Determining the Right Sample Size for Your Test: Theory and Application

Huairui Guo, Edward Pohl, & Athanasios Gerokostopoulos

Huairui Guo, Ph. D.
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ, 85710, USA
e-mail: Harry.guo@reliasoft.com

Edward Pohl, Ph. D.
Industrial Engineering
University of Arkansas
4207 Bell Engineering Center
Fayetteville, AR, 72701, USA
e-mail: epohl@uark.edu

Athanasios Gerokostopoulos
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ, 85710, USA
Email:Athanasisos.Gerokostopoulos@reliasoft.com

SUMMARY & PURPOSE

Determining the right sample size in a reliability test is very important. If the sample size is too small, not much information can be obtained from the test in order to draw meaningful conclusions; on the other hand, if it is too large, the information obtained through the tests will be beyond that needed, thus time and money are wasted. This tutorial explains several commonly used approaches for sample size determination.

Huairui Guo, Ph.D., CRE, CQE, CRP

Huairui Guo is the **Director of Theoretical Development** at ReliaSoft Corporation. He received his Ph.D. in Systems and Industrial Engineering from the University of Arizona. He has conducted consulting projects for over 20 companies from various industries, including renewable energy, oil and gas, automobile, medical devices and semi-conductors. As the leader of the theory team, he is deeply involved in the development of Weibull++, ALTA, DOE++, RGA, BlockSim, Lambda Predict and other products from ReliaSoft. Dr. Guo is a member of SRE, IIE and ASQ. He is a Certified Reliability Professional (CRP), a Certified Reliability Engineer (CRE) and a Certified Quality Engineer (CQE).

Edward Pohl, Ph.D.

Edward A. Pohl is an Associate Professor in the Department of Industrial Engineering at the University of Arkansas. Ed currently serves as the Director of the Operations Management Program and the Director of the ReliaSoft Risk, Reliability and Maintainability Research Alliance at the University of Arkansas. He has participated and led several risk and reliability related research efforts at the University of Arkansas. Before coming to Arkansas, Ed spent twenty years in the United States Air Force where he served in a variety of engineering, analysis, and academic positions during his career. Ed received his Ph.D. in Systems and Industrial Engineering from the University of Arizona. His primary research interests are in supply chain risk analysis, decision making, engineering optimization, quality and reliability. Ed is a Senior Member of IEEE, IIE, ASQ and SRE.

Athanasios Gerokostopoulos, CRE, CQE, CRP

Athanasios Gerokostopoulos is a Reliability Engineer at ReliaSoft Corporation. He is involved in the development of ReliaSoft's software products and the delivery of training seminars and consulting projects in the field of Reliability and Quality Engineering. His areas of interest include Reliability Program Plans, Design for Reliability, System Reliability, and Reliability Growth Analysis. Mr. Gerokostopoulos holds an M.S. degree in Reliability Engineering from the University of Arizona and an MBA from the Eller College of Management at the University of Arizona. He is a Certified Reliability Professional (CRP), an ASQ Certified Reliability Engineer and a Certified Quality Engineer.

Table of Contents

1. INTRODUCTION

In reliability testing, determining the right sample size is often times critical since the cost of tests is usually high and obtaining prototypes is often not easy. If the sample size used is too small, not much information can be obtained from the test, limiting one's ability to draw meaningful conclusions; on the other hand, if it is too large, information obtained through the tests may be beyond what's needed, thus incurring unnecessary costs.

Unfortunately, the majority of time, the reliability engineer does not have the luxury to request how many samples are needed but has to create a test plan based on the budget or resource constrains that are in place for the project. However, more often than not, when a reliability test design is solely based on resource constraints, the results are not very useful, often yielding a reliability estimate with a very large amount of uncertainty. Therefore, test designs always involve a trade-off between resource expenditure and confidence in the results. One needs to have a good understanding of what amount of risk is acceptable when calculating a reliability estimate in order to determine the necessary sample size. This tutorial will provide an overview of the methods that are available to help reliability engineers determine this required sample size.

In general, there are two methods for determining the sample size needed in a test. The first one is based on the theory of confidence intervals, which is referred to in this tutorial as *the estimation approach*, while the other is based on controlling Type I and Type II errors and is referred to as *the risk control approach*. The second one is also called *power and sample size* in the design of experiments (DOE) literature since power is defined as 1 minus the Type II error. The theory and applications of these two methods are given in the following sections.

2. DETERMINING SAMPLE SIZE BASED ON THE ESTIMATION APPROACH

2.1 An Introductory Example

In statistical analysis, as the sample size increases the confidence interval of an estimated statistic becomes narrower. To illustrate this point, assume we are interested in estimating the percentage, *x*, of black marbles in a pool filled with black and red marbles.

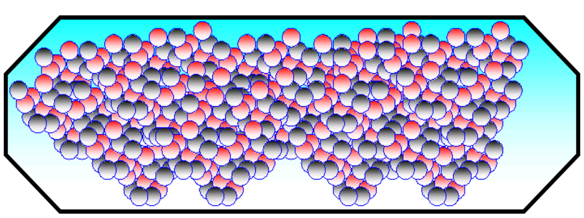


Figure 1. A Pool with Black and Red Marbles

Assume that a sample of 20 marbles was taken from the pool. 5 are black and 15 are red. The estimated percentage is 5/20 = 25%. Now assume that another sample of 20 marbles was taken and the estimated percentage is 35%. If we repeated the above process over and over again, we might find out that this estimate is usually between x_1% and x_2%, and we can assign a percentage to the number of times our estimate falls between these two limits. For example, we might notice that 90% of the time the percentage of the black marbles is between 5% and 35%. Then the confidence interval at a confidence level of 90% is 5% and 35%. In other words, we are 90% confident that the percentage of black marbles in the pool lies between 5% and 35%, or if we take a sample of 20 marbles and estimate the percentage of the black marbles, there is a 10% chance the estimated value is outside of those limits. That 10% is called the *risk level* or *significance level*. The relationship between the risk level α and the confidence level *CL* is:

$$\alpha = 1 - CL \tag{1}$$

Now, let's increase the sample size and get a sample of 200 marbles. From this sample, we found 40 of them are black and 160 are red. The estimated percentage is 40/200 = 20%. Take another sample of 200 and the estimated *x* is 15%. If we repeat this over and over again, we may observe that 90% of the time, the percentage of black marbles is between 15% and 25% which is narrower than the intervals we obtained when taking a sample size of 20 marbles.

Therefore, the larger the sample size, the narrower the confidence interval will become. This is the basic idea for determining sample size based on the requirement placed on the confidence interval of a given statistic. This requirement is usually given in terms of the bound ratio as given below:

$$B = \frac{x_{U,CL}}{x_{L,CL}} \tag{2}$$

or in terms of the bound width as:

$$B = x_{U,CL} - x_{L,CL} \tag{3}$$

where $x_{U,CL}$ and $x_{L,CL}$ are the upper bound and lower bound of a statistic at confidence level *CL*. If the variable of interest *x* follows a normal distribution, the bound width of Eq.(3) is used. If the variable is assumed to follow a lognormal distribution, then the bound ratio of Eq.(2) is usually used.

For instance, assume that the required width of the bounds is 0.10 for the above marble example. A sample size of 20 will not be adequate since the bound width was found to be 0.35-0.05=0.30 which is too wide. Therefore, we have to increase the sample size until the bound width is less than or equal to 0.10.

The above procedure can be done using simulation. If we assume the true value of the percentage of black marbles is μ_x =0.20, we can generate 20 observations based on this value. To do this, we first generate a random number between 0 and 1. If this random number is less than or equal to 0.2, then the observation is a black marble. Otherwise, it is red. Doing this 20 times we can generate 20 observations and estimate the percentage of black marbles *x*. Repeating this for many simulation runs, the bounds of the percentage at a given confidence level can be obtained. These bounds are sometimes called simulation bounds.

If we assume that *x* follows a certain distribution, the bounds can even be calculated analytically. For example, assume that the percentage x is normally distributed. Then its upper and lower bounds (two-sided) are:

$$x_{U,CL} = \mu_x + k\sigma_x = \mu_x + k_{\alpha/2} \times \sqrt{\frac{\mu_x(1-\mu_x)}{n}};$$
$$x_{L,CL} = \mu_x - k\sigma_x = \mu_x - k_{\alpha/2} \times \sqrt{\frac{\mu_x(1-\mu_x)}{n}} \quad (4)$$

where $k_{\alpha/2}$ is the (1+CL)/2 percentile of the standard normal distribution and n is the sample size. The definition and relationship between α and CL is given in Eq.(1). The bound width is:

$$B = 2k_{\alpha/2} \times \sqrt{\frac{\mu_x(1-\mu_x)}{n}} \quad (5)$$

Given a desired confidence level of 90%, $k_{\alpha/2}$ is 1.645. If the required bound width is 8%, then from Eq.(5), the necessary sample size can be solved. In this case it is:

$$n = \frac{\mu_x(1-\mu_x)}{B^2} \times 4k_{\alpha/2}^2 = \frac{0.2(1-0.2)}{0.08^2} \times 4(1.645)^2 = 270.55 \quad (6)$$

Therefore, 271 samples are needed in order to have a bound width of 8% for the estimated percentage of black marbles.

2.2 Determining Sample Size for Life Testing

When it comes to reliability testing, the logic for determining the necessary sample size is the same as with the marble example. The sample size can be determined based on the requirement of the confidence interval of reliability metrics such as, reliability at a given time, B10 life (time when reliability is 90%), or the mean life. Usually, reliability metrics are assumed to be log-normally distributed since they must be positive numbers. Therefore the requirement for the confidence interval is determined using the *bound ratio*. The sample size can either be determined using simulation or analytically.

2.2.1 Analytical Solution

Using the Weibull distribution as an example, the reliability function is:

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}} \quad (7)$$

Assuming that the estimated reliability from a data set is log-normally distributed, the bounds for the reliability at t are:

$$\ln(R_{U,CL}) = \ln(\hat{R}) + k_{\alpha/2}\sqrt{Var\left(\ln(\hat{R})\right)}$$
$$\ln(R_{L,CL}) = \ln(\hat{R}) - k_{\alpha/2}\sqrt{Var\left(\ln(\hat{R})\right)} \quad (8)$$

The bound ratio is:

$$B = R_{U,CL} / R_{L,CL} = 2k_{\alpha/2}\sqrt{Var\left(\ln(\hat{R})\right)} \quad (9)$$

where $\hat{R}$ is the estimated reliability and $Var(\ln(\hat{R}))$ is the variance of the logarithm transformation of the estimated reliability. $Var(\ln(\hat{R}))$ is a function of sample size and can be obtained from the Fisher information matrix [1, 2]. If the bound ratio B is given, we can use Eq.(9) to calculate $Var(\ln(\hat{R}))$ and get the necessary sample size.

2.2.2 Simulation

The process of using simulation to design a reliability test is similar to the one described in the introductory example. First, the required input to the simulation is an assumed failure distribution and its associated parameters. The next step is to generate a uniform random number between 0 and 1. Using the reliability equation of the chosen failure distribution, we can substitute this generated number for the reliability and calculate a time to failure. This can be repeated multiple times in order to obtain a desired sample size of times to failure. Then, this sample is used to re-estimate the parameters of the chosen distribution. Repeating this whole process for multiple simulation runs, we can obtain multiple sets of estimated parameters. Finally, by ranking those parameters in an ascending order, we can obtain the upper and lower bounds at a given confidence level of the parameters and any metrics of interest. More information on the simulation process is provided in [2].

An example that illustrates how to determine the required sample size for a reliability test using simulation is given below.

Example 1: From historical information, an engineer knows a component's life follows a Weibull distribution with β = 2.3 and η = 1,000. This information is used as the input for the simulation. The engineer wants to determine the required sample size based on the following estimation requirement: based on the failure data observed in the test, the expected bound ratio of the estimated reliability at time of 400 should be less than 1.2.

Solution for Example 1: In order to perform simulation the SimuMatic tool in Weibull++ was used. The bound ratio for different sample sizes was obtained through simulation and is given in the Table below.

Table 1. Bound Ratio (90% 2-sided bound) for Different Sample Sizes

Sample Size	Upper Bound	Lower Bound	Bound Ratio
5	0.9981	0.7058	1.4143
10	0.9850	0.7521	1.3096
15	0.9723	0.7718	1.2599
20	0.9628	0.7932	1.2139
25	0.9570	0.7984	1.1985
30	0.9464	0.8052	1.1754
35	0.9433	0.8158	1.1563
40	0.9415	0.8261	1.1397

From the above table, we see that the sample size should be at least 25 in order to meet the bound ratio requirement. Clearly, with samples above 25, we will be even more confident on the result because the bound ratio becomes smaller. The effect of sample size can be seen in Figure 2.

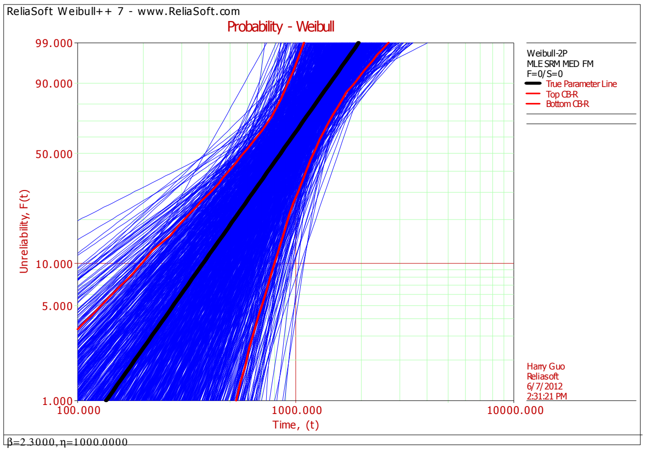


Figure 2(a). Simulation Result for Sample Size of 5

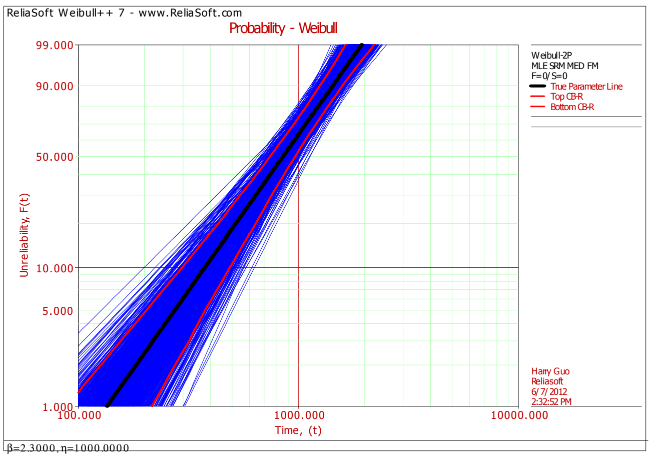


Figure 2(b). Simulation Result for Sample Size of 40

From Figure 2 we see that the confidence interval for the estimated reliability is much wider at a sample size of 5 than it is for a sample size of 40.

2.3 Determining Sample Size for Accelerated Life Testing

The estimation approach is also widely used for designing accelerated life testing (ALT). For the case of ALT, in addition to determining the total number of samples, we also need to determine the appropriate stress levels and how the samples should be allocated at each stress level. Similar to the test design for life test data, a test plan can be created using either an analytical solution or simulation. However for the case of ALT, simulation is rather cumbersome since besides the necessary assumptions regarding the parameters of the life-stress relationship and the failure distribution one needs to determine the optimal stress levels and the allocated units at each level. For that reason the analytical solution is more widely used in designing an ALT. In this section we'll provide an overview of the available analytical test plans, illustrate their use through an example and use simulation to validate the results of the test plan.

Many optimal testing plan methods have been proposed based on the estimation approach. "Optimal" refers to the fact that if the sample size is given, the optimal test plans will result in the minimal variance for an estimated reliability metric such as the B(X) life under different constraints.

For single stress, the most popular test plans are [6]:

- *The 2 Level Statistically Optimum Plan*. The plan will recommend two stress levels. One will be the maximum allowable stress and the second will be computed so that the variance of the B(X) life is minimized.
- *The 3 Level Best Standard Plan*. The plan will recommend three equally spaced stress levels with equal allocations. One stress will be the maximum allowable stress and the other two stresses will be computed so that the variance of the B(X) life is minimized.
- *The 3 Level Best Compromise Plan*. The plan will recommend three equally spaced stress levels using the same approach as the 3 Level Best Standard Plan. The difference is that the proportion of the units to be allocated to the middle stress level is defined by the user.
- *The 3 Level Best Equal Expected Number Failing Plan*. The plan will recommend three equally spaced stress levels using the same approach as the 3 Level Best Standard Plan. The difference is that the proportion of units allocated to each stress level is calculated such that the number of units expected to fail at each level is equal.

Let's use an example to show how to design an optimal accelerated life test.

Example 2: A reliability engineer is asked to design an accelerated life test for a new design of an electronic component. Initial HALT tests have indicated that temperature is the major stress of concern. The temperature at use condition is 300K, while the design limit was 380K. Looking at historical data of the previous design, he finds that after 2 years of operation (6,000 hours of usage) approximately 1% of the units had failed. He also knows the beta parameter of the Weibull distribution was 3. Given that the failure mode of the new design is expected to be similar, he feels that this is a good approximation for beta. Finally, previous tests have indicated that an acceleration factor of 30 can be achieved at temperature levels close to the design limit. He has 2 months or 1,440 hours and 2 chambers for the test. He wants to determine

- the appropriate temperature that should be set at each test chamber, and
- the number of units that should be allocated to each chamber.

Using the failure data obtained from the test, the failure distribution for the component can be estimated. Assume it is required to have a bound ratio of 2 with a confidence level of 80% for the estimated B10 life at the usage temperature.

Solution for Example 2: An optimal test plan can be found using either the simulation method or the analytical method. First, a test plan based on the analytical method will be generated and the results will be validated using simulation. The steps to generate a test plan are:

Step 1: Determine the eta parameter of the Weibull distribution at the normal use temperature. From Eq.(7), we know

$$R(t)=e^{-\left(\frac{t}{\eta}\right)^{\beta}} \Rightarrow 0.99=e^{-\left(\frac{6000}{\eta_1}\right)^{3}} \Rightarrow \eta_1=27802.96 \qquad (10)$$

Since the acceleration factor is believed to be 30, the eta value at the design limit of the temperate can be calculated as follows: 27802.96/30 = 926.765.

Step 2: Calculate the expected probability of failure at the usage stress level and at the design limit by the end of the test.

$$P_1=1-e^{-\left(\frac{1440}{27802.96}\right)^{3}}=0.00014\,;$$

$$P_2=1-e^{-\left(\frac{1440}{926.765}\right)^{3}}=0.97651$$

Step 3: Using a software package such as ALTA PRO from *ReliaSoft*, one can calculate the optimal design. The following screenshots are from ALTA PRO.

Test Plan	
Number of Simultaneous Stresses	1
Test Plan Type	2 Level Statistically Optimum Plan
BX% Life Estimate Sought	10
Available Test Time	1440
Unit Allocation	Show Allocations as %
Lifetime Distribution	Weibull
Beta	3
Stress 1	
Life-Stress Relationship	Arrhenius
Use Stress Value	300
Maximum Stress Value	380
Probabilities of Failure (%) at Time=1440	
P(Time, Use Stress)	0.014
P(Time, Maximum Stress)	97.651

Figure 3. Inputs for the Optimal Test Plan

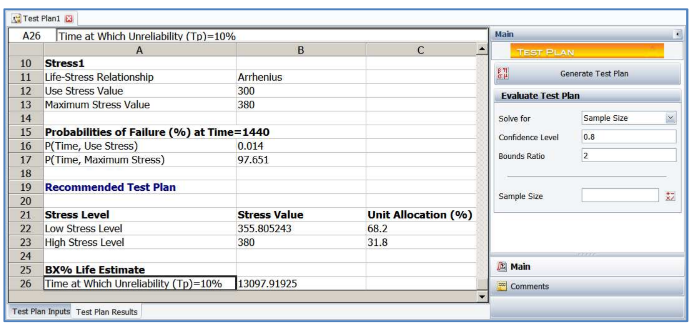


Figure 4. Outputs for the Optimal Test Plan

The results show that 68.2% of the test units should be tested at a temperature setting of 355.8 and 31.8% of the units should be tested at a temperature of 380. This test plan will give the minimal variance for the estimated B10 life at the usage stress level.

Step 4: Get the sample size based on the required bounds ratio. Based on the optimal test plan obtained in Step 3, we can use the design evaluation tool in ALTA PRO to get the necessary sample size.

Therefore, as shown above, 222 units should be tested in order to attain a bound ratio of 2. Therefore, $222\times 68.2\%\approx 151$ units should be tested at 355.8K and the remaining units, $222\times 31.8\%\approx 71$, should be tested at 380K.

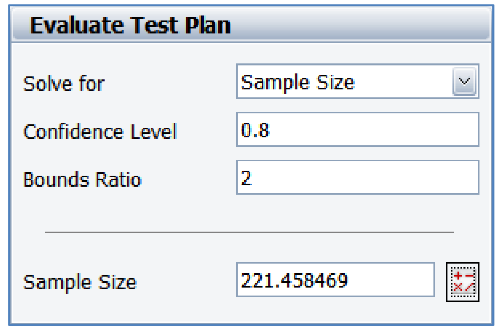


Figure 5. Obtain the Sample Size Based on Bound Ratio

The analytical solution that was used above can be validated using simulation. Some of the required inputs to run the simulation are the parameters of the life-stress relationship and the failure distribution. In this example, the Arrhenius model was used as the life-stress relationship, which is:

$$\eta = C\cdot e^{B/S} \qquad (11)$$

Since η_1 is calculated as 27802.96 at 300K and η_2 as 926.765 at 380K in this example, we can use this information in Eq.(11) to get the parameters of the Arrhenius model. They are B = 4846.7069 and C=0.002677. Finally, note that the beta parameter of the Weibull distribution was assumed to be 3.

Using the above calculated parameters and the optimal test plan that was obtained from the analytical solution (stress levels and units allocated at each level) as inputs to the simulation, the simulated upper bound for the B10 life is found to be 19262.67 and the lower bound 9532.1649. The simulated bounds ratio is 19262.67/9532.1649 =2.02 which is very close to the requirement. The plot of the simulation results is given below.

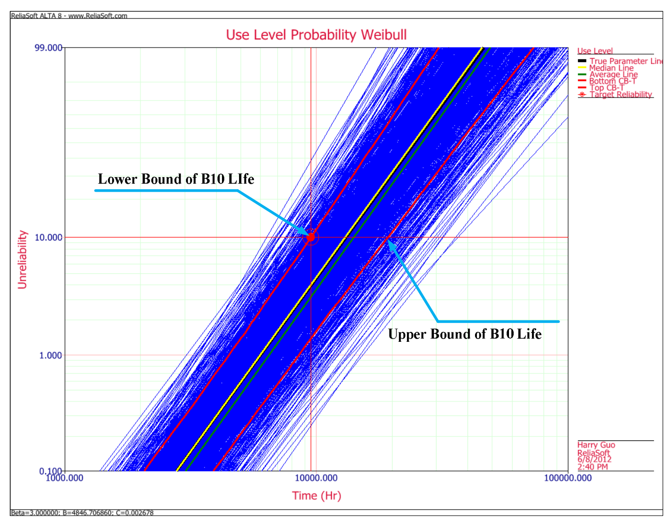


Figure 6. Use Simulation to Validate the Analytical Solution

It should be noted that it is rather cumbersome to use simulation in order to obtain an optimal test plan for accelerated life testing since both the optimal stress levels and the optimal sample size at each stress level need to be determined by the user before running the simulation. Therefore, the analytical method is generally preferred in ALT.

However, for regular life tests, since the sample size is the only variable that needs to be determined, we can use either the simulation method or the analytical method. Many applications of this estimation approach can be found in the literature. For example, it was used to design a test for repairable systems [3] and used to design an effective test for detecting a specific difference between the lives of two products [4]. The math behind the analytical solution of an ALT test plan is beyond the scope of this tutorial. Readers are referred to Reference [1].

The estimation approach discussed so far requires assuming a failure distribution and estimating the model parameters. In order to estimate parameters, failures are needed. If there are no failures in a test (when the test duration is short), the estimation approach cannot be used. However, we can still get some information regarding reliability from a zero failure test. For example, one would expect that testing 1,000 samples without failure would result in a higher demonstrated reliability than testing 10 samples without failure. Therefore the question that needs to be answered is how many samples are needed in order to demonstrate a required reliability in a zero failure test scenario. This question can be answered by using the method presented in Section 3.

3. DETERMINING SAMPLE SIZE BASED ON THE RISK CONTROL APPROACH

The risk control approach is usually used in the design of reliability demonstration tests. In reliability demonstration tests, there are two types of risks. The first, Type I risk, is the probability that although the product meets the reliability requirement it does not pass the demonstration test. This is also called producer's risk, α error, or false negative. Second, there is Type II risk, which is the probability that although the product does not meet the reliability requirement it passes the demonstration test. This is also called the consumer's risk, β error, or false positive. 1-β is usually called the power of the test. The following table summarizes the above statements.

Table 2. Summary of Type I and Type II Errors

	When H0 is true	When H1 is true
Do not reject H0	correct decision (probability = 1- α)	Type II error (probability = β)
Reject H0	Type I error (probability = α)	correct decision (power = 1 - β)

Note that H0 is the null hypothesis which is that the product meets the reliability requirement and H1 is the alternative hypothesis which is that the product does not meet the reliability requirement. With an increase of sample size, both Type I and Type II risks will decrease. A sample size can be determined based on controlling Type I risk (It also determines the confidence interval for the null hypothesis), Type II risk, or both.

3.1 An Introductory Example

<u>*Example 3*:</u> The voltage reading of a small device made from a manufacturing line is a critical reliability parameter. If it is higher than 12 volts, the device is defined as having failed. It was found that the readings follow a normal distribution with a mean of 10 and a standard deviation of 1. Based on this information, the reliability of this device is:

$$R = \Pr(x < 12) = 0.977$$

Due to the degradation of the manufacturing line, the mean value of the voltage reading will increase with time, leading to a decrease in the device reliability. Therefore, samples should be taken every day to monitor the manufactured devices. If one of the samples fails, the manufacturing line should be recalibrated. How many samples should be taken in order to have a 90% chance of detecting the problem once the reliability is below 0.95?

<u>*Solution for Example 3*:</u> In this example, the requirement of 90% for the detection power is given. The sample size should be determined based on this requirement. Assume n samples are taken for testing. If the reliability has decreased to 0.95, the probability of detecting this decrease is:

$$P = 1 - R_L^n = 1 - 0.95^n \tag{10}$$

From the stated risk requirement, we know that P is 90%. Therefore, from Eq.(10), solving for n we find that the required sample size 45. In other words if 45 samples are taken, the Type II error will be 0.1. For this test plan, the Type I risk also can be evaluated. It is:

$$\alpha = 1 - R^n = 1 - 0.977^n = 0.649 \tag{11}$$

As it can be seen, the Type I error is too large. What that means is that a lot of false alarms will be generated with this testing plan. Therefore, an appropriate sampling plan (sample size and allowed number of failures) should be defined based on the requirement for both the Type I and the Type II error.

3.2 Non-Parametric Binomial Reliability Demonstration Test

Similar to what was shown in example 3, in reliability demonstration tests one wants to determine the sample size needed in order to demonstrate a specified reliability at a given confidence level. In cases where the available test time is equal to the demonstration time, the following non-parametric binomial equation is widely used in practice:

$$1 - CL = \sum_{i=0}^{f} \binom{n}{i} (1-R)^i R^{n-i} \tag{12}$$

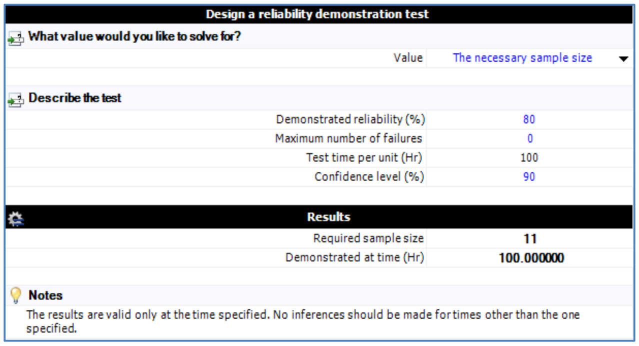
where CL is the confidence level, f is the number of failures, n is the sample size, and R is the demonstrated reliability. Given any three of them, the remaining one can be solved for. 1-CL is the probability of passing the demonstration test. Depending on the value of R used in Eq. (12), the probability of passing the test would be either the Type II error or the value of 1-Type I error, as illustrated in Eq.(10) and (11). Example 4 illustrates how this formula can be used.

<u>*Example 4*:</u> A reliability engineer wants to design a zero failure demonstration test in order to demonstrate a required reliability of 80% at a 90% confidence level. What is the required sample size?

Solution for Example 4: By substituting f = 0 (since it a zero failure test) Eq.(12) becomes:

$$1 - CL = R^n$$

So now the required sample size can be easily solved for any required reliability and confidence level. The result of this test design was obtained using Weibull++ and is:

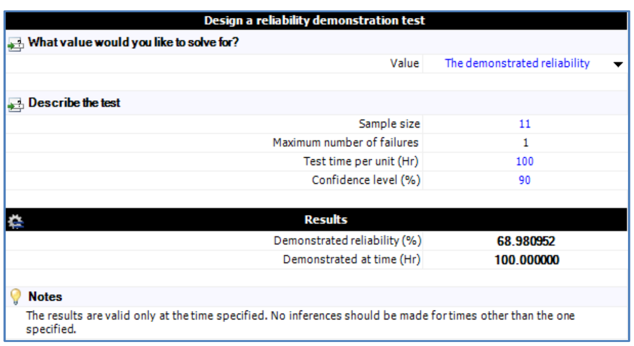


Design a reliability demonstration test	
What value would you like to solve for?	
Value	The necessary sample size
Describe the test	
Demonstrated reliability (%)	80
Maximum number of failures	0
Test time per unit (Hr)	100
Confidence level (%)	90
Results	
Required sample size	11
Demonstrated at time (Hr)	100.000000
Notes	
The results are valid only at the time specified. No inferences should be made for times other than the one specified.	

Figure 7. Sample Size Using the Non-Parametric Binomial

The above result shows that 11 samples are needed. Note that the "Test time per unit" input is the same as the "Demonstrated at time" value in this example. If those 11 samples are run for the required demonstration time and no failures are observed, then a reliability of 80% with a 90% confidence level has been demonstrated. If the reliability of the system is less than or equal to 80%, the chance of passing this test is less than or equal to 1-CL = 0.1, which is the Type II error. Therefore, Eq.(12) determines the sample size by controlling for the Type II error.

If 11 samples are used and one failure is observed by the end of the test, then the demonstrated reliability will be less than required. The demonstrated reliability is 68.98% as shown below.



Design a reliability demonstration test	
What value would you like to solve for?	
Value	The demonstrated reliability
Describe the test	
Sample size	11
Maximum number of failures	1
Test time per unit (Hr)	100
Confidence level (%)	90
Results	
Demonstrated reliability (%)	68.980952
Demonstrated at time (Hr)	100.000000
Notes	
The results are valid only at the time specified. No inferences should be made for times other than the one specified.	

Figure 8. Demonstrated Reliability with One Failure

The following figure shows how the demonstrated reliability changes with different numbers of failures and for different sample sizes.

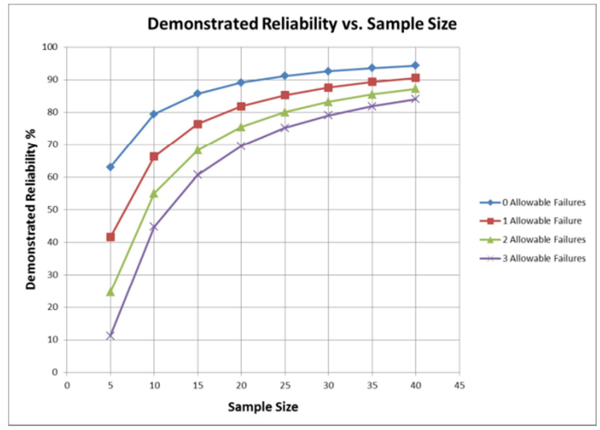


Figure 9. Demonstrated Reliability vs. Sample Size

3.3 Parametric Binomial Reliability Demonstration Test

As can be seen in Eq.(12), for the non-parametric binomial equation, a specific "time" parameter is not included. Usually this equation is used for one-shot systems, where time is not a factor or for cases where the available test time is equal to the required demonstration test time. However, in most cases the available test time is lower than the demonstration time. In those cases some assumptions need to be made regarding the failure distribution of the product in order to introduce time as part of the test plan. In the following example we use the case of the Weibull distribution in order to illustrate the parametric binomial approach.

Example 5: Let's assume that the failure times of a component follow a Weibull distribution with a beta of 2. We need to determine the required number of samples for the test in order to demonstrate a reliability of 80% at 2,000 hours at a confidence level of 90%. The available test duration is 1,500 hours and the maximum number of allowed failures is 1.

Solution for Example 5:

Step 1: Determine the Weibull scale parameter η using Eq.(7).

$$\eta = \frac{t}{\left(-\ln\left(R_t\right)\right)^{1/\beta}} = \frac{2000}{\left(-\ln\left(0.8\right)\right)^{0.5}} = 4233.87$$

Step 2: Calculate the reliability at the available test duration of T = 1,500 using Eq.(7).

$$R(T) = e^{-\left(\frac{T}{\eta}\right)^{\beta}} = e^{-\left(\frac{1500}{4233.87}\right)^{2}} = 0.882$$

Step 3: Use the calculated $R(T)$ determine the required sample size using the equations below:

$$1 - CL = \sum_{i=0}^{f} \binom{n}{i} (1-R)^i R^{n-i} \Rightarrow 0.1 = \sum_{i=0}^{1} \binom{n}{i} (1-0.882)^i \, 0.882^{n-i}$$

The result from Weibull++ is:

The result in the figure above shows that at least 32 samples are needed in order to demonstrate the required reliability.

The following plot shows how the sample size changes as a function of the available test time and number of failures allowed in the test for example 5.

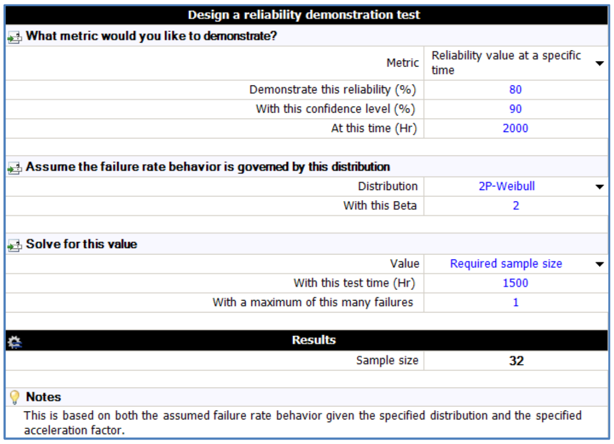


Figure 10. Sample Size Using the Parametric Binomial Method

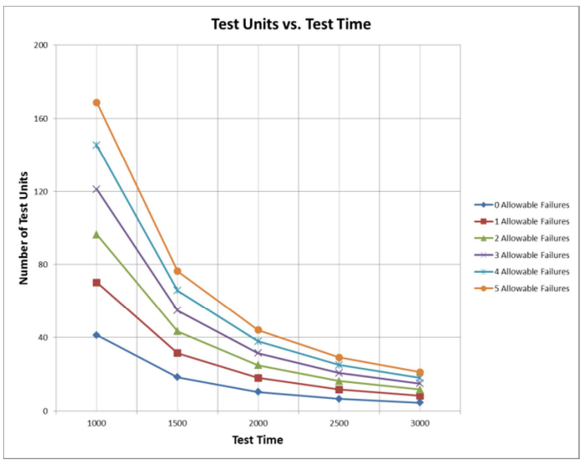


Figure 11. Sample Size and Test Time Relation for Parametric Binomial Method

Users can choose the right test plan using the above figure.

3.4 Exponential Chi-Squared Demonstration Test

Since the exponential distribution is frequently used and has a unique feature of being "memory less" (i.e. constant failure rate assumption), the following Chi-squared method is used to design test for systems following the exponential distribution.

$$\chi^2_{1-CL,2f+2} = \frac{2n \cdot T}{MTTF} \tag{13}$$

where:

$\chi^2_{1-CL,2f+2}$:	the 1-*CL* percentile of a Chi-squared distribution with 2*f*+2 degrees of freedom
f:	number of failures
n:	test samples
T:	test duration
MTTF:	mean time to failure

Eq.(13) includes 5 variables: *CL*, *f*, *n*, *T* and *MTTF*. Knowing any four of them, one can solve for the other. For the case of the exponential distribution, demonstrating the reliability is the same as demonstrating the MTTF since:

$$MTTF = -t / \ln(R) \tag{14}$$

<u>*Example 6*:</u> We desire to design a test to demonstrate a reliability of 85% at time t = 500 hours with a 90% confidence level while only allowing two failures during the test. The test duration is set at 300 hours. How many samples are needed?

<u>*Solution for Example 6*:</u> First, we need to convert the required reliability to *MTTF* using equation (14). It is:

$$MTTF = -t / \ln(R) = -500 / \ln(0.85) = 3076.57$$

Using Eqn.(13), the required sample size is:

$$n = MTTF \cdot \chi^2_{1-CL,2f+2} / (2T) = 3076.76 \times 10.6446 / 600 = 54.58$$

Therefore, we need at least 55 samples in the test. The total accumulated test time is $T_a = n \times T = 54.58 \times 300 = 16374.46$. This result is shown in the figure below.

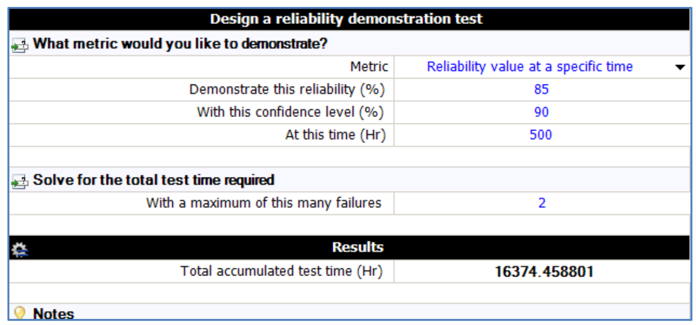


Figure 12. Result for the Chi-Squared Test Design

As we know when beta equals 1, the Weibull distribution becomes the exponential distribution. All the previously discussed binomial equations also work for the exponential distribution. So, what are the differences between the Chi-squared method and the binomial method? When there are 0 failures in the test, they give exactly the same results. This is because

$$\chi^2_{1-CL,2} = -2\ln(1-CL)$$

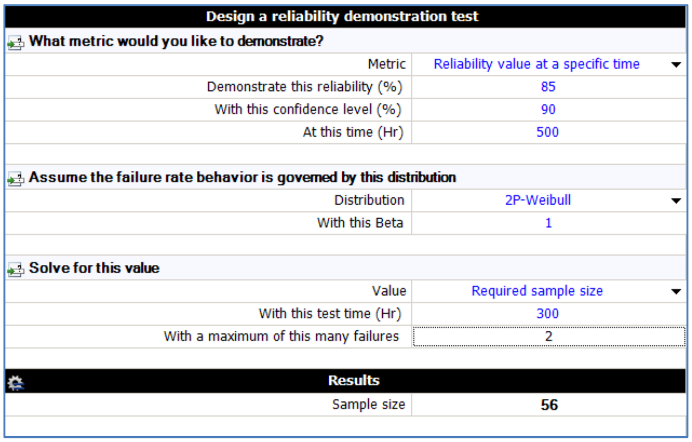
The binomial equation becomes:

$$1 - CL = R^n \Rightarrow \chi^2_{1-CL,2} = \frac{2n \cdot T}{MTTF}$$

When there are failures, they will give different results since the Chi-squared method replaces failed items during the test while the binomial method does not.

<u>*Example 7*:</u> Use the Parametric Binomial method using the same input provided in example 5. Check if the calculated sample size is close to the one provided by the Chi-squared method.

<u>*Solution for Example 7*</u>: The result from Weibull++ is given in the figure below.



Design a reliability demonstration test	
What metric would you like to demonstrate?	
Metric	Reliability value at a specific time
Demonstrate this reliability (%)	85
With this confidence level (%)	90
At this time (Hr)	500
Assume the failure rate behavior is governed by this distribution	
Distribution	2P-Weibull
With this Beta	1
Solve for this value	
Value	Required sample size
With this test time (Hr)	300
With a maximum of this many failures	2
Results	
Sample size	56

Figure 13. Using Parametric Binomial for the Case of an Exponential Distribution

The calculated sample size is 56, which is close to the value of 55 that was obtained from the Chi-squared method.

3.5 Non-Parametric Bayesian Test

In reliability tests, in order to better estimate the reliability, Bayesian methods have been used, especially when limited samples are available. If there is prior information on the system's reliability, or from previous subsystem tests, this information can be utilized in order to design better reliability demonstration tests given the sample constraints. It has been proven by [5] that the reliability, *R*, is a random variable following a beta distribution in the binomial equation Eq.(15).

$$1 - CL = Beta\left(R, n - f, f + 1\right) \tag{15}$$

In general, when a beta distribution is used as the prior distribution for reliability *R*, the posterior distribution obtained from the Eq.(12) is also a beta distribution. For example, assuming the prior distribution is Beta (R, α_0, β_0), the posterior distribution for *R* is:

$$1 - CL = Beta(R, n - f + \alpha_0, f + \beta_0) \tag{16}$$

Therefore, Eq.(16) can be used for Bayesian reliability demonstration test design. For a random variable x with beta distribution $Beta(x, \alpha_0, \beta_0)$, its mean and variance are:

$$E(x) = \frac{\alpha_0}{\alpha_0 + \beta_0}; \qquad Var\left(x\right) = \frac{\alpha_0 \beta_0}{\left(\alpha_0 + \beta_0\right)^2 \left(\alpha_0 + \beta_0 + 1\right)} \tag{17}$$

If the expected value and the variance are known, the parameters α_0 and β_0 in the beta distribution can be solved by:

$$\begin{aligned} \alpha_0 &= E\left(x\right)\left[\frac{E(x) - E^2(x)}{Var(x)} - 1\right] \\ \beta_0 &= \left(1 - E\left(x\right)\right)\left[\frac{E(x) - E^2(x)}{Var(x)} - 1\right] \end{aligned} \tag{18}$$

<u>*Example 8*</u>: According to the history of a system, it is known that

- The lowest possible reliability is: *a* = 0.87
- The most likely reliability is: *b* = 0.90
- The highest possible reliability is: *c* = 0.99

We need to design a test to demonstrate that the reliability is 90% at a confidence level of 80%. Assume 1 failure is allowed in the test. What is the necessary sample size for the test?

<u>*Solution for Example 8:*</u> First, based on the available prior information, we can approximate the mean and the variance of the system reliability. They are:

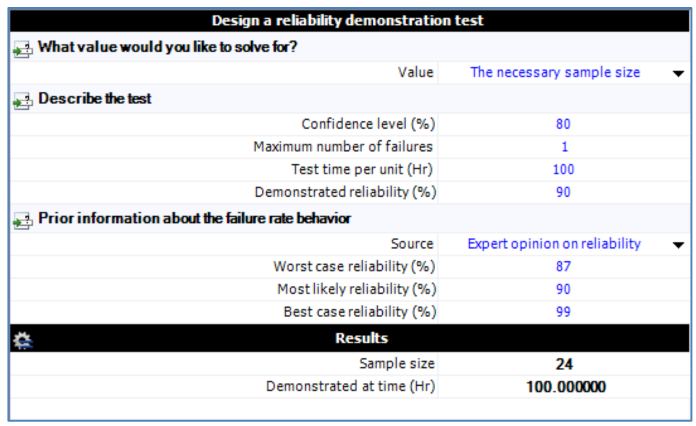
$$E(R) = \frac{a + 4b + c}{6} = 0.91, \; Var(R) = \left(\frac{c - a}{6}\right)^2 = 0.0004$$

Using the above two values in Eq.(18) we can get the prior distribution for *R*. It is a beta distribution $Beta(R, \alpha_0, \beta_0)$ with:

$$\alpha_0 = E\left(R_0\right)\left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right] = 185.4125$$

$$\beta_0 = \left(1 - E\left(R_0\right)\right)\left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right] = 18.3375$$

Using Eq.(16), we can solve for the required sample size n since *CL*, and *f* are given. The result given by Weibull++ is:



Design a reliability demonstration test	
What value would you like to solve for?	
Value	The necessary sample size
Describe the test	
Confidence level (%)	80
Maximum number of failures	1
Test time per unit (Hr)	100
Demonstrated reliability (%)	90
Prior information about the failure rate behavior	
Source	Expert opinion on reliability
Worst case reliability (%)	87
Most likely reliability (%)	90
Best case reliability (%)	99
Results	
Sample size	24
Demonstrated at time (Hr)	100.000000

Figure 14. Determining Sample Size Using a Bayesian Method Based on Prior System Reliability

Given the prior information, we need at least 24 samples in the test to demonstrate the required reliability.

When test results for subsystems are available, they can also be integrated in the system reliability demonstration test design [5]. Assume a system has k subsystems. For each subsystem i in a system, its reliability can also be modeled using a beta distribution. If there are f_i failures out of n_i test samples, R_i is a beta distribution with the cumulative distribution function as:

$$1 - CL = Beta(R_i, s_i, f_i + 1) \tag{19}$$

where $s_i = n_i - f_i$ represents the number of successes. Therefore, the expected value and the variance for R_i are given by:

$$E(R_i) = \frac{s_i}{n_i + 1}; \qquad Var\left(R_i\right) = \frac{s_i\left(n_i + 1 - s_i\right)}{\left(n_i + 1\right)^2\left(n_i + 2\right)} \tag{20}$$

Assuming that all the subsystems form a series configuration, then the expected value and the variance of the system's reliability *R* can then be calculated as follows:

$$E(R)=\prod_{i=1}^{k}E(R_i)\;;\qquad Var(R)=\prod_{i=1}^{k}\left[E^2(R_i)+Var(R_i)\right]-\prod_{i=1}^{k}\left[E^2(R_i)\right] \tag{21}$$

With the mean and variance in Eq.(21), we can get α_0 and β_0 for the prior distribution of R using Eq.(18). Once the prior distribution for the system reliability is obtained, we can use Eq.(16) to design the test.

Example 9: Assume a system of interest is composed of three subsystems A, B, and C. Prior information from tests for these three subsystems is given in the table below.

Table 3. Subsystem Test Results

Subsystem	Number of Units (*n*)	Number of Failures (*f*)
A	20	0
B	30	1
C	100	4

Given the above information, in order to demonstrate a system reliability of 90% at a confidence level of 80%, how many samples are needed in the test? Assume the allowed number of failures is 1.

Solution for Example 9: First, using Eq.(20) we can get the mean and the variance for the reliability of each subsystem; then using Eq.(21), we can get the mean and the variance for the system reliability. Once these two values are obtained, Eq.(18) can be used to get α_0 and β_0 for the prior distribution of R. The result obtained from Weibull++ is given below:

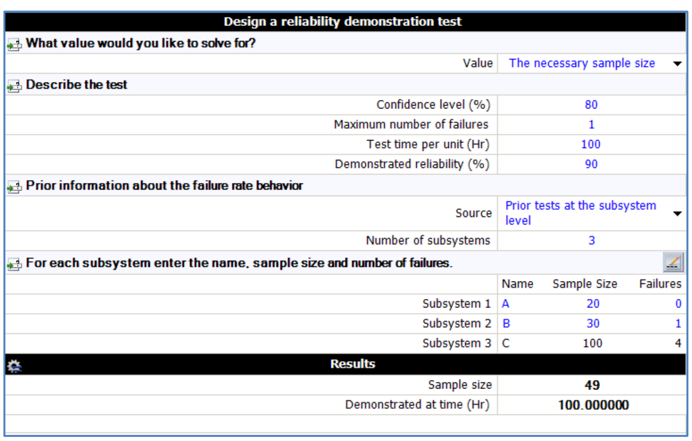


Figure 15. Determining Sample Size Using Bayesian Methods Based on Subsystem Tests

As seen in the above figure, 49 samples are needed in order to demonstrate the desired reliability.

If the estimated reliability based on the prior information either from expert opinion, previous tests, or subsystem tests is high, the Bayesian method will require fewer samples than the simple non-parametric binomial method. If the estimated reliability based on prior information is low, then more samples will be needed to demonstrate a high reliability. *Therefore, using the Bayesian method will not always lead to a reduced number of required samples. One must be cautious when applying Bayesian methods in data analysis since the validity of the prior information can have a significant effect on the calculated results.*

4. CONCLUSIONS

In this tutorial, the issue of sample size determination for reliability testing was discussed. Two commonly used approaches for determining sample size were introduced: the estimation approach and the risk controlling approach. It has been proven that these two methods lead to similar results under certain conditions [5]. If the purpose of a test is to accurately estimate a given reliability measure, using the estimation approach is recommended. If the purpose of a test is to demonstrate a specific reliability, then using the risk controlling method is advisable. In the case of the risk controlling method, the sample size can be determined either based on the Type I error, Type II error, or both. In this tutorial, the focus was on controlling the Type II error. Examples of applying different methods are illustrated throughout the tutorial.

6. REFERENCES

1. Meeker, W. Q., and Escobar, L.A. *Statistical Methods for Reliability Data*, John Wiley & Sons, Inc., New York, 1998.
2. ReliaSoft Corp., (2012), *Life Data Analysis Reference*, available online at: http://www.reliawiki.com/index.php/ReliaSoft_Books .
3. Guo, H., Pan, R., (2008), "On Determining Sample Size and Testing Duration of Repairable System Test," *RAMS*, Las Vegas, NV, Jan. 28-31, 2008.
4. Guo, H., Niu, P., Mettas, A., Ogden, D., "On Planning Accelerated Life Tests for Comparing Two Product Designs," *RAMS*, Reno, NV, Jan. 23-26, 2012.
5. Guo, H. and Liao, H., "Methods of Reliability Demonstration Testing and Their Relationships," *IEEE Transactions on Reliability*, vol. 61, no. 1, 2012.
6. ReliaSoft Corp., (2012), *Accelerated Life Testing Data Analysis Reference*, available online at: http://www.reliawiki.com/index.php/ReliaSoft_Books