Applied Occupational and Environmental Hygiene

ISSN: 1047-322X (Print) 1521-0898 (Online) Journal homepage: http://www.tandfonline.com/loi/uaoh20

Simple Procedures for Calculating Confidence Intervals around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data

Paul Hewett & Gary H. Ganser

Simple Procedures for Calculating Confidence Intervals Around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data

Paul Hewett[A] *and Gary H. Ganser*[B]

[A]National Institute for Occupational Safety and Health, 1095 Willowdale Road, Morgantown, West Virginia 26505-2888;
[B]Department of Mathematics, West Virginia University, Morgantown, West Virginia 26505-6310

The determination that a work environment is acceptable or unacceptable is often based upon the calculation of the mean exposure and/or exceedance fraction [i.e., the fraction of measurements expected to exceed an occupational exposure limit (OEL)] for an employee or exposure group. The purpose of this article is to introduce simplified procedures for calculating accurate estimates of the 95 percent lower confidence limit (LCL) and 95 percent upper confidence limit (UCL) for (1) the sample arithmetic mean and (2) the sample exceedance fraction for lognormally distributed exposure data. The procedure for the arithmetic mean is adapted from Land's procedure for calculating exact confidence intervals around the mean of lognormally distributed data. The procedure for the exceedance fraction is based on the Odeh and Owen confidence limits for a proportion in one tail of the normal distribution. Both procedures are graphically based and require no hard-to-find references. The two 95 percent confidence limits can be used to form a 90 percent confidence interval (5% is in each tail). The 90 percent confidence intervals indicate the degree of uncertainty in the point estimates of the mean and exceedance fraction. (This uncertainty is a function of both the inherent environmental variability in exposures and the sample size.) The resulting confidence limits can also be used for hypothesis tests by simply comparing the UCL or LCL to target values such as a long-term average OEL or an acceptable exceedance fraction. These procedures are simple, accurate, and can be used in lieu of (1) the means test procedures of the American Industrial Hygiene Association (AIHA 1991 monograph on exposure assessment) and Rappaport and Selvin (Am. Ind. Hyg. Assoc. J. 48:374–379, 1987) and (2) the exceedance fraction tests described by Tuggle (Am. Ind. Hyg. Assoc. J. 43:338–346, 1982) and Leidel and Busch (Patty's, Vol. 3A, 1994). © *1997 AIH.* Hewett, P.; Ganser, G.H.: Simple Procedures for Calculating Confidence Intervals Around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data. Appl. Occup. Environ. Hyg. 12(2):132–142; 1997.

There are several statistics that can be used to characterize a lognormal exposure distribution. Two statistics often discussed, particularly when compliance is an issue, are the mean exposure (averaged over n days) and the exceedance fraction (calculated from n measurements). The mean exposure is often used as an estimate of the long-term mean exposure, or the true average exposure, for an employee or group of employees over some extended interval, say 6 months to a year. The exceedance fraction is defined as the fraction of measurements expected to exceed some value, usually an occupational exposure limit (OEL), over this same interval.

The long-term mean is of interest (1) in occupational epidemiology where an estimate of the long-term mean is applied to all employees in an exposure group, or (2) in those instances where the OEL is defined as the limit for the long-term mean for each employee. There are few federal standards for toxic substances that are defined as long-term means. In most cases, a long-term average (LTA) OEL will be an internal standard (a company-specific standard), devised for substances for which there is no officially recognized OEL. For chronic disease agents with a time-weighted average (TWA) OEL, an LTA OEL can be devised from an existing OEL for daily TWA exposures. For example, the American Industrial Hygiene Association suggested that one could reasonably set an LTA OEL at approximately one-third the TWA OEL.(1) Roach and Rappaport,(2) after considering the uncertainty in the American Conference of Governmental Industrial Hygienists (ACGIH) Threshold Limit Values (TLV®), suggested that an LTA OEL of one-tenth to one-fourth of the TWA OEL may be appropriate. Similar recommendations have been presented by others.

In summary, there are instances where one may be interested in estimating the interval in which the true mean exposure is expected to lie. Armstrong(3) evaluated the ability of several simple procedures to generate accurate confidence limits around a sample mean for exposure data drawn from a lognormal distribution. Armstrong used as a gold standard the confidence limits computed using Land's(4) exact confidence interval procedure. He concluded that the "poor performance of the approximate methods . . . [for small sample sizes and geometric standard deviations (GSDs) larger than 2] suggests caution in their use," and in these instances recommended the use of Land's exact procedure. However, Armstrong(3) observed that Land's exact confidence limits are laborious to calculate and recommended the modified Cox procedure as the best of the approximate methods. Recently Lyles and Kupper(5) proposed a new procedure for calculating the upper confidence limit (UCL) for a sample mean of lognormally distributed data. Their computer simulations showed that their

 1047-322X/97/1202-132$17.00/11 · PII S1047-322X(96)00329-4

UCL will be similar to that calculated using Land's procedure, but that Land's procedure is slightly preferable for sample sizes less than ten or where the GSD is large. The procedure introduced here permits easy, yet accurate calculation of the preferred Land's exact confidence interval. This confidence interval can then be used to (1) gauge the uncertainty in the point estimate of the true mean and (2) evaluate the hypothesis that the true mean is greater than (or less than) the long-term mean OEL.

In those instances where one is dealing with either a TWA OEL or a short-term exposure limit, one often wishes to determine if the true exceedance fraction is less than some value. For example, Tuggle[6] recommended the use of the upper tolerance limit test and the complementary K′ test (pronounced "K prime") for testing whether or not the true exceedance fraction is significantly less than or greater than some target value, respectively. The alternative procedure described here was adapted from a procedure described by Odeh and Owen[7] for placing a confidence interval around the point estimate of the fraction in the tail of a normal distribution. A confidence interval around the exceedance fraction allows one, at a glance, to gauge the uncertainty in the point estimate and to compare it with some target value. A target value that has gained some acceptance is an exceedance fraction of 0.05, or 5 percent.[6,8]

Underlying both of these procedures are the following assumptions: (1) the true distribution is lognormal, or at least reasonably so; (2) the processes that give rise to the exposures will remain reasonably stable until the next scheduled survey; and (3) the measurements are reasonably random, independent observations. The last assumption implies that the following procedures are best suited for the situation where a single measurement is collected from each of n randomly selected workers per exposure group or n measurements are collected on randomly selected days for a specific worker.

Simplified Procedures

Confidence Interval About a Sample Mean

There are several statistics that can be used to estimate the true mean (μ) of a lognormally distributed variable from a sample of size n: (1) the simple arithmetic mean ($\bar{x}$), (2) the maximum likelihood estimate (MLE), and (3) the minimum variance unbiased estimator (MVUE). As pointed out by Attfield and Hewett[9] and Selvin and Rappaport,[10] there is often little difference between these estimators, particularly for large sample sizes. However, the MVUE is the preferred point estimate, particularly when the sample size is small and/or the sample GSD is large, and probably should be included along with the simple arithmetic mean when calculating summary statistics for any data set.

The simple arithmetic mean is the simplest and most obvious estimator of the true mean. It is unbiased, but is slightly more variable than the other estimators. A less variable estimator of μ is the MLE, which is best calculated from a large sample estimate of the true mean and variance of the log-transformed data:

$$\hat{\mu} = \exp\left(\hat{\mu}_y + \frac{1}{2}\hat{\sigma}_y^2\right) \tag{1}$$

where $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ are the mean and variance of the (natural) log-transformed concentration data [$y = \ln(c)$]. Note that for this equation the variance is calculated with n in the denominator rather than $n - 1$.

The MLE is a minimum variance estimator, but has a slight positive bias for small sample sizes, particularly when the underlying GSD is large. On average this results in overestimates of the true mean. The MVUE is both an unbiased and minimum variance estimator of the true mean:

$$\hat{\mu} = \exp(\bar{y})\Psi\left(\frac{s_y^2}{2}\right) \tag{2}$$

where $\bar{y}$ and s_y are the sample mean and standard deviation for $y = \ln(c)$. The Ψ function is defined[11,12] for any argument g as:

$$\Psi(g) = \left[1 + \frac{(n-1)g}{n} + \frac{(n-1)^3g^2}{n^2(n+1)2!} + \frac{(n-1)^5g^3}{n^3(n+1)(n+3)3!} + \frac{(n-1)^7g^4}{n^4(n+1)(n+3)(n+5)4!} + \cdots\right] \tag{3}$$

Equation 3 is easily calculated using a programmable calculator or personal computer. Calculation to at least five terms is usually sufficient.

The central limit theorem states that, with few exceptions, the distribution of sample means drawn from any distribution tends to approach a normal distribution as the sample size increases. This concept allows the use of t-values to construct reasonably accurate and symmetric confidence intervals about any sample mean for sample sizes greater than 30. However, the distribution of sample means from a lognormal distribution can be highly skewed for realistic sample sizes and for typically encountered GSDs. Consequently, the correct LCL and UCL for small sample sizes will not be symmetric about the true mean.

In response to this problem, Land[4] devised a set of equations and tables that allows one to calculate the exact UCL or LCL for the mean of a lognormal distribution. (These limits are exact in the sense that the actual confidence level equals the nominal confidence level.) For the lognormal distribution, Land's equations reduce to a familiar form for calculating confidence limits:

$$CL = \exp\left[\ln(\hat{\mu}) + C\frac{s_y}{\sqrt{n-1}}\right] \tag{4}$$

where:

$$\begin{aligned} CL &= \text{the UCL or LCL, depending upon the choice of } C \\ \hat{\mu} &= \exp\left(\bar{y} + \tfrac{1}{2}s_y^2\right) \\ C &= C(s_y;\, n,\, 1-\alpha) \text{ for the UCL where } \alpha = 0.05 \\ &= C(s_y;\, n,\, \alpha) \text{ for the LCL where } \alpha = 0.05 \end{aligned} \tag{5}$$

To use Land's procedure, one must use the equation given above for estimating the distribution mean (rather than Equa-

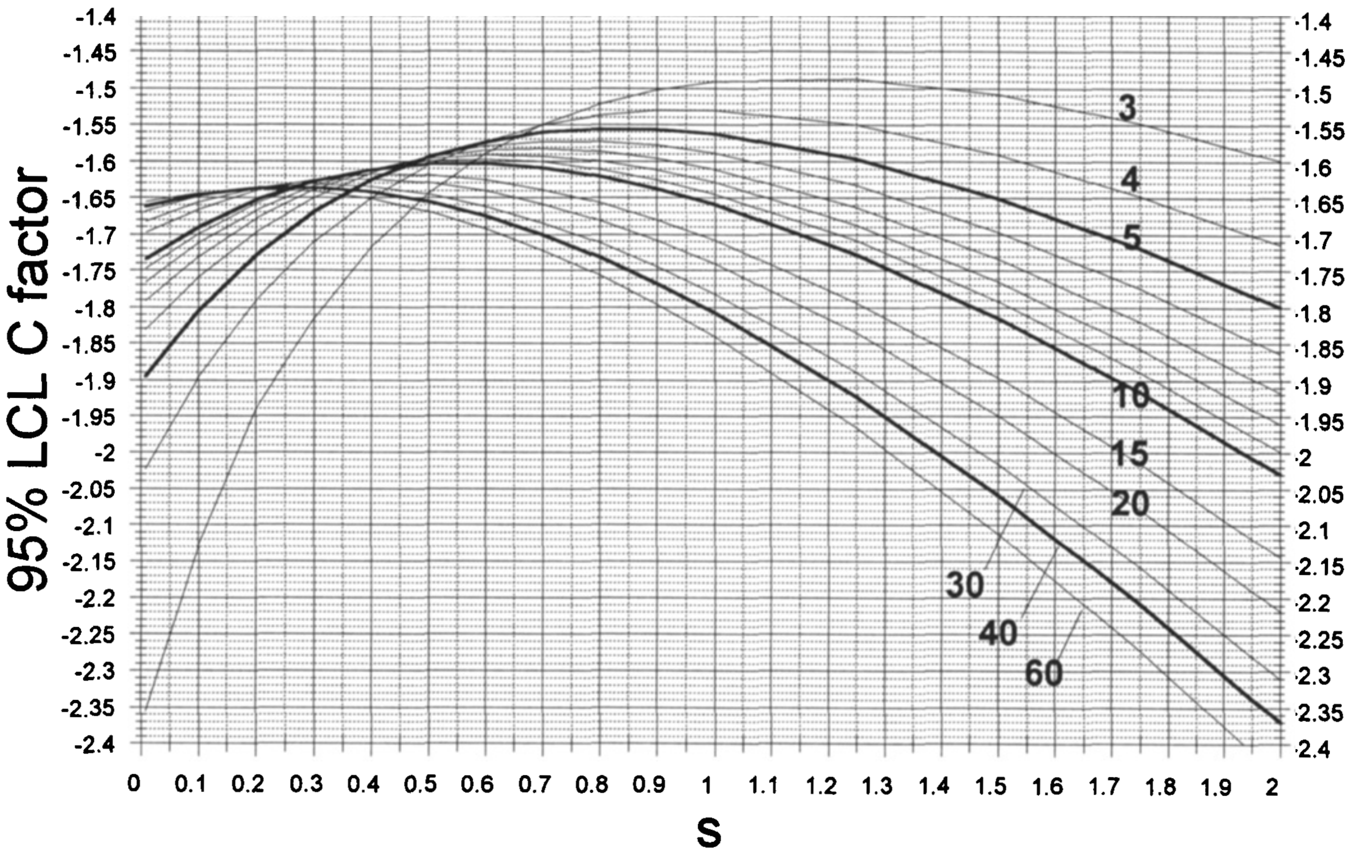

FIGURE 1. Land's LCL C-factor, as a function of S and sample size. C(S; n, α) may be read from the y axis for each combination of S and sample size and used to calculate the 95 percent LCL for the sample mean of lognormally distributed data.

tion 1). However, when reporting the point estimate of the true mean, one should use either the simple arithmetic mean or the preferred MVUE (Equation 2). Note that Land's C-factor and $\sqrt{(n-1)}$ replaces the familiar t-value and $\sqrt{n}$ found in the usual confidence limit formula. Also, unlike t-values, the absolute values of the UCL and LCL C-factors are not identical. Land's C-factors account for the variability in both the sample estimates of the true mean and variance of the log-transformed data used in the calculation of the MLE. Consequently, the minimum sample size for using Land's procedure is $n = 3$ (two parameters are estimated, resulting in the loss of two degrees of freedom).

Calculation of Land's exact 95 percent LCL and 95 percent UCL involves three steps:

1. Using Equation 5, calculate $\hat{\mu}$.
2. Obtain the C-factor for the LCL and UCL: $C(s_y; n, \alpha)$ and $C(s_y; n, 1-\alpha)$, respectively. These factors can be obtained from Land's[4] tables. [Land's tables are arranged by degrees of freedom ($\nu = n - 1$) rather than sample size.] However, Land's tables almost always require interpolation between the tabular values of s_y, and sometimes double interpolation as well (for those sample sizes for which there is no table one must interpolate both between and within tables). An alternative procedure recommended here is to simply read the appropriate C-factors from Figures 1 and 2.
3. Use Equation 4 to calculate the 95 percent LCL and 95 percent UCL for the true mean.

Taken together, the 95 percent LCL and 95 percent UCL comprise a 90 percent confidence interval. This interval can be interpreted to imply that if 100 sets of measurements were collected and analyzed, the true mean would lie within the 90 percent confidence interval 90 times out of 100.

Confidence Interval About the Exceedance Fraction

A point estimate of the true exceedance fraction (f) for lognormally distributed data can be calculated using the following equation:

$$\hat{f} = P(c > OEL) = P\left(Z > \frac{\ln OEL - \bar{y}}{s_y}\right)$$

where OEL is the limit for daily TWA exposures and $\bar{y}$ and s_y are the sample mean and sample standard deviation of the log-transformed data calculated from a sample of n measurements. Use a Z-table to determine the area under the standard normal curve to the right of this z-value. This area corresponds to the exceedance fraction.

The following procedure for calculating the 90 percent confidence interval around the exceedance fraction was

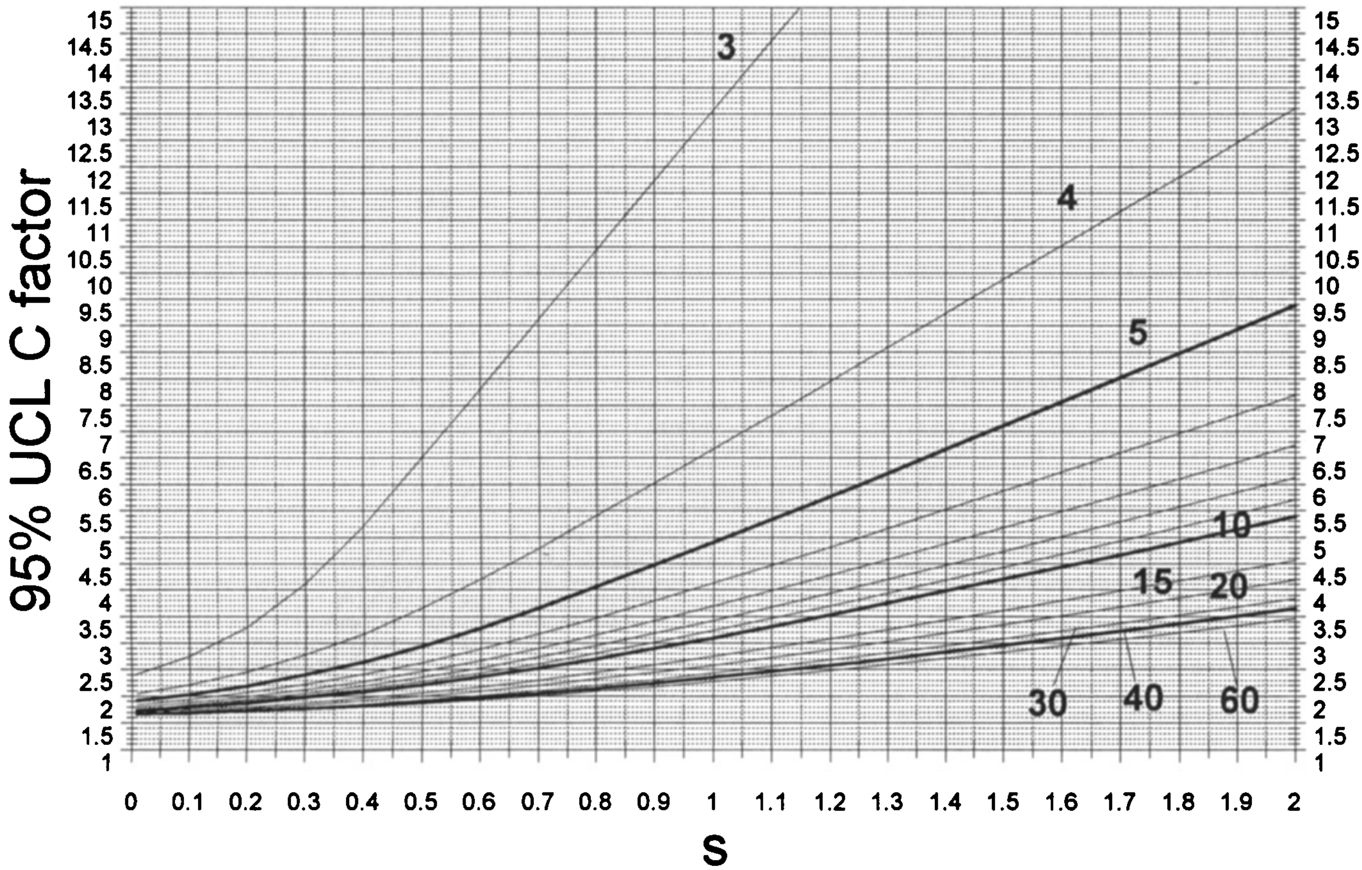

FIGURE 2. Land's UCL C-factor, as a function of S and sample size. C(S; n, 1 − α) may be read from the y, axis for each combination of S and sample size and used to calculate the 95 percent UCL for the sample mean of lognormally distributed data.

adapted from Table 7.4 of Odeh and Owen.[7] The procedure consists of two steps:

1. Calculate z (the same value used above to obtain the point estimate of the exceedance fraction):

$$z = \frac{\ln \mathrm{OEL} - \bar{y}}{s_y}$$

2. Using z and the sample size, n, read the confidence limit from Figure 3. This represents the 95 percent LCL for $\hat{f}$. Using negative z and n, read the confidence limit from Figure 3. The 95 percent UCL for $\hat{f}$ is the complement of this value (complement = 1 − value).

For most data sets Figures 4 and 5 should be used. These figures cover only a portion of Figure 3 and will often result in more accurate estimates of the LCL and UCL, respectively. Note that in Figure 5 the 95 percent UCL is read directly from the y axis. Taken together, the 95 percent LCL and 95 percent UCL form a 90 percent confidence interval for the true exceedance fraction.

Example

The following example illustrates these procedures. The data are real and represent measurements of exposure to total welding fume on five consecutive days for a single welder.

DATA. 4.25, 1.38, 3.11, 2.20, and 2.82 mg/m^3. The point estimates of the true arithmetic mean are 2.75 mg/m^3 (simple arithmetic mean) and 2.76 mg/m^3 (MVUE).

CONFIDENCE LIMITS ON THE MEAN.

Step 1: Using Equation 5, $\hat{\mu}$ is calculated to be 2.8133. (It is wise to retain several extra digits for all intermediate calculations, particularly for the estimates of variance, and round the last calculated value to the number of significant digits in the original data.)

Step 2: Read C(s_y; n, α) and C(s_y; n, 1 − α) from Figures 1 and 2, respectively. For s_y = 0.42 and n = 5 the C-factor for the 95 percent LCL calculation is approximately −1.62, while the C-factor for the 95 percent UCL calculation is approximately 2.7.

Step 3: Using Equation 4, C(0.42; 5, 0.05), s_y, and n, the 95 percent LCL is approximately 2.00. Using Equation 4, C(0.42; 5, 0.95), s_y, and n, the 95 percent UCL is approximately 4.97.

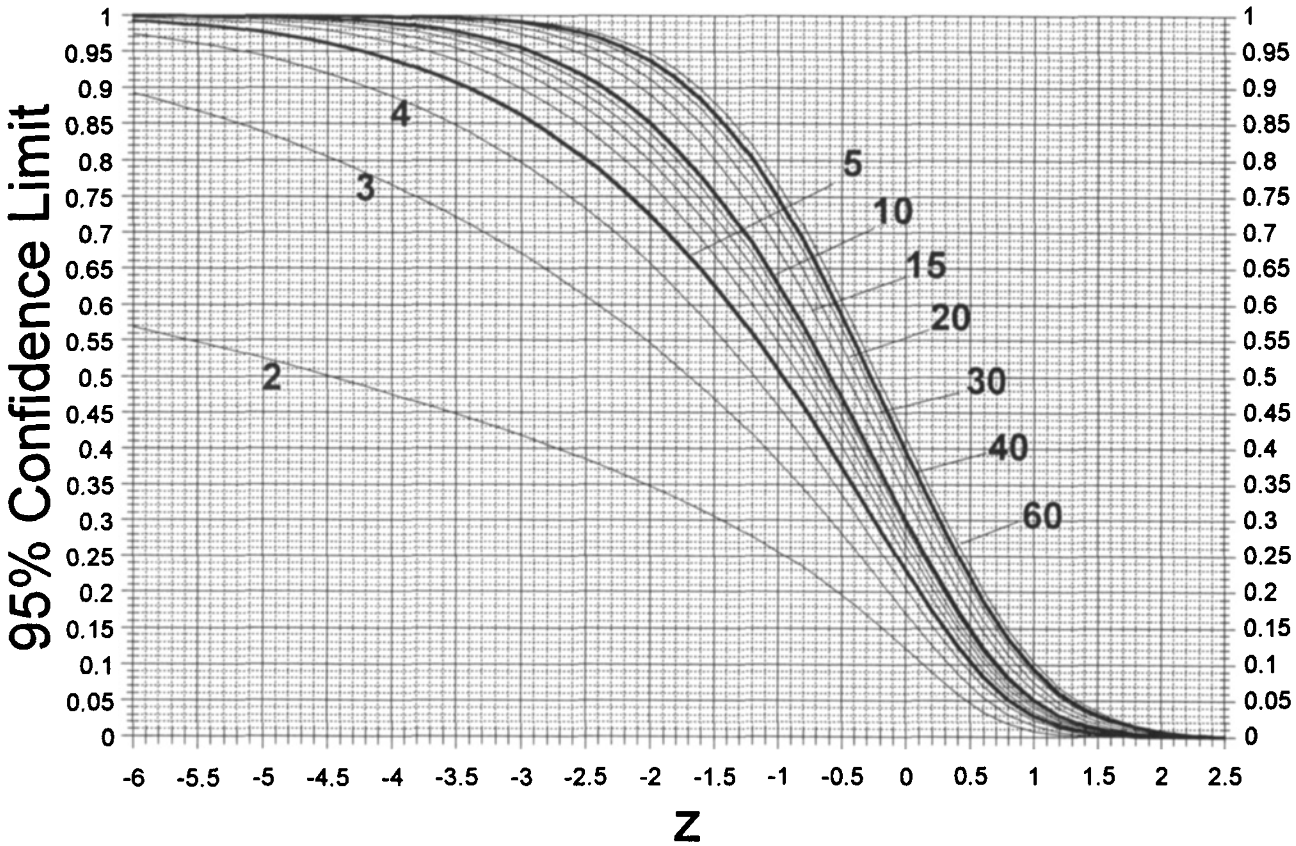

FIGURE 3. Confidence limit for the exceedance fraction (f) versus the calculated z-value. Using z and the sample size, read the 95 percent LCL from the y axis. Using negative z and the sample size, read the corresponding y axis value. The 95 percent UCL is the complement of this value (complement = 1 − value).

The resulting 90 percent confidence interval (2.00 to 4.97) compares favorably with that determined using Land's exact procedure: (2.0014 to 4.9757).

CONFIDENCE LIMITS ON THE EXCEEDANCE FRACTION.

Step 1: The sample geometric mean and GSD are 2.5748 and 1.5234, respectively. The ACGIH TLV for mild steel welding fumes is 5 mg/m^3. Calculate the z-value needed to determine the LCL and UCL for the exceedance fraction (relative to the TLV):

$$z = \frac{\ln 5 - \ln 2.5748}{\ln 1.5234} = 1.5766$$

[This value can also be used to determine the point estimate of the exceedance fraction: $\hat{f} = P(c > OEL) = P(Z > z)$. From a Z-table, the probability of a z-value of 1.5766 or greater is approximately 0.057.]

Step 2: Either Figure 3 or Figure 4 can be used to determine the 95 percent LCL for the point estimate of the exceedance fraction. Either Figure 3 or Figure 5 can be used to determine the 95 percent UCL. The LCL is determined by finding the curve for a sample size of $n = 5$ and reading the confidence limit corresponding to $z \approx 1.58$. For this example, a 95 percent LCL of about 0.005 can be read from Figure 4. Using Figure 3 and a negative z, −1.58 in this case, the UCL is determined by first reading the confidence limit and then subtracting this value from 1. From Figure 3, a confidence limit of about 0.64 can be read, resulting in a 95 percent UCL of 1 − 0.64, or 0.36. A similar value of 0.355 can be read from Figure 5.

The resulting 90 percent confidence interval (0.005 to 0.355) compares favorably with that determined by six-point Lagrange interpolation of the values in Odeh and Owen's[7] Table 7.4: (0.0038 to 0.3555).

Discussion

The procedure for estimating the confidence interval about the mean is based on Land's exact procedure, obviating the need to use the often inaccurate[3] approximate procedures, and results in a confidence interval which is, for practical purposes, nearly identical to that determined using Land's tables. The point estimate of the mean, whether it is the simple arithmetic mean or the MVUE, indicates the acceptability of the work environment (relative to the LTA OEL), while the confidence interval indicates the degree of uncertainty in the point estimate. The 90 percent confidence interval can be used for hypothesis testing. If the UCL is less than the LTA OEL, one is justified in concluding, with 95 percent confidence, that the

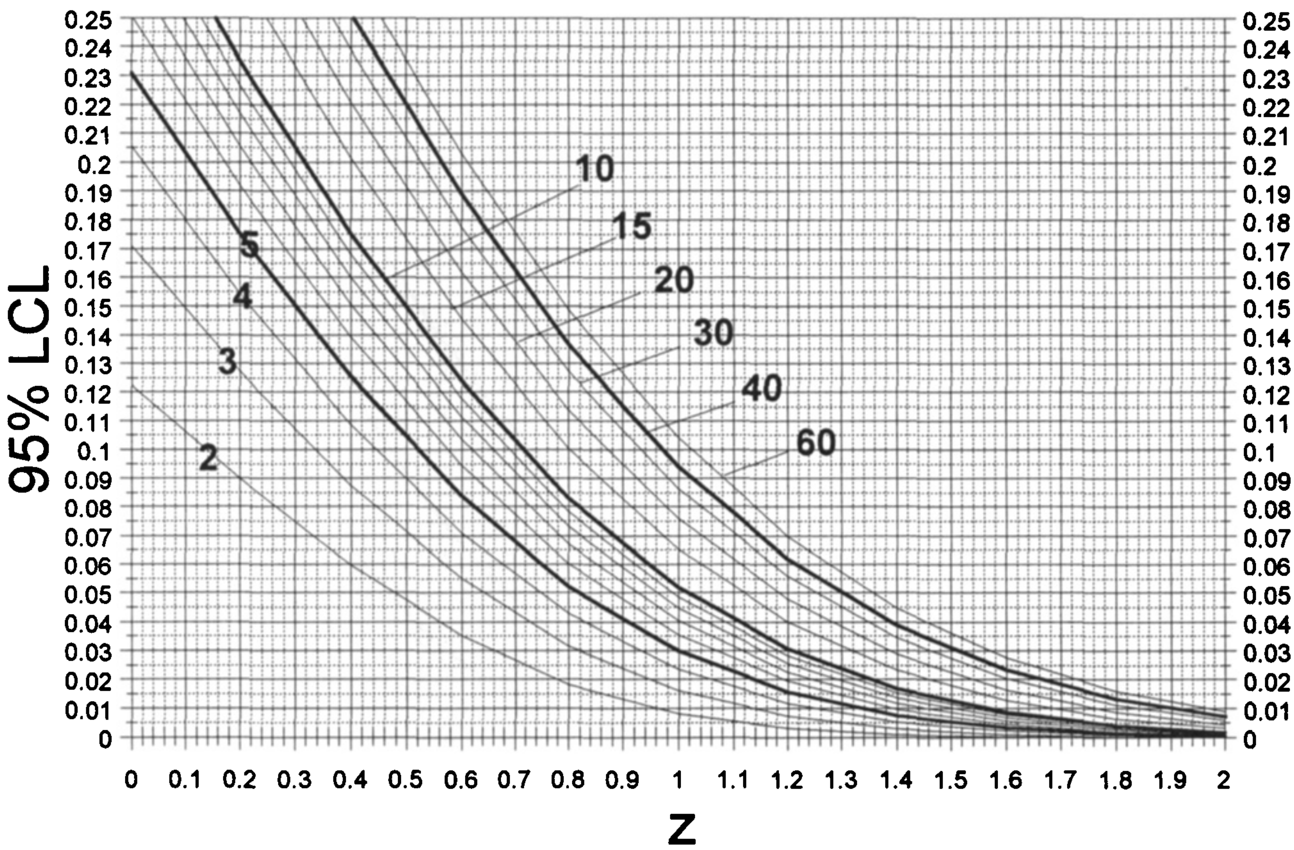

FIGURE 4. 95 percent LCL for the exceedance fraction (f) versus the calculated z-value. Using z and the sample size, read the 95 percent LCL from the y axis.

true mean is less than the LTA OEL and that the work environment is acceptable (relative to the LTA OEL). Conversely, if the LCL is greater than the LTA OEL, then one is justified in concluding, with 95 percent confidence, that the true mean is greater than LTA OEL and that additional controls are truly necessary.

The procedure for estimating the confidence interval about the exceedance fraction is an adaptation of an established procedure for estimating the UCL or LCL for the point estimate of a proportion in one tail of the normal distribution. Since, for a lognormal distribution, the log-transformed data are normally distributed, this procedure is easily be adapted for use in the occupational setting. The resulting confidence limits are reasonably accurate, closely approaching those determined from interpolation between the Odeh and Owen[7] confidence limits, and can be used for hypothesis testing. If the UCL for the exceedance fraction is less than some criterion, say 0.05 (or 5%), then one is justified in concluding, with 95 percent confidence, that the true exceedance fraction is less than 0.05 (or 5%). If the LCL for the exceedance fraction is greater than some criterion, say 0.05 (or 5%), then one is justified in concluding, with 95 percent confidence, that the true exceedance fraction is greater than 0.05 (or 5%). These two tests are analogous to the 95 percent upper tolerance limit test and the K′ tolerance limit test, respectively, described by Tuggle.[6]

The problem of estimating the uncertainty in the point estimate of the exceedance fraction was addressed recently by Leidel and Busch.[13] However, their procedure often yields confidence limits considerably different than those derived using the procedure of Odeh and Owen.[7] For the example above, their procedure yields a 90 percent confidence interval of 0.00 to 0.26. Leidel and Busch's procedure can be shown to be inaccurate by considering two hypothetical cases where the UCL or LCL should be a certain value. Consider a situation where the sample size is 5, the sample geometric mean is 1, and the sample GSD is 2. Set the OEL to the 95 percent one-sided upper tolerance limit for 95 percent of future values, which in this case is 18.4. Now calculate the 95 percent UCL for the exceedance fraction. Leidel and Busch's procedure yields a 95 percent UCL of 0.01. Using Odeh and Owen's procedure, one first calculates a z-value of approximately 4.2. Using $n = 5$, $-z = -4.2$, and Figure 5, the 95 percent UCL is approximately 0.05, which is the expected value. Consider the same situation but calculate the K′ tolerance limit described by Tuggle,[6] which in this case is 1.76. Set the OEL to 1.76 and calculate the 95 percent LCL for the exceedance fraction. As before, the expected value is 0.05. Leidel and Busch's procedure yields a 95 percent LCL of 0.007. Using Odeh and Owen's procedure, the corresponding z-value is approximately 0.82. Using $n = 5$, $z = 0.82$, and Figure 4, the 95 percent LCL is approximately 0.05. In light of these results the Odeh and Owen procedure is clearly preferable to that de-

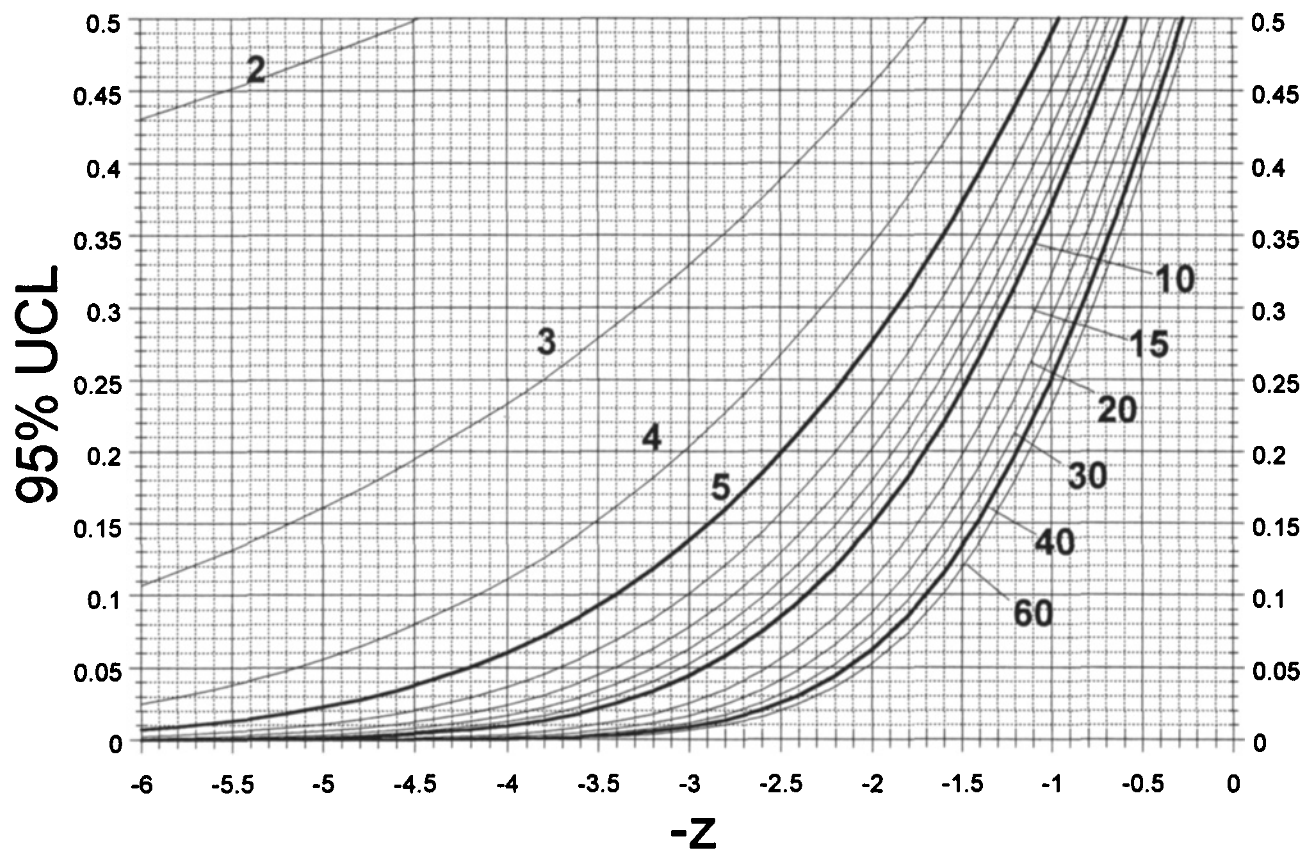

FIGURE 5. 95 percent UCL for the exceedance fraction (f) versus the negative of the calculated z-value. Using −z and the sample size, read the 95 percent UCL from the y axis.

scribed by Leidel and Busch. Furthermore, this exercise shows that the one-sided tolerance limit tests (the upper tolerance limit test and Tuggle's K′ test) and Odeh and Owen's confidence interval procedure, when used for testing hypotheses regarding the exceedance fraction, are equivalent. In fact, Tuggle's K′ factors and the one-sided upper tolerance limit K factors can be determined from Figures 4 and 5, respectively.

Conclusions

These procedures have several advantages: (1) they are simple to use, (2) they result in reasonably accurate estimates of the 95 percent LCL and UCL for two parameters often used to determine the acceptability of a work environment, (3) the resulting confidence intervals indicate the uncertainty in the point estimates, and (4) the LCL and UCL can be used for hypothesis testing.

A side benefit of these procedures is that one can use the figures to easily estimate the number of additional measurements needed to lower the UCL for either the mean or exceedance fraction to less than the critical value. This, of course, assumes that the exposure distribution is fairly stable and that future measurements will be similar to those already obtained. One should be aware, however, that the actual confidence level will be slightly less than the nominal confidence level for two-stage (sequential) sampling plans.[14,15]

Note

After this article was accepted, the authors developed a system of approximation functions that permits the automatic calculation of Land's confidence limits for the mean of lognormally distributed data and the Odeh and Owen confidence limits for the exceedance fraction. These functions and associated coefficients are listed in Appendices A and B.

References

1. Hawkins, N.C.; Norwood, S.K.; Rock, J.C.: A Strategy for Occupational Exposure Assessment. American Industrial Hygiene Association, Fairview, VA (1991).
2. Roach, S.A.; Rappaport, S.M.: But They Are Not Thresholds: A Critical Analysis of the Documentation of Threshold Limit Values. Am. J. Ind. Med. 17:727–753 (1990).
3. Armstrong, B.G.: Confidence Intervals for Arithmetic Means of Lognormally Distributed Exposures. Am. Ind. Hyg. Assoc. J. 53:481–485 (1992).
4. Land, C.E.: Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance. In: Selected Tables in Mathematical Statistics, Vol. III, pp. 385–419. American Mathematical Society, Providence, RI (1975).
5. Lyles, R.H.; Kupper, L.L.: On Strategies for Comparing Occupational Exposure Data to Limits. Am. Ind. Hyg. Assoc. J. 57:6–15 (1996).
6. Tuggle, R.M.: Assessment of Occupational Exposure Using

One-sided Tolerance Limits. Am. Ind. Hyg. Assoc. J. 43:338–346 (1982).
7. Odeh, R.E.; Owen, D.B.: Statistics: Textbooks and Monographs Series, Vol. 32—Tables for Normal Tolerance Limits, Sampling Plans, and Screening. (1980).
8. Leidel, N.A.; Busch, K.A.; Lynch, J.R.: Occupational Exposure Sampling Strategy Manual. National Institute for Occupational Safety and Health Pub. No. 77-173 (available from the National Technical Information Service, Pub. No. PB274792) (1977).
9. Attfield, M.D.; Hewett, P.: Exact Expressions for the Bias and Variance of Estimators of the Mean of a Lognormal Distribution. Am. Ind. Hyg. Assoc. J. 53:432–435 (1992).
10. Selvin, S.; Rappaport, S.M.: A Note on the Estimation of the Mean Value from a Lognormal Distribution. Am. Ind. Hyg. Assoc. J. 50:627–630 (1989).
11. Aitchison, J.; Brown, J.A.C.: The Lognormal Distribution with Special Reference to Its Uses in Economics. Cambridge University Press, Cambridge, UK (1957).
12. Finney, D.J.: On the Distribution of a Variate Whose Logarithm Is Normally Distributed. Journal of the Royal Statistical Society, Supplement 7:155–161 (1941).
13. Leidel, N.A.; Busch, K.A.: Statistical Design and Data Analysis Requirements. In: Patty's Industrial Hygiene and Toxicology, 3rd ed., Vol. 3, Part A, R.L. Harris, L.J. Cralley, and L.V. Cralley, Eds. (1994).
14. Montgomery, D.C.: Introduction to Statistical Quality Control, 2nd ed. John Wiley & Sons, New York (1991).
15. Selvin, S.; Rappaport, S.; Spear, R.; et al.: A Note on the Assessment of Exposure Using One-sided Tolerance Limits. Am. Ind. Hyg. Assoc. J. 48:89–93 (1987).

Appendix A: Approximation of Land's 95 Percent LCL and 95 Percent UCL C-Factors

The authors developed a set of equations that permit the accurate approximation of Land's $\alpha = 0.05$ and $\alpha = 0.95$ C-factors for all sample sizes between 3 and 1001 and all values of s_y between 0.01 and 4 (corresponding to GSDs of 1.01 to 54.5). These equations can be adapted for use in computer programs or spreadsheets to automatically calculate the 95 percent LCL and 95 percent UCL for the mean of a lognormally distributed data set. The percent difference between a 95 percent LCL C-factor estimated using this procedure and the published C-factors[4] is between −3 and +2 percent, depending upon the combination of sample size and s_y. Similarly, the percent difference for an approximate 95 percent UCL C-factor is between −3 and +5 percent. The percent differences for values of s_y and/or n not represented in the published tables are expected to be in a similar range for both the LCL and UCL C-factors.

Range Check

If $(n < 3)$ or $(n > 1001)$, then this procedure cannot be applied.

If $(s_y < 0.01)$, then let $s_y = 0.01$.

The estimated GSD is nearly 1.00 and can conservatively be set at 1.01 ($s_y = \ln(1.01) = 0.01$).

If $(s_y > 4)$, then let $s_y = 4$.

The estimated GSD is greater than 54.5; s_y should be set to 4 in order to estimate an approximate LCL and UCL. Realistically this data set should first be critically examined to determine why such a large sample GSD occurred.

If $(s_y \geq 0.01)$ and $(s_y \leq 4)$, then proceed.

Coefficients Needed to Estimate Land's 0.05 (95% LCL) C-Factor

$a = -0.85033767$

$b = -0.5258052$

$c = 0.92416176$

$d = 3.3298209$

$e = 0.94348568$

$f = 1.3213281$

$g = 0.8155562$

$h = -1.018148$

$i = 0.25248895$

Intermediate Equations

$F1 = a/(n-2)^c + b$

$F2 = d/(n-2)^f + e$

$F3 = g/(n-2)^i + h$

$F4 = -0.6226/(n-2)^{0.2426} + 1.1470$

Final Equation

$$C(s_y; n, 0.05) = -0.74295/(n-2) - 1.64765 + s_y \cdot [F3 + F2 \cdot \exp(F1 \cdot s_y^{F4})]$$

Coefficients Needed to Estimate Land's 0.95 (95% UCL) C-Factor

$a = 0.76766658$

$b = 3.8716869$

$c = 0.80598919$

$d = 6.0321019$

$e = 0.89998154$

$f = 2.012669$

$g = 0.21978875$

$h = 0.41575588$

$i = 0.29258276$

Intermediate Equations

$F1 = s_y \cdot [i + 1/(n-2)^c]$

$F2 = b + d/(n - 2)^c$

$F3 = F1 \cdot [1 - e \cdot \exp(-f \cdot F1)]$

$F4 = [1 + g \cdot \exp(-h \cdot F1)]$

$F5 = F2 \cdot F3/F4$

Final Equation

$C(s_y; n, 0.95) = 1.645 + a/(n - 2) + F5$

Appendix B: Approximation of the Exceedance Fraction 95 Percent LCL and 95 Percent UCL

The authors developed a set of equations that permit the accurate approximation of the confidence limits for the proportion in the tail of a normal (or lognormal) distribution for all sample sizes between 2 and 1000 and all values of z (or −z) between −6 and 3. These equations can be used in computer programs or spreadsheets to automatically calculate the 95 percent LCL and 95 percent UCL for the exceedance fraction for a lognormally distributed data set. The 95 percent LCL and 95 percent UCL resulting from this procedure will be within ±0.001 of those determined using the procedure given in Reference 7 and the z-values and sample sizes given in Table 7.4 of Reference 7. For intermediate values of z and/or n the authors expect that the accuracy of the following approximation equations will also be accurate to within ±0.001 of the exact confidence limit.

Range Check

If ($n < 2$) or ($n > 1000$), then this procedure cannot be applied.

Store the Original Value of z to Be Used Later

z_orig = z

Range Check for the LCL Calculation

If ($z > 3$), then return "<0.001."
If ($z < -6$), then $z = -6$.

Range Check for the UCL Calculation

$z = -z$ (The negative of z is used to calculate the UCL.)
If ($z > 3$), then return ">0.999."
If ($z < -6$), then $z = -6$.

Coefficients Needed for Sample Sizes of n = 2 to n = 9

If n = 2, then

a = 0.122639

b = −1.183888

c = −0.30512797

d = 1.025662

e = 0.30466959

f = −0.32067478

g = −0.1511292

h = 0.061852348

i = 0.036997109

j = −0.0023723692

k = −0.0035624416

Else if n = 3, then

a = 0.17093583

b = −0.36372227

c = −0.28537457

d = 0.34851212

e = 0.16073371

f = 0.054071371

g = −0.025966434

h = 0.002995818

i = −0.0050019868

j = 0.0016331608

k = 0.0014215249

Else if n = 4, then

a = 0.20527676

b = −0.3756623

c = −0.33810835

d = 0.36738495

e = 0.20373584

f = 0.018923968

g = −0.049770351

h = 0.011064939

i = 0.0024343744

j = 0.00085720372

k = 0.00050349027

Else if n = 5, then

a = 0.23087629

b = −0.32948211

c = −0.36170947

d = 0.35945667

e = 0.21196252

f = 0.019458266

g = −0.053579348

h = 0.015386764

i = 0.0042734939

j = 0.00075015011

k = 0.00019427452

Else if n = 6, then

a = 0.25087349

b = −0.30146272

c = −0.3787192

d = 0.3687124

e = 0.22058761

f = 0.016283799

g = −0.060450757

h = 0.020005908

i = 0.0074082789

j = 0.00039892655

k = −0.00027967341

Else if n = 7, then

a = 0.26701314

b = −0.26555197

c = −0.38705351

d = 0.38401727

e = 0.22384309

f = 0.015667728

g = −0.066696572

h = 0.024962675

i = 0.010898192

j = −0.000042375686

k = −0.00083411655

Else if n = 8, then

a = 0.28035886

b = −0.21322137

c = −0.38608727

d = 0.40265771

e = 0.22000682

f = 0.018042944

g = −0.071663981

h = 0.030244939

i = 0.014749941

j = −0.00060041754

k = −0.0014920275

Else if n = 9, then

a = 0.29158511

b = −0.14267075

c = −0.37596357

d = 0.41758212

e = 0.2072291

f = 0.023635477

g = −0.073405376

h = 0.03496479

i = 0.018251034

j = −0.0011924843

k = −0.0021651986

End

Coefficients Needed for Samples Sizes of n = 10 to n = 1000

If (10 ≤ n and n ≤ 1000), then
If (0 ≤ z and z ≤ 3), then

a = 0.82705588

b = 0.47348658

c = −1.5959026

d = −0.35885104

e = −0.3762358

f = −0.0059178009

g = 0.052553183

h = 0.26488682

i = −0.075413666

j = −2.4186812

k = 0.54932221

l = −1.0076516

Else if (−6.0 ≤ z and z < 0), then

a = 0.9159042

b = 0.46749901

c = −1.5978413

d = −0.21867454

e = −0.16711347

f = −0.055872009

g = 0.078057826

h = 0.28495322

i = −0.073167856

j = −2.4074152

k = 0.54932221

l = −1.0076516

End

End

If (2 ≤ n and n ≤ 9), then

$$CL = (a + c \cdot z + e \cdot z^2 + g \cdot z^3 + i \cdot z^4 + k \cdot z^5)/(1 + b \cdot z + d \cdot z^2 + f \cdot z^3 + h \cdot z^4 + j \cdot z^5)$$

Else if (10 ≤ n and n ≤ 1000), then

$$F1 = a/(n-1)^b + c$$

$$F2 = d \cdot (n-1)^e \cdot \exp\{f \cdot [\ln(n-1)]^2\}$$

$$F3 = g/(n-1)^h + i$$

$$F4 = j/(n-1)^k + 1$$

$$CL = \{1 + \exp[-(F1 \cdot z + F2 \cdot z^2 + F3 \cdot z^3)]\}^{F4}$$

End

For the LCL Calculation

If (CL < 0.001), then

LCL = "<0.001"

Else if (CL > 0.999), then

LCL = ">0.999"

Else if (z_orig < −6), then

LCL = ">(CL value)"

Else

LCL = CL (rounded to the nearest 0.001)

End

For the UCL Calculation

If (1 − CL < 0.001), then

UCL = "<0.001"

Else if (1 − CL > 0.999), then

UCL = ">0.999"

Else if (z_orig > 6), then

UCL = "<(1 − CL value)"

Else

UCL = 1 − CL (rounded to the nearest 0.001)

End