2}\sigma^2}, \tag{2}$$

$$\delta^2 = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right). \tag{3}$$

To determine an estimation for the moments, Finney supposes that a sample of $n$ objects is taken from the population. The sufficient statistics for the estimation of the parameters of the transformed distribution (Finney, 1941) are:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{4}$$

$$s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \overline{y})^2, \tag{5}$$

where

$$y_i = \text{sample points},$$

$$\overline{y} \quad = \quad \text{sample mean.}$$

Finney determined, through some derivation, that efficient estimates of the mean $m$ and variance $v$ of the original data are

$$\begin{aligned} m &= e^{\overline{x}} g\left(\frac{1}{2}s^2\right), & (6) \\ v &= e^{2\overline{x}}\left(g(2s^2) - g\left(\frac{n-2}{n-1}s^2\right)\right), & (7) \\ &= e^{2\overline{x}} g_1(s^2), & (8) \end{aligned}$$

where

$$\begin{aligned} g(t) &= 1 + \sum_{k=1}^{\infty} \frac{t^k (n-1)^{2k-1}}{k! n^k \prod_{j=2}^{k}(n+2j-3)}, & (9) \\ g_1(t) &= g(2t) - g\left(\frac{n-2}{n-1}t\right). & (10) \end{aligned}$$

Tables of $g(t)$ and $g_1(t)$ have been calculated (Aitchison and Brown, 1957) and abbreviated versions are given in Appendix A. The series $g(t)$ converges slowly, utilizing the available computing power at that time, making it unsuitable for computational purposes except for small values of $t$ (Finney, 1941). To overcome this obstacle, a more suitable expansion of $g(t)$ is

$$g(t) = e^t\left(1 - \frac{t(t+1)}{n} + \frac{t^2(3t^2+22t+21)}{6n^2} + \ldots\right), \tag{11}$$

yielding the following approximations to the efficient estimates, which are to the order of $n^{-2}$ (Finney, 1941):

$$\begin{aligned} m &= e^{\overline{x}+\frac{1}{2}s^2}\left(1 - \frac{s^2(s^2+2)}{4n} + \frac{s^4(3s^4+44s^2+84)}{96n^2}\right), & (12) \\ v &= e^{\overline{x}^2+s^2}\left[e^{s^2}\left(1 - \frac{2s^2(2s^2+1)}{n} + \frac{2s^4(12s^4+44s^2+21)}{3n^2}\right)\right. \\ &\quad \left. - \left(1 - \frac{s^2(s^2+2)}{n} + \frac{s^4(3s^4+28s^2+42)}{6n^2}\right)\right]. & (13) \end{aligned}$$

Finney proposes that a sample size of $n > 50$ in (12) and $n > 100$ in (13), both clearly not small samples, are safe limits for estimation (Finney, 1941).

*2.2 The Lognormal Distribution*

Aitchison and Brown define the lognormal distribution as "the distribution of a variate whose logarithm obeys the normal law of probability" (Aitchison and Brown, 1957). In *The Lognormal Distribution* (Aitchison and Brown, 1957), credit is given to D. McAlister for explicitly defining the theory of the lognormal distribution. In a paper presented in 1879 to the Royal Society of London, McAlister "gave expressions for the mean, median, mode and the second moment of the distribution, together with the quartiles and octiles" (Aitchison and Brown, 1957). The history of the lognormal distribution, as described by Aitchison and Brown, was precarious and sporadic having "remained the Cinderella of distributions" (Aitchison and Brown, 1957). A detailed history of the lognormal distribution is outlined in *The Lognormal Distribution* (Aitchison and Brown, 1957).

Let X $(0 < \mathrm{x} < \infty)$ be a positive random variable such that $\mathrm{Y} = \log \mathrm{X}$ is normally distributed with mean $\mu$ and variance $\sigma^2$. Then X is lognormally distributed denoted a $\Lambda$-variate. The distribution functions of X and Y are denoted by $\Lambda(x|\mu, \sigma^2)$ and $N(y|\mu, \sigma^2)$ respectively.

Before determining which estimation procedures yield better estimates, it is necessary to define what properties a good estimator is expected to possess.

> The three main criteria usually suggested are the following, of which the first two are theoretical, and the third is practical in nature (Aitchison and Brown, 1957):
>
> 1. The estimator should be unbiased [when the expected value, with respect to $\theta$, of a point estimate W of a parameter $\theta = \theta$ $(\mathrm{E}_\theta \mathrm{W} = \theta)$], or, when only large samples are in question, asymptotically unbiased (consistent).
> 2. The variance of the estimator should be as small as possible.
> 3. The calculations involved should be reasonable and within the capabilities of the available computing machinery (Aitchison and Brown, 1957).

The first two criteria combine to make up the mean squared error (MSE) of an estimator (Casella and Berger, 2002:330). Thus, the estimation procedure should offer a small MSE and should be reasonable to implement and compute.

Aitchison and Brown (1957) examined the method of maximum likelihood, the method of moments, the method of quantiles, and a graphical method in determining which estimation procedure was better for determining estimates of the mean and variance of a lognormal distribution. The method developed by Finney, which is equivalent to the method of maximum likelihood, was found to be the most desirable and is the procedure recommended by Aitchison and Brown. Established theory provides no means of obtaining exact confidence intervals for the mean and variance of a lognormal distribution (Aitchison and Brown, 1957).

### *2.3 Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance*

The tables provided by Land "define exact confidence intervals for linear functions of the normal mean and variance, and approximate confidence intervals for nonlinear functions" (Land, 1975). Land reemphasizes the idea that data transformation permits inferences to be made easily in terms of means in the transformed (normal) scale, but inferences about means in the original, untransformed scale are difficult and non-trivial (Land, 1975). These difficulties arise due to the means of the original variates being functions of both the means and variances of their normal transforms (Land, 1975). The tables provided by Land provide an exact and optimal solution when X is lognormal (Land, 1975).

The tables consist of factors C such that $\hat{\mu} + \frac{1}{2}\hat{\sigma}^2 + \hat{\sigma}\nu^{-\frac{1}{2}}$ is an exact one-sided confidence limit for $\mu + \frac{1}{2}\sigma^2$ based on the mean $\hat{\mu}$ and variance $\hat{\sigma}^2$ from a normal $(\mu, \sigma^2)$ random sample of size $\nu + 1$ (Land, 1975). The factors $\mathrm{C} = \mathrm{C}(\mathrm{S}; \nu, 1-\alpha)$ consist of the arguments

$$\begin{aligned} \mathrm{S} &= \hat{\sigma} \text{ times an appropriate multiplier,} \\ \nu &= \text{degrees of freedom for } \hat{\sigma}^2, \\ 1-\alpha &= \text{confidence level.} \end{aligned}$$

"Exact confidence limits for a linear function $\mu + \lambda\sigma^2(\lambda \neq 0)$ can be obtained, based on a $N\left(\mu, \frac{\sigma^2}{\gamma^2}\right)$ estimate $\hat{\mu}$ of $\mu$ [where $\gamma$ is a function of a constant (possibly $n$)] and a statistically independent $\frac{\sigma^2\chi^2(\nu)}{\nu}$ estimate $\hat{\sigma}^2$ of $\sigma^2$ " (Land, 1975). For example, if $\hat{\mu} = \overline{x}$, then the confidence limits will be based on $N\left(\mu, \frac{\sigma^2}{n}\right)$ estimates. Land provides the following exact one-sided upper confidence limit of level $1-\alpha$ for $\mu + \lambda\sigma^2$:

$$\mathrm{Q}_\lambda = \hat{\mu} + \lambda\hat{\sigma}^2 + \mathrm{kS}\nu^{-\frac{1}{2}}\mathrm{C}(\mathrm{S}; \nu, 1-\alpha^*), \tag{14}$$

where

$$\begin{aligned} S &= \left(\frac{2\lambda\hat{\sigma}^2}{k}\right)^{\frac{1}{2}}, \\ k &= \frac{\frac{1}{2}(\nu+1)}{(\lambda\gamma^2)}, \\ \alpha^* &= \begin{cases} \alpha & \text{if } \lambda > 0 \\ 1-\alpha & \text{if } \lambda < 0. \end{cases} \end{aligned}$$

This limit also provides an exact one-sided level $\alpha$ lower confidence limit for $\mu + \lambda\sigma^2$. Two-sided limits, of level $1 - \frac{1}{2}\alpha$, with equal tail probabilities can be obtained in pairs.

As an example of this method (Land, 1975), let $\hat{\mu} = 1.6$ and $\hat{\sigma}^2 = 0.81$ be the sample estimates of the mean and variance of a lognormal variate. Consider a simple model in which $\hat{\sigma}^2$ has 15 degrees of freedom and Var $\hat{\mu} = \frac{\sigma^2}{16}$. Then, $\lambda = \frac{1}{2}$, k = 1, and S = 0.9. The values of the arguments for C are provided in the tables and inputting these values yields $C(0.9; 15, 0.95) = 2.554$ (Land, 1975). From (12), the one-sided upper confidence limit of level 0.95 is 2.598, whose exponential, 13.44, is the corresponding confidence limit for E[X] (Land, 1975). From the tables, $C(0.9; 15, 0.05) = -1.686$, from which the one-sided level 0.95 lower limit is obtained and is equal to 1.613, whose exponential, 5.019, is the corresponding limit for E[X]. Also, the equi-tailed two-sided confidence interval of level 0.90 is (1.613, 2.598), and the interval of the exponentials, (5.019, 13.44), is the corresponding confidence interval for E[X] (Land, 1975).

Land (1975) indicates that for values of $S, \nu$ not listed in tables, $C = C(S; \nu, 1-\alpha)$ must be obtained by interpolation. The table for the above example can be found in Appendix B; a complete table for the values C can be found in *Selected Tables in Mathematical Statistics, Volume III* published in *American Mathematical Society* in 1975.

### *2.4 Calculating Confidence Intervals for the Mean of a Lognormally Distributed Variable*

T. B. Parkin, S. T. Chester, and J. A. Robinson conducted a study to report efficacy of different methods for "constructing confidence intervals about the mean of a lognormally distributed variable" (Parkin and others, 1990). Performance was assessed by identifying the proximity of the calculated probability levels for the confidence limits to the actual probability levels.

Three of these methods provided close approximations with one in particular providing exact levels. The three methods are: a method devised by Cox (see Land, 1972); a quantile method developed by the authors of this study; and an exact method developed by Land. The method proposed by Cox gives the limits,

$$\exp\left(\overline{\mu} + \frac{\hat{\sigma}^2}{2} \mp t\left(\frac{\hat{\sigma}}{\sqrt{n}}\right)\sqrt{1 + \frac{\hat{\sigma}^2 n}{2(n+1)}}\right), \tag{15}$$

where $t$ is the critical value from the Student's $t$ distribution with $n-1$ degrees of freedom.

The quantile method "is based on the quantile corresponding to the mean $p$ of the lognormal distribution," (Parkin and others, 1990) defined by

$$p = \mathrm{P}[x \leq \mathrm{E}[X]] = \Phi\left(\frac{\sigma}{2}\right), \tag{16}$$

where X is distributed as a lognormal random variable and $\Phi$ is the cumulative distribution function of a normal distribution. An estimate $\hat{p}$ can be obtained from (16) using $\hat{\sigma}$ (the positive square root of the variance of the log-transformed values). The confidence limits are estimated by selecting the appropriate order statistics,

$$\mathrm{LCL} = x(\mathrm{r}), \tag{17}$$

$$\mathrm{UCL} = x(\mathrm{s}), \tag{18}$$

where $x(r)$ and $x(s)$ are the $r$th and $s$th order statistics (r < s) (Parkin and others, 1990).

$$\sum_{i=0}^{\mathrm{r}-1} \binom{n}{i} \hat{p}^i (1-\hat{p})^{n-i} \geq 0.05 \tag{19}$$

$$\sum_{i=0}^{\mathrm{s}-1} \binom{n}{i} \hat{p}^i (1-\hat{p})^{n-i} \leq 0.95 \tag{20}$$

Precisely, r is the smallest integer such that (19) holds and s is the largest integer such that (20) holds (Parkin and others, 1990).

For $n > 20$, values for r and s can be defined by

$$\mathrm{r} = n\hat{p} - z_{0.95}\sqrt{n\hat{p}(1-\hat{p})}, \tag{21}$$

$$\mathrm{s} = n\hat{p} + z_{0.95}\sqrt{n\hat{p}(1-\hat{p})}, \tag{22}$$

where $z_{0.95}$ is the critical value from the standard normal distribution with $\alpha = 0.95$. Since (21) and (22) rarely yield integer solutions, r and s are obtained by rounding the results to the next highest integer (Parkin and others, 1990).

The exact method developed by Land (1971), provides the following lower and upper confidence limits:

$$\mathrm{LCL} = \exp\left(\overline{\mu} + \frac{\hat{\sigma}^2}{2} + \frac{\hat{\sigma}\mathrm{C_L}}{\sqrt{n-1}}\right), \tag{23}$$

$$\mathrm{UCL} = \exp\left(\overline{\mu} + \frac{\hat{\sigma}^2}{2} + \frac{\hat{\sigma}\mathrm{C_U}}{\sqrt{n-1}}\right), \tag{24}$$

where $\mathrm{C_L}$ and $\mathrm{C_U}$ are calculated from a function depending on $n$ (the number of observations), $\sigma$ (the standard deviation of the log-transformed values) and the $\alpha$ level selected (Parkin and others, 1990). Land developed an algorithm for computing these factors (Land, 1988).

Land's method is preferred over the other methods, as it provides exact coverage at the stated probability level for every lognormal population evaluated (Parkin and others, 1990). A small sample size ($n < 20$) posed problems for every method with the exception of the method proposed by Land. The quantile method developed by Parkin, Chester, and Robinson provided accurate results for $n > 20$. For $60 \leq n \leq 100$, Cox's method yielded almost exact coverage. Cox's method performed better (by providing nearly exact coverage) than the quantile method for highly skewed distributions. Several conclusions came out of this study.

1. Land's method is the preferred method for all situations, as it provides exact confidence limits.
2. With large sample sizes ($n > 60$), Cox's method is a suitable alternative, since it provides reasonably accurate coverage and it is simple to implement.
3. The quantile method developed by Parkin, Chester, and Robinson has applications for medium to large sample sizes ($n = 40 - 100$) (Parkin and others, 1990).

*2.5 Confidence Intervals for the Log-normal Mean*

In a simulation study conducted by Zhou and Gao (1997), four main methods for the construction of confidence intervals of lognormal means were evaluated for three criteria: (i) coverage error, the absolute value of the difference between the nominal level of coverage and the actual coverage probability; (ii) interval width; and (iii) relative bias, a measeure of the magnitude of the bias. The four methods evaluated include a naive approach, Cox's method (Land, 1972), Angus's conservative method (Angus, 1988), and a parametric bootstrap method (Angus, 1994).

First, Zhou and Gao made the same initial assumptions as Finney. That is, X has a two-parameter lognormal distribution and Y = log X is normally distributed with mean $\mu$ and variance $\sigma^2$. Then, equations (1), (2), (4), and (5) hold true. Zhou and Gao note that the naive approach is the most commonly applied approach; Land's exact method is computationally complicated and the numerical algorithms are sometimes unstable; and the three main approximate methods are Cox's method and two different methods proposed by Angus (Zhou and Gao, 1997).

The naive method involves two steps. First, a confidence interval for $\mu$ is constructed using the normal theory. Second, an antilogarithm function is applied to these limits to transform the limits back to the original scale. However, this confidence interval is for $e^{\mu}$ rather than for $\theta = e^{\mu+\frac{\sigma^2}{2}}$ and is therefore biased for large $\sigma^2$ (Zhou and Gao, 1997).

Cox's method is based upon complete sufficient statistics and uniformly minimum variance unbiased (UMVU) estimators. Inferences on $\log\theta$ based on $(\overline{y}, s^2)$ can be made since the statistic $(\overline{y}, s^2)$ is complete sufficient for $(\mu, \sigma^2)$ (Zhou and Gao, 1997). The UMVU estimator of $\log\theta$ and the corresponding variance are $\overline{y} + \frac{s^2}{2}$ and $\frac{s^2}{n} + \frac{s^4}{2(n-1)}$, respectively (Zhou and Gao, 1997). Cox, in a personal communication to Land (1971), proposed, with the above observations, to construct a confidence interval for $\log\theta$ by

$$\overline{y} + \frac{s^2}{2} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\left(\frac{s^2}{n} + \frac{s^4}{2(n-1)}\right)}. \qquad (25)$$

Because the probability that $\log\theta$ lies in the bounded region is at least $1-\alpha$, Angus (1988) proposed a conservative method based on the pivotal statistics,

$$\mathrm{V}(\theta) = \frac{\left(\overline{y} + \frac{s^2}{2} - \log\theta\right)\sqrt{n}}{\sqrt{s^2\left(1+\frac{s^2}{2}\right)}}. \tag{26}$$

From this, Angus derived a conservative lower and upper $1-\alpha$ limit as

$$\mathrm{L}_{1-\alpha}\left(\overline{y}, s^2\right) = \overline{y} + \frac{s^2}{2} - \frac{t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}}\sqrt{s^2\left(1+\frac{s^2}{2}\right)}, \tag{27}$$

$$\mathrm{U}_{1-\alpha}\left(\overline{y}, s^2\right) = \overline{y} + \frac{s^2}{2} - \frac{q_{\frac{\alpha}{2}}(n-1)}{\sqrt{n}}\sqrt{s^2\left(1+\frac{s^2}{2}\right)}, \tag{28}$$

where $t$ is a $t$-distribution with $n-1$ degrees of freedom and

$$q_{\alpha}(n-1) = \sqrt{\frac{n}{2}\left(\frac{n-1}{\chi^2_{\alpha}(n-1)} - 1\right)}, \tag{29}$$

with $\chi_{\alpha}(n-1)$ denoting the $\alpha$-percentile of the chi-square distribution with $n-1$ degrees of freedom.

In addition, Angus (1994) described a bootstrap method applied to (26). Letting $t_0$ and $t_1$ be the lower and upper limits of $\mathrm{V}(\theta)$, respectively, a theoretical $1-\alpha$ level confidence interval for $\log\theta$ (Zhou and Gao, 1997) is

$$\mathrm{I} = \left(\overline{y} + \frac{s^2}{2} - t_1\sqrt{\frac{s^2}{n}\left(1+\frac{s^2}{2}\right)}, \overline{y} + \frac{s^2}{2} - t_0\sqrt{\frac{s^2}{n}\left(1+\frac{s^2}{2}\right)}\right). \tag{30}$$

The unknown quantiles $t_0$ and $t_1$ can be estimated by a parametric bootstrap sample (Zhou and Gao, 1997).

The simulation study consisted of two-sided confidence intervals with equal tail probabilities, three sample sizes of $n = 11$, 101, and 400, and variance ranging from 0.1 to 20.0. The results of this study can be viewed in the tables of Appendix C. This simulation study concluded that the naive method is inappropriate for constructing the desired confidence intervals, Cox's method is recommended for moderate to large sample sizes, and the bootstrap method is recommended for small sample sizes (Zhou and Gao, 1997).

*2.6 Confidence Intervals for the Mean of a Log-Normal Distribution*

Ulf Olsson (2005) examines five different methods for calculating confidence intervals about the mean of a lognormal distribution. The same assumptions hold here as used by Finney (1941) and Zhou and Gao (1997). That is, a logarithmic transformation is applied to the original variable X and inferences are based on the transformed variable Y = log X.

In comparing the five methods for calculating the confidence intervals for $E[X] = \theta$, Olsson generated a numerical example using SAS® (1997) software. A sample of $n = 40$ observations was generated from a lognormal distribution with mean $\mu = 5$ and standard deviation $\sigma = 1$. The population mean of X was calculated to be $\theta = 244.69$. A log-transform was then applied to the data and is summarized in Table 1 (Olsson, 2005). Using the sample data provided in Table 1, confidence intervals about the mean of $x$ were

Table 1 Summary statistics for sample data

| Variable | Mean | Median | Standard Deviation |
|---|---|---|---|
| $x$ | 274.963 | 177.350 | 310.343 |
| $y = \log x$ | 5.127 | 5.170 | 1.004 |

calculated using several methods, including a naive method, Cox's method, a modified version of Cox's method, a method motivated by large sample theory, and a method based on generalized confidence intervals.

The naive approach, as described previously, resulted in a point estimate for $\theta$ of $\widehat{\theta} = e^{5.127} = 168.51$. A 95% confidence interval for $\mu$ was calculated to be [4.806, 5.448] which resulted in confidence limits for $\theta$ of [122.24, 232.29]. Olsson notes that the naive approach confidence interval is biased, in that it covers neither the population mean, 244.69, nor the sample mean, 274.963 (Olsson, 2005).

Using Cox's method, the resulting point estimate was $\widehat{\theta} = 279.22$. The 95% confidence interval for log X is [5.248, 6.016] resulting in [190.24, 409.82] as the 95% confidence interval for $\theta$. The modified version of the Cox method consists of replacing $z$, the critical value for the standard normal distribution, with $t$, the critical value of the Student's $t$ distribution with degrees of freedom based on the degrees of freedom for the estimate of $\sigma^2$ (Olsson, 2005). Two reasons for this modification is that a confidence interval for $\mu$ would be based on $t$, and the resulting confidence interval coverage is closer to the nominal level

(Olsson, 2005). This modified version yielded a 95% confidence interval for $\log \mathrm{X}$ of [5.237, 6.027]. Taking the antilogarithm results in a 95% confidence interval for $\theta$ of [188.0, 414.7].

Another method (Krishnamoorthy and Mathew, 2003:108) for computing a confidence interval for the mean of a lognormal distribution is as follows:

1. Calculate $\overline{y}$ and $s^2$ from the data.
2. For $i = 1$ to $m$ (where $m$ is large, i.e. $m = 10000$)
    - Generate $Z \sim N(0,1)$ and $U^2 \sim \chi^2_{(n-1)}$
    - For each $i$, calculate $T_{2i} = \overline{y} - \frac{Z}{U/\sqrt{(n-1)}}\frac{s}{\sqrt{n}} + \frac{1}{2}\frac{s^2}{U^2/(n-1)}$
3. (end $i$ loop)

"For a 95% confidence interval, the 2.5% and 97.5% percentiles for $T_2$ are calculated from the 10000 simulated values" (Olsson, 2005). These percentiles form the limits in a confidence interval for $\log \theta$. Thus, a 95% confidence interval for the lognormal mean is calculated as $[\exp(t_{2;0.025}), \exp(t_{2;0.975})]$ (Olsson, 2005).

The last method is based on the untransformed data. The Central Limit Theorem (CLT) gives us that if $n$ is reasonably large, the distribution of a sample mean $\overline{x}$ can be approximated with a normal distribution for a large class of distributions (Olsson, 2005). Provided the sample is large, the confidence interval can be calculated as

$$\overline{x} \pm z\sqrt{\frac{s_x^2}{n}}. \tag{31}$$

With the sample data provided, (31) yields the confidence interval [178.84, 371.16].

After exploring each method, Olsson performed a simulation study comparing the five methods with respect to the percentage of intervals that covered, were below, or were above the true parameter value. The results of this simulation, summarized in Appendix D, demonstrated that the modified Cox method and the generalized confidence interval method provided better results. Both of the aforementioned methods yielded intervals covering $\theta$ close to the desired 95% level for all sample sizes. "[T]he confidence intervals based on the modified Cox method work well for practical purposes . . . a small disadvantage [of the generalized confidence interval approach] is that it requires a computer to simulate the sampling distribution" (Olsson, 2005).

### *2.7 Comparison Between Prior Research and Current Research*

All of the research reviewed in this chapter involves a static process. That is, the mean of the lognormal distribution is determined from values, including the mean and variance, corresponding to transformed data. This data, transformed with the natural logarithm, takes on the form of a normal distribution. Values are then calculated in normal space and equations are given to determine the mean of the untransformed data.

The method proposed in the following chapter defines a more dynamic process. While the mean remains of utmost importance, it is the mean of a sample of data points rather than the mean of the entire data set that is of interest. The alternative approach works in a regression setting, in which the sample mean is a function of the sample data points. The sample mean varies over the range of the sample data values, depending upon the data points chosen for the random sample. The dynamics lie in the fact that removing, adding, or changing a data point changes the output; namely, the sample mean. Therefore, different samples result in different points along the regression line to be analyzed. The confidence intervals of interest are those intervals about the sample mean.

## *III. Methodology & Results*

### *3.1 Background & Scope*

Common practice involves transforming a lognormal distribution to a normal distribution, by taking the natural logarithm, to create symmetry, homoscedasticity, and linearity. The transformed data is easily analyzed, as the normal distribution is well known. However, given a point estimate and the associated confidence interval in the normal space, the transformation back to log space poses problems. The width of the confidence intervals, especially in the upper percentile region of the distribution, appear to be large. In some cases, the intervals tend to be somewhat irrelevant due to their large size. Also, the confidence itself comes into question as the confidence does not appear to have a one-to-one relationship upon back-transformation. This research strives to increase the confidence and decrease the width of the confidence interval upon back-transformation from the normal space to the log space.

A simulation approach aims to answer the research questions. Transforming data that follows or closely resembles a lognormal distribution to a normal distribution by taking the natural logarithm is common, due in part to its simplicity. Thus, to stay in this framework, simulation begins by generating a normal distribution, taking random samples, and back-transforming to log space. Therefore, simulation, rather than another method, such as bootstrapping, was chosen to allow the research to examine this very common process.

### *3.2 Simulation Approach*

Without loss of generality (WLOG), simulation occurs from a standard N(0,1) distribution. Each simulation consists of 100,000 runs conducted at seven different sample sizes: 5, 10, 15, 20, 25, 30, and 100. Note that simulation utilizing the $t$-statistic, generally used for small samples, excludes $n = 100$ and thus has only six different sample sizes. Furthermore, simulation utilized five different percentiles of the distribution: 0.10, 0.25, 0.50, 0.75, and 0.90. Thus, sample sizes generated at five different percentiles allow for both sample size and the percentile of the distribution from which the sample was drawn from to be analyzed.

In order to use simulation, the numerical value of each of these percentiles in both the log space and the normal space had to be identified. The LOGINV function in Microsoft® Excel 2000 (1999) calculated the corresponding values in log space. The random variables ln[X] and X are related to each other by either ln[·] or exp[·], depending on the direction of the transformation (Burmaster and Hull, 1997). Therefore, the percentiles are related by the same transforms (Burmaster and Hull, 1997). Thus, finding the corresponding values in normal space involved taking the natural logarithm of the values in log space. Table 2 summarizes the findings.

Table 2 Percentiles and associated values in log space and normal space

| percentile | log space | normal space |
|---|---|---|
| 0.10 | 0.2776 | -1.2816 |
| 0.25 | 0.5094 | -0.6745 |
| 0.50 | 1.0000 | 0.0000 |
| 0.75 | 1.9630 | 0.6745 |
| 0.90 | 3.6022 | 1.2816 |

### *3.3 Naive Approach*

The first step involved using the known naive approach to establish a baseline for the back-transformation of point estimates and their associated confidence intervals. As stated previously, the naive approach works by taking the natural exponent and raising it to the lower and upper confidence limit and the point estimate itself. Performing simulation using both the $z$-statistic and the $t$-statistic offers more comparison and takes into consideration the fact that small sample sizes generally utilize the $t$-statistic for interval estimation. The simulation code for the naive approach, using the $z$-statistic, is shown in Appendix E. The simulation code for the naive approach using the $t$-statistic is slightly different. The difference is in how L_transformed and U_transformed are computed. The 1.96 corresponding to $z_{0.95}$ is replaced with the appropriate value for $t_{n-1,\alpha/2}$ given in Table 3.

The code works by creating a 100,000 by $n$ matrix, (where $n$ is the sample size), with entries from a normal random number generator. The percentile from which the samples are drawn is dictated by the variable $j$ (see Table 2). The mean and standard deviation are then recorded for each row of the matrix. The lower and upper 95% confidence limits for

Table 3 Values of $t_{n-1,\alpha/2}$ where $n$ = sample size and $\alpha = 0.95$

| n | $t_{n-1,\alpha/2}$ |
|---|---|
| 5 | 2.77645 |
| 10 | 2.26216 |
| 15 | 2.14479 |
| 20 | 2.09302 |
| 25 | 2.06390 |
| 30 | 2.04523 |

the transformed data (normal space) are then recorded and back-transformed to log space. The interval (Upper Confidence Limit - Lower Confidence Limit) for each of the 100,000 runs is recorded, the mean width is computed, and the total number of means falling in that interval is determined. Counting the number of times that $i$, calculated as $e^j$, falls between the lower and upper confidence limit provides the empirical confidence level. The results for the simulation of the naive approach using the $z$-statistic and the $t$-statistic are given in Tables 4 and 5 respectively.

Table 4 Naive Approach Simulation Results – $z$-statistic

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| percentile | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| 0.10 | interval | 0.5143 | 0.3552 | 0.2874 | 0.2478 | 0.2208 | 0.2011 | 0.1092 |
| | confidence | 0.84483 | 0.90354 | 0.92068 | 0.92904 | 0.93321 | 0.93559 | 0.94646 |
| 0.25 | interval | 0.9387 | 0.6537 | 0.5269 | 0.4546 | 0.4049 | 0.3694 | 0.2005 |
| | confidence | 0.84410 | 0.90441 | 0.92114 | 0.92852 | 0.93265 | 0.93651 | 0.94595 |
| 0.50 | interval | 1.8518 | 1.2796 | 1.0363 | 0.8918 | 0.7953 | 0.7248 | 0.3934 |
| | confidence | 0.84452 | 0.90378 | 0.92138 | 0.92944 | 0.93584 | 0.93616 | 0.94707 |
| 0.75 | interval | 3.6295 | 2.5089 | 2.0337 | 1.7527 | 1.5615 | 1.4228 | 0.7722 |
| | confidence | 0.84610 | 0.90381 | 0.92016 | 0.92821 | 0.93369 | 0.93567 | 0.9454 |
| 0.90 | interval | 6.6483 | 4.6199 | 3.7343 | 3.2180 | 2.8613 | 2.6080 | 1.4165 |
| | confidence | 0.84499 | 0.90586 | 0.92118 | 0.92787 | 0.93310 | 0.93824 | 0.94615 |

The results from the naive approach demonstrate a wider interval and lower confidence for smaller sample sizes with smaller intervals and greater confidence as the sample sizes increase. As expected, the intervals at the lower percentiles are smaller than the intervals for the higher percentiles. This observation stems from the lognormal distribution being right-skewed. The observations suggest that the 95% confidence interval desired will be attained as $n$ approaches infinity.

Table 5 Naive Approach Simulation Results – $t$-statistic

| | | Sample Size | | | | | |
|---|---|---|---|---|---|---|---|
| percentile | | 5 | 10 | 15 | 20 | 25 | 30 |
| 0.10 | interval | 0.8258 | 0.4187 | 0.3358 | 0.2657 | 0.2331 | 0.2105 |
| | confidence | 0.93183 | 0.93897 | 0.95297 | 0.94543 | 0.94542 | 0.94512 |
| 0.25 | interval | 1.5121 | 0.7691 | 0.5828 | 0.4877 | 0.4279 | 0.3860 |
| | confidence | 0.93106 | 0.93984 | 0.94258 | 0.94350 | 0.94500 | 0.94577 |
| 0.50 | interval | 2.9751 | 1.5055 | 1.1434 | 0.9572 | 0.8408 | 0.7583 |
| | confidence | 0.93329 | 0.94017 | 0.94402 | 0.94577 | 0.94559 | 0.94662 |
| 0.75 | interval | 5.8301 | 2.9654 | 2.2417 | 1.8824 | 1.6504 | 1.4867 |
| | confidence | 0.93179 | 0.93931 | 0.94315 | 0.94542 | 0.94593 | 0.94557 |
| 0.90 | interval | 10.6495 | 5.4290 | 4.1142 | 3.4503 | 3.0245 | 2.7286 |
| | confidence | 0.93128 | 0.93966 | 0.94184 | 0.94423 | 0.94657 | 0.94616 |

### *3.4 Alternative Approaches*

The naive approach calculation is based upon both the mean and the standard deviation in normal space. We propose a series of alternative approaches in which it is not necessary to utilize the mean and the standard deviation. Our suggested method relies on only the point estimate itself. The series of approaches posed in the upcoming sections have the form $\hat{y} \pm c\sqrt{\hat{y}}$. In all simulations, $\hat{y} = e^y$, where $\hat{y}$ is the point estimate in log space and $y$ is the point estimate in normal space. WLOG, the variance has been standardized to 1. Standardizing the variance removes the possibility of having to deal with different formulas corresponding to different variances. We will now present our series of approaches and compare these with the naive approach.

*3.4.1 Approach 1: $\hat{y} \pm \sqrt[m]{\hat{y}}$.* In attempting to decrease the interval width and increase the confidence, the first approach involved analyzing $\hat{y} \pm \sqrt[m]{\hat{y}}$ with $m$ ranging from 2 to 100. Originally, $m$ ranged from 2 to 10. However, there appeared to be a trend in the interval width further supported by taking roots up to 100. The data demonstrated that the interval width seemed to approach 2 with larger $m$. The interval width has the form $2\sqrt[m]{\hat{y}}$, which converges to 2 since

$$\lim_{m\to\infty} 2\sqrt[m]{\hat{y}} = \lim_{m\to\infty} 2e^{(\ln \hat{y}/m)} = 2e^0 = 2.$$

The simulation code can be seen in Appendix E and the results of the simulation can be seen in Tables 20-24 in Appendix F.

*3.4.2 Approach 2: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$.* Looking specifically at $\hat{y} \pm \sqrt{\hat{y}}$, the confidence levels appeared quite large with many extremely close to or at 100% confidence. This observation led to a second approach, $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$ with $m$ ranging from 2 to 10. Again, the code can be found in Appendix E and the results of the simulation are in Tables 25-29 in Appendix F.

*3.4.3 Approach 3: $\hat{y} \pm \frac{1}{\sqrt{\frac{n}{a}}}\sqrt{\hat{y}}$.* From these results, it appeared at first glance that $\sqrt{n}$ worked better than $m$. That is, it appeared that $\hat{y} \pm \frac{1}{\sqrt{n}}\sqrt{\hat{y}}$ offered better results. However, with an increase in percentile it was necessary to divide $n$ by a variable $a$ to obtain desirable results. The values attempted for $a$ ranged from 1 to 15 depending on the percentile being tested. In testing the 10th percentile, $a = 1$ provided superior results. For the 25th percentile, values of 1 and 2 were used for $a$. Values of 3 and 4 were tried in testing the 50th percentile, values of 5 through 8 for the 75th percentile, and values of 8 through 15 for testing the 90th percentile. The code utilized for these tests is in Appendix E and the results are in Tables 30-34 in Appendix F.

*3.4.4 Approach 4: $\hat{y} \pm \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\hat{y}}$.* The equation offering better results, a smaller interval width and increased confidence, appears to have the form $\hat{y} \pm \frac{c}{\sqrt{n}}\sqrt{\hat{y}}$ where $c$ is some constant. Through simulation, a numerical value for $c$ at each percentile could be found that offered desirable results. These values are given in Table 6.

Table 6 Values of $c$ providing superior results

| percentile | c |
|---|---|
| 0.10 | 1 |
| 0.25 | 1.35 |
| 0.50 | 1.95 |
| 0.75 | 2.7 |
| 0.90 | 3.65 |

With these values, the Microsoft® Excel 2000 (1999) chart wizard plotted the data, added an exponential trendline, and provided the equation of the trendline. The equation of the trendline, $y = 0.8821e^{1.5532p}$ closely resembles $y = \frac{8}{9}\left(\frac{3\pi}{2}\right)^p$, where $p$ is the percentile. The second equation was approximated using Microsoft® Excel 2000 (1999) with $\frac{8}{9} = 0.8889$ approximating 0.8821 and $\frac{3\pi}{2} = 4.7124$ approximating $e^{1.5532} = 4.7266$. The second equation was also plotted using the Microsoft® Excel 2000 (1999) chart wizard. Figure 3 shows the result of the chart wizard.

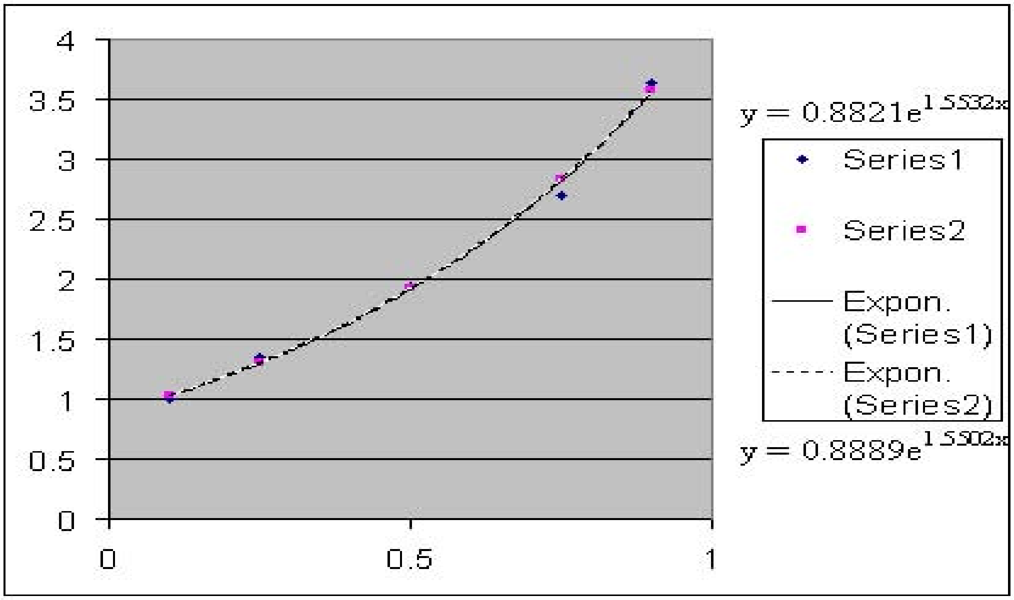


Figure 3 Equation trendlines

Series 1 represents the original equation and Series 2 represents the approximated equation. Figure 3 demonstrates that the two equations are very similar.

Utilizing the second equation as the coefficient $c$, $\hat{y} \pm \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\hat{y}}$ will be referred to as the $p$ equation. The two equations are compared at each percentile in Table 7. Thus, the

Table 7 Equation comparison at percentile $p$

| percentile | $y = 0.8821e^{1.5532p}$ | $y = \frac{8}{9}\left(\frac{3\pi}{2}\right)^p$ |
|---|---|---|
| 0.10 | 1.0303 | 1.0379 |
| 0.25 | 1.3006 | 1.3097 |
| 0.50 | 1.9177 | 1.9296 |
| 0.75 | 2.8277 | 2.8430 |
| 0.90 | 3.5695 | 3.5873 |

$p$ equation should offer desirable results in the form of reduced interval width, increased confidence, or both. The simulation code is shown in Appendix E and Table 8 shows the results.

Table 8 Percentile Dependent Equation Simulation Results

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| percentile | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| Naive($z$) 0.10 | interval | 0.5143 | 0.3552 | 0.2874 | 0.2478 | 0.2208 | 0.2011 | 0.1092 |
| | confidence | 0.84483 | 0.90354 | 0.92068 | 0.92904 | 0.93321 | 0.93559 | 0.94646 |
| Naive($t$) 0.10 | interval | 0.8258 | 0.4187 | 0.3358 | 0.2657 | 0.2331 | 0.2105 | |
| | confidence | 0.93183 | 0.93897 | 0.95297 | 0.94543 | 0.94542 | 0.94512 | |
| $p$ eqn 0.10 | interval | 0.5018 | 0.3502 | 0.2845 | 0.2461 | 0.2198 | 0.2006 | 0.1095 |
| | confidence | 0.94452 | 0.94720 | 0.94876 | 0.95013 | 0.94863 | 0.94863 | 0.95067 |
| Naive($z$) 0.25 | interval | 0.9387 | 0.6537 | 0.5269 | 0.4546 | 0.4049 | 0.3694 | 0.2005 |
| | confidence | 0.8441 | 0.90441 | 0.92114 | 0.92852 | 0.93265 | 0.93651 | 0.94595 |
| Naive($t$) 0.25 | interval | 1.5121 | 0.7691 | 0.5828 | 0.4877 | 0.4279 | 0.3860 | |
| | confidence | 0.93106 | 0.93984 | 0.94258 | 0.94350 | 0.94500 | 0.94577 | |
| $p$ eqn 0.25 | interval | 0.8581 | 0.5982 | 0.4869 | 0.4206 | 0.3758 | 0.3427 | 0.1872 |
| | confidence | 0.92547 | 0.92889 | 0.93023 | 0.92976 | 0.93093 | 0.93175 | 0.93172 |
| Naive($z$) 0.50 | interval | 1.8518 | 1.2796 | 1.0363 | 0.8918 | 0.7953 | 0.7248 | 0.3934 |
| | confidence | 0.84452 | 0.90378 | 0.92138 | 0.92944 | 0.93584 | 0.93616 | 0.94707 |
| Naive($t$) 0.50 | interval | 2.9751 | 1.5055 | 1.1434 | 0.9572 | 0.8408 | 0.7583 | |
| | confidence | 0.93329 | 0.94017 | 0.94402 | 0.94577 | 0.94559 | 0.94662 | |
| $p$ eqn 0.50 | interval | 1.7699 | 1.2351 | 1.0043 | 0.8688 | 0.7758 | 0.7075 | 0.3864 |
| | confidence | 0.93936 | 0.94171 | 0.94447 | 0.94552 | 0.94482 | 0.94518 | 0.94703 |
| Naive($z$) 0.75 | interval | 3.6295 | 2.5089 | 2.0337 | 1.7527 | 1.5615 | 1.4228 | 0.7722 |
| | confidence | 0.8461 | 0.90381 | 0.92016 | 0.92821 | 0.93369 | 0.93567 | 0.9454 |
| Naive($t$) 0.75 | interval | 5.8301 | 2.9654 | 2.2417 | 1.8824 | 1.6504 | 1.4867 | |
| | confidence | 0.93179 | 0.93931 | 0.94315 | 0.94542 | 0.94593 | 0.94557 | |
| $p$ eqn 0.75 | interval | 3.6519 | 2.5501 | 2.0741 | 1.7927 | 1.6010 | 1.4604 | 0.7977 |
| | confidence | 0.95158 | 0.95460 | 0.95474 | 0.95554 | 0.95677 | 0.95620 | 0.95624 |
| Naive($z$) 0.90 | interval | 6.6483 | 4.6199 | 3.7343 | 3.2180 | 2.8613 | 2.6080 | 1.4165 |
| | confidence | 0.84499 | 0.90586 | 0.92118 | 0.92787 | 0.93310 | 0.93824 | 0.94615 |
| Naive($t$) 0.90 | interval | 10.6495 | 5.4290 | 4.1142 | 3.4503 | 3.0245 | 2.7286 | |
| | confidence | 0.93128 | 0.93966 | 0.94184 | 0.94423 | 0.94657 | 0.94616 | |
| $p$ eqn 0.90 | interval | 6.2421 | 4.3580 | 3.5441 | 3.0627 | 2.7362 | 2.4959 | 1.3633 |
| | confidence | 0.93409 | 0.93818 | 0.93926 | 0.93837 | 0.94015 | 0.94026 | 0.94026 |

### *3.5 Naive vs. Proposed Approach*

Comparing the $p$ equation to the naive approach using the $z$-statistic, the results from Table 8 demonstrate that the $p$ equation provides smaller interval width and increased confidence in most cases. At the 25th percentile, sample sizes of 25, 30, and 100 offer smaller intervals; however, the confidence level falls slightly lower than that of the naive approach. Similarly, a sample size of 100 at the 90th percentile offers a smaller interval with a slightly smaller confidence level. Unfortunately, at the 75th percentile this approach yields an increased interval width by approximately 0.04 units across the different sample sizes. The benefit however, lies with the confidence level. At the 75th percentile, the confidence level is greater than 95% at all sample sizes. In fact, the greatest benefit of this approach is the higher confidence levels, especially at the smaller sample sizes. The naive approach offers approximately 85%, 90%, and 92% confidence at sample sizes of 5, 10, and 15 respectively, while the new approach offers confidence levels greater than 92.5% in all cases with many near or above the desired 95% level.

Comparing the $p$ equation to the naive approach using the $t$-statistic, the results demonstrate that the $p$ equation provides smaller interval width, much smaller at the small sample sizes, with a confidence level generally within $\pm 1\%$ of the naive approach confidence level. The greatest benefit of the $p$ equation over the $t$-statistic naive approach can be seen at the small sample sizes, $n = 5, 10, 15$, towards the tail, $p = 0.75, 0.90$, of the distribution. From the given results, an educated conjecture would be that as $p \to 1$, the benefit of the $p$ equation would further increase.

The main drawback of this approach is the dependency upon the percentile $p$. Often times, the user will not know $p$. Without knowing the percentile, the $p$ equation can not be utilized. However, with some work, the user is able to determine $p$. The question that remains is whether the amount of work necessary to determine $p$ is worth the improved results of this approach.

The original desired result involved an equation free from dependency upon the percentile. However, due to the nature of the lognormal distribution and the increased variability of the estimates at higher percentiles, it became apparent that this method required a knowledge of $p$. Methods for finding $p$ are examined in the following chapter.

## *IV. Conclusion*

Starting with a lognormal distribution, or a data set that closely resembles a lognormal distribution, a natural logarithm transformation converts the data to a normal distribution. In the normal distribution, data is easily analyzed with respect to regression modeling, and the confidence intervals are narrow with the desired level of confidence. However, upon back-transformation to the original space, the confidence levels either drop to a level less than that of the normal space or the interval width increases to a sometimes impractical width. This phenomenon is more apparent with small sample sizes and towards the right tail of the distribution, i.e. the higher percentiles of the distribution. This is demonstrated for sample sizes of 5, 10, and 15 at the 50th, 75th, and 90th percentile in Table 9. To view full results, see Table 8. Apparent from Table 9 is that the naive approach using

Table 9 Small Sample and High Percentile Confidence Level and Interval Width

| | | Sample Size | | |
|---|---|---|---|---|
| percentile | | 5 | 10 | 15 |
| Naive($z$) 0.50 | interval | 1.8518 | 1.2796 | 1.0363 |
| | confidence | 0.84452 | 0.90378 | 0.92138 |
| Naive($t$) 0.50 | interval | 2.9751 | 1.5055 | 1.1434 |
| | confidence | 0.93329 | 0.94017 | 0.94402 |
| Naive($z$) 0.75 | interval | 3.6295 | 2.5089 | 2.0337 |
| | confidence | 0.8461 | 0.90381 | 0.92016 |
| Naive($t$) 0.75 | interval | 5.8301 | 2.9654 | 2.2417 |
| | confidence | 0.93179 | 0.93931 | 0.94315 |
| Naive($z$) 0.90 | interval | 6.6483 | 4.6199 | 3.7343 |
| | confidence | 0.84499 | 0.90586 | 0.92118 |
| Naive($t$) 0.90 | interval | 10.6495 | 5.4290 | 4.1142 |
| | confidence | 0.93128 | 0.93966 | 0.94184 |

the $z$-statistic offers a lower confidence level [than the 95% desired level] and the naive approach using the $t$-statistic provides much wider intervals.

Table 9 demonstrates some disturbing properties. The naive approach using the $z$-statistic increases the risk factor in making a decision by approximately threefold. A 95% confidence interval in normal space correlates to an 85% confidence interval in the original space. This implies that a decision-maker believes the risk of being incorrect is only 5%, but the risk is actually 15%. Hence, the risk after back-transformation is three times greater. In addition, the naive approach using the $t$-statistic generates much wider interval widths. Especially in the tail of the distribution, these interval widths do not provide a reliable measure, as they are too large to be of practical use. Providing a method that offers equal and/or higher confidence levels with equal and/or smaller intervals define the overall goal of this thesis. That is, one wishes not only to decrease the risk factor involved in making a decision, but also provide practical interval widths upon back-transformation.

The ad hoc simulation approach adopted permits the common framework of the natural logarithm transformation to be explored. That is, the simplicity and the commonality of the natural logarithm transformation created a desire to explore a method that utilized this common approach. The first approach, using $\hat{y} \pm \sqrt[m]{\hat{y}}$, created a starting point. From this reference point, the data painted a path to follow. Ultimately, the $p$ equation: $\hat{y} \pm \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\hat{y}}$ demonstrated promising results. The results for sample sizes of 5, 10, and 15 are shown in Table 10.

Table 10 Comparison of Naive Approaches with $p$ equation

| | | Sample Size | | |
|---|---|---|---|---|
| percentile | | 5 | 10 | 15 |
| Naive($z$) 0.10 | interval | 0.5143 | 0.3552 | 0.2874 |
| | confidence | 0.84483 | 0.90354 | 0.92068 |
| Naive($t$) 0.10 | interval | 0.8258 | 0.4187 | 0.3358 |
| | confidence | 0.93183 | 0.93897 | 0.95297 |
| $p$ eqn 0.10 | interval | 0.5018 | 0.3502 | 0.2845 |
| | confidence | 0.94452 | 0.94720 | 0.94876 |
| Naive($z$) 0.25 | interval | 0.9387 | 0.6537 | 0.5269 |
| | confidence | 0.8441 | 0.90441 | 0.92114 |
| Naive($t$) 0.25 | interval | 1.5121 | 0.7691 | 0.5828 |
| | confidence | 0.93106 | 0.93984 | 0.94258 |
| $p$ eqn 0.25 | interval | 0.8581 | 0.5982 | 0.4869 |
| | confidence | 0.92547 | 0.92889 | 0.93023 |
| Naive($z$) 0.50 | interval | 1.8518 | 1.2796 | 1.0363 |
| | confidence | 0.84452 | 0.90378 | 0.92138 |
| Naive($t$) 0.50 | interval | 2.9751 | 1.5055 | 1.1434 |
| | confidence | 0.93329 | 0.94017 | 0.94402 |
| $p$ eqn 0.50 | interval | 1.7699 | 1.2351 | 1.0043 |
| | confidence | 0.93936 | 0.94171 | 0.94447 |
| Naive($z$) 0.75 | interval | 3.6295 | 2.5089 | 2.0337 |
| | confidence | 0.8461 | 0.90381 | 0.92016 |
| Naive($t$) 0.75 | interval | 5.8301 | 2.9654 | 2.2417 |
| | confidence | 0.93179 | 0.93931 | 0.94315 |
| $p$ eqn 0.75 | interval | 3.6519 | 2.5501 | 2.0741 |
| | confidence | 0.95158 | 0.95460 | 0.95474 |
| Naive($z$) 0.90 | interval | 6.6483 | 4.6199 | 3.7343 |
| | confidence | 0.84499 | 0.90586 | 0.92118 |
| Naive($t$) 0.90 | interval | 10.6495 | 5.4290 | 4.1142 |
| | confidence | 0.93128 | 0.93966 | 0.94184 |
| $p$ eqn 0.90 | interval | 6.2421 | 4.3580 | 3.5441 |
| | confidence | 0.93409 | 0.93818 | 0.93926 |

From Table 10, the $p$ equation provides interval widths comparable to those using the $z$-statistic naive approach and confidence levels comparable to those using the $t$-statistic naive approach. Also, the $p$ equation provides confidence levels of 92.5% or greater. Thus, the threefold increase in risk when using the $z$-statistic naive approach is virtually erased. The risk only slightly increases (from 5% to a maximum of 7.5%) upon back-transformation to the original space. Since small sample confidence intervals are generally evaluated using the $t$-statistic, comparing the $p$ equation to the $t$-statistic method demonstrates the true benefits. The intervals are much smaller when evaluated by the $p$ equation than when evaluated by the $t$-statistic naive approach. When the sample size is 5 ($n = 5$), the intervals provided by the $p$ equation are approximately 40% smaller than the intervals generated by the $t$-statistic naive approach. Similarly, when the sample size is 10 ($n = 10$), there is an approximate 20% decrease in interval width and when the sample size is 15 ($n = 15$), there is an approximate 10-15% decrease in interval width. Because the interval widths are wider at higher percentiles, it appears that the $p$ equation offers even greater benefits, in the form of total numeric decrease in interval width, as the percentile approaches 1. This is beneficial since the interval widths of impractical width are generally at higher percentiles where the variance is higher.

Obviously, the $p$ equation offers better results, especially at small sample sizes and higher percentiles. However, there remains some work for the user to accomplish to be able to utilize the $p$ equation. The percentile of the sample mean must be known. Finding the percentile proves to be fairly simple. After the transformation to normal space, the range ($R$) can be calculated as $R$ = maximum observation - minimum observation. Next, calculating the standard deviation ($s$) via the empirical rule follows, using $s = \frac{R}{4}$ (Mendenhall and others, 1999:66-67). Dividing the range by 4 stems from using small sample sizes; large sample sizes would warrant dividing the range by 6. Determining the standard z-score utilizes the equation,

$$\mathrm{z} = \frac{x - \bar{x}}{s},$$

where

$$x = \text{sample mean},$$

$$\bar{x} = \text{normal mean,}$$
$$s = \text{standard deviation.}$$

From the standard z-score, a conversion to percentile uses a table such as the one in Appendix G (Appendix A: The Conversion Table: How to Use it For Converting Scores, 2000-06). Figure 4 (The Empirical Rule, 2006) demonstrates approximate percentile and z-score conversion. With $p$ known, the user can utilize the $p$ equation to find more desirable results.

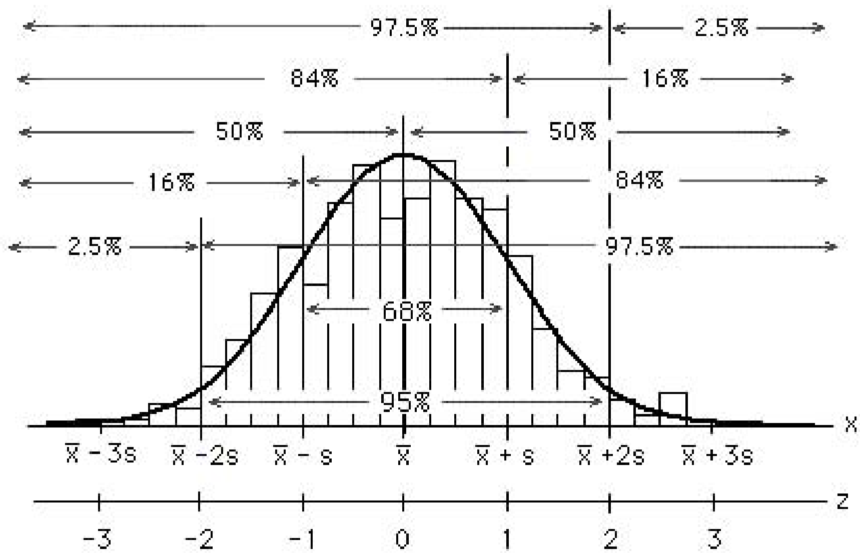


Figure 4 Percentile Relative to Standard z-score

Working within the normal distribution, the probability content within 1, 2, or 3 standard deviations of the mean is (Casella and Berger, 2002:104)

$$P(|X-\mu| \leq \sigma) = P(|Z| \leq 1) = .6826$$
$$P(|X-\mu| \leq 2\sigma) = P(|Z| \leq 2) = .9544$$
$$P(|X-\mu| \leq 3\sigma) = P(|Z| \leq 3) = .9974$$

where

$$X \sim n(\mu, \sigma^2)$$
$$Z \sim n(0, 1).$$

Since it is known that the $p$ equation works well for percentiles above 50, if the sample mean is greater than 1 standard deviation above the mean of the normal distribution, then the $p$ equation will definitely offer better results than the naive approach.

The $p$ equation definitely offers better results than the naive approach at small sample sizes and high percentiles. With the percentile being fairly easy to calculate, this alternative approach is simple to implement. While the $p$ equation provides promising results everywhere, the dramatic results fall under small sample size.

Several questions arise when considering the $p$ equation. With the demonstrated benefits and results of the $p$ equation, is the approach suggested here the only approach or the best approach? The answer to this question is probably not. The $p$ equation is not provided to be the ultimate answer to our original dilemma. The $p$ equation is offered to provide an alternative approach to the commonly applied naive approach. In addition, we have demonstrated that our $p$ equation performs better than the naive approach. Also, one could explore the analytical properties of the $p$ equation to attempt to determine exactly why it performs better than the naive approach.

The $p$ equation offers an alternative to the naive approach. Easy to implement, the $p$ equation outperforms the naive approach, especially at small sample sizes. The results demonstrate a considerable decrease in interval width (40% at a sample size of 5) when compared to the $t$-statistic naive approach and a dramatic decrease in risk (at least 50%) when compared to the $z$-statistic naive approach. Possibly not the ultimate equation, the $p$ equation offers very promising results and provides an example for further research.

Table 11 The function $g(t)$

| t \ n | 10 | 20 | 30 | 40 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| 0.05 | 1.0458 | 1.0485 | 1.0494 | 1.0499 | 1.0502 | 1.0507 | 1.0510 | 1.0512 | 1.0512 |
| 0.10 | 1.0934 | 1.0992 | 1.1012 | 1.1022 | 1.1028 | 1.1040 | 1.1046 | 1.1049 | 1.1050 |
| 0.15 | 1.1427 | 1.1521 | 1.1553 | 1.1569 | 1.1579 | 1.1598 | 1.1608 | 1.1614 | 1.1616 |
| 0.20 | 1.1938 | 1.2072 | 1.2118 | 1.2142 | 1.2156 | 1.2185 | 1.2199 | 1.2208 | 1.2211 |
| 0.25 | 1.2468 | 1.2648 | 1.2710 | 1.2742 | 1.2761 | 1.2800 | 1.2820 | 1.2832 | 1.2836 |
| 0.30 | 1.3018 | 1.3248 | 1.3329 | 1.3370 | 1.3395 | 1.3446 | 1.3472 | 1.3488 | 1.3493 |
| 0.35 | 1.3587 | 1.3874 | 1.3976 | 1.4028 | 1.4060 | 1.4124 | 1.4157 | 1.4177 | 1.4184 |
| 0.40 | 1.4177 | 1.4527 | 1.4652 | 1.4716 | 1.4756 | 1.4836 | 1.4877 | 1.4902 | 1.4910 |
| 0.45 | 1.4788 | 1.5207 | 1.5359 | 1.5437 | 1.5485 | 1.5582 | 1.5632 | 1.5663 | 1.5673 |
| 0.50 | 1.5421 | 1.5917 | 1.6097 | 1.6191 | 1.6248 | 1.6366 | 1.6426 | 1.6463 | 1.6475 |
| 0.55 | 1.6076 | 1.6657 | 1.6869 | 1.6980 | 1.7048 | 1.7188 | 1.7259 | 1.7303 | 1.7318 |
| 0.60 | 1.6754 | 1.7428 | 1.7676 | 1.7806 | 1.7886 | 1.8050 | 1.8135 | 1.8186 | 1.8204 |
| 0.65 | 1.7457 | 1.8231 | 1.8519 | 1.8670 | 1.8763 | 1.8955 | 1.9054 | 1.9115 | 1.9135 |
| 0.70 | 1.8184 | 1.9068 | 1.9399 | 1.9574 | 1.9681 | 1.9904 | 2.0019 | 2.0090 | 2.0114 |
| 0.75 | 1.8936 | 1.9940 | 2.0319 | 2.0519 | 2.0643 | 2.0900 | 2.1033 | 2.1115 | 2.1142 |
| 0.80 | 1.9714 | 2.0848 | 2.1279 | 2.1508 | 2.1650 | 2.1944 | 2.2098 | 2.2192 | 2.2223 |
| 0.85 | 2.0519 | 2.1794 | 2.2283 | 2.2542 | 2.2703 | 2.3040 | 2.3215 | 2.3323 | 2.3360 |
| 0.90 | 2.1352 | 2.2779 | 2.3330 | 2.3624 | 2.3807 | 2.4189 | 2.4389 | 2.4512 | 2.4554 |
| 0.95 | 2.2214 | 2.3804 | 2.4424 | 2.4755 | 2.4962 | 2.5395 | 2.5622 | 2.7075 | 2.5809 |
| 1.00 | 2.3104 | 2.4872 | 2.5565 | 2.5938 | 2.6170 | 2.6659 | 2.6916 | 2.5762 | 2.7129 |

Table 12 The function $g(t)$

| t \ n | 10 | 20 | 30 | 40 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| 1.05 | 2.4025 | 2.5984 | 2.6757 | 2.7174 | 2.7435 | 2.7985 | 2.8275 | 2.8454 | 2.8515 |
| 1.10 | 2.4977 | 2.7141 | 2.8002 | 2.8467 | 2.8759 | 2.9376 | 2.9702 | 2.9904 | 2.9973 |
| 1.15 | 2.5961 | 2.8345 | 2.9300 | 2.9818 | 3.0144 | 3.0834 | 3.1200 | 3.1427 | 3.1504 |
| 1.20 | 2.6978 | 2.9597 | 3.0655 | 3.1231 | 3.1594 | 3.2363 | 3.2773 | 3.3027 | 3.3114 |
| 1.25 | 2.8028 | 3.0901 | 3.2069 | 3.2707 | 3.3110 | 3.3967 | 3.4424 | 3.4709 | 3.4806 |
| 1.30 | 2.9114 | 3.2257 | 3.3544 | 3.4250 | 3.4696 | 3.5649 | 3.6158 | 3.6476 | 3.6584 |
| 1.35 | 3.0235 | 3.3668 | 3.5084 | 3.5862 | 3.6356 | 3.7412 | 3.7978 | 3.8332 | 3.8453 |
| 1.40 | 3.1393 | 3.5135 | 3.6689 | 3.7547 | 3.8092 | 3.9260 | 3.9889 | 4.0282 | 4.0417 |
| 1.45 | 3.2589 | 3.6661 | 3.8364 | 3.9307 | 3.9908 | 4.1199 | 4.1895 | 4.2332 | 4.2481 |
| 1.50 | 3.3824 | 3.8247 | 4.0111 | 4.1146 | 4.1807 | 4.3231 | 4.4001 | 4.4485 | 4.4650 |
| 1.55 | 3.5099 | 3.9897 | 4.1933 | 4.3068 | 4.3793 | 4.5361 | 4.6212 | 4.6747 | 4.6930 |
| 1.60 | 3.6415 | 4.1612 | 4.3832 | 4.5074 | 4.5870 | 4.7594 | 4.8532 | 4.9124 | 4.9326 |
| 1.65 | 3.7774 | 4.3394 | 4.5813 | 4.7171 | 4.8042 | 4.9935 | 5.0968 | 5.1621 | 5.1844 |
| 1.70 | 3.9176 | 4.5247 | 4.7878 | 4.9360 | 5.0313 | 5.2389 | 5.3525 | 5.4244 | 5.4490 |
| 1.75 | 4.0623 | 4.7173 | 5.0031 | 5.1646 | 5.2687 | 5.4961 | 5.6209 | 5.7000 | 5.7271 |
| 1.80 | 4.2116 | 4.9174 | 5.2275 | 5.4034 | 5.5170 | 5.7657 | 5.9027 | 5.9896 | 6.0194 |
| 1.85 | 4.3657 | 5.1253 | 5.4614 | 5.6527 | 5.7764 | 6.0482 | 6.1984 | 6.2938 | 6.3265 |
| 1.90 | 4.5246 | 5.3413 | 5.7052 | 5.9129 | 6.0477 | 6.3444 | 6.5087 | 6.6134 | 6.6493 |
| 1.95 | 4.6885 | 5.5657 | 5.9592 | 6.1847 | 6.3312 | 6.6547 | 6.8345 | 6.9491 | 6.9886 |
| 2.00 | 4.8575 | 5.7988 | 6.2239 | 6.4684 | 6.6276 | 6.9800 | 7.1764 | 7.3019 | 7.3451 |

Table 13 The function $g_1(t)$

| t \ n | 10 | 20 | 30 | 40 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| 0.05 | 0.0527 | 0.0533 | 0.0535 | 0.0536 | 0.0536 | 0.0538 | 0.0538 | 0.0539 | 0.0539 |
| 0.10 | 0.1112 | 0.1135 | 0.1143 | 0.1148 | 0.1151 | 0.1156 | 0.1159 | 0.1161 | 0.1162 |
| 0.15 | 0.1757 | 0.1812 | 0.1833 | 0.1844 | 0.1851 | 0.1865 | 0.1873 | 0.1877 | 0.1879 |
| 0.20 | 0.2468 | 0.2573 | 0.2613 | 0.2634 | 0.2648 | 0.2675 | 0.2690 | 0.2697 | 0.2701 |
| 0.25 | 0.3249 | 0.3423 | 0.3491 | 0.3527 | 0.3550 | 0.3597 | 0.3622 | 0.3635 | 0.3642 |
| 0.30 | 0.4105 | 0.4372 | 0.4477 | 0.4534 | 0.4569 | 0.4644 | 0.4682 | 0.4705 | 0.4714 |
| 0.35 | 0.5041 | 0.5428 | 0.5582 | 0.5666 | 0.5718 | 0.5828 | 0.5887 | 0.5920 | 0.5935 |
| 0.40 | 0.6063 | 0.6599 | 0.6817 | 0.6935 | 0.7010 | 0.7167 | 0.7250 | 0.7299 | 0.7320 |
| 0.45 | 0.7176 | 0.7897 | 0.8194 | 0.8356 | 0.8459 | 0.8676 | 0.8792 | 0.8861 | 0.8888 |
| 0.50 | 0.8386 | 0.9332 | 0.9726 | 0.9943 | 1.0081 | 1.0374 | 1.0531 | 1.0625 | 1.0662 |
| 0.55 | 0.9699 | 1.0916 | 1.1429 | 1.1713 | 1.1894 | 1.2281 | 1.2490 | 1.2616 | 1.2664 |
| 0.60 | 1.1123 | 1.2660 | 1.3317 | 1.3683 | 1.3917 | 1.4420 | 1.4692 | 1.4858 | 1.4921 |
| 0.65 | 1.2664 | 1.4579 | 1.5408 | 1.5873 | 1.6170 | 1.6815 | 1.7166 | 1.7380 | 1.7461 |
| 0.70 | 1.4329 | 1.6687 | 1.7720 | 1.8303 | 1.8678 | 1.9494 | 1.9939 | 2.0214 | 2.0317 |
| 0.75 | 1.6128 | 1.8999 | 2.0273 | 2.0996 | 2.1463 | 2.2485 | 2.3046 | 2.3395 | 2.3523 |
| 0.80 | 1.8067 | 2.1531 | 2.3088 | 2.3978 | 2.4554 | 2.5822 | 2.6522 | 2.6959 | 2.7120 |
| 0.85 | 2.0155 | 2.4301 | 2.6189 | 2.7274 | 2.7981 | 2.9541 | 3.0408 | 3.0951 | 3.1150 |
| 0.90 | 2.2402 | 2.7329 | 2.9600 | 3.0915 | 3.1774 | 3.3681 | 3.4746 | 3.5417 | 3.5662 |
| 0.95 | 2.4817 | 3.0634 | 3.3350 | 3.4932 | 3.5969 | 3.8285 | 3.9586 | 4.0410 | 4.0709 |
| 1.00 | 2.7410 | 3.4239 | 3.7467 | 3.9359 | 4.0605 | 4.3401 | 4.4981 | 4.5986 | 4.6349 |

Table 14 One-sided (upper) confidence limits - 15 degrees of freedom

| S | .0025 | .005 | .01 | .025 | .05 | .10 |
|---|---|---|---|---|---|---|
| 0.10 | -3.057 | -2.753 | -2.442 | -2.012 | -1.663 | -1.278 |
| 0.20 | -2.959 | -2.675 | -2.383 | -1.974 | -1.639 | -1.267 |
| 0.30 | -2.883 | -2.616 | -2.339 | -1.949 | -1.625 | -1.261 |
| 0.40 | -2.828 | -2.575 | -2.310 | -1.934 | -1.618 | -1.262 |
| 0.50 | -2.791 | -2.548 | -2.293 | -1.928 | -1.620 | -1.269 |
| 0.60 | -2.769 | -2.535 | -2.288 | -1.931 | -1.629 | -1.280 |
| 0.70 | -2.759 | -2.533 | -2.292 | -1.942 | -1.643 | -1.296 |
| 0.80 | -2.761 | -2.540 | -2.304 | -1.959 | -1.662 | -1.315 |
| 0.90 | -2.774 | -2.557 | -2.324 | -1.983 | -1.686 | -1.338 |
| 1.00 | -2.794 | -2.581 | -2.351 | -2.012 | -1.715 | -1.364 |
| 1.25 | -2.878 | -2.670 | -2.443 | -2.104 | -1.803 | -1.441 |
| 1.50 | -2.997 | -2.790 | -2.563 | -2.218 | -1.909 | -1.533 |
| 1.75 | -3.144 | -2.935 | -2.704 | -2.351 | -2.029 | -1.634 |
| 2.00 | -3.311 | -3.099 | -2.862 | -2.496 | -2.162 | -1.746 |
| 2.50 | -3.693 | -3.468 | -3.216 | -2.821 | -2.452 | -1.987 |
| 3.00 | -4.118 | -3.878 | -3.605 | -3.174 | -2.767 | -2.248 |
| 3.50 | -4.572 | -4.314 | -4.019 | -3.547 | -3.099 | -2.522 |
| 4.00 | -5.047 | -4.769 | -4.449 | -3.935 | -3.443 | -2.805 |
| 4.50 | -5.536 | -5.237 | -4.891 | -4.332 | -3.794 | -3.095 |
| 5.00 | -6.037 | -5.716 | -5.343 | -4.738 | -4.153 | -3.390 |
| 6.00 | -7.062 | -6.694 | -6.264 | -5.564 | -4.882 | -3.989 |
| 7.00 | -8.109 | -7.692 | -7.204 | -6.404 | -5.624 | -4.599 |
| 8.00 | -9.170 | -8.702 | -8.154 | -7.254 | -6.374 | -5.213 |
| 9.00 | -10.24 | -9.721 | -9.113 | -8.111 | -7.129 | -5.833 |
| 10.00 | -11.32 | -10.75 | -10.08 | -8.972 | -7.888 | -6.455 |

Table 15 One-sided (upper) confidence limits - 15 degrees of freedom

| S | .90 | .95 | .975 | .99 | .995 | .9975 |
|---|---|---|---|---|---|---|
| 0.10 | 1.325 | 1.743 | 2.130 | 2.618 | 2.978 | 3.337 |
| 0.20 | 1.361 | 1.800 | 2.212 | 2.737 | 3.130 | 3.525 |
| 0.30 | 1.406 | 1.871 | 2.311 | 2.880 | 3.312 | 3.749 |
| 0.40 | 1.460 | 1.954 | 2.428 | 3.047 | 3.523 | 4.010 |
| 0.50 | 1.524 | 2.050 | 2.562 | 3.239 | 3.763 | 4.307 |
| 0.60 | 1.596 | 2.160 | 2.712 | 3.453 | 4.032 | 4.638 |
| 0.70 | 1.677 | 2.280 | 2.879 | 3.687 | 4.326 | 4.998 |
| 0.80 | 1.765 | 2.412 | 3.059 | 3.940 | 4.642 | 5.384 |
| 0.90 | 1.861 | 2.554 | 3.251 | 4.209 | 4.976 | 5.791 |
| 1.00 | 1.963 | 2.704 | 3.454 | 4.491 | 5.325 | 6.215 |
| 1.25 | 2.242 | 3.109 | 3.998 | 5.240 | 6.249 | 7.332 |
| 1.50 | 2.544 | 3.544 | 4.579 | 6.034 | 7.223 | 8.502 |
| 1.75 | 2.862 | 4.000 | 5.183 | 6.857 | 8.228 | 9.707 |
| 2.00 | 3.191 | 4.470 | 5.804 | 7.699 | 9.254 | 10.93 |
| 2.50 | 3.870 | 5.435 | 7.078 | 9.415 | 11.34 | 13.43 |
| 3.00 | 4.565 | 6.422 | 8.376 | 11.17 | 13.47 | 15.96 |
| 3.50 | 5.271 | 7.422 | 9.689 | 12.93 | 15.61 | 18.51 |
| 4.00 | 5.983 | 8.429 | 11.01 | 14.71 | 17.76 | 21.07 |
| 4.50 | 6.699 | 9.442 | 12.34 | 16.49 | 19.92 | 23.64 |
| 5.00 | 7.418 | 10.46 | 13.67 | 18.28 | 22.09 | 26.22 |
| 6.00 | 8.862 | 12.50 | 16.35 | 21.87 | 26.43 | 31.39 |
| 7.00 | 10.31 | 14.55 | 19.03 | 25.46 | 30.78 | 36.56 |
| 8.00 | 11.77 | 16.60 | 21.72 | 29.06 | 35.14 | 41.74 |
| 9.00 | 13.22 | 18.65 | 24.41 | 32.67 | 39.51 | 46.93 |
| 10.00 | 14.68 | 20.71 | 27.10 | 36.28 | 43.87 | 52.12 |

Table 16 Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 11$

| $\sigma^2$ | Methods | Coverage probability | Coverage error | Length | % CI $> \log\theta$ | % CI $< \log\theta$ | Relative bias |
|---|---|---|---|---|---|---|---|
| 0.1 | naive | 0.8134 | 0.0866 | 0.3053 | 0.0242 | 0.1624 | 0.7406 |
| | conservative | 0.9582 | 0.0582 | 0.5166 | 0.0374 | 0.0044 | 0.7895 |
| | parametric B | 0.8996 | 0.0004 | 0.3557 | 0.0510 | 0.0494 | 0.0159 |
| | Cox's method | 0.8636 | 0.0364 | 0.3144 | 0.0508 | 0.0856 | 0.2551 |
| 0.5 | naive | 0.6442 | 0.2558 | 0.6828 | 0.0042 | 0.3516 | 0.9764 |
| | conservative | 0.9660 | 0.0660 | 1.2687 | 0.0260 | 0.0080 | 0.5294 |
| | parametric B | 0.9108 | 0.0108 | 0.9513 | 0.0438 | 0.0454 | 0.0179 |
| | Cox's method | 0.8664 | 0.0336 | 0.7783 | 0.0344 | 0.0992 | 0.4850 |
| 1.0 | naive | 0.4758 | 0.4242 | 0.9656 | 0.0008 | 0.5234 | 0.9969 |
| | conservative | 0.9744 | 0.0744 | 1.9748 | 0.0140 | 0.0116 | 0.0938 |
| | parametric B | 0.9170 | 0.0170 | 1.5901 | 0.0380 | 0.0450 | 0.0843 |
| | Cox's method | 0.8638 | 0.0362 | 1.2195 | 0.0240 | 0.1122 | 0.8238 |
| 2.0 | naive | 0.2404 | 0.6596 | 1.3655 | 0.0002 | 0.7594 | 0.9997 |
| | conservative | 0.9768 | 0.0768 | 3.2401 | 0.0054 | 0.0178 | 0.5345 |
| | parametric B | 0.9334 | 0.0334 | 2.8540 | 0.0230 | 0.0436 | 0.3093 |
| | Cox's method | 0.8614 | 0.0386 | 2.0167 | 0.0108 | 0.1278 | 0.8442 |
| 5.0 | naive | 0.0246 | 0.8754 | 2.1591 | 0.0000 | 0.9754 | 1.0000 |
| | conservative | 0.9706 | 0.0706 | 6.8054 | 0.0012 | 0.0282 | 0.9184 |
| | parametric B | 0.9506 | 0.0506 | 6.7812 | 0.0094 | 0.0400 | 0.6194 |
| | Cox's method | 0.8514 | 0.0486 | 4.2774 | 0.0024 | 0.1462 | 0.9677 |
| 20.0 | naive | 0.0000 | 0.9000 | 4.3182 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9576 | 0.0576 | 24.1736 | 0.0000 | 0.0424 | 1.0000 |
| | parametric B | 0.9632 | 0.0632 | 27.4496 | 0.0044 | 0.0324 | 0.7609 |
| | Cox's method | 0.8376 | 0.0624 | 15.3278 | 0.0004 | 0.1620 | 0.9951 |

Table 17 Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 101$

| $\sigma^2$ | Methods | Coverage probability | Coverage error | Length | % CI $> \log\theta$ | % CI $< \log\theta$ | Relative bias |
|---|---|---|---|---|---|---|---|
| 0.1 | naive | 0.5102 | 0.3898 | 0.1032 | 0.0008 | 0.4890 | 0.9967 |
| | conservative | 0.9306 | 0.0306 | 0.1181 | 0.0446 | 0.0248 | 0.2853 |
| | parametric B | 0.9076 | 0.0076 | 0.1093 | 0.0460 | 0.0464 | 0.0043 |
| | Cox's method | 0.8946 | 0.0054 | 0.1058 | 0.0454 | 0.0600 | 0.1385 |
| 0.5 | naive | 0.0280 | 0.8720 | 0.2308 | 0.0000 | 0.9720 | 1.0000 |
| | conservative | 0.9298 | 0.0298 | 0.2883 | 0.0386 | 0.0316 | 0.0997 |
| | parametric B | 0.9288 | 0.0288 | 0.2861 | 0.0348 | 0.0364 | 0.0225 |
| | Cox's method | 0.8964 | 0.0036 | 0.2585 | 0.0408 | 0.0628 | 0.2124 |
| 1.0 | naive | 0.0002 | 0.8998 | 0.3264 | 0.0000 | 0.9998 | 1.0000 |
| | conservative | 0.9284 | 0.0284 | 0.4468 | 0.0360 | 0.0356 | 0.0056 |
| | parametric B | 0.9408 | 0.0408 | 0.4683 | 0.0272 | 0.0320 | 0.0811 |
| | Cox's method | 0.8982 | 0.0018 | 0.4009 | 0.0366 | 0.0652 | 0.2809 |
| 2.0 | naive | 0.0000 | 0.9000 | 0.4616 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9306 | 0.0306 | 0.7300 | 0.0308 | 0.0386 | 0.1124 |
| | parametric B | 0.9538 | 0.0538 | 0.8140 | 0.0192 | 0.0270 | 0.1688 |
| | Cox's method | 0.9000 | 0.0000 | 0.6555 | 0.0314 | 0.0686 | 0.3720 |
| 5.0 | naive | 0.0000 | 0.9000 | 0.7298 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9314 | 0.0314 | 1.5274 | 0.0250 | 0.0436 | 0.2711 |
| | parametric B | 0.9682 | 0.0682 | 1.8319 | 0.0118 | 0.0200 | 0.2579 |
| | Cox's method | 0.9028 | 0.0028 | 1.3729 | 0.0252 | 0.0720 | 0.4815 |
| 20.0 | naive | 0.0000 | 0.9000 | 1.4596 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9302 | 0.0302 | 5.4161 | 0.0230 | 0.0468 | 0.3410 |
| | parametric B | 0.9728 | 0.0728 | 6.9070 | 0.0098 | 0.0174 | 0.2794 |
| | Cox's method | 0.8962 | 0.0038 | 4.8731 | 0.0236 | 0.0802 | 0.5453 |

Table 18 Coverage probabilities, coverage errors, length and relative biases of two-sided 90% CI for various methods with $\mu = -\frac{\sigma^2}{2}$ and $n = 400$

| $\sigma^2$ | Methods | Coverage probability | Coverage error | Length | % CI $> \log\theta$ | % CI $< \log\theta$ | Relative bias |
|---|---|---|---|---|---|---|---|
| 0.1 | naive | 0.0670 | 0.8330 | 0.0520 | 0.0000 | 0.9330 | 1.0000 |
| | conservative | 0.9154 | 0.0154 | 0.0560 | 0.0490 | 0.0356 | 0.1584 |
| | parametric B | 0.9092 | 0.0092 | 0.0546 | 0.0482 | 0.0426 | 0.0617 |
| | Cox's method | 0.9012 | 0.0012 | 0.0532 | 0.0494 | 0.0494 | 0.0000 |
| 0.5 | naive | 0.0000 | 0.9000 | 0.1162 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9136 | 0.0136 | 0.1366 | 0.0472 | 0.0392 | 0.0926 |
| | parametric B | 0.9286 | 0.0286 | 0.1426 | 0.0368 | 0.0346 | 0.0308 |
| | Cox's method | 0.8992 | 0.0008 | 0.1299 | 0.0474 | 0.0534 | 0.0595 |
| 1.0 | naive | 0.0000 | 0.9000 | 0.1643 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9144 | 0.0144 | 0.2117 | 0.0442 | 0.0414 | 0.0327 |
| | parametric B | 0.9388 | 0.0388 | 0.2330 | 0.0306 | 0.0306 | 0.0000 |
| | Cox's method | 0.9006 | 0.0006 | 0.2013 | 0.0446 | 0.0548 | 0.1026 |
| 2.0 | naive | 0.0000 | 0.9000 | 0.2324 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9170 | 0.0170 | 0.3457 | 0.0410 | 0.0420 | 0.0120 |
| | parametric B | 0.9506 | 0.0506 | 0.4041 | 0.0228 | 0.0266 | 0.0040 |
| | Cox's method | 0.9010 | 0.0010 | 0.3289 | 0.0412 | 0.0578 | 0.1677 |
| 5.0 | naive | 0.0000 | 0.9000 | 0.3674 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9140 | 0.0140 | 0.7230 | 0.0394 | 0.0466 | 0.0837 |
| | parametric B | 0.9662 | 0.0662 | 0.9052 | 0.0134 | 0.0204 | 0.2071 |
| | Cox's method | 0.8968 | 0.0032 | 0.6881 | 0.0398 | 0.0634 | 0.2287 |
| 20.0 | naive | 0.0000 | 0.9000 | 0.7348 | 0.0000 | 1.0000 | 1.0000 |
| | conservative | 0.9138 | 0.0138 | 2.5635 | 0.0354 | 0.0508 | 0.1787 |
| | parametric B | 0.9746 | 0.0746 | 3.3932 | 0.0098 | 0.0156 | 0.2283 |
| | Cox's method | 0.8960 | 0.0040 | 2.4402 | 0.0356 | 0.0684 | 0.3154 |

Table 19 Percent of all intervals that cover the true parameter value

| | Naive approach | | | Cox method | | | Modified Cox method | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | Below | Covering | Above | Below | Covering | Above | Below | Covering | Above |
| 5 | 13.5 | 86.2 | 0.3 | 10.6 | 87.2 | 2.2 | 5.9 | 93.5 | 0.6 |
| 10 | 31.3 | 68.5 | 0.0 | 8.2 | 91.1 | 0.7 | 5.9 | 93.9 | 0.2 |
| 20 | 54.8 | 45.2 | 0.0 | 4.8 | 94.2 | 1.0 | 3.6 | 95.7 | 0.7 |
| 30 | 75.9 | 24.1 | 0.0 | 6.5 | 92.6 | 0.9 | 5.4 | 93.9 | 0.7 |
| 50 | 94.3 | 5.7 | 0.3 | 4.0 | 95.4 | 0.6 | 3.9 | 95.5 | 0.6 |
| 100 | 99.9 | 0.1 | 0.0 | 3.3 | 95.5 | 1.2 | 3.2 | 95.7 | 1.1 |
| 200 | 100.0 | 0.0 | 0.0 | 2.6 | 95.2 | 2.2 | 2.6 | 95.2 | 2.2 |
| 500 | 100.0 | 0.0 | 0.0 | 3.0 | 95.1 | 1.9 | 3.0 | 95.1 | 1.9 |
| 1000 | 100.0 | 0.0 | 0.0 | 3.3 | 94.4 | 2.3 | 3.3 | 94.4 | 2.3 |

| | Large sample approach | | | Generalized CI | | |
|---|---|---|---|---|---|---|
| $n$ | Below | Covering | Above | Below | Covering | Above |
| 5 | 16.8 | 83.0 | 0.2 | 1.3 | 94.1 | 4.6 |
| 10 | 16.4 | 83.6 | 0.0 | 2.2 | 93.7 | 4.1 |
| 20 | 12.0 | 87.9 | 0.1 | 1.9 | 95.2 | 2.9 |
| 30 | 14.0 | 85.6 | 0.4 | 2.1 | 94.6 | 3.3 |
| 50 | 9.4 | 90.4 | 0.2 | 2.2 | 95.0 | 2.8 |
| 100 | 7.6 | 92.1 | 0.3 | 2.9 | 93.7 | 3.4 |
| 200 | 6.5 | 92.2 | 1.3 | 1.3 | 95.9 | 2.8 |
| 500 | 4.9 | 94.0 | 1.1 | 2.8 | 94.2 | 3.0 |
| 1000 | 4.8 | 93.8 | 1.4 | 2.3 | 95.8 | 1.9 |

```
X = zeros(100000, n);
for k = 1 : 100000
    X(k, :) = j + randn(1, n);
end
y = mean(X', 1);
s = std(X', 1);
L_transformed = y − 1.96s/√n;
U_transformed = y + 1.96s/√n;
L = exp(L_transformed);
U = exp(U_transformed);
interval = U − L;
int = mean(interval', 1)
ininterval = length(find(i ≤ U  &  i ≥ L))
percent = ininterval/100000
```

where

$$n = 5, 10, 15, 20, 25, 30, 100$$
$$i = 0.2776, 0.5094, 1.0000, 1.9630, 3.6022$$
$$j = -1.2816, -0.6745, 0.0000, 0.6745, 1.2816$$

Simulation Code: $\hat{y} \pm \sqrt[m]{\hat{y}}$

$X = \text{zeros}(100000, n);$
**for** $k = 1 : 100000$
$\quad X(k,:) = j + \text{randn}(1, n);$
**end**
$y = \text{mean}(X', 1);$
$s = \text{std}(X', 1);$
**for** $m = 2 : 100$
$\quad L = \exp(y) - \sqrt[m]{\exp(y)};$
$\quad U = \exp(y) + \sqrt[m]{\exp(y)};$
$\quad \text{interval} = U - L;$
$\quad int = \text{mean}(\text{interval}', 1)$
$\quad \text{ininterval} = \text{length}(\text{find}(i \leq U \;\; \& \;\; i \geq L))$
$\quad \text{percent} = \text{ininterval}/100000$
**end**

where

$$
\begin{aligned}
n &= 5, 10, 15, 20, 25, 30, 100 \\
i &= 0.2776, 0.5094, 1.0000, 1.9630, 3.6022 \\
j &= -1.2816, -0.6745, 0.0000, 0.6745, 1.2816
\end{aligned}
$$

Simulation Code: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

$X = \text{zeros}(100000, n);$
`for` $k = 1 : 100000$
  $X(k, :) = j + \text{randn}(1, n);$
`end`
$y = \text{mean}(X', 1);$
$s = \text{std}(X', 1);$
`for` $m = 2 : 10$
  $L = \exp(y) - \frac{1}{m}\sqrt{\exp(y)};$
  $U = \exp(y) + \frac{1}{m}\sqrt{\exp(y)};$
  $\text{interval} = U - L;$
  $int = \text{mean}(\text{interval}', 1)$
  $\text{ininterval} = \text{length}(\text{find}(i \leq U \;\; \& \;\; i \geq L))$
  $\text{percent} = \text{ininterval}/100000$
`end`

where

$$\begin{aligned} n &= 5, 10, 15, 20, 25, 30, 100 \\ i &= 0.2776, 0.5094, 1.0000, 1.9630, 3.6022 \\ j &= -1.2816, -0.6745, 0.0000, 0.6745, 1.2816 \end{aligned}$$

Simulation Code: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

$X = \text{zeros}(100000, n);$
for $k = 1 : 100000$
$\quad X(k,:) = j + \text{randn}(1, n);$
end
$y = \text{mean}(X', 1);$
$s = \text{std}(X', 1);$
$L = \exp(y) - \sqrt{\frac{a\exp(y)}{n}};$
$U = \exp(y) + \sqrt{\frac{a\exp(y)}{n}};$
$\text{interval} = U - L;$
$int = \text{mean}(\text{interval}', 1)$
$\text{ininterval} = \text{length}(\text{find}(i \leq U \;\;\&\;\; i \geq L))$
$\text{percent} = \text{ininterval}/100000$

where

$$
\begin{aligned}
n &= 5, 10, 15, 20, 25, 30, 100 \\
a &= 1, 2, \ldots, 15 \\
i &= 0.2776, 0.5094, 1.0000, 1.9630, 3.6022 \\
j &= -1.2816, -0.6745, 0.0000, 0.6745, 1.2816
\end{aligned}
$$

Simulation Code: $\exp(y) \pm \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\exp(y)}$

$X = \text{zeros}(100000, n);$
`for` $k = 1 : 100000$
    $X(k,:) = j + \text{randn}(1, n);$
`end`
$y = \text{mean}(X', 1);$
$s = \text{std}(X', 1);$
$L = \exp(y) - \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\exp(y)};$
$U = \exp(y) + \frac{\frac{8}{9}\left(\frac{3\pi}{2}\right)^p}{\sqrt{n}}\sqrt{\exp(y)};$
$\text{interval} = U - L;$
$int = \text{mean}(\text{interval}', 1)$
$\text{ininterval} = \text{length}(\text{find}(i \leq U \;\; \& \;\; i \geq L))$
$\text{percent} = \text{ininterval}/100000$
`end`

where

$$
\begin{aligned}
n &= 5, 10, 15, 20, 25, 30, 100 \\
p &= 0.10, 0.25, 0.50, 0.75, 0.90 \\
i &= 0.2776, 0.5094, 1.0000, 1.9630, 3.6022 \\
j &= -1.2816, -0.6745, 0.0000, 0.6745, 1.2816
\end{aligned}
$$

Table 20 10th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.5143 | 0.3552 | 0.2874 | 0.2478 | 0.2208 | 0.2011 | 0.1092 |
| | confidence | 0.84483 | 0.90354 | 0.92068 | 0.92904 | 0.93321 | 0.93559 | 0.94646 |
| naive($t$) | interval | 0.8258 | 0.4187 | 0.3358 | 0.2657 | 0.2331 | 0.2105 | |
| | confidence | 0.93183 | 0.93897 | 0.95297 | 0.94543 | 0.94542 | 0.94512 | |
| 2 | interval | 1.0813 | 1.0672 | 1.0624 | 1.0602 | 1.0590 | 1.0578 | 1.0554 |
| | confidence | 0.99988 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 3 | interval | 1.3200 | 1.3121 | 1.3094 | 1.3082 | 1.3076 | 1.3068 | 1.3056 |
| | confidence | 0.99990 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 4 | interval | 1.4614 | 1.4564 | 1.4546 | 1.4539 | 1.4535 | 1.4543 | 1.4524 |
| | confidence | 0.99986 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 5 | interval | 1.5545 | 1.5510 | 1.5498 | 1.5493 | 1.5490 | 1.5486 | 1.5483 |
| | confidence | 0.99982 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 6 | interval | 1.6203 | 1.6177 | 1.6168 | 1.6164 | 1.6162 | 1.6159 | 1.6157 |
| | confidence | 0.99981 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 7 | interval | 1.6692 | 1.6672 | 1.6665 | 1.6662 | 1.6661 | 1.6658 | 1.6657 |
| | confidence | 0.99979 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 8 | interval | 1.7070 | 1.7054 | 1.7048 | 1.7046 | 1.7045 | 1.7043 | 1.7042 |
| | confidence | 0.99979 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 9 | interval | 1.7370 | 1.7357 | 1.7352 | 1.7350 | 1.7350 | 1.7348 | 1.7348 |
| | confidence | 0.99978 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 10 | interval | 1.7615 | 1.7604 | 1.7600 | 1.7598 | 1.7598 | 1.7596 | 1.7596 |
| | confidence | 0.99978 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 20 | interval | 1.8765 | 1.8760 | 1.8758 | 1.8760 | 1.8760 | 1.8760 | 1.8759 |
| | confidence | 0.99978 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 60 | interval | 1.9578 | 1.9577 | 1.9577 | 1.9578 | 1.9578 | 1.9577 | 1.9577 |
| | confidence | 0.99977 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 100 | interval | 1.9746 | 1.9745 | 1.9745 | 1.9746 | 1.9745 | 1.9745 | 1.9745 |
| | confidence | 0.99977 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Table 21  25th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.9387 | 0.6537 | 0.5269 | 0.4546 | 0.4049 | 0.3694 | 0.2005 |
| | confidence | 0.8441 | 0.90441 | 0.92114 | 0.92852 | 0.93265 | 0.93651 | 0.94595 |
| naive($t$) | interval | 1.5121 | 0.7691 | 0.5828 | 0.4877 | 0.4279 | 0.3860 | |
| | confidence | 0.93106 | 0.93984 | 0.94258 | 0.94350 | 0.94500 | 0.94577 | |
| 2 | interval | 1.4645 | 1.4468 | 1.4401 | 1.4367 | 1.4339 | 1.4322 | 1.4295 |
| | confidence | 0.99649 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 3 | interval | 1.6157 | 1.6072 | 1.6037 | 1.6020 | 1.6003 | 1.5993 | 1.5984 |
| | confidence | 0.99636 | 0.99992 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 4 | interval | 1.7007 | 1.6957 | 1.6936 | 1.6925 | 1.6913 | 1.6907 | 1.6903 |
| | confidence | 0.99541 | 0.99989 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 5 | interval | 1.7550 | 1.7517 | 1.7503 | 1.7495 | 1.7486 | 1.7482 | 1.7481 |
| | confidence | 0.99475 | 0.99986 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 6 | interval | 1.7926 | 1.7904 | 1.7893 | 1.7887 | 1.7880 | 1.7877 | 1.7877 |
| | confidence | 0.99432 | 0.99985 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 7 | interval | 1.8203 | 1.8186 | 1.8178 | 1.8173 | 1.8168 | 1.8165 | 1.8166 |
| | confidence | 0.99405 | 0.99984 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 8 | interval | 1.8414 | 1.8401 | 1.8395 | 1.8391 | 1.8386 | 1.8384 | 1.8385 |
| | confidence | 0.99381 | 0.99982 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 9 | interval | 1.8581 | 1.8571 | 1.8565 | 1.8563 | 1.8558 | 1.8556 | 1.8558 |
| | confidence | 0.99368 | 0.99981 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 10 | interval | 1.8716 | 1.8708 | 1.8703 | 1.8701 | 1.8697 | 1.8695 | 1.8697 |
| | confidence | 0.99355 | 0.99980 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 20 | interval | 1.9343 | 1.9340 | 1.9338 | 1.9338 | 1.9337 | 1.9337 | 1.9338 |
| | confidence | 0.99320 | 0.99978 | 0.99999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 60 | interval | 1.9777 | 1.9777 | 1.9777 | 1.9777 | 1.9776 | 1.9776 | 1.9777 |
| | confidence | 0.99268 | 0.99966 | 0.99999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 100 | interval | 1.9866 | 1.9866 | 1.9866 | 1.9866 | 1.9865 | 1.9865 | 1.9866 |
| | confidence | 0.99261 | 0.99966 | 0.99999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Table 22 50th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 1.8518 | 1.2796 | 1.0363 | 0.8918 | 0.7953 | 0.7248 | 0.3934 |
| | confidence | 0.84452 | 0.90378 | 0.92138 | 0.92944 | 0.93584 | 0.93616 | 0.94707 |
| naive($t$) | interval | 2.9751 | 1.5055 | 1.1434 | 0.9572 | 0.8408 | 0.7583 | |
| | confidence | 0.93329 | 0.94017 | 0.94402 | 0.94577 | 0.94559 | 0.94662 | |
| 2 | interval | 2.0510 | 2.0256 | 2.0155 | 2.0129 | 2.0106 | 2.0100 | 2.0025 |
| | confidence | 0.96845 | 0.99753 | 0.99981 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 3 | interval | 2.0226 | 2.0155 | 2.0066 | 2.0058 | 2.0048 | 2.0048 | 2.0011 |
| | confidence | 0.96572 | 0.99600 | 0.99932 | 0.99994 | 0.99997 | 1.0000 | 1.0000 |
| 4 | interval | 2.0127 | 2.0065 | 2.0036 | 2.0033 | 2.0028 | 2.0029 | 2.0006 |
| | confidence | 0.96091 | 0.99403 | 0.99901 | 0.99986 | 0.99995 | 0.99998 | 1.0000 |
| 5 | interval | 2.0082 | 2.0042 | 2.0022 | 2.0021 | 2.0018 | 2.0020 | 2.0004 |
| | confidence | 0.95739 | 0.99257 | 0.99864 | 0.99978 | 0.99993 | 0.99998 | 1.0000 |
| 6 | interval | 2.0057 | 2.0029 | 2.0015 | 2.0015 | 2.0013 | 2.0015 | 2.0003 |
| | confidence | 0.95470 | 0.99144 | 0.99843 | 0.99969 | 0.99992 | 0.99997 | 1.0000 |
| 7 | interval | 2.0042 | 2.0022 | 2.0010 | 2.0011 | 2.0010 | 2.0012 | 2.0002 |
| | confidence | 0.95228 | 0.99072 | 0.99826 | 0.99959 | 0.99992 | 0.99997 | 1.0000 |
| 8 | interval | 2.0032 | 2.0017 | 2.0007 | 2.0009 | 2.0008 | 2.0010 | 2.0002 |
| | confidence | 0.95052 | 0.99017 | 0.99814 | 0.99953 | 0.99991 | 0.99997 | 1.0000 |
| 9 | interval | 2.0025 | 2.0013 | 2.0006 | 2.0007 | 2.0006 | 2.0008 | 2.0001 |
| | confidence | 0.94903 | 0.98958 | 0.99805 | 0.99945 | 0.99988 | 0.99997 | 1.0000 |
| 10 | interval | 2.0021 | 2.0011 | 2.0004 | 2.0006 | 2.0005 | 2.0007 | 2.0001 |
| | confidence | 0.94797 | 0.98921 | 0.99795 | 0.99937 | 0.99987 | 0.99996 | 1.0000 |
| 20 | interval | 2.0003 | 2.0003 | 2.0001 | 2.0001 | 2.0001 | 2.0003 | 2.0000 |
| | confidence | 0.94518 | 0.98737 | 0.99705 | 0.99926 | 0.99979 | 0.99994 | 1.0000 |
| 60 | interval | 2.0000 | 2.0000 | 2.0000 | 2.0000 | 2.0000 | 2.0001 | 2.0000 |
| | confidence | 0.94207 | 0.98622 | 0.99669 | 0.99911 | 0.99974 | 0.99992 | 1.0000 |
| 100 | interval | 2.0000 | 2.0000 | 2.0000 | 2.0000 | 2.0000 | 2.0000 | 2.0000 |
| | confidence | 0.94151 | 0.98597 | 0.99662 | 0.99908 | 0.99969 | 0.99992 | 1.0000 |

Table 23 75th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 3.6295 | 2.5089 | 2.0337 | 1.7527 | 1.5615 | 1.4228 | 0.7722 |
| | confidence | 0.8461 | 0.90381 | 0.92016 | 0.92821 | 0.93369 | 0.93567 | 0.9454 |
| naive($t$) | interval | 5.8301 | 2.9654 | 2.2417 | 1.8824 | 1.6504 | 1.4867 | |
| | confidence | 0.93179 | 0.93931 | 0.94315 | 0.94542 | 0.94593 | 0.94557 | |
| 2 | interval | 2.8723 | 2.8387 | 2.8263 | 2.8183 | 2.8177 | 2.8134 | 2.8052 |
| | confidence | 0.88117 | 0.97273 | 0.99296 | 0.99820 | 0.99959 | 0.99987 | 1.0000 |
| 3 | interval | 2.5317 | 2.5189 | 2.5139 | 2.5103 | 2.5106 | 2.5086 | 2.5053 |
| | confidence | 0.84061 | 0.95158 | 0.98320 | 0.99378 | 0.99786 | 0.99909 | 1.0000 |
| 4 | interval | 2.3819 | 2.3753 | 2.3725 | 2.3705 | 2.3709 | 2.3696 | 2.3679 |
| | confidence | 0.82097 | 0.93768 | 0.97491 | 0.98970 | 0.99549 | 0.99785 | 1.0000 |
| 5 | interval | 2.2978 | 2.2938 | 2.2921 | 2.2907 | 2.2911 | 2.2902 | 2.2891 |
| | confidence | 0.80917 | 0.92879 | 0.96886 | 0.98632 | 0.99335 | 0.99679 | 1.0000 |
| 6 | interval | 2.2440 | 2.2414 | 2.2402 | 2.2391 | 2.2396 | 2.2389 | 2.2381 |
| | confidence | 0.80128 | 0.92239 | 0.96442 | 0.98360 | 0.99176 | 0.99594 | 1.0000 |
| 7 | interval | 2.2066 | 2.2048 | 2.2039 | 2.2031 | 2.2035 | 2.2030 | 2.2024 |
| | confidence | 0.79554 | 0.91816 | 0.96154 | 0.98164 | 0.99058 | 0.99518 | 0.99999 |
| 8 | interval | 2.1792 | 2.1779 | 2.1772 | 2.1765 | 2.1769 | 2.1764 | 2.1760 |
| | confidence | 0.79161 | 0.91460 | 0.95926 | 0.98013 | 0.98958 | 0.99448 | 0.99999 |
| 9 | interval | 2.1582 | 2.1572 | 2.1566 | 2.1561 | 2.1564 | 2.1560 | 2.1557 |
| | confidence | 0.78839 | 0.91183 | 0.95719 | 0.97890 | 0.98858 | 0.99392 | 0.99999 |
| 10 | interval | 2.1416 | 2.1408 | 2.1403 | 2.1399 | 2.1402 | 2.1398 | 2.1396 |
| | confidence | 0.78599 | 0.90980 | 0.95554 | 0.97803 | 0.98770 | 0.99327 | 0.99999 |
| 20 | interval | 2.0690 | 2.0689 | 2.0690 | 2.0687 | 2.0686 | 2.0687 | 2.0686 |
| | confidence | 0.77584 | 0.89988 | 0.94872 | 0.97244 | 0.98422 | 0.99115 | 0.99999 |
| 60 | interval | 2.0226 | 2.0227 | 2.0227 | 2.0226 | 2.0226 | 2.0226 | 2.0226 |
| | confidence | 0.76920 | 0.89360 | 0.94390 | 0.96886 | 0.98139 | 0.98925 | 0.99999 |
| 100 | interval | 2.0135 | 2.0136 | 2.0136 | 2.0135 | 2.0135 | 2.0135 | 2.0135 |
| | confidence | 0.76767 | 0.89256 | 0.94290 | 0.96815 | 0.98080 | 0.98890 | 0.99999 |

Table 24 90th Percentile: $\hat{y} \pm \sqrt[m]{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 6.6483 | 4.6199 | 3.7343 | 3.2180 | 2.8613 | 2.6080 | 1.4165 |
| | confidence | 0.84499 | 0.90586 | 0.92118 | 0.92787 | 0.93310 | 0.93824 | 0.94615 |
| naive($t$) | interval | 10.6495 | 5.4290 | 4.1142 | 3.4503 | 3.0245 | 2.7286 | |
| | confidence | 0.93128 | 0.93966 | 0.94184 | 0.94423 | 0.94657 | 0.94616 | |
| 2 | interval | 3.8942 | 3.8422 | 3.8279 | 3.8192 | 3.8458 | 3.8106 | 3.7998 |
| | confidence | 0.75639 | 0.90047 | 0.95618 | 0.98036 | 0.99040 | 0.99581 | 1.0000 |
| 3 | interval | 3.1012 | 3.0822 | 3.0773 | 3.0741 | 3.0731 | 3.0709 | 3.0671 |
| | confidence | 0.65825 | 0.82045 | 0.89841 | 0.94047 | 0.96370 | 0.97934 | 0.99995 |
| 4 | interval | 2.7733 | 2.7634 | 2.7611 | 2.7595 | 2.7591 | 2.7578 | 2.7559 |
| | confidence | 0.60957 | 0.77455 | 0.86017 | 0.91173 | 0.94092 | 0.96207 | 0.99962 |
| 5 | interval | 2.5952 | 2.5891 | 2.5878 | 2.5868 | 2.5866 | 2.5857 | 2.5846 |
| | confidence | 0.58178 | 0.74656 | 0.83467 | 0.89081 | 0.92383 | 0.94926 | 0.99924 |
| 6 | interval | 2.4835 | 2.4794 | 2.4786 | 2.4779 | 2.4778 | 2.4771 | 2.4764 |
| | confidence | 0.56364 | 0.72720 | 0.81714 | 0.87628 | 0.91174 | 0.93878 | 0.99877 |
| 7 | interval | 2.4071 | 2.4040 | 2.4035 | 2.4030 | 2.4029 | 2.4024 | 2.4019 |
| | confidence | 0.55110 | 0.71286 | 0.80393 | 0.86492 | 0.90272 | 0.93038 | 0.99834 |
| 8 | interval | 2.3514 | 2.3491 | 2.3487 | 2.3483 | 2.3483 | 2.3479 | 2.3475 |
| | confidence | 0.54139 | 0.70220 | 0.79421 | 0.85650 | 0.89582 | 0.92419 | 0.99798 |
| 9 | interval | 2.3092 | 2.3073 | 2.3070 | 2.3067 | 2.3067 | 2.3064 | 2.3061 |
| | confidence | 0.53398 | 0.69387 | 0.78661 | 0.84950 | 0.88963 | 0.91939 | 0.99767 |
| 10 | interval | 2.2759 | 2.2744 | 2.2742 | 2.2740 | 2.2740 | 2.2737 | 2.2735 |
| | confidence | 0.52781 | 0.68757 | 0.77997 | 0.84422 | 0.88467 | 0.91543 | 0.99735 |
| 20 | interval | 2.1329 | 2.1325 | 2.1326 | 2.1325 | 2.1324 | 2.1324 | 2.1323 |
| | confidence | 0.49820 | 0.65858 | 0.75563 | 0.81857 | 0.86331 | 0.89540 | 0.99509 |
| 60 | interval | 2.0432 | 2.0432 | 2.0432 | 2.0432 | 2.0432 | 2.0432 | 2.0432 |
| | confidence | 0.48166 | 0.63981 | 0.73651 | 0.80125 | 0.84712 | 0.88073 | 0.99338 |
| 100 | interval | 2.0258 | 2.0258 | 2.0258 | 2.0258 | 2.0258 | 2.0258 | 2.0258 |
| | confidence | 0.47805 | 0.63585 | 0.73254 | 0.79746 | 0.84389 | 0.87757 | 0.99289 |

Table 25 10th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.5143 | 0.3552 | 0.2874 | 0.2478 | 0.2208 | 0.2011 | 0.1092 |
| | confidence | 0.84483 | 0.90354 | 0.92068 | 0.92904 | 0.93321 | 0.93559 | 0.94646 |
| naive($t$) | interval | 0.8258 | 0.4187 | 0.3358 | 0.2657 | 0.2331 | 0.2105 | |
| | confidence | 0.93183 | 0.93897 | 0.95297 | 0.94543 | 0.94542 | 0.94512 | |
| 2 | interval | 0.5401 | 0.5336 | 0.5313 | 0.5302 | 0.5293 | 0.5293 | 0.5275 |
| | confidence | 0.95944 | 0.99640 | 0.99960 | 0.99995 | 0.99999 | 1.0000 | 1.0000 |
| 3 | interval | 0.3601 | 0.3557 | 0.3542 | 0.3535 | 0.3529 | 0.3529 | 0.3517 |
| | confidence | 0.83515 | 0.95003 | 0.98381 | 0.99516 | 0.99808 | 0.99932 | 1.0000 |
| 4 | interval | 0.2701 | 0.2668 | 0.2656 | 0.2651 | 0.2647 | 0.2647 | 0.2638 |
| | confidence | 0.70581 | 0.86311 | 0.93015 | 0.96525 | 0.98103 | 0.98992 | 1.0000 |
| 5 | interval | 0.2161 | 0.2134 | 0.2125 | 0.2121 | 0.2117 | 0.2117 | 0.2110 |
| | confidence | 0.60029 | 0.76629 | 0.85416 | 0.91095 | 0.94040 | 0.96149 | 0.99985 |
| 6 | interval | 0.1800 | 0.1779 | 0.1771 | 0.1767 | 0.1764 | 0.1764 | 0.1758 |
| | confidence | 0.51849 | 0.68001 | 0.77687 | 0.84429 | 0.88532 | 0.91587 | 0.99833 |
| 7 | interval | 0.1543 | 0.1524 | 0.1518 | 0.1515 | 0.1512 | 0.1512 | 0.1507 |
| | confidence | 0.45433 | 0.60644 | 0.70511 | 0.77647 | 0.82387 | 0.86110 | 0.99279 |
| 8 | interval | 0.1350 | 0.1334 | 0.1328 | 0.1325 | 0.1323 | 0.1323 | 0.1319 |
| | confidence | 0.40343 | 0.54493 | 0.64174 | 0.71504 | 0.76503 | 0.80426 | 0.98145 |
| 9 | interval | 0.1200 | 0.1186 | 0.1181 | 0.1178 | 0.1176 | 0.1176 | 0.1172 |
| | confidence | 0.36191 | 0.49327 | 0.58579 | 0.65829 | 0.70939 | 0.75017 | 0.96338 |
| 10 | interval | 0.1080 | 0.1067 | 0.1063 | 0.1060 | 0.1059 | 0.1059 | 0.1055 |
| | confidence | 0.32829 | 0.44936 | 0.53717 | 0.60663 | 0.65865 | 0.69992 | 0.94036 |

Table 26  25th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.9387 | 0.6537 | 0.5269 | 0.4546 | 0.4049 | 0.3694 | 0.2005 |
| | confidence | 0.8441 | 0.90441 | 0.92114 | 0.92852 | 0.93265 | 0.93651 | 0.94595 |
| naive($t$) | interval | 1.5121 | 0.7691 | 0.5828 | 0.4877 | 0.4279 | 0.3860 | |
| | confidence | 0.93106 | 0.93984 | 0.94258 | 0.94350 | 0.94500 | 0.94577 | |
| 2 | interval | 0.7305 | 0.7228 | 0.7197 | 0.7179 | 0.7174 | 0.7167 | 0.7148 |
| | confidence | 0.87604 | 0.97091 | 0.99217 | 0.99783 | 0.99950 | 0.99984 | 1.0000 |
| 3 | interval | 0.4870 | 0.4819 | 0.4798 | 0.4786 | 0.4783 | 0.4778 | 0.4766 |
| | confidence | 0.69915 | 0.85835 | 0.92792 | 0.96180 | 0.97973 | 0.98864 | 0.99999 |
| 4 | interval | 0.3653 | 0.3614 | 0.3599 | 0.3589 | 0.3587 | 0.3584 | 0.3574 |
| | confidence | 0.56424 | 0.72922 | 0.82524 | 0.88234 | 0.91900 | 0.94374 | 0.99945 |
| 5 | interval | 0.2922 | 0.2891 | 0.2879 | 0.2871 | 0.2870 | 0.2867 | 0.2859 |
| | confidence | 0.46720 | 0.62126 | 0.72260 | 0.78907 | 0.83846 | 0.87510 | 0.99480 |
| 6 | interval | 0.2435 | 0.2409 | 0.2399 | 0.2393 | 0.2391 | 0.2389 | 0.2383 |
| | confidence | 0.39636 | 0.53723 | 0.63640 | 0.70373 | 0.75686 | 0.79981 | 0.98036 |
| 7 | interval | 0.2087 | 0.2065 | 0.2056 | 0.2051 | 0.2050 | 0.2048 | 0.2042 |
| | confidence | 0.34404 | 0.47051 | 0.56368 | 0.62983 | 0.68240 | 0.72683 | 0.95392 |
| 8 | interval | 0.1826 | 0.1807 | 0.1799 | 0.1795 | 0.1794 | 0.1792 | 0.1787 |
| | confidence | 0.30269 | 0.41884 | 0.50491 | 0.56830 | 0.61747 | 0.66199 | 0.92034 |
| 9 | interval | 0.1623 | 0.1606 | 0.1599 | 0.1595 | 0.1594 | 0.1593 | 0.1589 |
| | confidence | 0.27038 | 0.37659 | 0.45451 | 0.51521 | 0.56130 | 0.60574 | 0.88150 |
| 10 | interval | 0.1461 | 0.1446 | 0.1439 | 0.1436 | 0.1435 | 0.1433 | 0.1430 |
| | confidence | 0.24501 | 0.34122 | 0.41388 | 0.47041 | 0.51445 | 0.55749 | 0.84129 |

Table 27 50th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 1.8518 | 1.2796 | 1.0363 | 0.8918 | 0.7953 | 0.7248 | 0.3934 |
| | confidence | 0.84452 | 0.90378 | 0.92138 | 0.92944 | 0.93584 | 0.93616 | 0.94707 |
| naive($t$) | interval | 2.9751 | 1.5055 | 1.1434 | 0.9572 | 0.8408 | 0.7583 | |
| | confidence | 0.93329 | 0.94017 | 0.94402 | 0.94577 | 0.94559 | 0.94662 | |
| 2 | interval | 1.0260 | 1.0126 | 1.0075 | 1.0062 | 1.0054 | 1.0044 | 1.0011 |
| | confidence | 0.73305 | 0.88225 | 0.94521 | 0.97316 | 0.98592 | 0.99352 | 1.0000 |
| 3 | interval | 0.6840 | 0.6751 | 0.6717 | 0.6708 | 0.6703 | 0.6696 | 0.6674 |
| | confidence | 0.54139 | 0.70583 | 0.80245 | 0.86225 | 0.90185 | 0.93036 | 0.99913 |
| 4 | interval | 0.5130 | 0.5063 | 0.5038 | 0.5031 | 0.5027 | 0.5022 | 0.5006 |
| | confidence | 0.42155 | 0.57205 | 0.66765 | 0.73339 | 0.78833 | 0.82640 | 0.98717 |
| 5 | interval | 0.4104 | 0.4051 | 0.4030 | 0.4025 | 0.4022 | 0.4017 | 0.4004 |
| | confidence | 0.34348 | 0.47323 | 0.56212 | 0.62677 | 0.68229 | 0.72424 | 0.95331 |
| 6 | interval | 0.3420 | 0.3375 | 0.3358 | 0.3354 | 0.3351 | 0.3348 | 0.3337 |
| | confidence | 0.28912 | 0.40251 | 0.48271 | 0.54220 | 0.59476 | 0.63536 | 0.90281 |
| 7 | interval | 0.2931 | 0.2893 | 0.2879 | 0.2875 | 0.2873 | 0.2870 | 0.2860 |
| | confidence | 0.24937 | 0.34900 | 0.42111 | 0.47563 | 0.52406 | 0.56278 | 0.84641 |
| 8 | interval | 0.2565 | 0.2532 | 0.2519 | 0.2516 | 0.2514 | 0.2511 | 0.2503 |
| | confidence | 0.21960 | 0.30724 | 0.37235 | 0.42294 | 0.46793 | 0.50376 | 0.78740 |
| 9 | interval | 0.2280 | 0.2250 | 0.2239 | 0.2236 | 0.2234 | 0.2232 | 0.2225 |
| | confidence | 0.19611 | 0.27450 | 0.33333 | 0.37954 | 0.42202 | 0.45492 | 0.73254 |
| 10 | interval | 0.2052 | 0.2025 | 0.2015 | 0.2012 | 0.2011 | 0.2009 | 0.2002 |
| | confidence | 0.17588 | 0.24902 | 0.30158 | 0.34402 | 0.38315 | 0.41348 | 0.68155 |

Table 28 75th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 3.6295 | 2.5089 | 2.0337 | 1.7527 | 1.5615 | 1.4228 | 0.7722 |
| | confidence | 0.8461 | 0.90381 | 0.92016 | 0.92821 | 0.93369 | 0.93567 | 0.9454 |
| naive($t$) | interval | 5.8301 | 2.9654 | 2.2417 | 1.8824 | 1.6504 | 1.4867 | |
| | confidence | 0.93179 | 0.93931 | 0.94315 | 0.94542 | 0.94593 | 0.94557 | |
| 2 | interval | 1.4362 | 1.4182 | 1.4122 | 1.4099 | 1.4085 | 1.4065 | 1.4026 |
| | confidence | 0.57163 | 0.74042 | 0.82902 | 0.88844 | 0.92361 | 0.94917 | 0.99948 |
| 3 | interval | 0.9575 | 0.9454 | 0.9415 | 0.9399 | 0.9390 | 0.9377 | 0.9350 |
| | confidence | 0.40417 | 0.54839 | 0.64136 | 0.70962 | 0.76376 | 0.80729 | 0.98205 |
| 4 | interval | 0.7181 | 0.7091 | 0.7061 | 0.7049 | 0.7042 | 0.7033 | 0.7013 |
| | confidence | 0.30994 | 0.42762 | 0.50834 | 0.57217 | 0.62670 | 0.67081 | 0.92617 |
| 5 | interval | 0.5745 | 0.5673 | 0.5649 | 0.5639 | 0.5634 | 0.5626 | 0.5610 |
| | confidence | 0.24949 | 0.34843 | 0.41780 | 0.47502 | 0.52451 | 0.56458 | 0.84760 |
| 6 | interval | 0.4787 | 0.4727 | 0.4707 | 0.4700 | 0.4695 | 0.4688 | 0.4675 |
| | confidence | 0.21053 | 0.29398 | 0.35456 | 0.40213 | 0.44886 | 0.48451 | 0.76690 |
| 7 | interval | 0.4103 | 0.4052 | 0.4035 | 0.4028 | 0.4024 | 0.4019 | 0.4007 |
| | confidence | 0.18114 | 0.25372 | 0.30687 | 0.34980 | 0.39184 | 0.42324 | 0.69244 |
| 8 | interval | 0.3590 | 0.3545 | 0.3531 | 0.3525 | 0.3521 | 0.3516 | 0.3506 |
| | confidence | 0.15888 | 0.22279 | 0.26907 | 0.30865 | 0.34630 | 0.37436 | 0.62831 |
| 9 | interval | 0.3192 | 0.3151 | 0.3138 | 0.3133 | 0.3130 | 0.3126 | 0.3117 |
| | confidence | 0.14165 | 0.19834 | 0.24015 | 0.27622 | 0.30985 | 0.33441 | 0.57376 |
| 10 | interval | 0.2871 | 0.2836 | 0.2824 | 0.2820 | 0.2817 | 0.2813 | 0.2805 |
| | confidence | 0.12780 | 0.17926 | 0.21674 | 0.24982 | 0.28079 | 0.30255 | 0.52705 |

Table 29 90th Percentile: $\hat{y} \pm \frac{1}{m}\sqrt{\hat{y}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 6.6483 | 4.6199 | 3.7343 | 3.2180 | 2.8613 | 2.6080 | 1.4165 |
| | confidence | 0.84499 | 0.90586 | 0.92118 | 0.92787 | 0.93310 | 0.93824 | 0.94615 |
| naive($t$) | interval | 10.6495 | 5.4290 | 4.1142 | 3.4503 | 3.0245 | 2.7286 | |
| | confidence | 0.93128 | 0.93966 | 0.94184 | 0.94423 | 0.94657 | 0.94616 | |
| 2 | interval | 1.9459 | 1.9220 | 1.9130 | 1.9094 | 1.9078 | 1.9066 | 1.9003 |
| | confidence | 0.44044 | 0.59592 | 0.69040 | 0.75962 | 0.81229 | 0.84996 | 0.99086 |
| 3 | interval | 1.2973 | 1.2813 | 1.2753 | 1.2729 | 1.2719 | 1.2711 | 1.2669 |
| | confidence | 0.30280 | 0.42230 | 0.50247 | 0.56479 | 0.62141 | 0.66433 | 0.91968 |
| 4 | interval | 0.9730 | 0.9610 | 0.9565 | 0.9547 | 0.9539 | 0.9533 | 0.9501 |
| | confidence | 0.23045 | 0.32373 | 0.39026 | 0.44344 | 0.49235 | 0.52900 | 0.81117 |
| 5 | interval | 0.7784 | 0.7688 | 0.7652 | 0.7637 | 0.7631 | 0.7626 | 0.7601 |
| | confidence | 0.18521 | 0.26310 | 0.31668 | 0.36188 | 0.40338 | 0.43665 | 0.70615 |
| 6 | interval | 0.6486 | 0.6407 | 0.6377 | 0.6365 | 0.6359 | 0.6355 | 0.6334 |
| | confidence | 0.15536 | 0.22065 | 0.26530 | 0.30404 | 0.34188 | 0.36972 | 0.61922 |
| 7 | interval | 0.5560 | 0.5491 | 0.5466 | 0.5455 | 0.5451 | 0.5447 | 0.5429 |
| | confidence | 0.13339 | 0.19011 | 0.22964 | 0.26211 | 0.29620 | 0.32000 | 0.54688 |
| 8 | interval | 0.4865 | 0.4805 | 0.4782 | 0.4773 | 0.4770 | 0.4766 | 0.4751 |
| | confidence | 0.11623 | 0.16620 | 0.20165 | 0.23037 | 0.26135 | 0.28210 | 0.48767 |
| 9 | interval | 0.4324 | 0.4271 | 0.4251 | 0.4243 | 0.4240 | 0.4237 | 0.4223 |
| | confidence | 0.10403 | 0.14805 | 0.17969 | 0.20506 | 0.23227 | 0.25216 | 0.43773 |
| 10 | interval | 0.3892 | 0.3844 | 0.3826 | 0.3819 | 0.3816 | 0.3813 | 0.3801 |
| | confidence | 0.09330 | 0.13309 | 0.16189 | 0.18443 | 0.20895 | 0.22766 | 0.39830 |

Table 30    10th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Approach | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.5143 | 0.3552 | 0.2874 | 0.2478 | 0.2208 | 0.2011 | 0.1092 |
| | confidence | 0.84483 | 0.90354 | 0.92068 | 0.92904 | 0.93321 | 0.93559 | 0.94646 |
| naive($t$) | interval | 0.8258 | 0.4187 | 0.3358 | 0.2657 | 0.2331 | 0.2105 | |
| | confidence | 0.93183 | 0.93897 | 0.95297 | 0.94543 | 0.94542 | 0.94512 | |
| a = 1 | interval | 0.4832 | 0.3374 | 0.2744 | 0.2371 | 0.2119 | 0.1932 | 0.1055 |
| | confidence | 0.93378 | 0.93892 | 0.93934 | 0.93951 | 0.94106 | 0.94023 | 0.94205 |

Table 31    25th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Approach | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 0.9387 | 0.6537 | 0.5269 | 0.4546 | 0.4049 | 0.3694 | 0.2005 |
| | confidence | 0.8441 | 0.90441 | 0.92114 | 0.92852 | 0.93265 | 0.93651 | 0.94595 |
| naive($t$) | interval | 1.5121 | 0.7691 | 0.5828 | 0.4877 | 0.4279 | 0.3860 | |
| | confidence | 0.93106 | 0.93984 | 0.94258 | 0.94350 | 0.94500 | 0.94577 | |
| a = 1 | interval | 0.6549 | 0.4569 | 0.3716 | 0.3213 | 0.2869 | 0.2616 | 0.1429 |
| | confidence | 0.83068 | 0.83666 | 0.83747 | 0.83799 | 0.83567 | 0.83854 | 0.83906 |
| a = 2 | interval | 0.9259 | 0.6464 | 0.5259 | 0.4544 | 0.4058 | 0.3701 | 0.2022 |
| | confidence | 0.94554 | 0.94863 | 0.95008 | 0.95035 | 0.94995 | 0.95135 | 0.95229 |

Table 32    50th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Approach | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 1.8518 | 1.2796 | 1.0363 | 0.8918 | 0.7953 | 0.7248 | 0.3934 |
| | confidence | 0.84452 | 0.90378 | 0.92138 | 0.92944 | 0.93584 | 0.93616 | 0.94707 |
| naive($t$) | interval | 2.9751 | 1.5055 | 1.1434 | 0.9572 | 0.8408 | 0.7583 | |
| | confidence | 0.93329 | 0.94017 | 0.94402 | 0.94577 | 0.94559 | 0.94662 | |
| a = 3 | interval | 1.5877 | 1.1092 | 0.9012 | 0.7793 | 0.6961 | 0.6352 | 0.3468 |
| | confidence | 0.90904 | 0.91183 | 0.91390 | 0.91535 | 0.91591 | 0.91552 | 0.91547 |
| a = 4 | interval | 1.8355 | 1.2802 | 1.0412 | 0.8999 | 0.8038 | 0.7332 | 0.4006 |
| | confidence | 0.94731 | 0.95283 | 0.95151 | 0.95330 | 0.95348 | 0.95444 | 0.95392 |

Table 33 75th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Approach | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 3.6295 | 2.5089 | 2.0337 | 1.7527 | 1.5615 | 1.4228 | 0.7722 |
| | confidence | 0.8461 | 0.90381 | 0.92016 | 0.92821 | 0.93369 | 0.93567 | 0.9454 |
| naive($t$) | interval | 5.8301 | 2.9654 | 2.2417 | 1.8824 | 1.6504 | 1.4867 | |
| | confidence | 0.93179 | 0.93931 | 0.94315 | 0.94542 | 0.94593 | 0.94557 | |
| a = 5 | interval | 2.8773 | 2.0067 | 1.6304 | 1.4101 | 1.2592 | 1.1486 | 0.6274 |
| | confidence | 0.88178 | 0.88396 | 0.88661 | 0.88780 | 0.88798 | 0.88760 | 0.89105 |
| a = 6 | interval | 3.1520 | 2.1982 | 1.7860 | 1.5447 | 1.3794 | 1.2582 | 0.6873 |
| | confidence | 0.91093 | 0.91525 | 0.91650 | 0.91702 | 0.91795 | 0.91818 | 0.92031 |
| a = 7 | interval | 3.4045 | 2.3744 | 1.9291 | 1.6684 | 1.4899 | 1.3590 | 0.7424 |
| | confidence | 0.93303 | 0.93662 | 0.93815 | 0.93889 | 0.93905 | 0.94022 | 0.94151 |
| a = 8 | interval | 3.6396 | 2.5383 | 2.0623 | 1.7836 | 1.5928 | 1.4529 | 0.7936 |
| | confidence | 0.94861 | 0.95268 | 0.95400 | 0.95448 | 0.95484 | 0.95530 | 0.95672 |

Table 34 90th percentile: $\hat{y} \pm \sqrt{\frac{a\hat{y}}{n}}$

| | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Approach | | 5 | 10 | 15 | 20 | 25 | 30 | 100 |
| naive($z$) | interval | 6.6483 | 4.6199 | 3.7343 | 3.2180 | 2.8613 | 2.6080 | 1.4165 |
| | confidence | 0.84499 | 0.90586 | 0.92118 | 0.92787 | 0.93310 | 0.93824 | 0.94615 |
| naive($t$) | interval | 10.6495 | 5.4290 | 4.1142 | 3.4503 | 3.0245 | 2.7286 | |
| | confidence | 0.93128 | 0.93966 | 0.94184 | 0.94423 | 0.94657 | 0.94616 | |
| a = 8 | interval | 4.9187 | 3.4345 | 2.7969 | 2.4143 | 2.1580 | 1.9679 | 1.0749 |
| | confidence | 0.85705 | 0.86080 | 0.86030 | 0.86172 | 0.86147 | 0.86443 | 0.86455 |
| a = 9 | interval | 5.2170 | 3.6429 | 2.9666 | 2.5608 | 2.2889 | 2.0873 | 1.1401 |
| | confidence | 0.87953 | 0.88297 | 0.88240 | 0.88376 | 0.88389 | 0.88623 | 0.88725 |
| a = 10 | interval | 5.4992 | 3.8399 | 3.1270 | 2.6993 | 2.4127 | 2.2002 | 1.2018 |
| | confidence | 0.89782 | 0.90171 | 0.90082 | 0.90241 | 0.90181 | 0.90434 | 0.90510 |
| a = 11 | interval | 5.7677 | 4.0273 | 3.2797 | 2.8310 | 2.5305 | 2.3076 | 1.2605 |
| | confidence | 0.91280 | 0.91666 | 0.91666 | 0.91786 | 0.91748 | 0.91933 | 0.91968 |
| a = 12 | interval | 6.0241 | 4.2064 | 3.4255 | 2.9569 | 2.6430 | 2.4102 | 1.3165 |
| | confidence | 0.92533 | 0.92894 | 0.92884 | 0.93033 | 0.93006 | 0.93126 | 0.93292 |
| a = 13 | interval | 6.2701 | 4.3782 | 3.5654 | 3.0777 | 2.7509 | 2.5086 | 1.3703 |
| | confidence | 0.93583 | 0.93939 | 0.93965 | 0.94074 | 0.94063 | 0.94134 | 0.94297 |
| a = 14 | interval | 6.5068 | 4.5435 | 3.7000 | 3.1938 | 2.8548 | 2.6033 | 1.4220 |
| | confidence | 0.94515 | 0.94827 | 0.94815 | 0.94949 | 0.94969 | 0.95054 | 0.95171 |
| a = 15 | interval | 6.7352 | 4.7029 | 3.8298 | 3.3059 | 2.9550 | 2.6946 | 1.4719 |
| | confidence | 0.95253 | 0.95538 | 0.95605 | 0.95761 | 0.95734 | 0.95770 | 0.95888 |

Table 35 Percent to Standard z-score Conversion

| % | z | % | z | % | z | % | z |
|---|---|---|---|---|---|---|---|
| 0 | -3.00 | | | | | | |
| 1 | -2.33 | 26 | -0.64 | 51 | 0.03 | 76 | 0.71 |
| 2 | -2.05 | 27 | -0.61 | 52 | 0.05 | 77 | 0.74 |
| 3 | -1.88 | 28 | -0.58 | 53 | 0.08 | 78 | 0.77 |
| 4 | -1.75 | 29 | -0.55 | 54 | 0.10 | 79 | 0.81 |
| 5 | -1.65 | 30 | -0.52 | 55 | 0.13 | 80 | 0.84 |
| 6 | -1.56 | 31 | -0.50 | 56 | 0.15 | 81 | 0.88 |
| 7 | -1.48 | 32 | -0.47 | 57 | 0.18 | 82 | 0.92 |
| 8 | -1.41 | 33 | -0.44 | 58 | 0.20 | 83 | 0.95 |
| 9 | -1.34 | 34 | -0.41 | 59 | 0.23 | 84 | 0.99 |
| 10 | -1.28 | 35 | -0.39 | 60 | 0.25 | 85 | 1.04 |
| 11 | -1.23 | 36 | -0.36 | 61 | 0.28 | 86 | 1.08 |
| 12 | -1.18 | 37 | -0.33 | 62 | 0.31 | 87 | 1.13 |
| 13 | -1.13 | 38 | -0.31 | 63 | 0.33 | 88 | 1.18 |
| 14 | -1.08 | 39 | -0.28 | 64 | 0.36 | 89 | 1.23 |
| 15 | -1.04 | 40 | -0.25 | 65 | 0.39 | 90 | 1.28 |
| 16 | -0.99 | 41 | -0.23 | 66 | 0.41 | 91 | 1.34 |
| 17 | -0.95 | 42 | -0.20 | 67 | 0.44 | 92 | 1.41 |
| 18 | -0.92 | 43 | -0.18 | 68 | 0.47 | 93 | 1.48 |
| 19 | -0.88 | 44 | -0.15 | 69 | 0.50 | 94 | 1.56 |
| 20 | -0.84 | 45 | -0.13 | 70 | 0.52 | 95 | 1.65 |
| 21 | -0.81 | 46 | -0.10 | 71 | 0.55 | 96 | 1.75 |
| 22 | -0.77 | 47 | -0.08 | 72 | 0.58 | 97 | 1.88 |
| 23 | -0.74 | 48 | -0.05 | 73 | 0.61 | 98 | 2.05 |
| 24 | -0.71 | 49 | -0.03 | 74 | 0.64 | 99 | 2.33 |
| 25 | -0.67 | 50 | 0.00 | 75 | 0.67 | 100 | 3.00+ |

*Bibliography*


[1] Aitchison, J. and Brown, J.A.C. *The Lognormal Distribution.* New York: The Syndics of the Cambridge University Press, 1957.

[2] Angus, J.E. "Bootstrap one-sided confidence intervals for the log-normal mean," *Statistician*, 43:395-401 (1994).

[3] —–. "Inferences on the lognormal mean for complete samples," *Communications in Statistics-Simulation and Computation*, 17:1307-1331 (1988).

[4] "Appendix A: The Conversion Table: How to Use it For Converting Scores," *North Central Association Commission on Accreditation and School Improvement (NCA CASI)*, n. pag. http://www.ncacasi.org/webservices/ssw/manual/apxa. 16 May 2006.

[5] Burmaster, D.E. and Hull, D.A. "Using LogNormal Distributions and LogNormal Probability Plots in Probabilistic Risk Assessments," *Human and Ecological Risk Assessment*, 3(2):235-255 (1997).

[6] Casella, G. and Berger, R.L. *Statistical Inference* (Second Edition). California: Duxbury, 2002.

[7] Dallal, G.E. "Logarithms." 28 November 2005. n. pag. www.tufts.edu/ gdallal/logs.htm. 30 November 2005.

[8] "The Empirical Rule," *Armstrong Atlantic State University Department of Mathematics*, n. Pag. http://www.math.armstrong.edu/statsonline/3/empirical_rule.html. 16 May 2006.

[9] Finney, D.J. "On the Distribution of a Variate whose Logarithm is Normally Distributed," *J. Royal Statis. Soc. Suppl*, 7:155-161 (1941).

[10] Krishnamoorthy, K. and Mathew, T. "Inferences on the means of lognormal distributions using generalized $p$-values and generalized confidence intervals," *Journal of statistical planning and inference*, 115:103-121 (2003).

[11] Land, C.E. "Confidence intervals for linear functions of the normal mean and variance," *Annals of Mathematical Statistics*, 42:1187-1205 (1971).

[12] —–. "An evaluation of approximate confidence interval estimation methods for lognormal means," *Technometrics*, 14:145-158 (1972).

[13] ——. "Hypothesis tests and interval estimates," In Crow, E.L. and Shimizu, K. (editors) *Lognormal Distributions: theory and applications*. New York: Marcel Dekker, 1988.

[14] ——. "Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance," p. 385-419. In *Selected Tables in Mathematical Statistics. Volume III. American Mathematical Society*, (1975).

[15] Mendenhall, W., Beaver, R.J., and Beaver, B.M. *Introduction to Probability and Statistics* (Tenth Edition). Cincinnati OH: Duxbury Press, 1999.

[16] *Microsoft Office 2000 Premium*. Computer software. Microsoft Corporation, 1999.

[17] Parkin, T.B., Chester, S.T., and Robinson, J.A. "Calculating Confidence Intervals for the Mean of a Lognormally Distributed Variable," *Soil Sci. Soc. Am. J.*, 54:321-326 (1990).

[18] Olsson, U. "Confidence Intervals for the Mean of a Log-Normal Distribution," *Journal of Statistics Education*, 13:n1 (2005).

[19] *SAS/STAT software: Changes and enhancements through Release 6.12*. Computer software. SAS Institute Inc., Cary NC, 1997.

[20] Schwarz, C.J. *Intermediate Sampling and Experimental Design and Analysis.* Department of Statistics and Actuarial Science, Simon Fraser University, March 2005.

[21] Simon, S. "Log Transformation" Website. http://www.cmh.edu/stats/linear/log.asp. 30 June 2005.

[22] Tracy, S.P. *Estimate at Completion: A Regression Approach to Earned Value.* MS Thesis, AFIT/GCA/ENC/05-04. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2005.

[23] "Transformations to Improve Fit," *NIST/SEMATECH Engineering Statistics Internet Handbook*, n. pag. http://www.itl.nist.gov/div898/handbook/pmd/section6 /pmd624.htm. 30 June 2005.

[24] Weerahandi, S. "Generalized confidence intervals," *Journal of the American Statistical Association*, 88:899-905 (1993).

[25] Zhou, X-H. and Gao, S. "Confidence intervals for the log-normal mean," *Statistics in Medicine*, 16:783-790 (1997).

*Vita*

Second Lieutenant Jason E. Tisdel enlisted in the United States Air Force on March 15, 2000. Upon completing Basic Military Training, he attended technical school in Biloxi, MS. There, Lt. Tisdel studied principles of electronics and intrusion detection systems. From Biloxi, he travelled to Ft. Meade, MD to study concepts of television and studio equipment. Lt. Tisdel's first assignment took him to Whiteman AFB, MO where he worked in the Communication Squadron as a Visual Imagery and Intrusion Detection Systems (VIIDS) apprentice. While at Whiteman AFB, Lt. Tisdel desired to finish his Bachelor's degree. With a successful application, he was selected to pursue a Bachelor's degree under the Airmen Education and Commissioning Program (AECP). Selecting Wright State University, Lt. Tisdel completed his Bachelor's degree in Mathematics, graduating Magna Cum Laude. Wishing to further his education, Lt. Tisdel applied for direct accession to AFIT with a focus in Statistics. Upon being commissioned as an officer on June 12th, 2004, AFIT was the next step in an exciting and rewarding career. The continuing journey takes Lt. Tisdel to the 46th Test Squadron at Eglin AFB.