

Sample Size and Number of Failure Requirements for Demonstration Tests with Log-Location-Scale Distributions and Type II Censoring

Scott W. McKane
3M Pharmaceuticals
3M Center, Building 0275-2W-14
St. Paul, MN 55144-1000

Luis A. Escobar
Dept. of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
luis@lsu.edu

William Q. Meeker
Dept. of Statistics
Iowa State University
Ames, IA 50011
wqmeeke@iastate.edu

March 2, 2002

Abstract

Reliability demonstration tests require demonstrating, with some level of confidence, that reliability exceeds a given standard. Demonstration tests can be expensive and time-consuming. Careful planning of sample size and test length are essential. This paper develops exact theoretical methods, based on pivotal quantities and confidence intervals, to aid in proper sample size selection and determining how long the test should be run (in terms of how many units must fail before the test's end) for demonstration tests with Type II censored data from log-location-scale (and the corresponding location-scale) distributions. The methods have been implemented in S-PLUS for the lognormal, Weibull, and loglogistic distributions to allow users to develop graphs depicting probability of successful demonstration as a function of actual reliability, a target reliability, sample size, and number of units failing for an assumed distribution.

Key words: Maximum Likelihood, Reliability, Simulation.

1 Introduction

1.1 Problem and motivation

Test planners frequently are interested in determining the sample size and test length for life tests to be used in demonstrating, with some level of confidence, that reliability exceeds a given standard.

For instance, in a larger-is-better application (e.g., strength) a supplier may be asked to demonstrate that the p quantile of the distribution, t_p , exceeds a target quantity (usually in the lower tail of the distribution), t^\dagger . A demonstration is successful if $t_p > t^\dagger$, where t_p is a lower confidence bound on t_p . Alternatively (but equivalently), a demonstration can be stated in terms of probabilities. For example, a manufacturer may be required by a customer to demonstrate that a product's reliability (probability of survival) at time t^\dagger , say q , is greater than a stated target reliability, q^\dagger . A demonstration is successful if $q > q^\dagger$, where q is a lower confidence bound on q . For smaller-is-better applications (e.g., operating noise level), there are similar demonstration statements.

This paper reviews statistical methods for planning demonstration tests involving Type II censored data, a type of censoring in which the test is run until a specified number of failures occurs. The paper presents methods for computing the probability of successful demonstration as a function of actual reliability, sample size, and number of units failing.

1.2 Related work

Lieberman and Resnikoff (1955) gave theory, tables, and charts for variables one and two-sided sampling plans for a normal distribution with complete data. In related work, Owen (1962, page 108–137) presented early theory and tables for constructing tolerance intervals and one-sided tolerance bounds. Faulkenberry and Weeks (1968) discussed methods for finding sample sizes such that the enclosure probability of a one-sided tolerance bound would not be too big. A one-sided tolerance bound is equivalent to a one-sided confidence bound for a particular quantile of the random variable Y . Odeh and Owen (1980, page 267–271) reviewed the theory in Owen (1962) and produced tables of constants to use in constructing one-sided tolerance bounds from complete (i.e., uncensored) data. Other related statistical intervals and sample size issues are also addressed in this book and the book provides other relevant references.

Hahn and Meeker (1991, page 150–168) developed curves to aid in sample size selection to achieve a desired probability of successful demonstration for demonstration tests involving normal data with no censoring. Figures 9.1a–n from Hahn and Meeker (1991, page 154–167) and Appendix Section A.3 of this paper rely on noncentral- t distribution theory discussed in Odeh and Owen (1980). In Section 5.1, the methods developed in this paper are used to replicate Figure 9.1k of Hahn and Meeker (1991). That figure is a special case, using complete data from a normal distribution, of the methodology given here.

Thoman, Bain, and Antle (1970) provide theory (and limited tables) for computing lower confidence bounds on survival probabilities with exact coverage for complete samples from a Weibull distribution. They also show how to use their tables to compute one-sided confidence bounds for quantiles. Billman, Antle, and Bain (1972) give limited tables for lower bounds on survival probabilities with failure censored samples from the Weibull distribution. See Lawless (1982, page 155) for other related references. Schneider (1986, 1989) describes sampling plan designs for failure-censored normal, lognormal, and Weibull data, based on large-sample approximations.

In this paper we extend previous work, providing *exact* theory and methods for planning Type II censored life tests for location-scale or log-location scale distributions.

1.3 Overview

The purpose of this paper is to assist test planners in selecting a sample size, n , and number of failures, r , needed to achieve a desired probability of successful demonstration for life tests using direct

theoretical results and simulation. The methods provide exact probabilities of successful demonstration using a pivotal quantity approach for location-scale based distributions with Type II censoring. Section 2 discusses the models and estimation procedures to be utilized. Details of constructing exact confidence intervals for a quantile based on pivotal quantities are given in Section 3. These confidence intervals provide the underlying theory for the calculation of the probability of successful demonstration in Section 4. Section 5 gives numerical examples and comparisons. The Appendix provides derivations and other technical details, including a demonstration that noncentral- t based methods for uncensored normal distributed data are a special case of the more general methods presented here.

2 Model and Estimation

2.1 Location-scale and log-location-scale distributions

The results of this paper apply to location-scale and log-location-scale distributions. A random variable Y , $-\infty < Y < \infty$, belongs to the location-scale family of distributions if its cdf can be expressed as

$$F(y; \mu, \sigma) = \Pr(Y \leq y) = \Phi\left(\frac{y - \mu}{\sigma}\right)$$

where μ ($-\infty < \mu < \infty$) is a location parameter, σ ($\sigma > 0$) is a scale parameter, and $\Phi(z)$ is a cdf that does not depend on any unknown parameters. $\Phi(z)$ is the cdf of $(Y - \mu)/\sigma$ and when $\mu = 0$ and $\sigma = 1$ then $\Phi(z)$ is the cdf of Y . The normal (NOR), the smallest extreme value (SEV), and the logistic distributions are location-scale distributions.

A positive random variable T belongs to the log-location-scale family distribution if $Y = \log(T)$ is a member of the location-scale family. The lognormal, the Weibull, and the loglogistic are among the important distributions of this family. For example, the cdf and pdf of the Weibull random variable T can be expressed as

$$\begin{aligned} F(t) &= F(t; \mu, \sigma) = \Pr(T \leq t) = \Phi_{\text{sev}}\left(\frac{\log(t) - \mu}{\sigma}\right) \\ f(t) &= f(t; \mu, \sigma) = \frac{dF(t)}{dt} = \frac{1}{\sigma t} \phi_{\text{sev}}\left(\frac{\log(t) - \mu}{\sigma}\right) \end{aligned}$$

where $\Phi_{\text{sev}}(z) = 1 - \exp(-\exp(z))$ and $\phi_{\text{sev}}(z) = \exp(z - \exp(z))$ are the standard smallest extreme value cdf and pdf, respectively. For the lognormal distribution, replace Φ_{sev} and ϕ_{sev} above with Φ_{nor} and ϕ_{nor} , the standard normal cdf and pdf, respectively.

2.2 Quantiles and probabilities

In life tests, inferences frequently focus on quantiles or probabilities. For example, estimation of a quantile, t_p ; probability of failure at a given time t , $F(t)$; or reliability at t , $1 - F(t)$, may be of interest. For ease of calculation when using the log-location-scale family distributions, the quantile t_p is often replaced with $y_p = \log(t_p)$. This convention will be followed throughout this paper. The p quantile, y_p , of the distribution of Y can be expressed as $y_p = \mu + \Phi^{-1}(p)\sigma$, where μ is the location parameter, σ is the scale parameter, and $\Phi^{-1}(p)$ is the p quantile (inverse function) of the standard cdf $\Phi(z)$.

2.3 Censored data

Life tests often result in censored data. Type I (time) censored data result when unfailed units are removed from test at a prespecified time, perhaps due to limited time for study completion. Type II (failure) censored data result when a test is terminated after a specified number of failures, say $2 \leq r \leq n$. If all units fail, the data are called “complete.” The results of this paper are exact for life tests involving complete or Type II censored data, but only approximate for tests involving Type I censored data.

2.4 Maximum likelihood estimation

For a Type II censored sample with r failures in n independent observations from a log-location-scale random variable T with cdf $\Phi[(\log(t) - \mu)/\sigma]$, the likelihood is

$$L(\mu, \sigma) = \frac{n!}{(n-r)!} \prod_{i=1}^r \left\{ \frac{1}{\sigma t_{[i]}} \phi \left[\frac{\log(t_{[i]}) - \mu}{\sigma} \right] \right\} \times \left\{ 1 - \Phi \left[\frac{\log(t_{[r]}) - \mu}{\sigma} \right] \right\}^{n-r}$$

where $t_{[1]} < \dots < t_{[r]}$ are the ordered failures. Standard computer software (e.g., JMP, Minitab, SAS, S-PLUS) provide maximum likelihood (ML) estimates of μ and σ . Denote these values by $\hat{\mu}$ and $\hat{\sigma}$, respectively. By the invariance property of ML estimates, the ML estimate of the p log quantile, $y_p = \log(t_p) = \mu + \Phi^{-1}(p)\sigma$, is

$$\hat{y}_p = \hat{\mu} + \Phi^{-1}(p) \hat{\sigma}$$

where $\Phi^{-1}(p)$ is the p quantile of the standardized cdf $\Phi(z)$.

3 Confidence Intervals for Quantiles and Reliabilities

Using a general pivotal quantity approach, computation of factors for computing confidence intervals, upper, and lower bounds for functions of model parameters is easily accomplished for location-scale based distributions.

3.1 Confidence intervals and bounds for quantiles

To obtain a confidence interval for the quantile y_p , we proceed as follows. Consider the pivotal quantity

$$K = K(\hat{\mu}, \hat{\sigma}, y_p) = \frac{\hat{\mu} - y_p}{\hat{\sigma}} = \frac{\hat{\mu} - \mu - \Phi^{-1}(p)\sigma}{\hat{\sigma}}. \quad (1)$$

For complete data or Type II censored data the distribution of K depends only on (p, n, r) and the assumed distribution $\Phi(z)$, but not on any unknown parameters, this result follows from Lawless (1982, page 534).

For $0 < \gamma < 1$, define $k_{(\gamma;p,n,r)}$ as the γ quantile of the distribution of K , then

$$\begin{aligned} 1 - \alpha &= \Pr \left(k_{(\alpha/2;p,n,r)} \leq \frac{\hat{\mu} - y_p}{\hat{\sigma}} \leq k_{(1-\alpha/2;p,n,r)} \right) \\ &= \Pr \left(\hat{\mu} - k_{(1-\alpha/2;p,n,r)} \hat{\sigma} \leq y_p \leq \hat{\mu} - k_{(\alpha/2;p,n,r)} \hat{\sigma} \right). \end{aligned}$$

Consequently, a two-sided $100(1 - \alpha)\%$ confidence interval for y_p based on K is

$$\left[\underline{y}_p, \tilde{y}_p \right] = \left[\hat{\mu} - k_{(1-\alpha/2;p,n,r)} \hat{\sigma}, \hat{\mu} - k_{(\alpha/2;p,n,r)} \hat{\sigma} \right]. \quad (2)$$

This confidence interval depends on the quantiles $k_{(1-\alpha/2;p,n,r)}$ and $k_{(\alpha/2;p,n,r)}$; details for the computation of these quantiles are given in Section 3.3. Appendix Section A.1 shows that the confidence interval in (2) is invariant to choices of the pivotal quantity within a large class of possible pivotal quantities, of which (1) is a special case.

A lower (upper) one-sided confidence bound for y_p can be obtained by using the appropriate side of the two-sided interval in (2) and adjusting for the confidence level. For example, a one-sided $100(1-\alpha)\%$ lower confidence bound on y_p is

$$\underline{y}_p = \hat{\mu} - k_{(1-\alpha;p,n,r)} \hat{\sigma}. \quad (3)$$

3.2 Confidence intervals and bounds for reliabilities

Consider the reliability (survival probability) at y^\dagger , $q = \Pr(Y > y^\dagger) = 1 - \Phi[(y^\dagger - \mu)/\sigma]$, and its ML estimate $\hat{q} = 1 - \Phi[(y^\dagger - \hat{\mu})/\hat{\sigma}]$. Denote by Q the random variable corresponding to these \hat{q} estimates. Appendix Section A.2 shows that the sampling distribution of Q , say $F_Q(x; q, n, r) = \Pr(Q \leq x)$, depends only on (q, n, r) and the standardized cdf $\Phi(z)$. Consequently, for specified values of (q, n, r) and $\Phi(z)$, one can obtain $F_Q(x; q, n, r)$ by simulation.

Using a standard method to compute confidence intervals on a single parameter (see for example, Casella and Berger, 2002, page 432), one obtains a two-sided $100(1-\alpha)\%$ confidence interval $[\underline{q}, \tilde{q}]$ for q by solving for \underline{q} and \tilde{q} the equations $F_Q(\hat{q}; \underline{q}, n, r) = 1 - \alpha/2$ and $F_Q(\hat{q}; \tilde{q}, n, r) = \alpha/2$. Similarly, a one-sided lower $100(1-\alpha)\%$ confidence bound \underline{q} for q is obtained from solving for \underline{q} in the equation $F_Q(\hat{q}; \underline{q}, n, r) = 1 - \alpha$. Appendix Section A.2 shows that this lower bound can be obtained numerically by solving either of the following equations for \underline{q}

$$\hat{q} = 1 - \Phi\left(-k_{(1-\alpha;1-\underline{q},n,r)}\right) \quad (4)$$

$$\frac{\hat{\mu} - y^\dagger}{\hat{\sigma}} = k_{(1-\alpha;1-\underline{q},n,r)}. \quad (5)$$

3.3 Computation of the quantiles of K

To compute the quantiles $k_{(\gamma;p,n,r)}$ needed in (2)–(5), we proceed as follows. From the definition of $k_{(\gamma;p,n,r)}$

$$\gamma = \Pr\left(\frac{\hat{\mu} - y_p}{\hat{\sigma}} \leq k_{(\gamma;p,n,r)}\right) = \Pr\left(\frac{\hat{\mu} - \mu - \Phi^{-1}(p)\sigma}{\hat{\sigma}} \leq k_{(\gamma;p,n,r)}\right).$$

Because $K = (\hat{\mu} - \mu - \Phi^{-1}(p)\sigma)/\hat{\sigma}$ is a pivotal quantity its distribution is invariant to the parameter values (μ, σ) of the distribution $\Phi[(\log(t) - \mu)/\sigma]$ that generated the data. In particular, K has the same distribution as $[\hat{\mu}^* - \Phi^{-1}(p)]/\hat{\sigma}^*$ where $(\hat{\mu}^*, \hat{\sigma}^*)$ denote the ML estimates of the parameters $(\mu = 0, \sigma = 1)$ from a Type II censored sample (r failures from n independent observations, and $1 \leq r \leq n$) from the standardized distribution $\Phi[\log(t)]$.

The distribution of K is obtained by simulating a large number B of ML estimates $(\hat{\mu}_j^*, \hat{\sigma}_j^*)$, $j = 1, \dots, B$. For the chosen value of p and each pair $(\hat{\mu}_j^*, \hat{\sigma}_j^*)$, compute $[\hat{\mu}_j^* - \Phi^{-1}(p)]/\hat{\sigma}_j^*$ and order these values from the smallest to the largest to obtain the order statistics $\{[\hat{\mu}^* - \Phi^{-1}(p)]/\hat{\sigma}^*\}_{[j]}$, $j = 1, \dots, B$. The plot of these order statistics (on the horizontal axis) versus j/B (on the vertical

axis) is the empirical cdf of K . The quantiles of K are obtained by interpolation in the empirical cdf. McCool (1970) gives tables of $k_{(\gamma;p,n,r)}$ for limited combinations of n and r from censored failure samples from the Weibull distribution.

4 Probability of Successful Demonstration

4.1 Quantile approach

To have a successful demonstration (SD) that $y_p > y^\dagger$, the demonstration test must yield the result $y_p > y^\dagger$, where y^\dagger is a specified value of y and y_p is the lower confidence bound defined in (3). Then for fixed μ and σ

$$\begin{aligned} \Pr(SD) &= \Pr\left(y_p > y^\dagger\right) = \Pr\left[\hat{\mu} - k_{(1-\alpha;p,n,r)}\hat{\sigma} > \mu + \Phi^{-1}(p_a)\sigma\right] \\ &= \Pr\left[\frac{\hat{\mu} - \mu - k_{(1-\alpha;p,n,r)}\hat{\sigma}}{\sigma} > \Phi^{-1}(p_a)\right] \end{aligned}$$

where the confidence interval factor $k_{(1-\alpha;p,n,r)}$ was obtained in Section 3.3 and $p_a = 1 - q$ is the actual proportion of the population failing before y^\dagger . $Z = Z(\hat{\mu}, \hat{\sigma}, p) = (\hat{\mu} - \mu - k_{(1-\alpha;p,n,r)}\hat{\sigma})/\sigma$ is pivotal (see Lawless, 1982, page 534) and it has the same distribution as $(\hat{\mu}^* - k_{(1-\alpha;p,n,r)}\hat{\sigma}^*)$. Thus

$$\Pr(SD) = \Pr[Z > \Phi^{-1}(p_a)] = \Pr[1 - \Phi(\hat{\mu}^* - k_{(1-\alpha;p,n,r)}\hat{\sigma}^*) < q].$$

Using simulation, $\Pr(SD)$ is obtained as follows. Compute a large number B of realizations of $1 - \Phi(\hat{\mu}^* - k_{(1-\alpha;p,n,r)}\hat{\sigma}^*)$ and order these to obtain the order statistics $\{1 - \Phi(\hat{\mu}^* - k_{(1-\alpha;p,n,r)}\hat{\sigma}^*)\}_{[j]}$, $j = 1, \dots, B$. Plot j/B (on the vertical axis) versus these order statistics (on the horizontal axis) for $j = 1, \dots, B$ to obtain graphical display of the empirical distribution corresponding to $\Pr(SD)$ as a function of q . The probability of successful demonstration for a given $q = 1 - p_a$ can be obtained by interpolation in the empirical distribution.

Usually, one is interested in a test plan, n and r , with a large value of $\Pr(SD)$. When, however, the actual reliability is equal to what one is attempting to demonstrate, $q = 1 - p_a = 1 - p$, (implying $p_a = p$), it is interesting to note that

$$\Pr(SD) = \Pr[Z > \Phi^{-1}(p_a)] = \Pr\left[\frac{\hat{\mu}^* - \Phi^{-1}(p_a)}{\hat{\sigma}^*} > k_{(1-\alpha;p,n,r)}\right] = \alpha$$

corresponding to the probability of a type 1 error for the test that $q = 1 - p_a = 1 - p$.

4.2 Reliability (survival probability) approach

The quantile statement that $y_p > y^\dagger$ is equivalent to the reliability statement that $q > q^\dagger$, where $q^\dagger = 1 - p$ and (as in Section 3.2) $q = 1 - \Phi[(y^\dagger - \mu)/\sigma]$. Note that the roles of p , q , and the \dagger change from the quantile statement to the reliability statement. In particular, the specified p in the quantile statement corresponds to $1 - q^\dagger$ in the reliability statement. Similarly, q in the reliability statement is the probability of exceeding y^\dagger in the quantile statement.

Appendix Section A.2 shows the equivalence of the quantile demonstration procedure (Section 4.1) in which the demonstration is successful if $y_p > y^\dagger$ and the similar reliability demonstration procedure in which the demonstration is successful if $q > q^\dagger$, where q is defined by (4). Appendix Section A.2 also shows the equivalence of corresponding methods to compute $\Pr(SD)$.

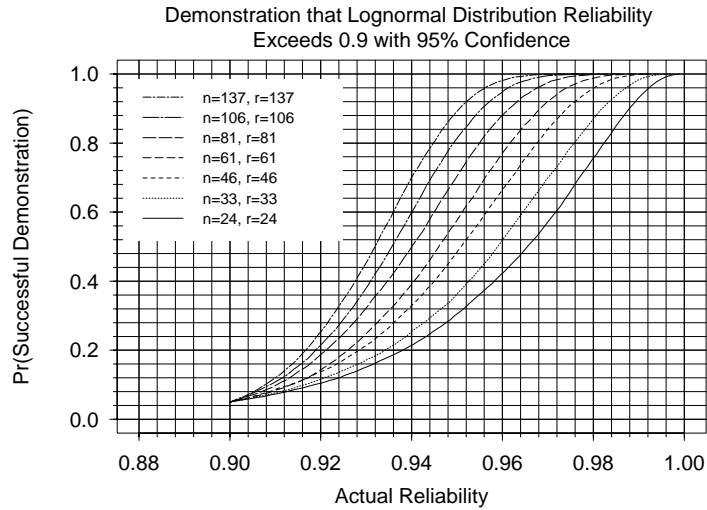


Figure 1: Probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$ as a function of actual reliability, q , with 95% confidence for the normal distribution and no censoring.

5 Examples, Comparisons, and Discussion

5.1 Normal (lognormal) distribution complete data case

Figures 9.1a–n in Hahn and Meeker (1991, page 154–167) were computed using a method based on the noncentral- t distribution outlined in Appendix Section A.3 of this paper for the case when the underlying distribution is normal (or lognormal) and there is no censoring (i.e., $r = n$). Figure 1, replicating Figure 9.1k of Hahn and Meeker (1991, page 164), was obtained by using the more general method described in Section 4.1. This figure contains curves illustrating the probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$, with 95% confidence, as a function of the actual reliability, q . Each curve represents a different sample size. Unless noted otherwise, the curves shown in this and subsequent figures in this paper are based on $B = 10,000$ simulations.

Figure 1 provides the required sample size to demonstrate, with $100(1 - \alpha)\% = 95\%$ confidence, that $q > 0.90$ for a desired probability of successful demonstration. If the actual reliability is $q = 0.96$, enter the horizontal scale at 0.96 and move up. Simultaneously enter the vertical scale at the desired probability of successful demonstration, say 0.95, and move right. After finding the point of intersection, interpolate for the needed value of n . In this case, the necessary sample size appears to be close to 106 units. Figure 2 (based on $B = 100,000$ simulations) provides a magnified view in the region of the actual probability (0.96) and the desired probability of successful demonstration (0.95) for sample sizes near 106. From this plot it is easy to see that the sample size must be 107 to achieve probability of successful demonstration of at least 0.95.

5.2 Smallest extreme value (Weibull) distribution complete data case

Figure 3, similar to Figure 1, gives probability of successful demonstration curves for the smallest extreme value (or Weibull) distribution. Comparing Figures 1 and 3 shows that the assumed distri-

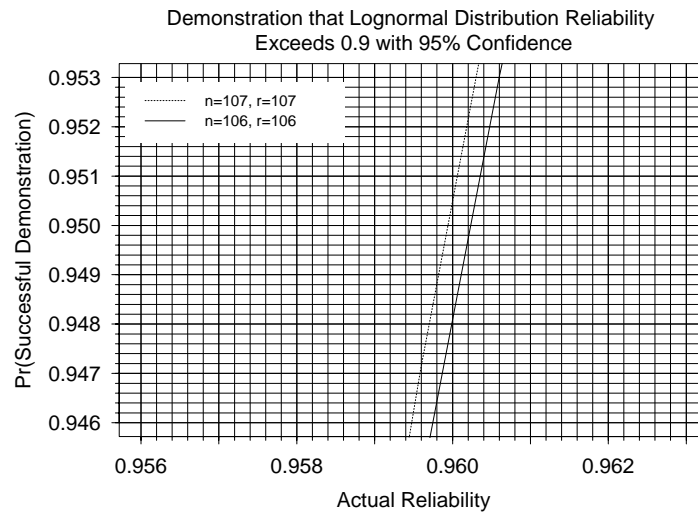


Figure 2: Magnified view of the probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$ as a function of actual reliability, q , with 95% confidence for the normal distribution and no censoring.

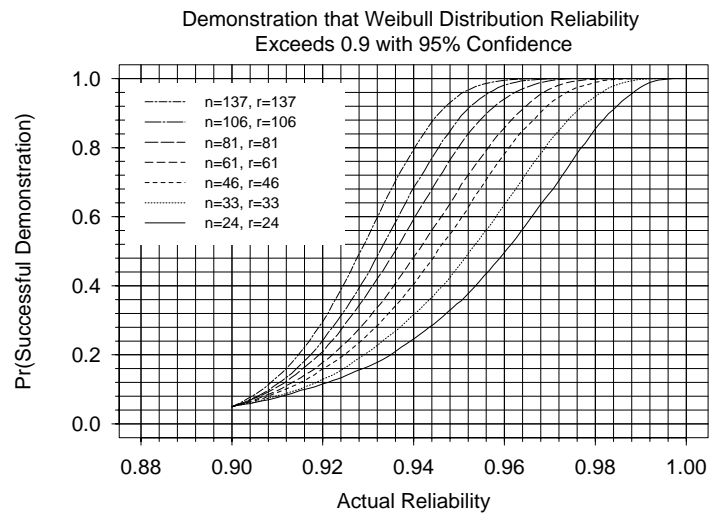


Figure 3: Probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$ as a function of actual reliability, q , with 95% confidence for the Weibull distribution and no censoring.

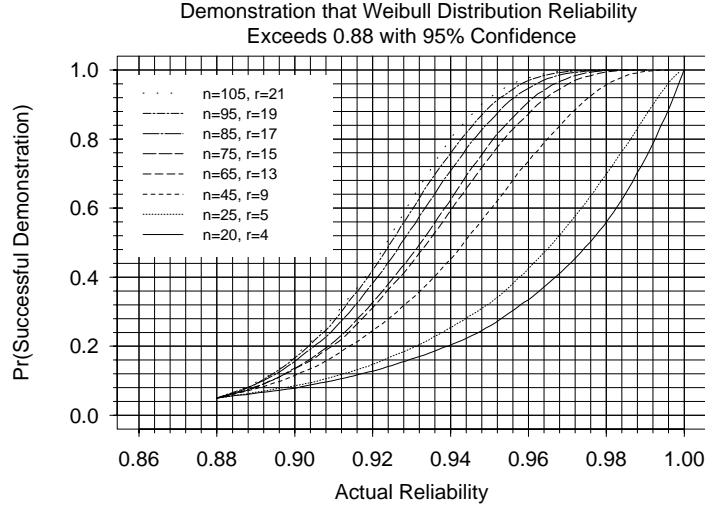


Figure 4: Probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.88$ as a function of actual reliability, q , with 90% confidence for the Weibull distribution and allowing 20% of the units to fail.

bution will affect the sample size needed to obtain a desired probability of successful demonstration. For the same reliability demonstration problem of Section 5.1, but assuming the smallest extreme value (Weibull) distribution instead of the normal (lognormal) distribution, a sample size of 81 units would yield the desired 0.95 probability of successful demonstration. For this example, the smallest extreme value (Weibull) distribution model results in smaller sample sizes than the normal (lognormal) distribution model. This is due to the heavier upper tail of the lognormal distribution, causing more variability in the ML estimate of q for a given value of q .

5.3 Censored life test example

A manufacturer is planning to conduct a life test of a newly developed insulation. The engineer in charge must plan a test to demonstrate that at least $q^\dagger = 88\%$ (i.e., $q \geq 0.88$) of the product will survive past $t^\dagger = 500$ hours, with 90% confidence. A probability of successful demonstration of 0.95 is desired. The test will be stopped after 20% of the units have failed. For purposes of test planning, the engineers will use a Weibull distribution to describe the failure time distribution and believe that the true reliability at $t^\dagger = 500$ is at least 0.95. Refer to Figure 4 to determine the necessary sample size to meet the requirements. Find the actual reliability $q = 0.95$ on the horizontal axis and follow the line up. Now find probability of successful demonstration 0.95 on the vertical axis and follow the line to the right. The point of intersection determines the needed sample size, between 85 and 95 units, with 17 and 19 failures, respectively.

Figure 5 displays a magnified view of the plot within the range of the actual probability and desired probability of successful demonstration for sample size/failure combinations that fall between the sample sizes indicated above. From this plot it is easy to see the sample size must be 93 units with 19 failures to achieve probability of successful demonstration 0.95. Figure 5 is based on $B = 100,000$

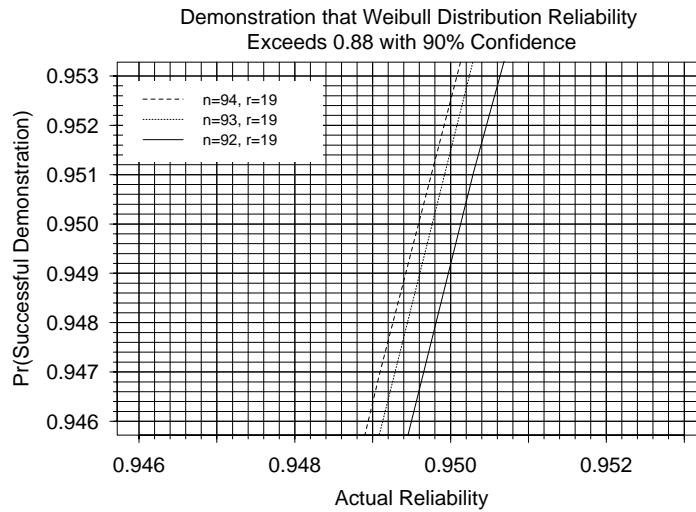


Figure 5: Magnified view of the probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.88$ as a function of actual reliability, q , with 90% confidence for the Weibull distribution and allowing 20% of the units to fail.

simulations. Note that since the test is run until 20% of the units fail, the number of failures must be rounded up to be an integer. For the example, 20% of 93 is 18.6, which must be rounded up to $r = 19$ failures. Thus, the required sample size is 93 units, waiting for 19 failures.

5.4 Effects of different q while holding the ratio r/n constant

The actual reliability, q , plays an important role in choosing the sample size and number of failures for a demonstration test plan. The example in Section 5.3 showed that if true reliability is 0.95, to demonstrate with $\Pr(SD) = 0.95$ that reliability exceeds 0.88 with 90% confidence, one requires an experiment allowing 19 of 93 units to fail. If true reliability is 0.96, however, it requires testing just 65 units and waiting for 13 of them to fail. Thus, the closer the actual reliability is to the target reliability, the larger the sample size needed to have the desired probability of successful demonstration.

5.5 Effects of varying r while holding n constant

Figure 6 illustrates the effect of varying the number of units failing, r , for a fixed sample size $n = 100$. The figure displays probabilities of successfully demonstrating that reliability exceeds 0.90, with 95% confidence, running the test until the proportion of units failing reaches 5%, 10%, 50% and 100% for Weibull data. At any given actual reliability, 0.96 for example, as the test length is increased (allow more units to fail) the probability of successful demonstration increases because more information is obtained. In this case, waiting for 5% of the units to fail results in $\Pr(SD) \approx 0.71$, 10% yields $\Pr(SD) \approx 0.87$, 50% gives $\Pr(SD) \approx 0.91$ and 100% has $\Pr(SD) \approx 0.98$. Note the diminishing returns as the proportion failing in the test increases. In particular, the probability of successful demonstration increases by smaller and smaller amounts as the test proportion failing exceeds the

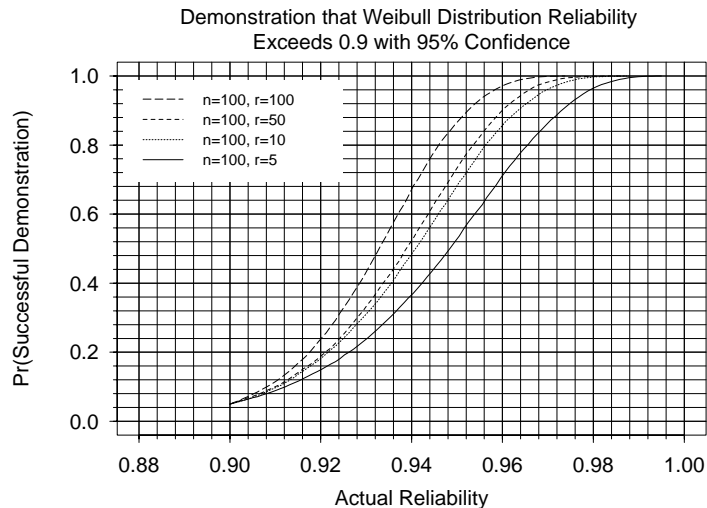


Figure 6: Probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$ as a function of actual reliability, q , with 95% confidence for the Weibull distribution and samples of size 100 with $r = 5, 10, 50, 100$.

proportion failing to be demonstrated ($1 - q$). This is related to the known fact that with censored data (failure or time censored), to estimate the p quantile y_p there is little gain in precision by extending the experiment far beyond $100p\%$ failures, see for example Meeker and Escobar (1998, page 244).

5.6 Weibull versus lognormal with censoring

Figure 7 is similar to Figure 6, but for the lognormal distribution. As in the comparison described in Section 5.2, the $\text{Pr}(SD)$ curves for the Weibull distribution are higher than the corresponding curves for the lognormal distribution. With more censoring, however, the differences diminish and for the smaller values of r/n , the curves are almost indistinguishable. This is not surprising, however. With heavy censoring, especially, estimates of small quantiles differ little from one distribution to the other.

6 Concluding Remarks and Possible Extensions

This paper gives methods for computing the exact probability of successful demonstration for life tests involving location-scale and log-location-scale distributions using complete or Type II censored data. The methods will be useful for planning life tests. The methods have been implemented in the S-PLUS computing language, using a collection of S-PLUS functions available from the authors.

Possible extensions of this work include:

1. Development of similar methods for planning life tests with Type I censored data.

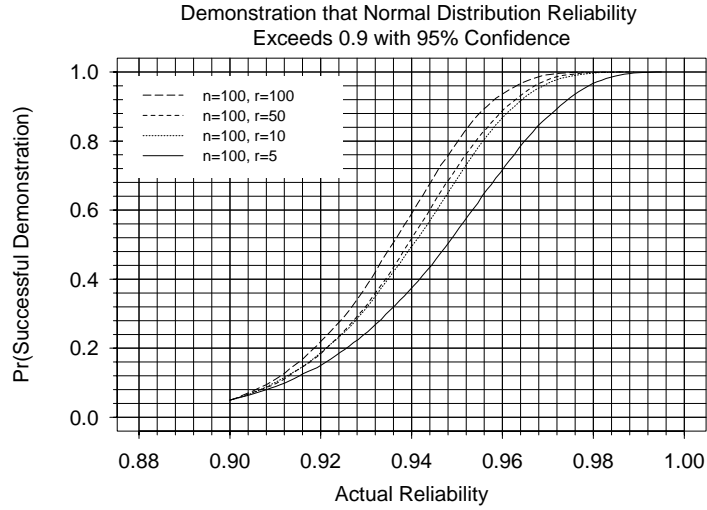


Figure 7: Probability of successfully demonstrating that reliability exceeds $q^\dagger = 0.90$ as a function of actual reliability, q , with 95% confidence for the lognormal distribution and samples of size 100 with $r = 5, 10, 50, 100$.

2. Development of similar methods for distributions that are not based on the location scale family of distributions.
3. Extension to determine sample size needs for demonstrations based on accelerated tests in which acceleration factors need to be estimated.
4. Extension to Bayesian models that will allow the use of prior information to reduce the needed sample size.

Some of these extensions will probably require the use of approximate methods because pivots for the construction of the confidence bounds used in the demonstrations may not be available.

Acknowledgments

We would like to thank Gerry Hahn, Wayne Nelson, and Yao Zhang for helpful comments on an earlier version of this paper.

A Appendix

A.1 Invariance to choice of a pivotal

The pivotal quantity used in (1) was chosen for convenience, but there are other choices for the pivotal quantity that lead into the same results obtained in Sections 3 and 4. Here we show that the results in Sections 3 and 4 are invariant to choice of the pivotal quantity within the class of monotone transformations of $(\hat{\mu} - y_p)/\hat{\sigma}$.

Result 1 Suppose that $h(x)$ is a known monotone function of x . Define the transformation

$$H = H(\hat{\mu}, \hat{\sigma}, y_p) = h\left(\frac{\hat{\mu} - y_p}{\hat{\sigma}}\right) = h(K). \quad (6)$$

Then

- (a) H is a pivotal quantity.
- (b) The $100(1 - \alpha)\%$ confidence intervals $\left[\underline{y}_p, \tilde{y}_p \right]$ for y_p derived using H and K are identical.
- (c) The $\Pr(SD) = \Pr\left(\underline{y}_p > y^\dagger\right)$ derived using H and K are identical.

An example of an alternative pivotal quantity H is as follows. Define $h(x) = x + \Phi^{-1}(p)$. Then

$$H = h\left(\frac{\hat{\mu} - y_p}{\hat{\sigma}}\right) = \frac{\hat{\mu} - y_p}{\hat{\sigma}} + \Phi^{-1}(p) = \frac{\hat{\mu} + \Phi^{-1}(p)\hat{\sigma} - y_p}{\hat{\sigma}} = \frac{\hat{y}_p - y_p}{\hat{\sigma}}.$$

To prove the result, notice that if H is a known function of a pivotal quantity, then its distribution does not depend of unknown parameters implying that H is also pivotal.

Let $g_{(\gamma;p,n,r)}$ be the γ quantile of the distribution of H and for the sake of the argument suppose that $h(x)$ is monotone increasing. Then

$$\gamma = \Pr\left[h\left(\frac{\hat{\mu} - y_p}{\hat{\sigma}}\right) \leq g_{(\gamma;p,n,r)}\right] = \Pr\left[\frac{\hat{\mu} - y_p}{\hat{\sigma}} \leq h^{-1}\left(g_{(\gamma;p,n,r)}\right)\right]. \quad (7)$$

Equation (7) implies that $k_{(\gamma;p,n,r)} = h^{-1}\left(g_{(\gamma;p,n,r)}\right)$. It follows that

$$\begin{aligned} 1 - \alpha &= \Pr\left[h^{-1}\left(g_{(\alpha/2;p,n,r)}\right) \leq \frac{\hat{\mu} - y_p}{\hat{\sigma}} \leq h^{-1}\left(g_{(1-\alpha/2;p,n,r)}\right)\right] \\ &= \Pr\left[\hat{\mu} - h^{-1}\left(g_{(1-\alpha/2;p,n,r)}\right)\hat{\sigma} \leq y_p \leq \hat{\mu} - h^{-1}\left(g_{(\alpha/2;p,n,r)}\right)\hat{\sigma}\right] \\ &= \Pr\left(\hat{\mu} - k_{(1-\alpha/2;p,n,r)}\hat{\sigma} \leq y_p \leq \hat{\mu} - k_{(\alpha/2;p,n,r)}\hat{\sigma}\right). \end{aligned}$$

This shows that a $100(1 - \alpha)\%$ confidence interval for y_p based on H is

$$\left[\hat{\mu} - k_{(1-\alpha/2;p,n,r)}\hat{\sigma}, \hat{\mu} - k_{(\alpha/2;p,n,r)}\hat{\sigma}\right].$$

This confidence interval is identical to the $100(1 - \alpha)\%$ confidence interval for y_p given in (2) and based on K .

Because y_p is invariant to the pivotal choice and $\Pr(SD) = \Pr(y_p > y^\dagger)$, it follows that $\Pr(SD)$ is also invariant to the pivotal choice.

A.2 Equivalence of the quantile and reliability demonstration methods

Using the definition of the random variable Q in Section 3.2, one gets

$$\begin{aligned} F_Q(x; q, n, r) &= \Pr(Q \leq x) = \Pr\left[1 - \Phi\left(\frac{y^\dagger - \hat{\mu}}{\hat{\sigma}}\right) \leq x\right] \\ &= \Pr\left[1 - \Phi\left(\frac{\mu - \hat{\mu} + \Phi^{-1}(1 - q)\sigma}{\hat{\sigma}}\right) \leq x\right] \\ &= \Pr\left[1 - \Phi\left(\frac{-\hat{\mu}^* + \Phi^{-1}(1 - q)}{\hat{\sigma}^*}\right) \leq x\right] \end{aligned} \quad (8)$$

where $\hat{\mu}^*, \hat{\sigma}^*$ are ML estimates of $\mu = 0, \sigma = 1$ in a failure censored sample (r failures and n observations) from $\Phi(z)$. Thus from (8) we get

$$\begin{aligned} F_Q(x; q, n, r) &= \Pr \left[\Phi^{-1}(1-x) \leq \frac{-\hat{\mu}^* + \Phi^{-1}(1-q)}{\hat{\sigma}^*} \right] \\ &= \Pr \{ \Phi [\hat{\mu}^* + \Phi^{-1}(1-x)\hat{\sigma}^*] \leq 1-q \} \end{aligned}$$

which shows that, for fixed x , $F_Q(x; q, n, r)$ is monotone decreasing in q .

Also from (8), one gets $F_Q(x; q, n, r) = \Pr[1 - \Phi(-K) \leq x]$, where $K = K(\hat{\mu}, \hat{\sigma}, y_{(1-q)})$ is defined as in (1). Then the quantiles $q_{(\gamma; q, n, r)}$ of $F_Q(x; q, n, r)$ and the quantiles $k_{(\gamma; 1-q, n, r)}$ of $K(\hat{\mu}, \hat{\sigma}, y_{(1-q)})$ are related by the equation

$$q_{(\gamma; q, n, r)} = 1 - \Phi(-k_{(\gamma; 1-q, n, r)}), \quad 0 < \gamma < 1, \quad 0 < q < 1. \quad (9)$$

Thoman et. al. (1970, page 370, equation (19)) give a special case of (9) for complete sample from a Weibull distribution.

Now we show that the $\Pr(SD)$ procedure based on quantiles is the same as the procedure based on reliabilities. Define the reliability lower bound \underline{q} by the solution of equation $F_Q(\hat{q}; \underline{q}, n, r) = 1 - \alpha$.

Then using (9), we get

$$\begin{aligned} 1 - p = q^\dagger < \underline{q} &\Leftrightarrow F_Q(\hat{q}; q^\dagger, n, r) > F_Q(\hat{q}; \underline{q}, n, r) = 1 - \alpha \\ &\Leftrightarrow \hat{q} > F_Q^{-1}(1 - \alpha; q^\dagger, n, r) \\ &\Leftrightarrow \hat{q} > 1 - \Phi(-k_{(1-\alpha; 1-q^\dagger, n, r)}) \\ &\Leftrightarrow \frac{y^\dagger - \hat{\mu}}{\hat{\sigma}} < -k_{(1-\alpha; p, n, r)} \\ &\Leftrightarrow y^\dagger < \underline{y}_p \end{aligned}$$

which shows the equivalence of the quantiles based and reliability based procedures for $\Pr(SD)$.

Finally, to compute \underline{q} , observe that $F_Q(\hat{q}; \underline{q}, n, r) = 1 - \alpha$ implies $\hat{q} = q_{(1-\alpha; \underline{q}, n, r)}$, then using (9), we get

$$\begin{aligned} \hat{q} &= 1 - \Phi \left(-k_{(1-\alpha; 1-\underline{q}, n, r)} \right) \\ \frac{\hat{\mu} - y^\dagger}{\hat{\sigma}} &= k_{(1-\alpha; 1-\underline{q}, n, r)}. \end{aligned}$$

Thus, \underline{q} can be obtained by solving numerically either of these last two equations.

A.3 Special case: complete normal data

When the data are complete (i.e., no censoring) from a normal random variable, there is an alternative approach, based on a noncentral- t pivotal quantity, to compute the confidence intervals for the quantiles and the probability of successful demonstration. In this appendix we show that the alternative approach yields the same confidence intervals as the general procedure given in Sections 3 and 4.

Assume that we have a random sample of size n from a $\text{NOR}(\mu, \sigma)$. The ML estimates of $(\hat{\mu}, \hat{\sigma})$ are $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = s\sqrt{(n-1)/n}$, where $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$. The ML estimate of the p quantile, $y_p = \mu + \Phi_{\text{nor}}^{-1}(p)\sigma$, of the distribution is $\hat{y}_p = \bar{x} + \Phi_{\text{nor}}^{-1}(p)\hat{\sigma}$.

A.3.1 Confidence interval for a quantile based on a noncentral- t pivotal quantity

A random variable $(U/\sqrt{V/\nu})$, where U and V are independent, $U \sim \text{NOR}(\delta, 1)$, and $V \sim \chi_\nu^2$, has a noncentral- t distribution with ν degrees of freedom and noncentrality parameter δ . Thus

$$T_{n-1}(\delta_p) = \frac{\frac{(\bar{x} - \mu)\sqrt{n}}{\sigma} - \Phi_{\text{nor}}^{-1}(p)\sqrt{n}}{\frac{s}{\sigma}} \quad (10)$$

is a pivotal quantity with a noncentral- t with $(n-1)$ degrees of freedom and noncentrality parameter $\delta_p = -\Phi_{\text{nor}}^{-1}(p)\sqrt{n}$. For the rest of this appendix, the degrees of freedom are $(n-1)$, and to simplify the notation, we use $T(\delta_p)$ to denote the nct random variable in (10). The cdf of $T(\delta_p)$ is $F_{\text{nct}}(t; n-1, \delta_p) = \Pr(T(\delta_p) \leq t)$ and the γ ($0 < \gamma < 1$) quantile of this cdf is denoted by $t'_{(\gamma)} = t'_{(\gamma, n-1, \delta_p)}$.

Starting with the definition of the quantiles of the distribution of the random variable in (10) it is easy to show that a two-sided $100(1-\alpha)\%$ confidence interval for the p quantile of Y , y_p is given by either of the following two equivalent expressions

$$[\bar{x} - g'(1-\alpha/2; n, p) s, \quad \bar{x} - g'(\alpha/2; n, p) s] \quad (11)$$

$$\left[\bar{x} - g'(1-\alpha/2; n, p) \sqrt{\frac{n}{n-1}} \hat{\sigma}, \quad \bar{x} - g'(\alpha/2; n, p) \sqrt{\frac{n}{n-1}} \hat{\sigma} \right] \quad (12)$$

where $g'(1-\alpha/2; n, p) = t'_{(1-\alpha/2)}/\sqrt{n}$ and $g'(\alpha/2; n, p) = t'_{(\alpha/2)}/\sqrt{n}$. The expression in (11) is given, for example, in Hahn and Meeker (1991, page 56).

A.3.2 Equivalence of the general pivotal and noncentral- t pivotal methods

For the normal (lognormal) complete data, the $100(1-\alpha)\%$ confidence interval from (2) is

$$\left[\underline{y}_p, \quad \tilde{y}_p \right] = \left[\hat{\mu} - k_{(1-\alpha/2; p, n, r)} \hat{\sigma}, \quad \hat{\mu} - k_{(\alpha/2; p, n, r)} \hat{\sigma} \right] \quad (13)$$

where as in Section 3.3, $k_{(\alpha/2; p, n, r)}$ and $k_{(1-\alpha/2; p, n, r)}$ are the quantiles of $[\hat{\mu}^* - \Phi^{-1}(p)]/\hat{\sigma}^*$ and $(\hat{\mu}^*, \hat{\sigma}^*)$ are the ML estimates of $(\mu^*, \sigma^*) = (0, 1)$ in complete random samples from a $\Phi_{\text{nor}}(z)$. It is known that, $\hat{\mu}^*$ and $\hat{\sigma}^*$ are independent, $\hat{\mu}^* \sim \text{NOR}(0, 1/\sqrt{n})$, and $n(\hat{\sigma}^*)^2 \sim \chi_{n-1}^2$. Thus

$$\frac{\sqrt{n-1} [\hat{\mu}^* - \Phi^{-1}(p)]}{\hat{\sigma}^*} = \frac{\sqrt{n}\hat{\mu}^* - \sqrt{n}\Phi^{-1}(p)}{\sqrt{n(\hat{\sigma}^*)^2/(n-1)}} \sim T_{n-1}(\delta_p). \quad (14)$$

Therefore, the quantiles for the left hand side of (14) are identical to the quantiles for the right hand side of (14), i.e., $\sqrt{n-1} k_{(\gamma; p, n, r)} = t'_{(\gamma; n-1, \delta_p)}$, $0 < \gamma < 1$, which implies $k_{(\alpha/2; p, n, r)} = t'_{(\alpha/2; n-1, \delta_p)}/\sqrt{n-1}$ and $k_{(1-\alpha/2; p, n, r)} = t'_{(1-\alpha/2; n-1, \delta_p)}/\sqrt{n-1}$. This shows that the confidence intervals in intervals (11) and (12) are equivalent to the confidence interval in (13).

An alternative proof of the equivalence of the intervals is as follows. First note that $T(\delta_p) = \sqrt{n-1} [(\hat{\mu} - y_p)/\hat{\sigma}]$ is a monotone transformation of the pivotal $(\hat{\mu} - y_p)/\hat{\sigma}$. Then **Result 1** implies that the confidence intervals (12) and (13) are identical.

A.3.3 Probability of successful demonstration based on the noncentral- t

A successful demonstration (SD) results if $y_p > y^\dagger$. Using the corresponding one-sided $100(1 - \alpha)\%$ lower bound obtained from (11) and $y^\dagger = \mu + \Phi^{-1}(p_a)\sigma$, the probability of successful demonstration is

$$\begin{aligned} \Pr(SD) &= \Pr\left(\tilde{y}_p > y^\dagger\right) = \Pr\left(\bar{x} - \frac{t'_{(1-\alpha)}s}{\sqrt{n}} > y^\dagger\right) \\ &= \Pr\left(\frac{\frac{(\bar{x} - \mu)\sqrt{n}}{\sigma} - \Phi_{\text{nor}}^{-1}(p_a)\sqrt{n}}{\frac{s}{\sigma}} > t'_{(1-\alpha, n-1, \delta_p)}\right) \\ &= \Pr\left[T(\delta_{p_a}) > t'_{(1-\alpha, n-1, \delta_p)}\right] = 1 - F_{\text{nct}}(t'_{(1-\alpha, n-1, \delta_p)}; n-1, \delta_{p_a}) \end{aligned}$$

where $\delta_{p_a} = -\Phi_{\text{nor}}^{-1}(p_a)\sqrt{n}$ and $\delta_p = -\Phi_{\text{nor}}^{-1}(p)\sqrt{n}$.

An interesting special case is when $p = p_a$, i.e., the reliability at t_e is exactly equal to the target reliability. In this situation

$$\Pr(SD) = 1 - F_{\text{nct}}(t'_{(1-\alpha, n-1, \delta_p)}; n-1, \delta_p) = 1 - (1 - \alpha) = \alpha$$

which shows, as expected, that the procedure to demonstrate reliability leads to the correct decision 100% of the times when the actual reliability is equal to that to be demonstrated.

References

- Billman, B., Antle, C., and Bain, L. J. (1972), Statistical inferences from censored Weibull samples, *Technometrics*, **14**, 831–840.
- Casella, G., and Berger, R. L. (2002), *Statistical Inference*. Second Edition. Duxbury: California.
- Escobar, L. A., and Meeker, W. Q. (1994), Fisher information matrix for the extreme value, normal and logistic distributions and censored data, *Applied Statistics*, **43**, 533–554.
- Faulkenberry, G. D., and Weeks, D. L. (1968), Sample size determination for tolerance limits, *Technometrics*, **10**, 343–348.
- Hahn, G. J., and Meeker, W. Q. (1991), *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons: New York.
- Lawless, J. F. (1982), *Statistical Models and Methods for Life Time Data*. John Wiley & Sons: New York.
- Lieberman, G. J. and Resnikoff, G. J. (1955), Sampling plans for inspection by variables, *Journal of the American Statistical Association*, **50**, 457–516.
- McCool, J. I. (1970), Inferences on Weibull percentiles and shape parameter for maximum likelihood estimates. *IEEE Transactions on Reliability*, **R19**, 2–9.
- Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*. John Wiley & Sons: New York.
- Nelson, W. (1982), *Applied Life Data Analysis*. John Wiley & Sons: New York.

- Owen, D. B. (1962), *Handbook of Statistical Tables*. Addison Wesley: Massachusetts.
- Odeh, R. E., and Owen, D. B. (1980), *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. Marcel Dekker: New York.
- Schneider, H. (1986), *Truncated and Censored Samples from Normal Populations*. Marcel Dekker: New York.
- Schneider, H. (1989), Failure-censored variables-sampling plans for lognormal and Weibull distributions, *Technometrics*, **31**, 199–206.
- Thoman, D. R., Bain, L. J., and Antle, C. E. (1969), Inferences on the parameters of the Weibull distribution, *Technometrics*, **11**, 445–460.
- Thoman, D. R., Bain, L. J., and Antle, C. E. (1970), Reliability and tolerance limits in the Weibull Distribution, *Technometrics*, **12**, 363–370.