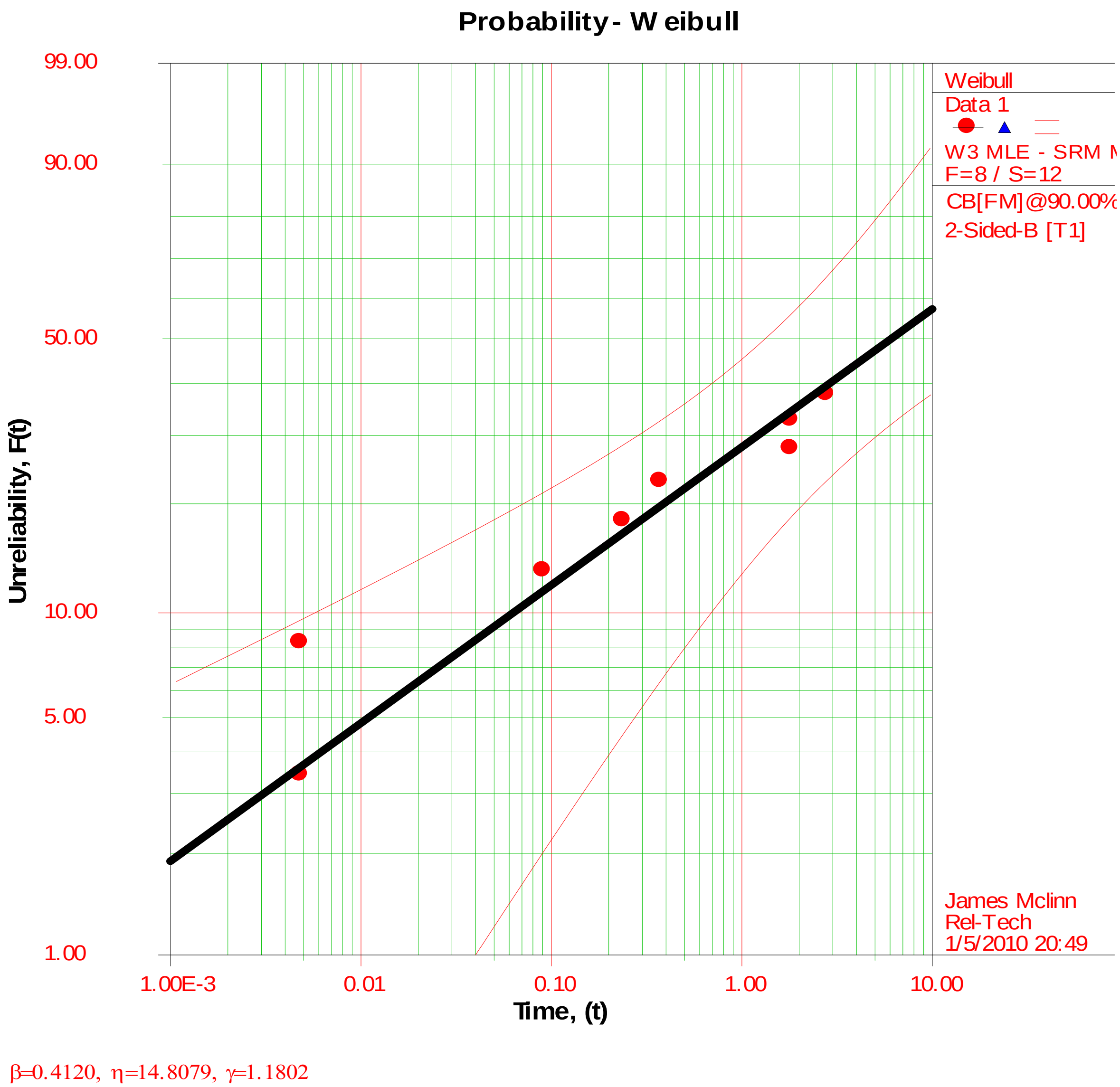
Practical Weibull Analysis Techniques - Fifth Edition

by James A. McLinn



Published by
The Reliability Division
of ASQ - January 2010
ISBN 0277-9633

Preface - Fifth Edition

This book was initially created out of response for the expressed desire by many engineers for a simple, introductory group of topics on the applications of the Weibull distribution. While several larger texts and book chapters on Weibull techniques already exist, these may be found by some to be difficult to follow. This book was designed to be short, simple and to the point as well as aimed for most reliability and quality engineers. No prior extensive Weibull or statistical knowledge is assumed. The use of statistics is kept to a minimum and the background formulas shown where it would be useful to the reader. A wide variety of simple and useful examples are presented in the text. Only the last example runs for more than two pages. Excessive use of or dependence on statistical tools have been avoided because most engineers find the language of statistics to be foreign and often obscures the engineering or business point being studied. Show this book to your co-workers and when you are ready move beyond it to the specific chapter of another book that provides many more details. A detailed bibliography was provided for this purpose. A reading list exists at the end for additional topics related to Weibull. These are primarily from recent journals and cover events happening in Europe. For the first time there is also list of reliability web sites that relate to Weibull. These should help people quickly find Weibull applications and sources of software help.

The author also wishes to thank the many people who commented and suggested improvements on earlier versions. It is sometimes hard to believe that this has evolved into a fourth edition. The articles that began this book were first published in December 1993. People who helped to stimulate changes and better understanding in this edition include Walt Thomas of NASA, John Berner of Minneapolis and Harold Williams of the Reliability Review. I thank them for their insightful comments and encouragement. The largest thanks go to my wife Connie McLinn who provided a thorough review of the manuscript and timely comments.

This edition has received a thorough scrutiny by a number of readers and helpful and suggestive changes. It also incorporates a number of improvements to the topics and additional details to calculations. I want to thank all of the readers who took the time to call or contact me with their questions and comments.

James McLinn
Hanover, Minnesota
January 2010
JMREL2@aol.com

Table of Contents - Fifth Edition

Area and Topic	Page Number

Table of Figures - Fifth Edition

Description of the Figure | Page Number

List of Tables - Fifth Edition

Simple Weibull Analysis

1.0 A History of Weibull

For a number of years, Walloddi Weibull was the main source of the development of the Weibull methods and was the sole person doing applications and development. This was because Walloddi published primarily in Swedish in European journals in the 1930s. He is credited with inventing the distribution in 1937 and claimed that the distribution applied to a wide range of problems. His early works showed examples that ranged from the strength of steel to the distribution of the height of males in the British Isles. Weibull had been born in 1887 and started his career as a midshipman in the Swedish navy in 1904. He rose in the ranks to become a Captain in 1916 and eventually a Major in 1940. He attended the Royal Institute of Technology in Sweden and graduated in 1924 and later obtained a Doctorate in 1932 from the University of Uppsala. His industry work was primarily in the development of ball and roller bearings in the German and Swedish industries. In 1951 he published his first paper for the ASME in English. It was titled "A Stastistical Distribution Function of Wide Applicability" By 1959 he had produced "Statistical Evaluation of Data from Fatigue and Creep Rupture Tests. Fundamental Concepts and General Methods" as a Wright Air Development Center Report 59-400 for the US military.
The 1950s was in the exciting and changing decade for reliability as the Weibull distribution became better known in the United States. In the next decade (1960s) a number of people began to use and contribute to the growth and development of the Weibull function, the common use of the Weibull graph, and the propagation of Weibull analysis methods and applications. In 1963 Weibull was a visitng professor at Columbia and there work with professors Gumbel and Freudenthal in the Institute for the Study of Fatigue and Reliability. While he was a consultant for the US Air Force Materials Laboratory he published a book on materials and fatigue testing in 1961. He died in France on October 12, 1979 [1]. No attempt has been made so far to credit all of these diverse people for their many contributions, both large and small, to the development of Weibull analysis around the world since the 1950s. I encourage the readers to look at Appendix N of the Fourth edition of The New Weibull Handbook by Dr. Robert Abernethy [1] for more historical details of Walloddi Weibull.

Any review of a Reliability And Maintainability Symposium (RAMS) notes will typically turn up two or three articles on applications or theory of Weibull almost every year.

A few of these people who helped develop Weibull are mentioned here. First Dorain Shainin who wrote an early booklet on Weibull in the late 1950s, while Leonard Johnson at General Motors who helped improve the plotting methods by suggesting median ranks and beta Binomial confidence bounds in 1964. Professor Gumbel demonstrated that the Weibull distribution is a Type III Smallest Extreme Value distribution. This is the distribution that describes a weakest link situation. Dr Robert Abernethy was an early adaptor at Pratt and Whitney and he developed a number of applications, analysis methods and corrections for the Weibull function.

I want to thank all of the prior people who shared their knowledge in the past and those today who continue to advance and spread knowledge of the many uses of the Weibull distribution and the variety of Weibull analysis techniques. They have made it possible for so many of us to easily analyze complex data.

1.1 Applications of Weibull

A number of interesting applications and methods for the Weibull distribution have been developed. These include:

1) Analysis of a group of time to failure data points for complete and incomplete data sets

2) Projecting future failures from information about past failures.

3) Determination of how many spares components or assemblies would be needed to replace failures. This projection would typically include optimum number of spares, costs for replacement, optimum replacement time to avoid failures and optimum cost for the process.

4) The estimation of the number of non-conforming parts or systems in a population and based upon testing of a sample.

5) The estimation of warranty costs based upon testing of a sample.

6) The sample size and length of test required to demonstrate at some confidence level, to demonstrate the minimum time to failure of a component or system.

7) The sample size and length of test required to demonstrate at some confidence level, to demonstrate that a failure mode has been corrected or is no longer important.

8) The demonstration and modeling of time to wear out situations.

9) Maintenance planning for systems to avoid unexpected shut-down situations.

10) Evaluating drift situations of systems and products.

11) Evaluation of time to failure for mixed samples or multiple failure modes.

12) Applications of Quality Control to estimate defective numbers in a population.

13) Reliability growth applications.

14) Compatible with Monte Carlo and other random techniques to extend the impact of Weibull analysis.

15) The analysis of field data situations. This includes when data is missing, ramp-up estimations and tests various repair policies.

16) Three parameter analysis (time or operating cycle offset situation) versus a two parameter model of time or cycles to failure.

17) Estimation of minimum life when no failures occur in test.

18) Compatible with Bayesian models to make a Weibayes analysis.

1.2 – Weibull 101 -

Weibull, through his numerous publications, presented three different types of formulas to describe the Weibull distribution [2]. Here, I will use the most common one throughout. We can start the many uses of the Weibull distribution with the simplest reliability formula. This is stated as Reliability plus unreliability totals 1.0. As a formula this becomes:

$$\mathbf{R(t) + F(t) = 1.0} \qquad \mathbf{(1)}$$

Here, the reliability is represented by R(t) and its complement, F(t), is called the cumulative failure distribution (cdf) or the cumulative distribution function and represent the unreliability. In general, we may write for the Weibull function, employing the most common symbols:

$$\mathbf{F(t) = 1 - e}^{-(\frac{t-\gamma}{\eta})^{\beta}} \qquad \mathbf{(2)}$$

This equation is also sometimes shown as

$$\mathbf{F(t) = 1 - e}^{-(\frac{t-\gamma}{\eta-\gamma})^{\beta}} \qquad \mathbf{(3)}$$

This alternate form reflects the fact that the offset impacts the characteristic life, η.

The probability density function, or pdf, is typically labeled **f(t)** and is the time derivative of the cumulative distribution function, **F(t)**. It can, in general, be written for the Weibull distribution or for any other distribution as the time derivative, $\frac{dF}{dt}$. This is:

$$\mathbf{f(t) = \beta\, \eta}^{-\beta}\, \mathbf{t}^{\beta-1}\ \mathbf{e}^{-(\frac{t-\gamma}{\eta})^{\beta}} \qquad \mathbf{(4)}$$

This is often controversial and sometimes mis-identified as the "hazard rate". In fact the hazard rate is related to this function through the reliability as shown Equation 5. Most basic reliability texts clarify these relationships [3] and they hold for any distribution.

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{R(t)} = \boldsymbol{\beta\, \eta}^{-\beta}\, \mathbf{(t-\gamma)}^{\beta-1} \qquad \mathbf{(5)}$$

The variables β, η and γ are one set of standard Weibull symbols in common use.

β (Beta) is usually called the "shape parameter" and it describes the shape of the time to failure

distribution. It is the **slope** when plotted on the Weibull graph itself. (Figure 1)

η (Eta) is called the "scale parameter" and is located where 63.2% of the test units would have failed. This unusual location is the same for all values of the shape parameter and sometimes causes confusion because it is not at 50% as true of some other distributions.

γ (Gamma) is a "location parameter" and indicates the place where the time to failure distribution actually begins. It can be a positive or negative and is typically treated as a time or cycles to life offset. This factor becomes very important in the strength of materials, as it represents a minimum break strength. Figure 1 shows the relationship of all these parameters.

The great strength of the Weibull distribution lies in the fact that it has wide applicability to many reliability, maintainability, test and quality problems as listed in section 1.1 These applications include materials and material strength analyses, determination of the rate of wear, measure of the time to wear out, evaluation of a variety of stress versus life examples and ability to model almost all increasing and decreasing failure rate behaviors for both electronic or mechanical components. For simple systems, Weibull may also be a good model, but there are limits to describing systems, because of the complexity and the competing failure modes often present. Much of the information to be presented here was initially developed in the 1960s and 1970s in the United States and in Europe. Thus, much of it was new to Weibull himself near the end of his career. The greatest advances in the use of Weibull probably came in the late 1980's when canned Weibull software began to be developed. This provided new and powerful opportunities and a variety of new applications for Weibull analysis. In the last 15 years additional great strides have occurred in use of Weibull. The development of computers with standard software permit many people to do basic Weibull analysis for themselves. It is even possible to write your own software routines for many of the Weibull functions, if you are patient and willing to dig into the details. One author, created a how to guide in the 1990s [4]. This book will mention some of these applications and show some of what can be done.

The Weibull distribution's value is greatly enhanced by the modern graphical and computer analysis techniques that quickly and easily solve problems and test the data. These techniques are more then strictly time and labor saving. They often extend the use and applications of Weibull's original distribution by permitting rapid and extensive testing of the assumptions underlying the models or coherence of the data. This is explored in later sections of the book. The use of computers also permits data to be “filled in”, “extrapolated”, “extended” and even created (Monte Carlo) when good or complete data is not available. Each of these examples will be discussed and methods presented to explain how to fill in or extrapolate data. Limits will also be identified.

Accompanying this book is a piece of generic "Weibull Graph" paper generated by the program. This graph paper is an example of probability plotting paper similar to that employed for the other statistical distributions such as Normal or Log-normal. This paper is commonly available through a variety of sources or can be easily generated yourself. The following section details how. There are several different Weibull graph paper formats and pick the one that best suits your needs. Weibull analysis offers an amazing number of possible uses as discussed in section 1.1.

The reason the Weibull graph paper works so well is that if we take the double natural log of the rearranged Equation **(2)** we can create the following convenient equation:

$$\mathbf{LnLn}\ [\frac{1}{1-F(t)}] = \beta\ \mathbf{Ln(t)} - \beta\ \mathbf{Ln}(\eta) \qquad \mathbf{(6)}$$

This equation now has the form of **Y = mx + b** which is familiar and easy to use. With this form we can easily create a graph where β is the slope. The Weibull graph plots the LnLn[1/1-F(t)] expressed as a percentage failure versus the Ln(operating time). Thus Equation 6 is the start of the graph.

1.3 – Simple Approaches to Weibull

The Weibull distribution can be used to analyze time to failure, cycles to failure and a variety of material strength situations. Modern computer tools and statistical tables now make this much easier then in the recent past. These later tools were developed primarily in the United States since the 1960s and are documented in a variety of places including textbooks and on the web. Various handbooks [1, 3, 4, 5] document some of the wide variety of reliability, maintainability and quality applications of Weibull analysis. Some examples of the use of the Weibull distribution and Weibull graph paper occur as articles in journals such as the “IEEE Transactions on Reliability” or the “Journal of Quality Technology” [6 - 9]. A few of the basic and a few of the more complex examples will be illustrated here to show how the basic principles and applications evolve.

The Weibull distribution, through the shape parameter beta, β, is flexible and time to failure distribution can take a number of different shapes. Figure 1 shows one of the many possible shapes. These shapes will be described by the value of β and they typically range from a low of about 0.5 to a high of about 10 in most applications. A value of β less than 1.0 has a distinct Weibull shape that is very different from one with β greater than 1.0. Figure 1 shows a shape of about 1.5 (i.e. $\beta \sim 1.5$) with a time offset labeled as γ. Three points are noted and these correspond to 7.2% cumulative failures, 50% cumulative failures and the location of the mean, here at 56% cumulative failures. This figure is really the depiction of the "probability density function", or pdf that is often used. The familiar bell-shaped curve is another example of a pdf that corresponds to the Normal distribution (here $\beta \sim 3.44$). Figure 1 shows the mean of a distribution which is not always found at the middle, and the median, which is also the middle “by count”. The location of the mean really depends upon the shape of the probability distribution or the value of β in the case of Weibull analysis. The location of the mean of a Weibull distribution may vary from 46% to 70% as β varies over the range of 0.5 to 5.0. Equation 7 shows that the general relationship between the characteristic life and the mean depends upon β through the Gamma function. This is:

$$\text{Mean} = \text{MTBF} = \eta\,\Gamma(1 + 1/\beta) \qquad \textbf{(7)}$$

The Gamma function, Γ, is dependent upon the value of $1 + 1/\beta$ and this number needs to be looked up in a special Gamma Function table. This correction number usually represents a small correction to the value of η. Most of the time this correction number varies between 0.88 and about 1.00. It is possible for the correction to be greater than 1.0 when β is less then one.

Figure 1 shows both the median point, at 50% and the mean, at 56% here and labeled as the MTBF. The 7.2% point is the place where the shaded area under the graph represents 7.2% of the total area. This corresponds to a reliability of $R = 1 - 0.072 = 0.928$ since no more than 7.2% of test units would have failed by this point. The location of the median can be determined through the number of units on test. If we had 31 units, the median value would be the time of the 16 th failure. At this point there are the same number before and after this point “by count”. The point at 63.2% is also known as the characteristic life, η. This is a unique point on a curve as shown in Equation 8 and by Figure 2. The last point of interest is the time offset, γ. In this case, there are no failures are expected before the value of γ. At this point in time is where the probability distribution really begins. Before this point, the probability of failure is so low, that no failures will occur over the time involved. The time offset can be positive or negative.

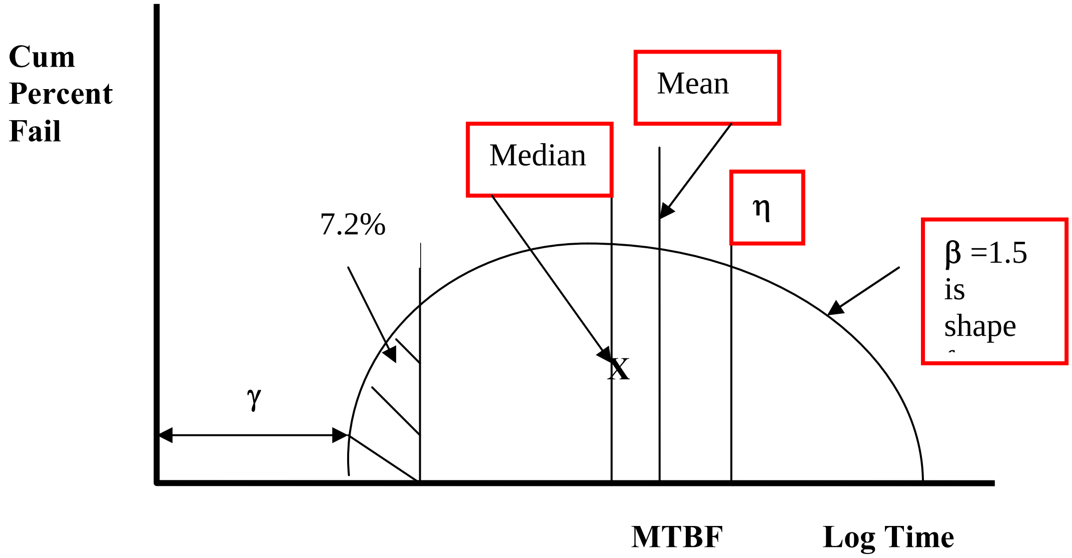


Figure 1 – The Typical Time Graph

When negative, the time to failure probability begins before the official time clock begins. This happens with some failure modes and mechanisms that are intrinsic to the nature of the components or to the material from which a component is constructed. Two simple examples are: positive offset represents a "failure free" time period. The first example is a positive offset and the second is a negative offset in time.

1) A new **car tire -** It takes some use before accumulated wear is evident and some time Longer before this wear mechanism accumulates enough to lead to a tire failure.

2) A group of electrolytic **capacitors -** A recent problem in the computer industry identified one Far East manufacturer who produced thousands of capacitors with leaky cases. These didn't become evident until after 18 to 24 months of continuous use. The time to failure distribution started before the computers were shipped to customers, hence it would be a negative offset

Either example shows a time or use offset even though none may have failed as a result of short term normal test or use conditions.

$$\mathbf{R(t)} = e^{-(\frac{t}{\eta})^{\beta}} = e^{-(\frac{\eta}{\eta})^{\beta}} = e^{-1} = 0.368 \qquad \mathbf{(8)}$$

Figure 2 shows that when the operating time equals the value of η, then for all values of β, the same probability is achieved.

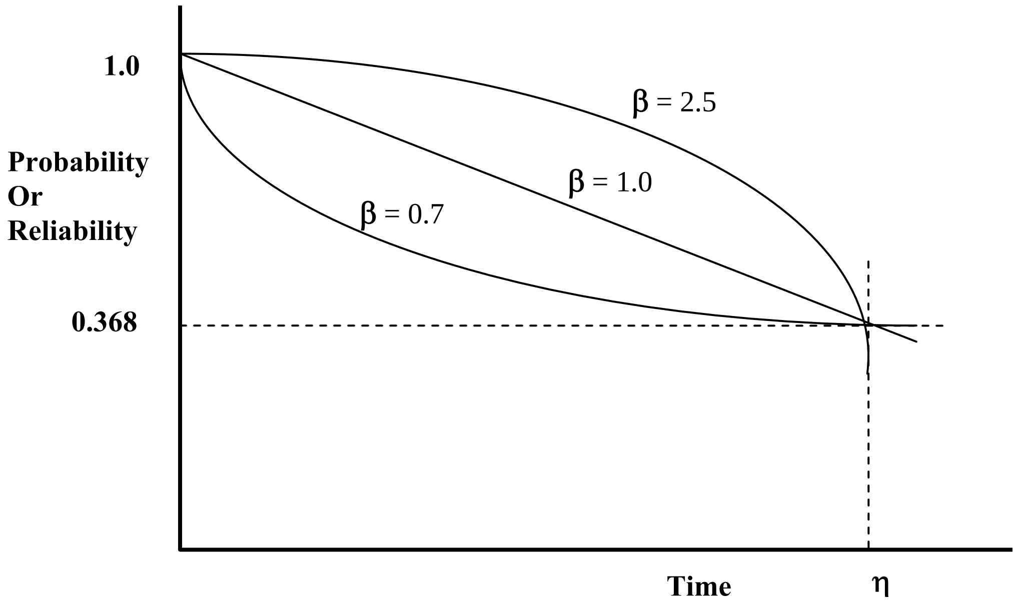


Figure 2 – Time Dependence of Various Values of Beta

1.4 - Simple Examples and Calculations

Example 1.1 – The following example is based upon a real life test of 10 mechanical products which will be labeled from A to J for identity. These products are dominated by the accumulation of wear if no early life failures occur. All units will be operated simultaneously with a constant load for each and driven by 120 VAC regulated power source. Before the test begins we need to define a "definition of failure" and determine what should be monitored. As the test progresses we can note the various "times to failure" as they occur. One definition of failure is the volume of air pumped into a fixed load decreases by 10% from earlier measures under the same test conditions. This implies the same event occurs periodically and normal variation from measurement to measurement is smaller than the definition of failure. This is an example of a performance measurement. If we change the definition of failure, then the times to failure might also change. It is also wise to establish a meaningful definition for the "End of Life" of all performance measures. Since the definition is based upon some accumulation of wear, it is likely that a third Weibull parameter may be needed in the analysis. In the case of these units, after 1200 hours of continuous operation for all samples, five failures were observed. Failed units are removed from test for further analysis when discovered. Note that the discovery time might not be the same as the failure time. If the units are monitored periodically (say once an hour), there could be a small time of uncertainty. If monitoring is once a day, this period of uncertainty could become important. This is called interval testing and some Weibull program handle this automatically. All good units remained on test. The test data at

1200 hours is summarized in Table 1. This table has already been filled out as if the data were to be plotted on a graph. These calculations and entries are explained in the following section.

Table 1 - Life Data

Failure Rank Number	Identification of Failure Unit	Failure Time	Calculation	Median Rank Location
1	C	203 hours	0.7/ 10.4 =	6.73%
2	J	422 hours	1.7/ 10.4 =	16.3%
3	F	607 hours	2.7/ 10.4 =	26.0%
4	G	767 hours	3.7/ 10.4 =	35.6%
5	A	982 hours	4.7/ 10.4 =	45.2%
6	B none – operated successfully to 1200 hours			
7	D none – operated successfully to 1200 hours			
8	E none – operated successfully to 1200 hours			
9	H none – operated successfully to 1200 hours			
10	I none – operated successfully to 1200 hours			

Half, or five of the test units have failed by 1000 hours while the remaining 5 operated to the end of test, at 1200 hours. The remaining five may have experienced some degradation of performance, but this extra data would not be used when calculating by hand or using Rank Regression in a canned program. We could improve the overall analysis by making a separate plot of the degradation versus time. What can we conclude from the data in Table 1? Before the data can be plotted on a Weibull graph, the median ranks need to be calculated. The Median Rank location for each failure is the Y ordinate on the Weibull graph. These positions were derived from the Median Rank formula, Equation (9). The X co-ordinate is typically the natural log of the time to failure. It is also known as the Erto formula and is:

$$\textbf{Failure Location \%} = \frac{FailureRankNumber - 0.3}{TestExposureSize + 0.4} \qquad \textbf{(9)}$$

This median rank formula helps us plot the data on the Weibull graph by providing "better estimates" of the cumulative distribution function through the cumulative percent failure. The formula always places the median failure of a sample at 50% cumulative failure where the median belongs. Some authors suggest that this formula also unbiases the percent axis. This topic has been covered by Dr. Abernethy in his New Weibull Handbook [1]. Median ranks are preferred to mean ranks and other possible plotting methods for Weibull. The first failure at 203 hours was calculated as follows.

$$\text{First Median Rank} = \frac{1-0.3}{10+0.4} = \frac{0.7}{10.4} = 0.0673 = 6.73\%$$

The second median rank point, at 422 hours, is done in a similar fashion. This is:

$$\text{Second Median Rank} = \frac{2-0.3}{10+0.4} = \frac{1.7}{10.4} = 0.163 = 16.3\%$$

Figure 3 shows the Weibull graph of this whole data set. Note that the unfailed units are not plotted on the graph and they are only represented through the value of N. This is why the data points stop about half-way on the vertical axis. Most Weibull programs give options on the data analysis to handle this and other common situations. Three typical options are data analysis by “Rank Regression on the X variable”, “Rank Regression on the Y variable” and a Maximum Likelihood Estimator. In the case of the two rank regression methods, a best fit single line is generated and compared to the data points only. The difference between the best fit line and the X axis location of all the failure points is then minimized by adjusting the β and η for determining the best fit line. For Rank regression on Y, the difference between the best fit line and the Y variable is minimized. These two approaches can give slightly different estimates for the parameters of the best fit line. Usually, a goodness-of-fit estimator is also provided with most canned software programs. The ideal fit is 1.0 and no fit is 0. One rule of thumb is to have this goodness-to-fit metric be at least 0.90. Normally, Rank Regression is used when all of the units on test have failed. When only a fraction of the units have failed, as in this case, the Maximum Likelihood Estimator (**MLE**) is the preferred choice. MLE is a more complex calculation, since it generates a formula that fits all of the failures and then also models possible “time-to-failure” for the unfailed units. Thus, Figure 3 is based upon the MLE approach for this data set. A goodness-to-fit measure also exists and the more negative this number is, the better the fit. One caution should be noted with MLE. If only a few failures exist for a large sample, the best fit line generated by this method need not always run through the failures. This situation does happen occasionally and can be difficult to explain to management or those not familiar with the method.

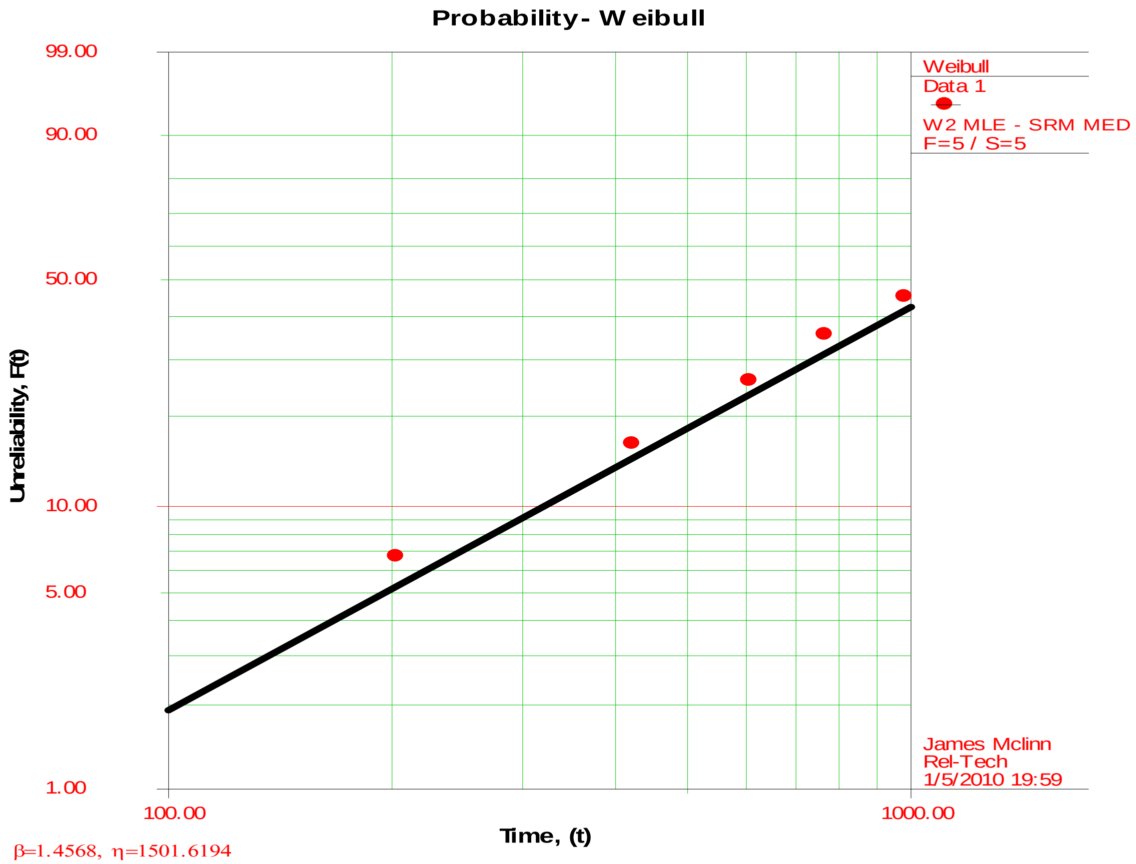


Figure 3 - The Basic Weibull Graph

The denominator, N, of the median rank formula depends upon the "test exposure size". This is a number that measures the "opportunity for failure" that is presented by all of the test units. It is sometimes simply the total test sample size, here N = 10 when the units are placed on test simultaneously and are not repaired or removed from test for reasons other than failure during the test period. For other types of test or situations this calculation can be more complex. The following is a short set of rules to can be used for estimating the "test exposure size" or "effective sample size" we call N.

Table 2 – Rules for Determining the Sample Size

Non-Repairable Components and Systems – Not replaced upon failure

The test exposure size, typically N, is the total number of units initially placed upon test.

Non-Repairable Components and Systems – Replaced upon failure

The test exposure size, typically N, is the total number of units placed upon test.

For Repairable Components and Systems

The test exposure size, typically N, is the total number of repairable units on test plus the number of repairs and replacements that occurred during test. This represents the total "opportunity for failure".

Or In the case of software or items that can be tested more than once, the total number of failures, N is the total of tests that were run.

For example, in the case of this last rule consider three sets of software that were tested, repaired and then retested as following:

Software A – tested 6 times

Software B – tested 12 times

Software C – tested 8 times

N = 6 + 12 + 8 = 26 opportunities

These three software programs provided 26 opportunities for test and failure, so we would set N at 26. Since the times to failure would be all mixed up, we would have to create a single table for all 26 failures by ordering the times from shortest to longest. The shortest time is the first failure, the second shortest time is the second failure and so on. Since an opportunity for failure occurs when a component or system is placed upon test or after a repair. Each time a system is repaired it will be treated ***as if*** it were new. Normally, this plan is followed unless some other counting scheme or purpose is desired. These three rules will cover almost all situations that can be encountered. Be careful, as these simple rules may cover 95% of the situations you encounter, they may not cover some unique situations.

The Weibull graph allows us to read (i.e. estimate) two (or three) important parameters that complete the Weibull formula. If more precision (or decimal places) is desired, then these parameters may be estimated from the two formulas shown below. The following two formulas **are for the MLE estimation** of β and **η.** The solutions to both are approximate because they often cannot be solved in closed form. In most cases, the approximate formulas provide relatively close and compatible answers to more precise approaches. The exception is when data has many censored entries (removed from test before failure) or has suspensions (units that did not fail during test). In these Rank Regression cases, equation 11 is the correct one. Both situations will be discussed later in more detail. The following two coupled equations describe β and η.

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N} Ln(\frac{t_{final}}{t_i})} \qquad \hat{\eta} = \left[\frac{\sum_{i=1}^{N} (t_i)^{\beta}}{N} \right]^{\frac{1}{\beta}} \qquad \textbf{(10)}$$

Here, N represents the total number of units on test

t_{final} represents the final test operating time.

t_i represents each failure time.

Be sure to count carefully for repairable systems. That is, more specifically, the time since the start of test that unit has accumulated. After each repair we restart the time clock on that unit. Be cautious about counting and keeping tracking of time. When a **suspension test** is run, these formulas in Equation 10 must be modified to a more complex form shown in Equations 11 and 12.

$$\frac{\sum_{i=1}^{r} r(t_i)^{\beta} + (N-r)(t_r)^{\beta} Ln(t_r)}{\sum_{i=1}^{r} (t_i)^{\beta} + (N-r)(t_r)^{\beta}} - \frac{1}{\beta} - \frac{1}{r}\sum_{i=1}^{r} Ln(t_i) \text{ solve for } \hat{\beta} \qquad \textbf{(11)}$$

$$\hat{\eta} = \left[\frac{\sum_{i=1}^{r} (t_i)^{\beta} + (N-r)(t_r)^{\beta}}{r} \right]^{\frac{1}{\beta}} \qquad \textbf{(12)}$$

Here, **N** represents the total number of units on test or the test exposure size, **r** is the number of failures out of the sample size N. Now t_r represents the final test operating time and t_i are the individual failure times. When we apply the Equation 11 formulas to the mechanical data of Table 1 we get as best estimates by hand calculation (hence the hat above the parameters) for β and η as:

$$\hat{\beta} = 1.467 \text{ and } \hat{\eta} = 1497.7 \text{ hours}$$

These estimated numbers are based upon calculating these estimates, based upon the **Maximum Likelihood Estimator, MLE,** by hand. This was done by estimating the value of β using a first pass through Equation 11 and then using Equation 12. Basic data is shown in Table 3, but the first equation needs to be solved by advanced methods such as Newton-Raphson. The initial estimate of β was 1.1 which led to the more precise estimate of β at 1.5 and then the best estimate of 1.467 with η of 1497.7 hours. These numbers were generated from this table after several passes through the table. Reference 10, page 296 has detailed examples and more precise approaches to solve these formulas. Canned software program are more precise and much faster.

Most of the time a canned software program will easily calculate the values of β and η. The values achieved by one software program were β = 1.457 and η = 1501.6 hours for the **MLE method** and β = 1.382 and η = 1428.6 hours with Rank Regression on Y. The MLE is shown in

Figure 3. Occasionally, a simple Weibull software program may employ the least-squares method representing another alternative. All three methods will usually provide similar answers when all units in test are run to complete failure. Round-off and the use of different data fitting formulas are responsible for the slightly different answers. When suspensions exist, it is common for the various methods to create values of β and η that may vary quite a bit. Watch the method employed by the software when calculating Weibull data. The MLE method counts the unfailed units as equally important as the failed units when creating a model and a "best fit" line for the data. The rank regression method and the least squares method focuses primarily upon creating a model that is primarily "best fit" for the failures only. When suspensions exist, the MLE is probably the best method for the data set. With small sample sizes (less than 10 or so) and suspensions present Abernethy has identified a correction factor for the rank regression method that provides answers closer to the MLE method [1].

Table 3 - The Details of the MLE Method for Calculating One Set of Weibull Parameters for the Mechanical System data

Failures	t_i	**Ln**(t_i)	$(t_i)^{\beta}$	$(t_i)^{\beta}$ **Ln**(t_i)
1	203 hrs.	5.313	345.3	1834.7
2	422 "	6.045	772.4	4669.2
3	607 "	6.409	1152.1	7383.3
4	767 "	6.642	1490.3	9899.3
5	982 "	6.890	1955.8	13,474.7
		--------	--------	------------
	Column Totals	31.299	25,756.2	178,937.22
6 to 10	1200 hrs.	7.090	5(2438.4)	5(17,262.7)

Abernethy's approach has begun the process of reconciling Weibull parameters calculated from the two most common methods when small sample sizes and suspensions are present. This new method is not yet considered a standard approach for calculating Weibull parameters, but may become one in the not too distant future. Using Equation 11 we have:

$$\frac{\sum_{i=1}^{r}(t_i)^{\beta}Ln(t_i)+(N-r)(t_r)^{\beta}Ln(t_r)}{\sum_{i=1}^{r}(t_i)^{\beta}+(N-r)(t_r)^{\beta}} - \frac{1}{r}\sum_{i=1}^{r}Ln(t_i)-\frac{1}{\beta} = 0$$

and filling in the numbers from Table 3 with the beta estimate of 1.1 gives:

$$\text{Difference} = \frac{123702}{17907.7} - \frac{1}{5}(31.299) - \frac{1}{1.1} =$$

$$\text{Difference} = 6.9078 - 6.2598 - 0.9091 = -0.261$$

Then the value of η was calculated and this completed the first pass through the hand calculation. The second pass, starting with a new beta estimate (1.50) would then eventually lead to the hand results cited earlier. Further passes through this equation set (Equations 11 and 12) could be performed and the results **might** come closer to the canned software MLE solution. The error here is small for both β and η. Since the canned software may employ more sophisticated methods to solve the complex equations, the two methods may sometimes be different.

1.5 - Non-Straight Lines on Weibull

Occasionally, when the data is plotted on the Weibull graph, we do not get a single straight line as the best fit approximation to the failures. Assuming that we have properly set up the test, run the test carefully and accounted for any "noisy" or irrelevant data points, then we usually expect to get a straight line. This line typically represent a single failure mode for a component, a dominant failure mode for a system or a mixed set of failure modes for a system. If the plotted failure data does not fit a single straight line we can try several approaches as part of the analysis to straighten the resultant line. These will be presented as a series of options in order of probability.

The first data correction or analysis method is to try a three-parameter Weibull model rather than a two parameter model. This suggests that the failure probability distribution did not start at the same time as the usual customer time clock. Either the failure probability started far before the customer time clock or far after the customer time clock. This was mentioned earlier with Figure 1.

One Example of a Cause for a Non-Straight Line

The following data set based upon the time ten systems on test illustrates one non-linear data possibility. If a three-parameter Weibull situation is suspected, then the data must be corrected before going on to calculate either β and η. The third Weibull parameter, γ, a time off-set, has a strong influence on the rest of the analysis and is always calculated first. Note, data points 7 through 10 are labeled right suspensions since the test was stopped before these units failed.

First, plot the original data on the Weibull graph and look at the graph. The plotted data shows a curve as in Figures 4 and 6. This curve could be up or down, the process is the same for either curve. A smooth curve is often an indication of the need for a three parameter Weibull. When a three parameter is suspected, then determine a correction factor based upon Equation 13. Figure 4 shows the plot of the original data shown in Table 4. A best-fit straight line drawn through the data points, even though it doesn't fit very well. Figure 5 shows the same data corrected for the third Weibull parameter and the straight line best-fit is a much better fit to the corrected data.

Table 4 - Non-Linear Failure Information

Failure Rank No.	Time to Failure	Median Rank Location
1	297 hours	6.73%
2	356 "	16.35%
3	545 "	25.96%
4	912 "	35.58%
5	1543 "	45.19%
6	3100 "	54.81%
7	3200 "	right suspension
8	3200 "	right suspension
9	3200 "	right suspension
10	3200 "	right suspension

The third Weibull parameter is typically calculated by a canned Weibull program as shown in Figure 5, but when calculating the correction by hand [3], Equation 13 may be used. Figure 6 shows the definitions of the terms $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$.

$$\gamma = \mathbf{t}_2 - \left[\frac{(t_3 - t_2)(t_2 - t_1)}{(t_3 - t_2) - (t_2 - t_1)} \right] \qquad \textbf{(13)}$$

Estimate the value of $\mathbf{t}_2$ and then fill in Equation 13 with the known values of $\mathbf{t}_1$ and $\mathbf{t}_3$. Equation 13 becomes with all three numbers:

$$\gamma = 485 - \{(3100\text{-}485)(485\text{-}297)/[(3100\text{-}485)\text{-}(485\text{-}297)]\}$$

$$\gamma = 485 - \{491{,}620/2427\} = 485 - 202.6$$

$\gamma = 282.4$

Now the correction factor needs to be applied to each data point in order to straighten out the line.

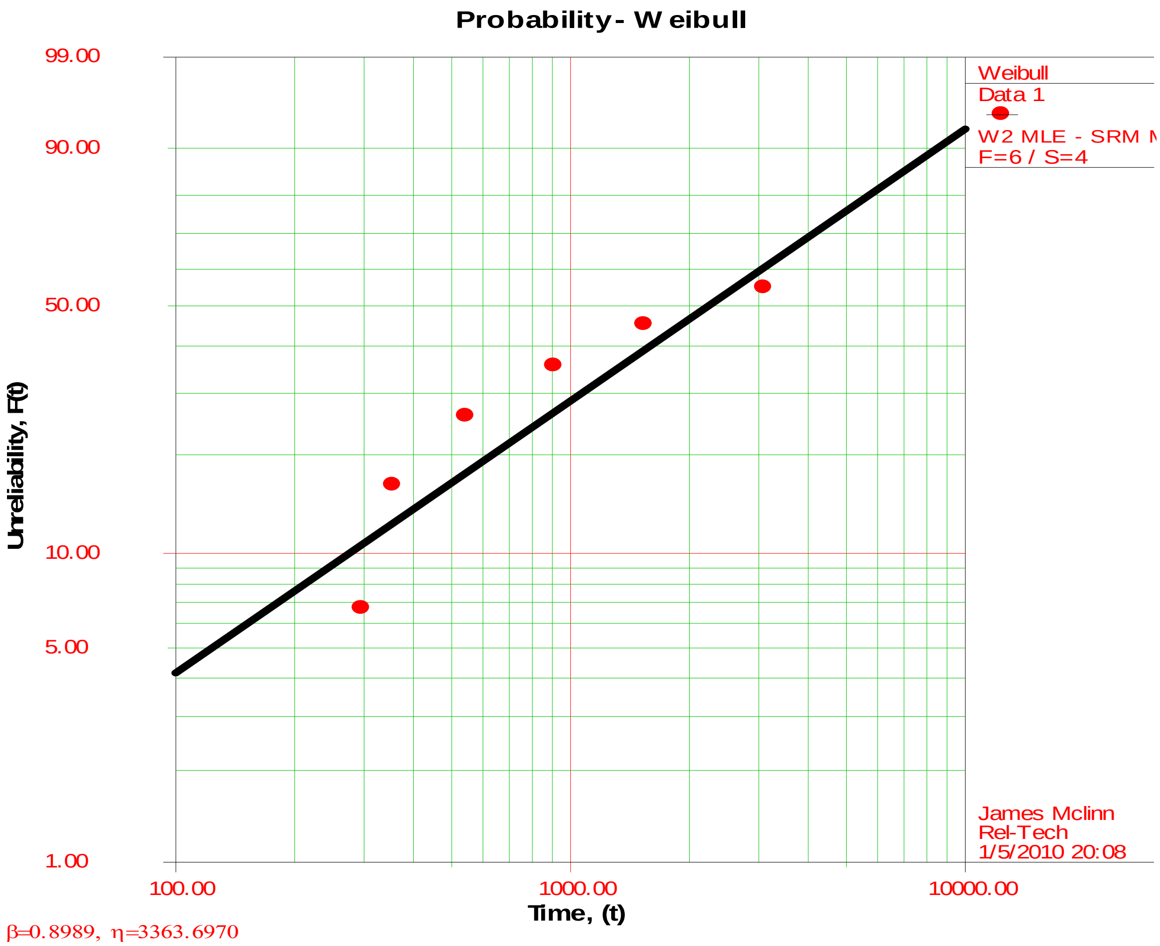


Figure 4 - The Curved Data on a Weibull

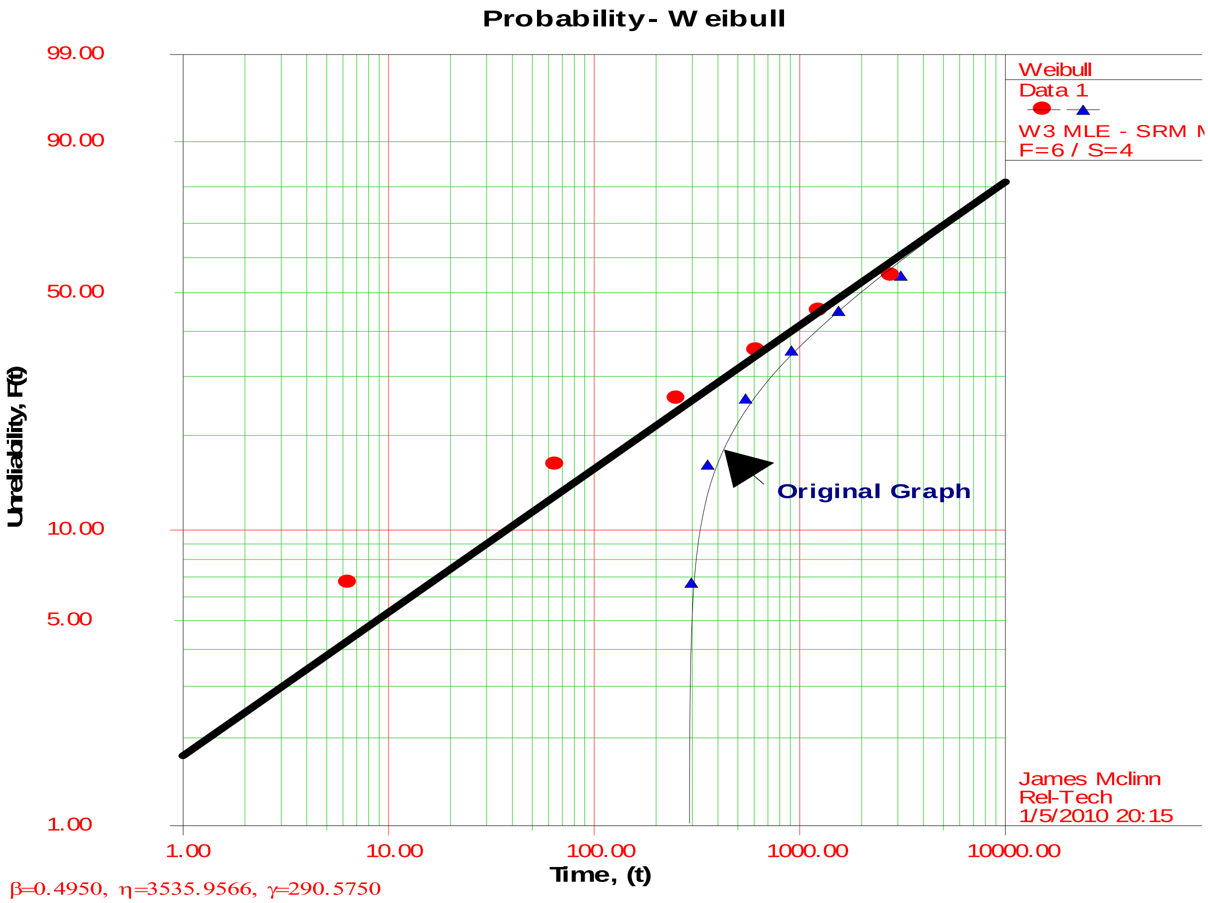


Figure 5 - Three Parameter Data Transformed

Figure 6 shows the relationship of the three times $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ that allow us to calculate γ. The initial estimation of $\mathbf{t}_2$ impacts slightly the calculation of the correction factor. With an initial estimate of $\mathbf{t}_2$ between 480 to 490 so a $\mathbf{t}_2$ of 485 hours was selected as a "best guess". The calculated result for γ became 282.4. Had $\mathbf{t}_2$ been selected as 480 or 490 instead, then the estimate of γ would be 283.3 or 281.6 respectively. Even $\mathbf{t}_2$ estimates as low as 400 or as high as 560 lead to γ estimates of 292.9 and 266.6. This is not a great range, being only about 9.8%, but this value does impact the estimate of β and η. Figure 4 shows the calculation of γ as 290.67 leading to values of β = 0.392 and η = 10,919 with the rank regression method (**RR**). The preferred **MLE** method (because of suspensions) would yield 291.3 for γ and 0.463 for β with 8210 for η. There is some variation between RR and MLE as expected. The biggest difference between RR and MLE is the characteristic life, which was about 30% while the slope was about 18%. It is possible to estimate γ directly from the graph, but the calculation with Equation 13 or a calculation from a canned Weibull program is usually more accurate than the simple eyeball estimate. Figure 5 shows that a canned program arrived at a value of γ as 290.6. Note, the value of γ is, in all cases, **approximately** the X axis intercept for the curve at about 0.1% cumulative percent defective. Typically the value of γ is less than the time to the first failure. Occasionally, a canned program will yield a large negative number for γ. In this case, one should reject the idea that a the third parameter is the best answer for a curved line on the Weibull graph.

To create Figure 5 from Figure 4 one subtracts the newly calculated value of γ from each of the original data points. Thus, we correct the "original data" by changing the time-to-failure only. A plot of the new data points using the same median rank locations, but with the corrected time is what we see in Figure 5. The canned program calculations are consistent with the simple formula and hand calculations shown earlier. Canned Weibull programs typically use a more elaborate approach than hand calculations. In some cases, they use a more elaborate search routine to obtain the best fit. Both the RR and MLE methods use different formulas for estimating this third Weibull parameter (see MLE section for details) and calculating β and η. It is surprising that with 40% suspensions only slightly different values were calculated by the two methods.

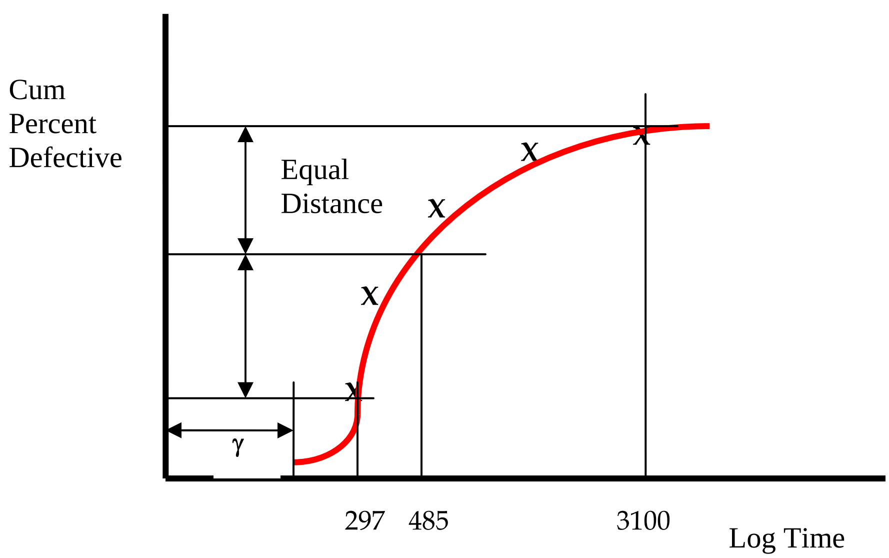


Figure 6 – Weibull Graph of Curved Line

Is It Positive of Negative?

The third parameter, γ, may be positive or negative. The existence of a γ really adjusts the start of the "time clock". When it is positive, this offset is sometimes said to represent a "failure free time". It is a time for which no failures are expected to occur, because the failure probability is so low as shown in Figure 7. When γ is a negative value, it **may imply** that failures can be expected at or before the customer start time or "t = 0". This point in time is usually measured as release to the customer or is at the end of assembly and test process. We usually describe these early failure units as "dead on arrivals". They represent failures that appear before a regular customer "start up" of the unit or appear as part of a formalized start up or acceptance process. Figure 7 shows both the before time zero failures and the early life failures. It is possible to run a small sample, such as 10 or 12 units and calculate a negative γ. This is suggestive that the larger population from which the sample data was derived contains these failures even though none were observed. We must look carefully at the meaning of the time offset, as it may imply "prior events or potential failure modes" as well as a "dead on arrival" interpretation. Sound engineering requires that the third parameter should be tied to a failure mode or understood failure mechanism to be considered a real time offset. Otherwise, a calculated time offset might be considered an artifact of noise in data taking or even “over-modeling”. As a rule, a three parameter model provides a better fit to most data sets than a two parameter model. This doesn’t mean the three parameter model is true.

As a last note, it is not reasonable to try and screen out or run a burn-in process or Stress Screen process on a sample with a negative time offset. This is like running a set of tires 10,000 miles to verify no early failures exist. The longer the screening time, the more failures are typically observed during the screen and also after the screen. Figure 7 shows this situation graphically. The real problem shown in Figure 7 is that the β value would often be a number between 1.5 and 5.0. This implies accumulative wear is a concern and that running the sample would find some failures, but would also accelerate the failure mechanisms on all the unfailed units.

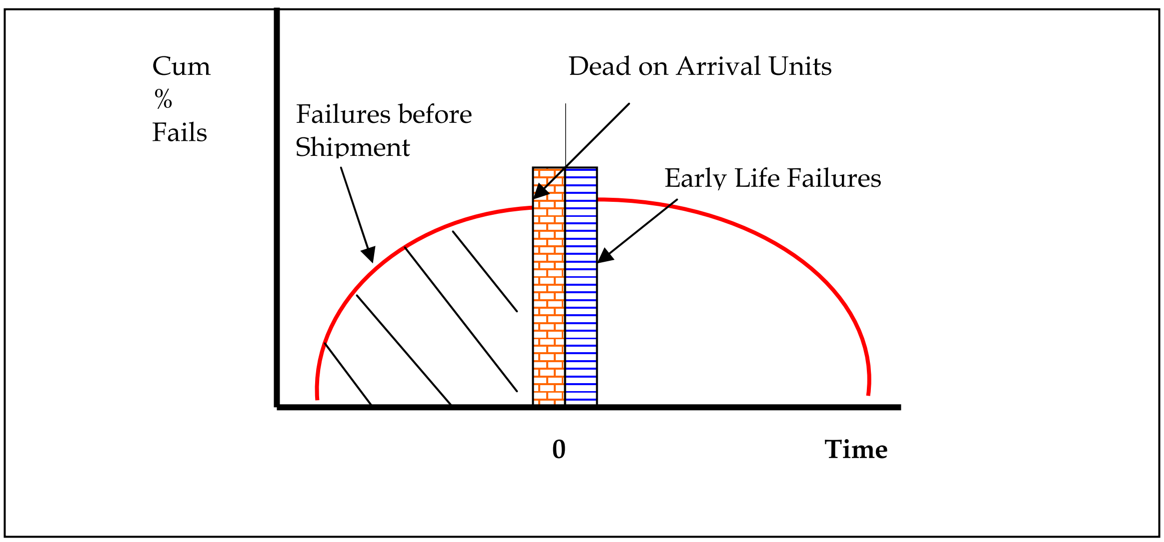


Figure 7 – A Negative Offset

The almost linear alignment of the new data points, as shown in Figure 5. This is one measure of the quality of the revised data fit for the three parameter Weibull. We expect that the correlation coefficient, called r^2 , for RR method or called the Likelihood Measure for MLE method. The correlation indicator, $\mathcal{L}$, will be a measure of the "goodness-of-fit" of the line to all of the data for MLE. The fit should always improve when we include the time correction, γ. One common "**rule of thumb**" is that we should reduce the difference of the RR method metric, r^2 from 1.00, which is the perfect fit. This difference should be reduced by at least half with this data transformation. If the old, two-parameter correlation factor for the RR fit was 0.931, then the three-parameter fit should be at least 0.967 to be considered valid. If this rule of thumb is not met and we have no reason to believe there are intrinsic failure modes, we should investigate other time-to-failure distributions such as Normal or Lognormal. These may provide an improved fit as a two parameter model. One common canned software program has an option to compare six different distributions to identify the one that is a best fit for the data. Additional reasons for the initial poor line fit to the raw data also exist. A list of the possible reasons or causes are covered in section 2.0 of the book. A rule of thumb exists for the MLE method. We would expect to improve the Likelihood fit indicator measure by at least 10% when going from a two parameter to a three parameter situation. In general, the larger (more negative) the likelihood number, $\mathcal{L}$, the better the fit to the data.

It is wise to use all of the information available and at hand before making a decision. This is a good engineering practice. Remember that the third parameter should correspond to a real failure mode or modes of the units under test. Engineering analysis of the situation or failure analysis of the failures may be required to support the proposed time offset. It is not unusual to think that some products have intrinsic off-sets. When investigating many material strength situations, a positive offset is typically expected. It is related to the minimum material strength that exists in many materials stressed at 50% or more below the fracture value. There are many time to failure situations that also fit this three parameter model. Any material with a slow accumulation of wear, any material that has a threshold to initiate failure or any material that has a slow propagation of defects that would eventually turn into a failure are all often modeled by a positive off-set. Not all canned software programs are able to evaluate the two and three parameter models for either the RR or MLE calculations. If one considers the four common distributions of Exponential, Weibull, Normal and Lognormal, these may not be described as two and three parameter with RR and MLE. A hand calculation may be required by the user to test this the three parameter option. One reason to look at the hand calculations here, is that the practical engineer can easily create several scenarios that postulate three parameter option, modify the existing data and try the new data set as a two parameter. Looking at the goodness-to-fit of these modified data sets shows if there are possibilities. I have done this myself using a spread sheet and then filling in the most likely modified data sets to look at the results and get the best fit. This lack of three parameter option may be a problem especially when conducting a multi-level life test. Both the Normal and Lognormal distributions can be evaluated with a third parameter as we do with Weibull.

1.6 - Other Possible Causes for Non-Linear Data

The data in Table 4 may **not be best modeled** by a Weibull distribution. We sometime need to consider other distributions such as Normal or Lognormal. It is possible that a Lognormal may fit the data equally well or better then the Weibull distribution. The Lognormal distribution has sufficient flexibility with up to three defining parameters (in some canned program), as does the Weibull distribution, so there are situations when either distribution may "be made to appear to be a good fit" to the data set. This is part of an ALT example at the end of the book. Since the original data is often from a small sample, such as 10 units with only 6 failures, it may be difficult to decide which of several distributions are really the best-fit for the data. More than one distribution may appear to be an "approximately equal" as a fit when measured by either the correlation coefficient or other "best fit" parameter. In these cases, we need to look carefully at the failure modes and the root failure causes to help make a good decision. Past history, engineering models, even common sense may also be considered to help us make this decision. If a best distribution cannot be easily determined by the initial failure information, then either extend the test to obtain a few more failures or increase the test sample size. This issue is addressed directly by Accelerated Life Test examples. Determining the best-fit distribution can be one of the most difficult problems associated with Weibull analysis, especially with small sample size. While some canned software manufacturers will present a matrix of 5 or more distribution choices and even perform an initial distribution comparison, it is left to the engineer to select the best distribution or follow one of the software recommendations.

When performing accelerated life tests, the problem of selecting the "best fit" reliability distribution typically increases. Nelson in ***Accelerated Testing*** [11] presents several examples of non-linear Weibull data derived from accelerated life testing. This only compounds the problem of selection and this area needs to be handled separately. The next section will aid in determining what is happening with non-linear data and how to decide when Weibull graphs are not clear. At the end of this book is an example to test your own skills by analyzing a difficult data set.

The following page is a blank Weibull graph. Use it when doing examples by hand or doing trial calculations for the offset. The reader can also generate Normal and Lognormal paper as needed. These probability plotting papers are readily available [12] from several sources.

"Non-Linear often seems to be the Nature of the World" - James McLinn

2.0 Messy Weibull Data

Not long after most people discover the value of the Weibull Analytical approach, they discover some data will not plot as a nice straight line. This statement allows for "noise" or scatter that always exists in the data. There may also be natural variability due to small sample sizes, the effects of imprecise time measurement and a variety of other causes of variability that typically lead to "rough looking data" on the Weibull graph. In addition to noisy or rough data, there are other definite and systematic reasons that strange and/or non-linear behavior may be exhibited on the Weibull graph. These causes range from multiple or competing failure modes, the use of the wrong modeling distribution for the data and extend to interval data.

A variety of options exist for plotting Weibull data. These range from hand plotting, which is seldom done unless the data (usually incomplete field data) can't be put into a format that can be plotted by a standard program. A few examples will be described here, as well as mentioning some standard software packages such as Weibull Easy *TM* , WinSmith *TM* , Weibull 7.0 *TM* , Excel *TM* and MiniTab *TM* . The last two are not Weibull packages, but programs that can do Weibull. These were designed for other statistical analysis purposes, can do some Weibull analysis, but do not have the range of capability or the ease of use of the first three packages. Like most things in life, when you buy a dedicated package you want the most useful features and ease of use that you can obtain. Other software packages exist and a longer list appears at the end of this book.

Table 5 starts with a collection of the reasons for strange or unusual behavior. The following eight reasons cover almost all of the possibilities that may commonly be encountered. Each reason requires some short explanation and an example for clarity.

Table 5 - The Eight Reasons for Non-Linear Weibull Behavior

1. **The Bath Tub Curve** - A sample in question may be exhibiting one or more of the stages of the "bath tub curve" while it operates on test. This situation applies equally well to electronic or mechanical components or systems. A variety of failure modes are typically exhibited through the life history. The early life failure stage might exhibit a Weibull slope (β) of 0.6, followed by a mid-life stage with a slope of about 1.0. The wear phase has a Weibull slope typically greater than 1.5 and sometimes as high as 5.0. Some of the modes associated with early life failures, or infant mortality type failures, but other failure modes are usually associated with middle-life or end-of-life failures, the data can't be linear. Further work and investigation is typically required to verify this situation, once a non-linear situation is recognized.

2. **A Mixed Population -** A sample in question may have been drawn from more than one sub-populations. These sub-populations often have distinct failure modes that are exhibited on a Weibull graph as an "S-shaped curve" on the Weibull graph. Not all of the “S curves” may be visible due to sample size restrictions or even short test times. An example of the S curve follows.

3. **Varying Environmental Conditions -** This possibility may occur when test units or field systems operate in different environmental conditions. This is normally inadvertent or accidental and is often not discovered until data is plotted upon the Weibull graph and questions asked. The situation leads typically to a bimodal or multimodal result on the Weibull graph.

4. **Mixed-Age Parts -** Since most parts and systems do not carry a time clock, we cannot easily tell how old or how aged a part or system may be just by looking at it or putting it on test. Imagine for a moment a collection of aged (already have operated about 500 hours) mixed with new parts and all placed on life test or in operation. The test results would probably look a lot like a mixed population Weibull graph. The difference here is that the age difference is the main reason for differing sub-populations.

5. **Three Parameter Weibull -** Some parts (or systems) seem to have a natural bias concerning time-to failure. Examples include many material strength situations, car tires, or telephone poles. This situation leads to non-straight lines because of the natural offset present. Once this offset is recognized and corrected by a software program, the curves often straighten out into one smooth line.

6. **Odd Distributions -** Some distributions do not appear as straight lines on a Weibull graph. The most prominent example is the LogNormal distribution, which often appears as two straight lines which join at or very near 50% cumulative failures. Test this possibility by plotting data on a Lognormal plot or another distribution.

7. **Mixed Failure Modes -** Very different failure modes, if operational during a test time or study period, may lead to unusual lines on the Weibull graph. Each line is usually associated with a dominant failure mode. When modes are separated, as is customarily done, each failure mode usually appears as a straight line.

8. **A Roll-Out Situation** – When a manufacturer ships increasing number of units into the field and then looks at the field failures on the Weibull graph, the results are usually not a straight line and may be confusing.

The following set of examples with show these possibilities. The data set in Table 6 will provide more extensive explanations.

2.1 - Three Parameter Weibull

Some non-linear data is best analyzed by the use of the three parameter Weibull function. The general reliability Weibull formula is shown by Equation 14. Other applications of this formula in mechanical may use an alternate form of this equation. The additional two formulas shown in Equation 15 will provide standard functional relationships that aid any reliability calculations.

$$\mathbf{R} = \mathbf{e}^{-(\frac{t-\gamma}{\eta})^{\beta}} \qquad \textbf{(14)}$$

$$\textbf{and} \quad \textbf{MTBF} = \eta[\,\Gamma(1+\frac{1}{\beta})] \quad \textbf{with } \sigma^2 = \eta^2[\,\Gamma(1+\frac{2}{\beta}) - \Gamma^2(1+\frac{1}{\beta})] \qquad \textbf{(15)}$$

Data that appears as a "curved line" on the Weibull graph may be due to one of two conditions that we will consider here. A smooth, down-turning curve, sometimes said to be a "positive correction", appears to cross the time axis before it crosses the cumulative failure axis was shown in Figures 3, 4 and 5. This period is sometimes referred to as a "Failure Free Period". It suggests there is a period of time before which no failures are expected, even for large sample sizes.

The other non-linear possibility occurs when the data forms a smooth curve that crosses the cumulative percent line at some point above the time line. This is sometimes said to be a "negative curve". This curve implies there are failures within a population, even at very short life times, at zero time or before zero time. Such a condition suggests that even reasonably small sample size may contain "already failed" units. Worse yet, the “screening or burn-in process" **may not help later performance** since increased operating time means increased failures will result. Said another way, the sample acts as if it has a short, finite life and screening has consumed a major part of this life. This unusual situation may be very perplexing the first time it is encountered.

Figure 8 shows an example of a real curve derived from a 4,000,000 cycle life test of 20 small coil springs. Table 6 contains the "raw data" that was employed to plot Figure 8. Figure 9 shows what happens when the offset (third parameter) correction formula was applied. The confidence bounds are also shown in Figure 9. Table 6 also shows the "corrected data" in the last column.

Be careful with the graph as shown in Figure 9. Remember there is still a time offset which must be included in all calculations. What happened after γ was calculated? This number correction was subtracted from each of the 8 original data points and new values of η and β were calculated. When using a computer program to calculate a best-fit value for γ, a series of iterations are often used to arrive at the best value. When calculating the correction factor by hand, we sometimes have to use the Equation 13 more than once to get the "best-fit". In this situation, we may discover that our first estimate of γ did not straighten the fit of the data when it is included. Work with the "new data", recalculate a new correction factor and replot the data. If the second calculation straightens the original line, then combine the two correction factors to get a best-fit γ value. Now calculate the new values of β and η. Try this on the coil spring data. The first hand estimate for the offset was 1.28 million cycles.

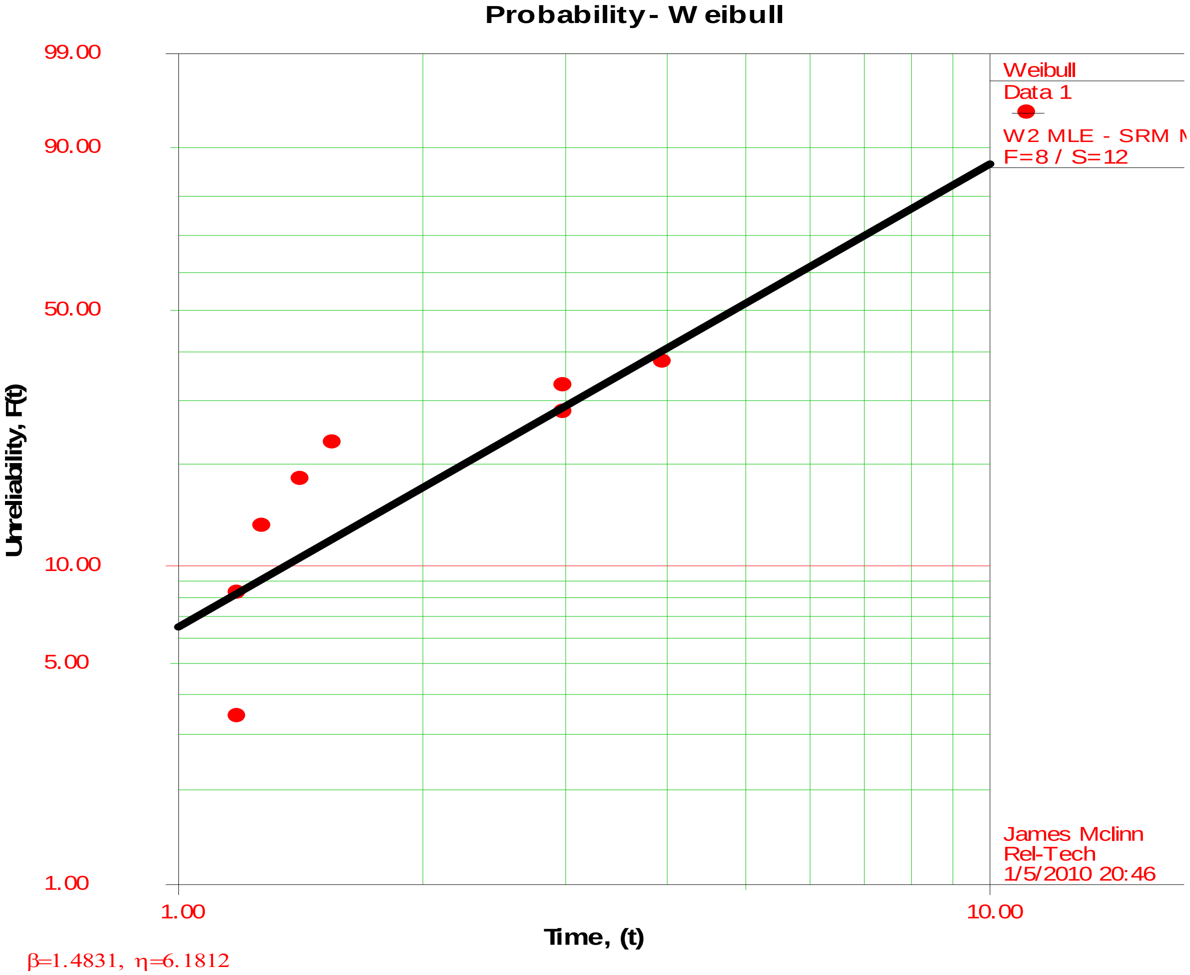
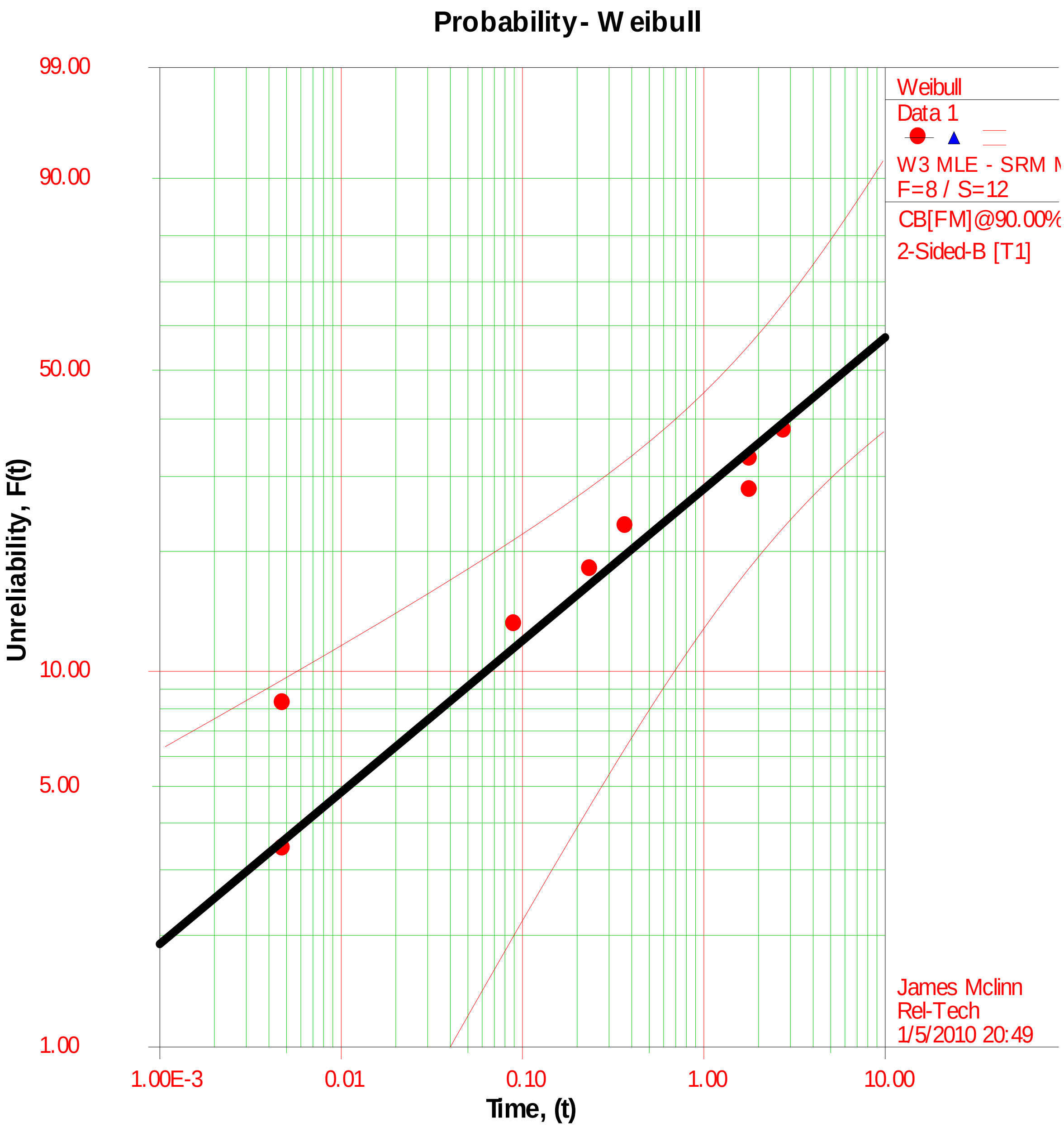


Figure 8 = Life Test of Coil Spring



β=0.4120, η=14.8079, γ=1.1802

Figure 9 – Three Parameter Analysis of Coil Spring Data

Table 6 - Coil Spring Life Data - MLE		
Failure Number	**Cycles to fail**	**Corrected data points - three parameter Weibull**
1	1.185 x 10^6	0.005 x 10^6
2	1.185 "	0.005 "
3	1.271 "	0.091 "
4	1.418 "	0.238 "
5	1.553 "	0.373 "
6	2.987 "	1.807 "
7	2.987 "	1.807 "
8	3.962 "	2.782 "
9 to 20	suspensions at 4,000,000 cycles	
two parameter, MLE	β = 1.483, η = 6.181 x 10^6 or	
three parameter, MLE	β = 0.412, η = 14.808 x 10^6, γ = 1.1802 x 10^6	

Remember the time offset may be positive or negative. In either case, not only do we move the location of the **start of the Weibull distribution**, but also the location of η and the shape of the distribution, hence β changes as well. Always remember, that we need to add back the correction factor when calculating the **real value** of the characteristic life. This approach is different from the way Weibull terms are defined in mechanical engineering applications of material strength.

In this example, we see that a two parameter MLE analysis provides us with a β = 1.483 which is consistent with the slow accumulation of wear and a finite limited life. The two parameter analysis suggests this life is not much more than 6 million cycles. The three parameter MLE analysis however yielded an offset of 1.18 million cycles with β = 0.412 and a characteristic life of 14.8 million cycles. This is very different, for it suggests that the slow accumulation of wear is not present since β is far less than 1.0. While the RR and MLE could be calculated, which analysis is correct? Both have some elements of fact. In theory, the MLE should provide a better estimate with the 60% suspensions present. We do expect a spring to have a limited life and even show some accumulation of wear, but we do not expect the spring to have a life of only 4 to 6 million cycles. Rather, mechanical engineering suggests this spring life is in excess of 20 million cycles. Why the vast differences in these 2 and 3 parameter models? Let us try another approach, perhaps the data is a combination of early and mid-life failures. This would be a mixture or bathtub curve situation. Such data should appear as an S-Curve or have some other non-linear behavior. Another possibility is that we don't understand the test results. It is an easy error to look at test data alone and assume it tells the story. In fact, a fraction of the time, a reliability engineer must look at the failed parts or systems and let the parts tell the story.

2.2 - The S-Curve on a Weibull Graph

Another reason for a non-linear behavior on the Weibull graph is the "S-Curve". Its cause is very different from that of the previous example. Often the S-Curve can be a result of a mixed population of components according to Abernethy [1] and others [13, 14, 15, 16]. On the time-to-failure graph, a simple linear graph of time and probability of failure, we may have something similar to Figure 11. This figure may be said to represent a mixture of two groups. The reasons for such a mixture may include a collection of components built at different manufacturing locations, components built at different times in the same manufacturing location, different stress operating environments on the components or even mixed age components. All can lead to a mixture when there is sufficient difference between the ages of the two or more groups. Mixtures are common as is the bimodal results. This model is also the basis for Burn-in or screening when it one lobe can be easily sorted from the other. [2, 17]

Table 7 - S-Curve Data

Fail Number	Time	Number	Time	Number	Time
1	0.2 hrs.	9	10 hrs.	17	32 hrs.
2	1.0 "	10	11 "	18	35 "
3	2.0 "	11	12 "	19	41 "
4	3.0 "	12	12 "	20	67 "
5	6.0 "	13	13 "	21	101 "
6	7.0 "	14	15 "	22	135 "
7	8.0 "	15	20 "	23	182 "
8	8.0 "	16	23 "	24	225 "

Items 25-32 are suspensions at 225 hours

These many mixed component possibilities may be more common especially during early product development, where adequate development and quality systems may not be fine-tuned. An engineering or quality investigation may be required to find true causes. The Weibull graph only points the way toward the follow-up questions that should be asked. The early time failures in this mixture may sometimes be incorrectly labeled "Infant Mortality" rather than recognized as a true mixture situation. These early failures are usually not typical of the failure modes of the long-lived group. The reliability engineer must look carefully if he suspects an S-Curve.

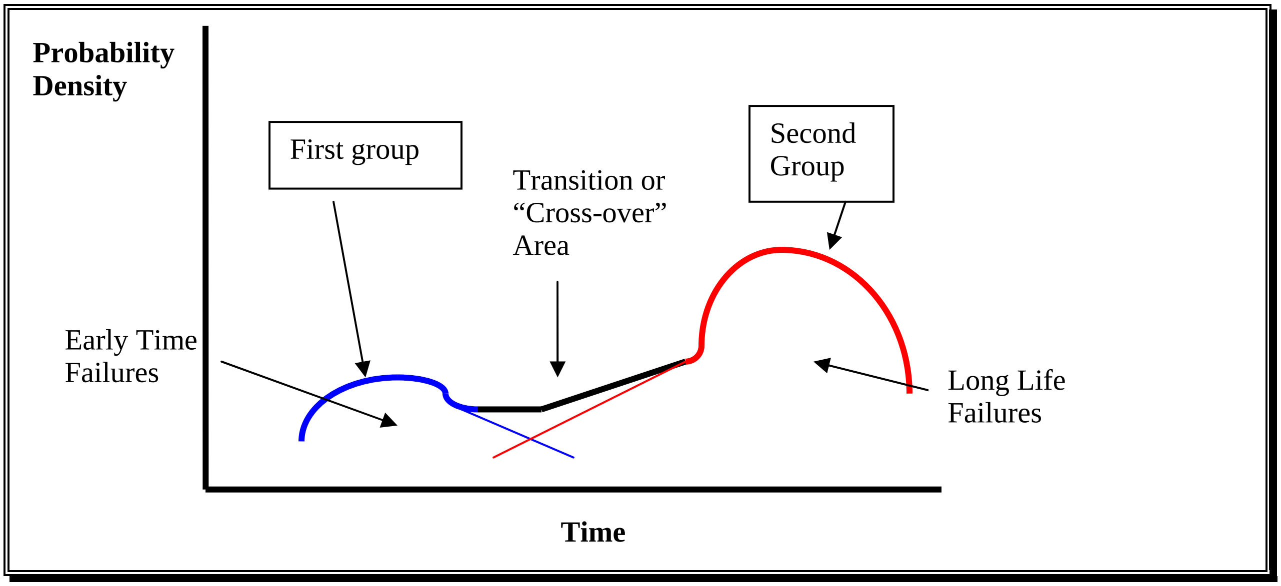


Figure 10 - A Linear Probability Graph

Table 7 contains a sample of S-Curve data generated from the life test of a semiconductor. Components often create an example of the S-Curve, but system examples are also possible, albeit more rare. Note, in Figure 11, eight units remain unfailed at the end of the life test at 225 hours. A MLE analysis is suggested for if this option is available. Be sure to consider the possibility of a 3 parameter Weibull and other distributions when analyzing. The following is a short summary of the possibilities to always consider when analyzing any suspicious curved data.

Possibility	**Analysis**
1) Two parameter Weibull or Normal	Not very likely for a curve on Weibull.
2) Two parameter Lognormal	Possible, look at cross-over, is it near 50%?
3) Three parameter Weibull	Possible, look at both the MLE and Rank Regression methods.
4) Three parameter Lognormal	Possible, try this for fit and corrected line.
5) Three parameter Normal	Possible, try this for fit and corrected line.
6) Mixture of 2 or more groups	Two groups seem most likely. Check failure modes for additional information.

The mixture situation is shown dramatically in Figure 10 with the two different distributions and their overlap forming the S Curve.

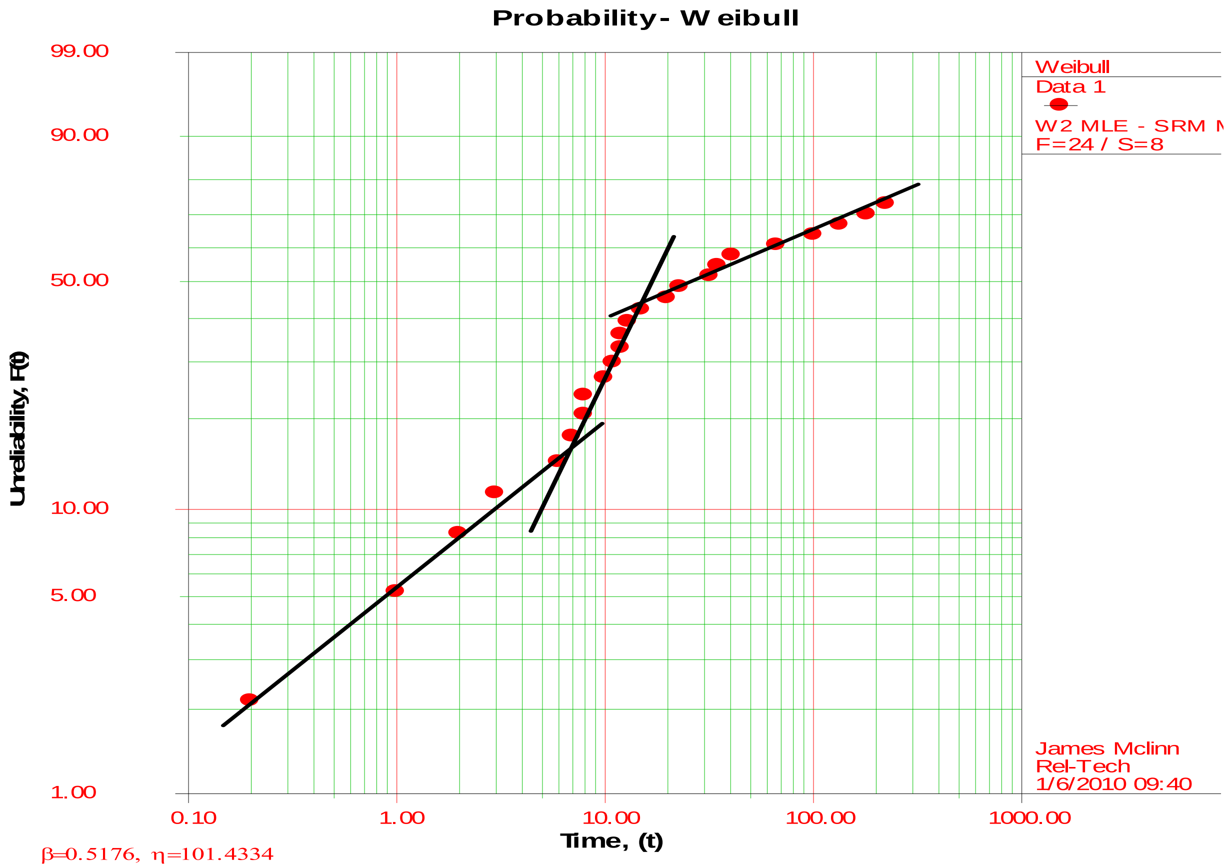


Figure 11 – The S Curve on Weibull

Figure 11 shows the complex S Curve graph plotted on Weibull paper. The average value of β is 0.52 with a η of 101.3 hours. If one breaks this graph into three pieces the following values for each piece are determined.

β1 = 1.12	η1 = 15.55 hours	portion = 22.6%
β2 = 5.88	η2 = 11.55 hours	portion = 14.4%
β3 = 0.45	η3 = 342.3 hours	portion = 63%

2.3 - Handling Censoring and Suspensions

The existence of censored and suspended data can lead to large changes in Weibull graphs through the way the suspensions strongly impact the appearance of the graph. See Nelson or Dodson [4, 11] for more examples on how to handle these situations. Figure 11 and Table 7 contained eight right censored points; these "items" were units that were still operating at the end of the test. This is a common situation, stopping a test before all of the samples have failed. Most canned software handles this situation directly and simply. When making a hand plot, the total number "exposed to test" includes the right censored (unfailed) items. Remember, however, that only failures are plotted on the Weibull graph.

Strictly speaking **suspended items** are those removed from test ***during the test itself***, not unfailed items remaining at the end of test, which are censored items. Suspended items really also appear as "holes" in the distribution on the Weibull graph. The following **example of suspension** shows how the data is manipulated and a corrected Weibull graph can be created. The term, "suspension" may also be applied to the situation when multiple failure modes exist. The second failure mode and any other subsequent failure modes may be treated ***as a suspension of the first failure mode.*** We may treat failure modes this way when they are independent of each other and/or have independent causes. There is no cause effect relationship between any of the failure modes. A corrected rank table should be created for clarity of the modes and for separating the modes.

Removing samples from test prior to failure of the sample or for reasons not having to do with failure is also a suspension. Plotting suspended data introduces a simple correction developed almost 50 years ago by Leonard Johnson. A correction to the distribution is required for each suspended item since the item in question could have failed, but was not permitted to do so when it was removed from test prematurely. The Johnson method is very simple, but not wholly realistic, because it does not matter at what time an item was suspended. This seems contrary to common sense as well as engineering principles. Recent alternative methods have been suggested that appear to be major improvements [18, 19]. In any case, the correction factors for suspensions will lead to non-integer median rank numbers and non-standard plotting percentages as shown in Table 9 and by Equation 16. The corrected number typically falls between the lowest possible probability of failure and the highest probability. The lowest choice reflects the possibility that no failure of the suspended sample would occur before a certain next event or period of time which is usually determined by another failure. The highest probability represents the possibility that the suspended sample would not have failed before the next known event or failure.

The following example shows why this is true, based on a probability argument. Later, a simple rule will be provided that also correctly calculates the correction factors. The following two tables (8 and 9) will be used to demonstrate how the process works and the Mean Rank Orders were calculated.

The first failure at 727 cycles was preceded by a suspension at 550 cycles. Had this suspension been left on test, it might have failed before 727 cycles. Thus, Failure 1 would have a Failure Rank Number between 1 and 2. It would be 1 if the suspension 1 would not fail by 727 cycles and would be Failure Rank Number 2 if the suspension would fail before 727 cycles. Next, we create two tables showing all of the possibilities. Table 8a shows the possibilities when Failure 1 is the first event and Table 9b shows the possibilities when Suspension 1 is the first event and Failure 1 is the second event.

Table 8 – Summary of Failures and Suspensions

Item Number	Status	Time to Event	Mean Rank Order

1	Suspension 1	550 cycles	----
2	Failure 1	727 cycles	1.20
3	Failure 2	1083 cycles	2.40
4	Suspension 2	1200 cycles	----
5	Failure 3	1452 cycles	4.22

Table 9a – Shows possibilities when F1 is the first event **Table 9b – F1 is second**

F1 event is in Position 1 (n = 1)									F1 in Position 2 (n= 2)	
1	2	3	4	5	6	7	8		1	2
F1	**F1**	**F1**	**F1**	**F1**	**F1**	**F1**	**F1**		S1	S1
S1	S1	F2	F2	F2	F2	F2	F2		**F1**	**F1**
F2	F2	S1	S1	S2	S2	F3	F3		F2	F2
S2	F3	S2	F3	S1	F3	S1	S2		S2	F3
F3	S2	F3	S2	F3	S1	S2	S1		F3	S2

Since there are eight ways with F1 in position one and two ways with F1 in position two. The Mean Rank Order Number (MRON) can be quickly calculated as:

$$MRON_1 = [(8)(1)+(2)(2)]/(8+2) = 12/10 = 1.20$$

Repeating this same calculation with the second failure provides the similar tables of possibilities as shown in 9c and 9d.

Table 9c – Shows possibilities when Fs is the second event **Table 9d – F2 is third event**

F2 event is in Position 2 (n = 2)							F2 event is in Position 3 (n = 3)			
F1	F1	F1	F1	F1	F1		S1	S1	F1	F1
F2	F2	F2	F2	F2	F2		F1	F1	S1	S1
S1	S2	S1	S2	F3	F3		F2	F2	F2	F2
S2	S1	F3	F3	S1	S2		S2	F3	S2	F3
F3	F3	S2	S1	S2	S1		F3	S2	F3	S2

Since there are six ways with F2 in position two and four ways with F2 in position three. The Mean Rank Order Number (MRON) can be quickly calculated as:

$$MRON_2 = [(6)(2)+(4)(3)]/(6+4) = 24/10 = 2.40$$

Repeating this same calculation with the third failure provides the same three tables of possibilities. The calculation of the Mean Rank Order becomes:

$$MRON_3 = [(2)(3)+(3)(4)+(4)(5)]/(2+3+4) = 38/9 = 4.22$$

These three MRON numbers would be used to calculate the Median Rank Percentages for locating the failures on the Weibull Graph. Figure 12 shows this situation on a probability plot, while Figure 13 shows a different set of suspended data from Table 10 plotted on the Weibull graph.

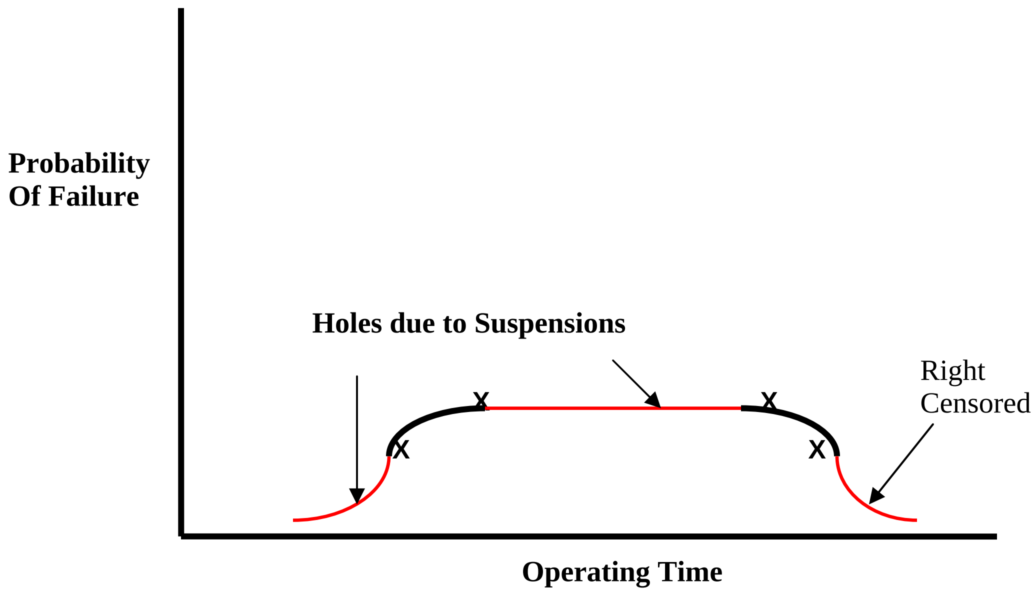


Figure 12- The Suspensions as Holes in the Probability Distribution

Table 10 shows a list of eight items in test. A suspension occurs at 10 hours, before the first failure at 31 hours. The failure at 31 hours and all later failures require a correction for the initial suspension. The first failure (at 31 hours) will be placed between 1 and 2 as a median rank number. After all, this first failure could have occurred at 11 hours (the lowest probability, or prior event) or as late as 44 hours (the next event in the table). The rank corrections continue to **add** as we move down Table 10 as you will see. This simple approach literally fills in the "holes" of the probability distribution that were caused by the presence of suspensions. Figure 12 shows how this works for the probability distribution. Table 10 shows the newly calculated median rank values and percentages. This process is performed by carefully counting down the "master table" and filling in the appropriate numbers in Equation 16 for each entry of Table 10. The detailed calculations for each suspension and failure are shown on the next page. Look at them carefully, as they are based upon simple counting schemes that correspond to the probability tables shown previously. This approach is actually easier then calculating all of the possibilities (as was shown in Tables 8 and 9), so this short hand approach is preferred.

New median rank numbers can be calculated from Equation 16, this is the **Rank Increment** formula. These increments are then employed to correct the initial table entries. The corrected entries are then used to calculate the new median rank percentages for plotting on the Weibull graph.

$$\textbf{Rank Increment} = \frac{(N+1) - \text{Pr}\,eviously\_Calculated\_Rank}{[1 + Number\_Beyond\_\text{Pr}\,ior\_Tabled\_Item]} \qquad \textbf{(16)}$$

Table 10 - Suspended Data Example

Initial Rank Order - Item	Time	Status	New Rank Numb.	New Median Rank
1	10 hrs.	suspension	---	---
2	31	failure	1.125	9.8%
3	45	suspension	---	---
4	49	failure	2.4375	25.4%
5	80	failure	3.75	41.1%
6	92	failure	5.0625	56.7%
7	103	failure	6.375	72.3%
8	110	suspension	---	---

$\beta = 2.74$, $\eta = 92.3$

Here N = 8 (exposures to test initially) and we must use the simple information in Table 10 to calculate the remaining correction numbers in order to update the table. The process begins by observing that the data in Table 10 shows suspensions at 10 hours and 45 hours. The first failure is at 31 hours, which is after the first suspension.

The first failure at 31 hours must be corrected since the first item in the table was a suspension at 10 hours. This is done by "counting out" the remainder of the table. Literally point to the item in question (here the first entry, a suspension) and then count to the end of the table. This gives seven as the number of items beyond the present suspension at 10 hours. The first increment becomes:

First New Rank Increment Correction = $\frac{(8+1)-0}{[1+7]}$ = 1.125

Likewise, the fourth item, or the second failure at 49 hours is also preceded by a suspension at 45 hours. Counting out the table from this point gives 5. Here, we have calculated a prior rank number of 1.125, so this is also entered into Equation 16. This second rank increment becomes:

Second Rank Increment Correction (fourth table entry) = $\frac{(8+1)-1.125}{[1+5]}$ = 1.3125

The **new corrected Median Rank of the second failure** is calculated by summing the two previously calculated corrections. This is:

Rank of second failure = 1.125 + 1.3125 = 2.4375.

The next failure at 80 hours is calculated similarly as:

Third Failure Rank Increment Correction = $\frac{(8+1)-2.4375}{[1+4]}$ = 1.3125

The **new Median Rank of the third failure** = 2.4375 + 1.3125 = 3.7500.

The last two failures are calculated in a similar fashion follows:

Fourth Rank Increment = $\frac{(8+1)-3.75}{[1+3]}$ = 1.3125, yielding 5.0625 as the Median Rank

Fifth Rank Increment = $\frac{(8+1)-5.0625}{[1+2]}$ = 1.3125, yielding 6.375 as the Median Rank

We can now plot all of the failure points on the Weibull graph as shown in Figure 13. The holes in the distribution are shown on the Weibull graph.

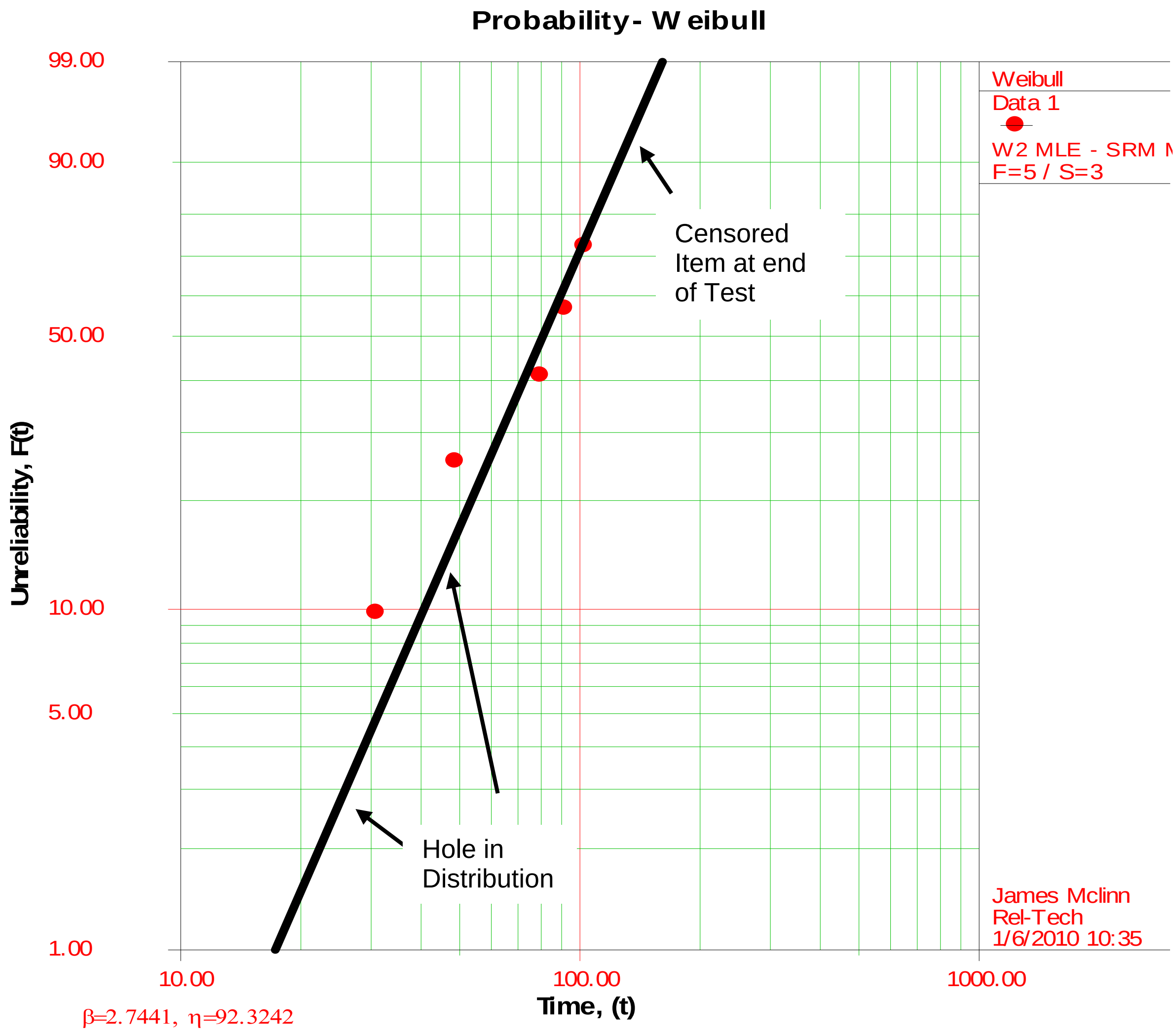


Figure 13 Showing a Distribution with Holes

2.4 - Unusual Distributions

There are some statistical distributions that do not appear as straight lines on the Weibull graph. The most common example of a distribution that is not straight is the Lognormal. Otherwise, most non-straight lines result from multiple failure modes (a mixture situation), a bimodal distribution, a Bathtub curve, different operating environments, mixed age components, or an S-Curve situation. The Lognormal distribution typically appears as two straight lines which merge **at about 50% cumulative** failures. The easiest way to test any data suspected of being Lognormal is to plot the data on Lognormal probability paper and determine if the data is a straight line. Most Weibull software packages will allow us to quickly do this and easily switch between the Weibull, Normal, Exponential and Lognormal distributions. It is often wise to try different data sets against the various distributions. If the first distribution does not seem to fit well, try another distribution. This is simple modeling of the data to a distribution. Getting a reasonable to good fit, one time, for a distribution does not prove the component, failure modes or system will always follow this distribution. Remember to use **all the data and information available** to select the best distribution for a given situation. This is a good balance of the use of statistics, engineering knowledge all combined with experience.

A Bimodal data set example is in Figure 14. No simple data transformation or switch to another distribution can straighten the data shown in the figure. The data was generated from a data set of the Time to Failure of RAMS. If various failure modes are identified, then individual modes may be separated and individually plotted on their own Weibull graphs. Plotting the data on the LogNormal or other distributions does not improve the fit.

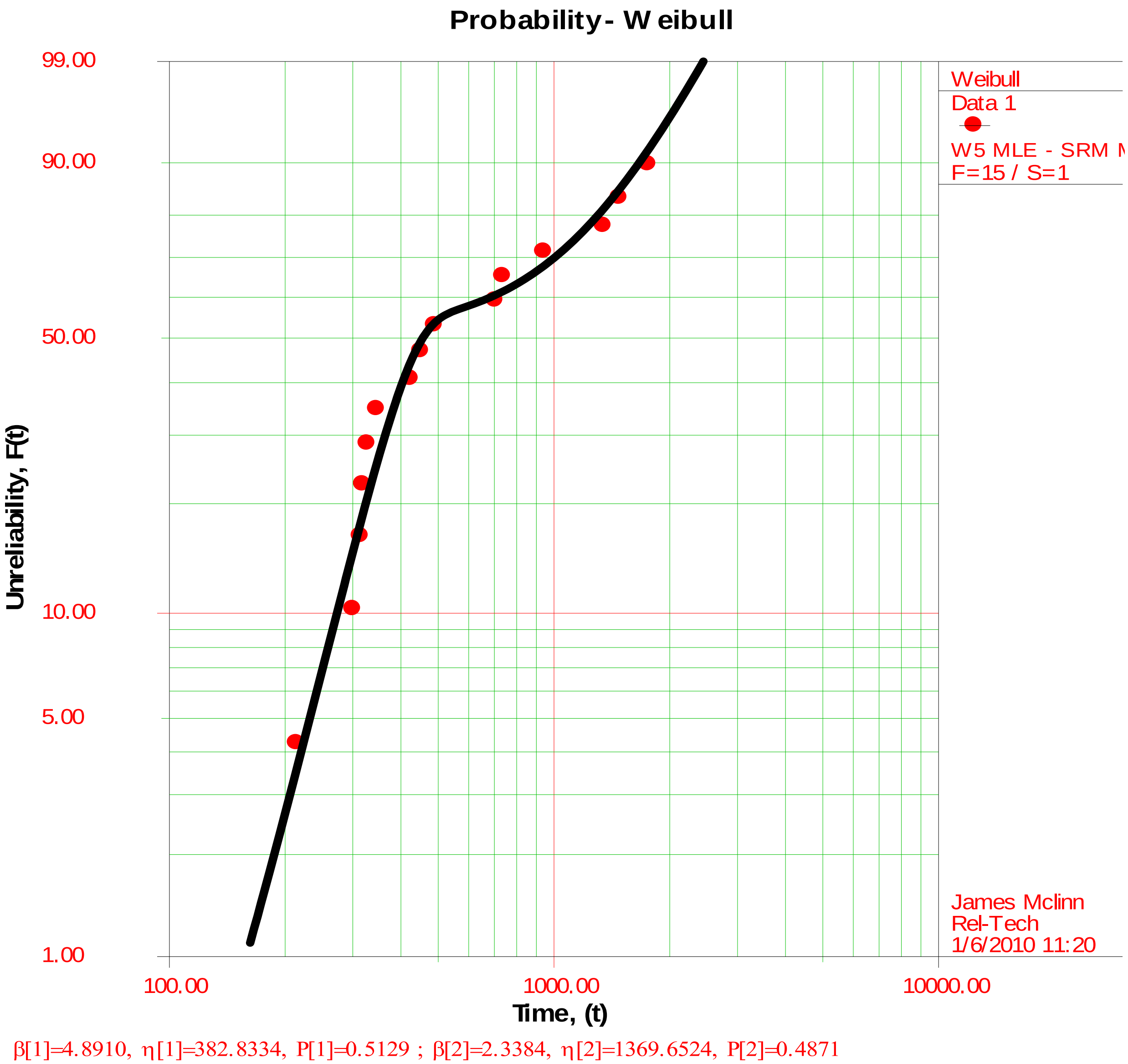


Figure 14 – Mixed Failure Modes from an ALT of RAMs

I have personally worked with a situation where one system data set actually contained three different hidden failure modes. Once the modes were identified and separated, the Noisy graph could be disassembled into three simple graphs. The first failure mode was actually a three parameter Weibull situation with a negative time offset. The second failure mode showed an S-Curve situation. The third mode was actually a simple straight line. This combination of three failure modes produced such a noisy data set that was easy to misinterpret what occurred if the failure mode information had not been included. How does one take apart these modes? Some competing failure modes analysis is required. Abernethy discusses this in great detail [1].

Example 2.1 - The following data set from Abernethy [1,p. 3-22] is left with the reader as an example of a curved Weibull. Be sure to try both the Rank Regression and Maximum Likelihood methods for data analysis while considering both 2 and 3 parameter analysis. You might also consider treating this as a LogNormal distribution. What is the best fit?

Times to failure are: 90, 130, 165, 220, 275, 370, 525, 1200 hours.
These results are interesting no matter which approach the analyst employs.

3.0 - Weibull Risk Forecasting

The Weibull function can be used to estimate numerous situations of risk. Here we use the distribution as a measure of the probability of future failure. Weibull risk forecasting can be combined with modern computer tools and probabilistic software programs for a variety of practical applications. Only a few of these will be covered in this book. The readers are recommended references [1, 4, 10, 16] for additional reading and other applications. The examples considered here include:

1. Making simple predictions about the "**Expected Number**" of failures that may occur in the future.
2. Looking at **Present Risk and/or Future Risk** through the use of the Weibull function.
3. Using **Monte Carlo Simulation combined with** the Weibull distribution.

Actually the Weibull Risk Analysis Method requires only a small amount of real information or data and with simple models this combination may be used to make predictions of the future. The existence of a valid risk model actually implies a lot of past information. Being able to manipulate data suggests this technique can be used to extend a small amount of existing data. This can be a valuable tool when filling in holes of a data set. As always, the more real and solid information, the better any risk analysis will turn out. More solid data usually permits more complex models to be created and possesses the ability for self-check of the model itself. We need not limit the models to simple functional relationships or familiar engineering equations. Monte Carlo Simulation may also be used to extend the data set or create a whole range of new data sets. This statistical approach has been a major help with developing some reliability topics such as confidence interval estimation when actual confidence distributions may not be known.

Table 11 - Information needed for Weibull Risk Analysis

1. **Weibull Failure Rate Distribution -** That is, an estimate of the Characteristic Life, η and the Weibull slope, β are needed to model many data situations.
2. **A Part or System Usage Rate per operating period -** This may be measured by the expected average usage per week, per month or per year as typically expressed in operating hours or cycles per a period of time. For example, each unit receives an average of 250 operating hours per calendar month or customer usage month, whichever is most appropriate. This implies detailed customer use distributions to create these types of models. See Weibull Easy™ for an example of canned software programs this includes this capability. Here, the estimate of the product reliability is combined with an estimate of a customer use distribution in order to project field failures.

3. **Introduction Rate of New Units into the field -** This number represents the known periodic introduction of new units into the field population. For example, 15 new units are shipped each month. One unit per month may be retired because of old age, by non-repairable failure or replacement with an alternate model. Then the expected field failure rate over the next period of time may then be estimated.

4. **The Age of Replacement Parts -** Since most parts do not carry a time clock, one cannot easily determine how many operating hours is on a part is just by observing it in a system. Therefore, knowing its age at the time the part was introduced into the field may be important for estimating the future of the field population.

5. The **Removal from Service** of systems that are removed for reasons other than failure. Examples include old age, lack of interest in using the system (obsolete unable to meet a customer need, but still functional), replacement by a new model, or sold to a new and undocumented user. Most of the time, these events remain unknown.

6. The **Wait Times** or **delays**, if any, to begin the use of the product. These may be storage times prior to introduction to a customer, warehouse times, **installation** time or **acceptance time** or other customer delays. Even though the systems are "in the field" they are not in active use during these periods or they are treated as not active.

7. **Maintenance times -** Estimates of the optimal period of time to perform preventative maintenance in order to minimize total cost or reduce the incidence of unexpected failure can be generated through Weibull risk analysis.

Entries 2, 3, 4 and 7 are pieces of information that may be said to describe or measure the **Field Exposure** of the population. Items 5 and 6 in Table 11 are corrections that are sometimes needed to obtain a more accurate operating picture of the exposed population.

In the simplest examples we could have N systems operating in the field, which are:

1. ***All of the same age***
2. **Operating the same average length of time** each time period (week, month or year)
3. Experiencing **no undocumented replacements or failures**
4. S**tatic** for all other operating conditions
5. Built from the **same manufacturing** population or are reasonably similar
6. Operate in the **same customer conditions** across a wide variety of customers and customer applications
7. Are **not subject to customer abuse** or unusual events such as ESD

3.1 Present Risk

The **Present Risk** is the probability that a failure will occur (or should have occurred) up to the present accrued operating time. This is the complement of reliability, so we can use F(t) to describe each of the N systems operating in the field. Should there be different operating times or customer use conditions, we usually include these entries separately unless an assumption about average operating time has being made. We have a simple measure of the Present Risk through the total probability for failure described as:

$$\textbf{Present Risk} = \sum_{j=1}^{N} F(t_j) \qquad \textbf{(17)}$$

Where **F**(t_j) represents the Weibull Cumulative Failure Distribution or the complement of Reliability. Each F(t) distribution will be calculated over the time frame and conditions in question.

Example 3.1 – A sample of new eight bearing system will be placed into the field. Prior history suggests that a good estimate of β is 2.4 and η is 8,800 hrs. When the eight new systems are placed in the field; they accumulate the total operating times as shown in Table 8. Based upon the operating conditions and similarity to the past, how many failures would we expect to have already occurred as a result of operating these systems in the field for the times shown? Note, since these are new systems, the failure probability begins at zero. If the eight systems had been in the field for some prior period, this would have to be known for each system in order to properly calculate the risk probabilities for each unit.

Table 12 - Present Risk Information

Number - N of systems at each age or time	**Age or Field operating time**	**F(t_j) Each Probability**	**N x F (t_j) Cum. Prob. at age**
1	960 hrs.	0.00489 X 1 =	0.00489
2	1270	0.00956 X 2 =	0.01911
1	1475	0.01366 X 1 =	0.01366
3	2350	0.04118 X 3 =	0.12355
1	3420	0.09832 X 1 =	0.09832
---	------		---------
8 units total	15,445 hrs. total		0.25953 cum. failure probability

That is, we would expect 0.25953 failures or 25.95% to have failed among the eight units in the field having operated for the field times shown. This type of fractional probability information can be used for planning for both **unexpected failures** and/or any necessary time dependent **maintenance activities**. If a different number of failures actually occurred over some longer period of time, we could test this actual outcome versus expected with a variety of goodness-of-fit tests. Individual field sites could be tested against each other if concerns about unusual customer results or failures exist. Example of a goodness-of-fit calculations are in references 1, 3 and 10. Many Weibull software programs may contain a variety of goodness-of-fit measures such as the Kolmogorov-Smirnov test, a Chi-square test or other similar tests. Since no real failures are expected by the example above, all failures are projected over the time frame in question.

3.2 Future Risk

The present data set does need not be the limit to accumulated operating times. We can actually look to the future for some period of time beyond the present. This time unit is labeled, **u**. Thus, any **Future Risk** calculation can be mathematically described as:

$$\textbf{Future Risk (FR)} = \sum_{j=1}^{N} \frac{F(t_j + u) - F(t_j)}{1 - F(t_j)} = \frac{Difference\_in\_Fail\_\mathrm{Pr}obability}{Normalizing\_Factor} \qquad \textbf{(18)}$$

The calculations of Future Risk may be done by hand, as shown here, or by the use of off-the-shelf software spreadsheet programs. Imagine we have a number of systems which are distributed primarily in two different locations of the United States. We will label then the Northeast (**NE**) and the Southwest (**SW**) to identify an apparent geographic-related performance difference. By analyzing the past history, we suspect that these same systems have different time-to-failure distributions and hence different reliability distributions. Prior field failure analysis has suggested two different Weibull distributions were present. For each region we have:

NE has β = 2.4 and η = 8,800 hrs. and **SW** has β = 1.6 and η = 11,500 hrs.

Example 3.2 - We wish to know the risk of failure that will occur over the **next month**, that is, during the next operating month of 250 hours customer hours (equal to **u**) will be accumulated on each system. Table 13 shows this data by location by location and by current operating age.

Table 13 - Future Risk Information
Current

Number	**Location**	**Age**	**Future Age**	**NxF(t_j + u)**	**N xF(t_j)**	**ΔFR $_j$**
1	NE	500	750 hrs.	0.002710.001020.00169		
2	NE	670	935	0.009190.004130.00506		
2	**SW**	**775**	**1045**	**0.042640.026540.01610**		
1	NE	910	1180	0.008020.004310.00371		
1	**SW**	**1110**	**1395**	**0.033640.023460.01018**		
1	**SW**	**1350**	**1600**	**0.042610.031940.01067**		
1	**SW**	**1630**	**1880**	**0.053660.042940.01072**		
2	NE	1835	2085	0.062130.045920.01621		
1	**SW**	**1990**	**2240**	**0.070390.058610.01178**		
1	NE	2115	2375	0.042220.032130.01009		
-----		------	------			---------
13 systems Total		16,165 hours	19,550 hours			0.09621 change

We can now calculate the probability of failure (a Future Risk, based upon **FR** $_j$) for each of the locations. We can do this first separately and then as a combined calculation. Totaling the numbers from the last column of Table 13 for each geographic group gives:

SW Future Risk = 0.05945 or a little less than 6% probability of failure among the six **SW** systems in the field over the next operating month. The average failure probability is just under 1% per 250 hour operating month.

NE Future Risk = 0.03676 or a little more than 3.56% probability of failure among the seven **NE** systems in the field over the next operating month. The average failure probability is just over 0.5% per 250 hour operating month.

Combining the two failure probabilities**,** we have the total field failure future risk for any system as:

Total Future Risk = 0.05945 + 0.03676 = 0.09621 or 9.62% probability of failure among the 13 systems in the field over the **next operating month of 250** hours. This represents a small fractional failure risk.

Example 3.3 - We could go back and recalculate the probability of failure over the next 6 months which consists of **1500 additional** hours of operation for the average system. This calculation assumes the use remains steady and equal for all of the 13 systems. No new systems are added to either geographic group and none are removed. Any maintenance activity occurs during the normal off time of the system and does not impact the failure probabilities. Said another way, any maintenance or repairs of failures leaves the system as it was. A summary of the calculations are presented below. Figure 15 shows the two different regional distributions as they would exist on a probability graph. The results of the calculation for the six months are:

SW Future Risk = 0.40658 = 40.66% probability of failure of one of the **SW** systems.
NE Future Risk = 0.35895 = 35.90% probability of failure of one of the **NE** systems.

The probability of a failure among any of the systems is 76.55% during this future time period. We can also perform similar calculations when maintenance activities are included in the possibilities.

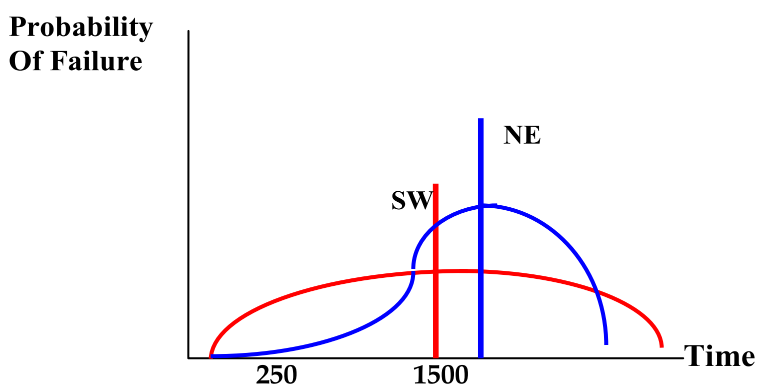


Figure 15 - A Comparison of the Distributions

3.3 Weibayes

This term is often applied to the use of Weibull Analysis in special reliability situations. Three common applications are:

1.) No Failures of the systems occur during the test period.

2.) Few Failures and few samples, when a good (often historical) β estimate is known.

3.) When prior information is known about a system. This involves special Bayes models.

The characteristic life at a given confidence according to Abernethy [1, page 5.2] is:

$$\eta = \left[\sum_{j=1}^{N} \left(\frac{(t_j)^{\beta}}{r} \right) \right]^{\frac{1}{\beta}} \qquad \textbf{(19)}$$

Where t_j is the operating time of each unit, component or system on test

N - is the total number of systems operating (both failures and non-failures)

η = characteristic life

r - measures the confidence desired and varies with failure versus no-failure test situations. It is expressed as a decimal, see Table 14 below.

β is an estimate of the slope of the Weibull graph. It often measures experiences of past similar designs and/or systems. This is the Bayes Contribution. Thus, β represents a best first estimate to the Weibull shape parameter.

The following two equations, 20 and 21 relate confidence, **C,** to the value of **r**. These cover two common situations encountered in reliability.

For the **no-failures** situation we have the equation that relates confidence and **r**:

$$- \ln(1 - C) = r \qquad (20)$$

The factor, **r,** depends directly upon the confidence as shown in Table 14.

When the Weibull data contains **failures**, Nelson [11] has suggested the value of **r** in Equation 20 be replaced by the following formula more complex formula:

$$r = \frac{1}{2} X^2_{(1-C,2f+2)} \qquad (21)$$

This number may be evaluated by any standard Chi-square table for the appropriate confidence, C and degrees of freedom. The two choices for degrees of freedom are ***2f*** for a failure-ended test and ***2f +2*** degrees of freedom for a time-ended test. Here, ***f*** represents the number of failures observed during test.

Table 14 - The Value of r as a Function of Confidence for Non-Failure Situations

Confidence	50%	60%	63.2%	70%	80%	90%	95%	99%	99.9%
Value of r	0.693	0.916	1.00	1.204	1.609	2.303	2.996	4.605	6.908

Example 3.4 - Five air conditioning systems are tested to estimate the life of an improved design. All of these five systems have operated failure-free to 870, 1230, 1550, 1625 and 1710 hours respectively. Past experiences of similar designs has provided a value of β = 1.3 is most likely. What is the best "estimate of life", at 95% confidence, at 80% confidence and at 60% confidence for the improved design?

Using Equation 21 we can sum up the test data at 95% confidence as:

$$\eta_{0.95} = \left[\frac{(870)^{1.3} + (1230)^{1.3} + (1550)^{1.3} + (1625)^{1.3} + (1710)^{1.3}}{2.996} \right]^{\frac{1}{1.3}} \tag{22}$$

$\eta_{0.95}$ = 2087.5 hrs at 95% confidence

This number represents a lower limit of demonstrated characteristic life based upon this failure free data and the prior information. Similar calculations at the **other confidence** values yield:

$\eta_{0.80}$ = 3366.7 hrs at 80% confidence demonstrated so far

and $\eta_{0.60}$ = 5192.8 hrs at 60% confidence

We see that the less confident we are about the value of η, the larger the estimated value becomes.

Having **a good prior estimate of β** is a great aid for accurately calculating η in Example 3.4. This piece of information, if accurate, is more valuable than having ten test points to estimate β because it is based on historical use data of typically more than 10 samples. One downside of the use of historical data is that the value of β may have changed with recent design improvements or customer applications. Example 3.7 shows one method to calculate the possible range of β when the data that generated the initial value of β is itself uncertain. **Monte Carlo Simulation** is a second method to make such estimates.

Example 3.5 - Imagine that we determine that we can safely accelerate the test of Example 3.4 by 20%. This acceleration of 1.20 allows us to run **only three systems in the accelerated life test**. The following new test times allow us to create a new estimate η. This is:

New Sample 1 - operated failure free to 1230 accelerated hours or 1476 "normal use" hours
New Sample 2 - operated failure free to 1550 accelerated hours or 1860 "normal use" hours
New Sample 3 - operated failure free 1710 accelerated hours or 2052 "normal use" hours

The new estimate of characteristic life at 95% confidence will be evaluated as follows:

$$\eta = \left[\frac{(1476)^{1.3} + (1860)^{1.3} + (2052)^{1.3}}{2.996} \right]^{\frac{1}{1.3}}$$

η = 1802.7 hours at 95% confidence

This value is lower than the prior calculation because fewer data points (only 3 versus 5) and lower total test time was provided by the accelerated test. The use of this approach for Weibull confidence normally provides a smaller confidence band for η than many other common methods [1, 13]. Other confidence approaches are described in the references [10, 11, 21].

Example 3.6 - A life test of a mechanical assembly generated the following data set. What is an estimate of the characteristic life at 99% confidence for the five samples? Failed units are not replaced or repaired. Failures occurred at 250 and 592 hours. Three remaining units operated failure-free to 564, 662 and 907 hours respectively. Past history suggests a best estimate value of β of 1.35. The current data (these 5 points) is not really sufficient to create a better estimate of β, so the prior information will be employed.

$$\eta_{0.99} = [\frac{(250)^{1.85} + (592)^{1.85} + (564)^{1.85} + (662)^{1.85} + (907)^{1.85}}{8.406}]^{\frac{1}{1.85}}$$

$$\eta_{0.99} = [88796.7]^{0.54054} = 472.9 \text{ hours}$$

The **Extreme Value** Distribution can also be used as a basis of estimating the Weibull function confidence interval [1, 10]. Two formulas for estimating the upper and lower confidence limits are shown below. Only the formulas for limits on β are shown here; similar formulas exist for upper and lower estimates on η.

$$\beta_L = \hat{\beta} \; [\frac{X^2_{[1-\gamma, c(n-1)]}}{cn}]^{\frac{1}{1+p^2}} \qquad \beta_U = \hat{\beta} \; [\frac{X^2_{[\gamma, c(n-1)]}}{cn}]^{\frac{1}{1+p^2}} \qquad \textbf{(23)}$$

where $\hat{\beta}$ is the best estimate of β as derived from prior information. This formula may be very important, because it may help reduce the uncertainty (confidence interval) of a result. In some ways these formulas are like Bayesian analysis by multiplying two functions to obtain a best estimate. The definitions for this extreme value formula are:

- **p** - is the fraction of the total that are operated to failure
- **c** - is a constant based on **p** and is a scale factor for probability shown in Table 15
- **n** - is total number operated to failure
- γ - measures the desired confidence

Note, that degrees of freedom for the extreme value distribution is **c**(**n**-1), not the usual value. Thus, it is possible to have non-integer values that require extrapolation in Chi-Square tables.

Table 15 - The values of c as a function of p

p =	1.0	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
c =	0.82	0.88	1.00	1.15	1.31	1.49	1.67	1.83	1.95	2.01

Example 3.7 - There are 10 systems on a life test that was stopped after only four failures. What are the best estimates for β and η? Estimate the upper and lower limits on β as well. The times to failure were observed as 387, 654, 817 and 1146 operating cycles and the test was stopped at 1200 cycles. We calculate that $\hat{\beta} = 2.089$ and $\hat{\eta} = 1645.2$ cycles as rough estimates, based upon the MLE method for ten samples and four failures. We have estimates for the upper and lower value of the slope, β from Equation 23.

$$\beta_U = \hat{\beta}\left[\frac{X^2_{[0.1,5.00)]}}{1.667(4)}\right]^{0.8621} = (2.089)\left[\frac{9.236}{6.667}\right]^{0.8621} = 2.767 \sim 2.77$$

$$\beta_L = \hat{\beta}\left[\frac{X^2_{[0.9,5.00]}}{1.667(4)}\right]^{0.8621} = (2.089)\left[\frac{1.610}{6.667}\right]^{0.8621} = 0.6137 \sim 0.62$$

with degrees of freedom being (1.67)(3) = 5.00

These are improved estimates representing the upper and lower 90% limits (i.e. 80% confidence interval) values of β.

Next, we must ask whether this range make sense? A typical value of β of 2.089 suggests a situation of a slowly increasing failure rate; that is, some **wear or degradation seems to be present** over the range of knowledge. The lower 90% limit estimate on β allows for the possibility of a broad range of decreasing failure rate possibilities. That is, the number of possibilities (about 18%) of test outcomes might yield a value of β less than 1.00. How do we resolve this all too common dilemma? We need more information in order to make a better judgment or narrow the range of β. There are several possible sources of this additional information. They include:

1. Look at the failure modes of the present four failures. Are they all the same or are some different for the failed systems?

2. Run the test longer to see if the trend toward a slow wear and/or other failure modes trends provide a trend in β for the system.

3. Consider past history of prior tests for this system, if it exists. Look at failure modes, β and the variation that led to the uncertainty about the mean value of β. Perhaps we can have more

data and then create an improved estimate of the limits. Remember, it is possible that another distribution such as the Normal may be the most appropriate one to describe the variability in β as determined by past history.

4. Consider past information about similar designs. This information is similar to item 3 but focuses upon similarity of design, while item 3 assumed the same design operated in similar test conditions.

5. Consider other sources of data of similar designs or similar customer situations such as data derived from field experience. This is less likely to be as close as desired, but can be the subject of meaningful studies.

Each of these possibilities may provide additional information that may allow us to make better judgments about the true range of, or limits on β. Classical statistic tools do not provide much more help in this situation, so no additional statistic tests or methods are suggested. The main problem is the small amount of information concerning failures. This dilemma is not an uncommon situation for people who run tests, especially during product development. Small sample size, few failures and short test length create a risk of not being sure about the results. We should supplement the current test with more information from some source before moving ahead. This situation is aggravated when the number of permitted failures in test is also small. The worst condition, of course, is a test with no failures.

Example 3.8 - Imagine we decide to continue the test of Example 3.7 by running it to a total of 6 failures. The fifth and sixth failures occur at 1402 and 1719 cycles respectively and the test is now stopped at 1800 cycles. These test points combined with the first four give us improved estimates on β as well as improved upper and lower limits of β. We have for this improved data set:

$$\hat{\beta} = 1.817 \text{ with } \qquad \hat{\eta} = 1873.7 \text{ cycles}$$

The new limits yield:

$$\beta_L = \hat{\beta}\left[\frac{X^2_{[0.9,6.55]}}{1.31(6)}\right]^{0.7353} = (1.817)\left[\frac{2.55}{7.86}\right]^{0.7353} = (1.817)(0.437) = 0.7939 \sim 0.79$$

$$\beta_U = \hat{\beta}\left[\frac{X^2_{[0.1,6.55)]}}{1.31(6)}\right]^{0.7353} = (1.817)\left[\frac{11.40}{7.86}\right]^{0.7353} = 2.388 \sim 2.39$$

from Equation 23 at 90% confidence, with degrees of freedom being (1.31)(5) = 6.55

What did we learn from these additional 600 cycles of test time and the two additional failures? We narrowed the range of the $\hat{\beta}$ uncertainty slightly from (2.77-0.62) to (2.39-0.79) while the initial estimate of $\hat{\beta}$ range declined from 2.09 to 1.82. The estimates of $\hat{\eta}$ increased from 1645 to 1873. These updates could be significant when projecting field failures or similar results.

Consider what would happen if we required 95% confidence or 99% confidence in Example 3.8. The range of values of β would also increase. It is left to the reader to demonstrate that at 95% confidence the range of β is (2.69-0.65) and at 99% the range of β is (3.30-0.42).

Imagine for a minute we could perform this test a number of times and so get a better estimate on the nominal value of β as well as the upper and lower limit estimates at 90% confidence. It is costly to perform such laboratory testing and does take time and extra manpower. There is an alternate way; this is through the use of a **Monte Carlo Simulation** program on a computer. This is one quick way to "improve what we know" by estimation rather than test. This statistical method can quickly run hundreds of simulated tests and provides an alternative to estimating confidence limits. Many canned Weibull programs contain this as a regular part of the program.

4.0 - Applications of the Weibull Distribution

Example 3.8 raised several issues concerning the best ways to extend a test in order to obtain more information. A variety of methods exist to extend data or fill in holes in data sets. We will look at Monte Carlo first. First create a "series of trials" with the information from Example 3.8 as the initial values. This series of trials would then be employed to determine upper and lower estimates of the values of β and η. Both are critical parameters for making predictions or estimating life. Standard Weibull software was used for creating this series of trials. The information from one set of trials is listed below. This series creates a better idea for what is a reasonable **range** for β and η and could serve as a means to estimate other parameters of the Weibull function. Otherwise, we would have used a statistical method for estimating confidence limits as was done in section 3.8.

Monte Carlo depends on a random number generator to help generate simulated life data from which statistical information might be generated. Therefore, the Monte Carlo approach might be compared to another approach called the "Bootstrap Method". This method is sometimes employed to extend the range of a small data set. Often this is the situation in reliability, few samples and even fewer data points (failures). See *Quality Progress*, June 1994, the Statistics Corner, for details on the Bootstrap Method. It is also possible to use Excel™ to generate random numbers and this method could also generate a data set. Dodson shows a variety of ways to generate confidence limits for Weibull [4] as does Abernethy [1] and Leemis [23].

The Monte Carlo data set was generated that was based upon Example 3.8. It is shown in Table 16 and can be improved by increasing the sample size and increasing the replicates for more precision. Here, eleven samples of 100 replicates each were generated. From these, the value of β and η were calculated for each sample. I limited the samples and replicates for simplicity. One Thousand samples of 1000 replicates each would create greater confidence in the results. Both Abernethy and Dodson identify situations where a million replicates are generated. The following is a brief example of how Monte Carlo can be used to supplement existing data. The seed numbers (from example 3.8) for the calculation of Monte Carlo were **η = 1873.7** and **β = 1.82**. Table 16 shows some of the Monte Carlo results obtained **as if we had rerun the same physical test 11 more times**. This data can be interpreted to reflect the scatter about the initial seed values of 1873.7 and 1.82 and employed to calculate new mean values for β and η as well as confidence limits. The Monte Carlo Method would suggest a more tightened confidence interval about the original estimation of the Weibull parameters than was shown in Example 3.8. Be sure to note the conclusions of the next example were drawn with a small sample size and a small number of replicates. Larger samples and replicates would likely narrow the estimates further.

Example 4.1 - Look at the range of values of β as shown in Table 16. They run from a low estimate of 1.609 to a high of 1.927 with an "average value" of 1.812. The value of the characteristic life has a similar situation with a range from 1761.8 to 2068.35 with an "average value" of 1901.31. This Monte Carlo range for confidence limits of β and η are very different from the values calculated in Examples 3.7 and 3.8. For better estimates we could repeat the Monte Carlo trials or even vary some of the initial values to look at the sensitivity of these numbers to the final results. Table 16 represents the summary of the eleven trials of 100 data points each. Since we can calculate the average of the values of β and η, we can treat these eleven numbers generated by Monte Carlo as if they are Normally distributed through the central limit theorem. Thus, the upper and lower estimates are at plus and minus three standard deviations.

Table 16 - Results of Monte Carlo Simulations at 90% Confidence

Trial	**η**	**β**	**Samples in Trial**
1	1761.8	1.866	100 points generated
2	1812.011.894	100 "	
3	1888.0	1.609	100 "
4	2049.181.875	100 "	
5	2047.731.927	100 "	
6	1818.081.725	100 "	
7	1788.111.899	100 "	
8	2068.351.693	100 "	
9	2022.861.883	100 "	
10	1868.641.782	100 "	
11	1789.631.780	100 "	
Average	**1901.311.812**		
Standard Dev.	**± 121.19**	**± 0.102**	
Upper Estimate at +3σ2264.882.118			
Lower Estimate at -3σ	**1537.741.506**		

Now calculating the 90% confidence we have as estimates for β and η:

β_L = 1.812 - 1.645 (0.102) = 1.644 ~ 1.64

β_U = 1.812 + 1.645 (0.102) = 1.97979 ~ 1.980

$\eta_{L.}$ = 1901.31 - 1.645 (121.19) = 1701.95 ~ 1702

$\eta_{U.}$ = 1901.31 + 1.645 (121.19) = 2100.67 ~ 2100.7

Table 17 - Comparison of the 90% Confidence Values of β

	Example 3.8 (10 points) total calculated limits	Table 16 Calculation (100 Points) eleven times calculated limits at 90%
β_L	0.6473	1.644
β_U	2.292	1.980

Note, the raw data used for generating entries in Table 16 could also have been used to generate quick 90% confidence estimates by looking at the 11 data sets of β and finding the 10th. and 90 th. percent points in each data set. These rough estimates (1788 and 2049 for η and 1.725 and 1.899 for β) are a lot quicker. The reader is referred to the ***New Weibull Handbook*** [1] and Dodson [4] for additional examples of calculating confidence limits with Monte Carlo methods. Monte Carlo can be a powerful tool for supplementing small data sets, filling holes and finding ways to extend data. Abernethy shows a method to unbias small data sets that he calls the Dauser shift and also uses the Reduced Bias Adjustment (RBA) for Normal and LogNormal distributions to unbias the calculation for the standard deviation [1] when using the MLE. This RBA may be expressed as:

$$\text{RBA}_{\sigma} = \frac{\sqrt{N(N-1)}}{C_4} \quad \text{where } C_4 = \left[\sqrt{\frac{2}{n-1}}\right]\frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)!} \qquad \textbf{(24)}$$

C_4 is the standard quality correction factor that relates σ and $\bar{s}$ (see Juran's Handbook ref [26]) for tables of this factor. Here N is the total number on test and n is the number of failures.
The correction to the standard deviation for Normal and LogNormal is expressed as

$$\sigma_{unb} = (\sigma)RBA_{\sigma}$$

For the Weibull distribution there is a correction for β when using the MLE. This is expressed as:

$$\beta_{unb} = (\hat{\beta})\mathbf{RBA}_{\beta} = (\hat{\beta})(C_4)^{3.5} \qquad \textbf{(25)}$$

If we had 10 units on test with 6 observed failures the correction factor C_4 would become

$$C_4 = [\sqrt{\frac{2}{n-1}}]\frac{(\frac{n-2}{2})!}{(\frac{n-3}{2})!} = [\sqrt{\frac{2}{5}}]\frac{(\frac{6-2}{2})!}{(\frac{6-3}{2})!} = [\sqrt{\frac{2}{5}}]\frac{(2)!}{(\frac{3}{2})!} = (0.623456)[(2)!/(3/2)!]$$

Non-Integer Factorials can be evaluated by

$(n)! = \Gamma(n+1)$ so $(3/2)! = \Gamma(3/2+1) = 3/2\ \Gamma(1.5) = (1.5)(0.88623) = 1.3293$

$C_4 = 1.2649/1.32935 = 0.95153$

And in the case of Weibull slope β we would have $(C_4)^{3.5} = 0.84038$ as the correction factor.

4.1 - Weibull Analysis Applications for Maintenance Issues

Maintenance issues range across almost every aspect of life. We perform maintenance upon automobiles and other personal property as well as houses. Institutions perform maintenance upon roads and bridges, airplanes, trains, buses, software, military equipment and all power generation facilities. Any product or service that is expected to operate for more than a few years, failure free, usually requires maintenance to ensure reliable operation. The Weibull function can be used to support common maintenance issues. Most Weibull approaches are similar to that presented in the section under risk forecasting. It might be helpful to briefly review that section before reading further. In order to use risk methods for maintenance, we only need a small amount of information. As always, the more information that is available, the better the outcome will be. The first maintenance example we will include the impact of a planned field replacement, prior to failure and briefly look at the cost tradeoff of field failures versus planned replacements.

Table 18 - Information needed for Weibull Maintenance Analysis

1. **Weibull Failure Rate Distribution -** That is, the Characteristic Life, η, and Weibull slope, β, are evaluated to model most situations. Identify whether an offset, γ, exists and relate it to some real physical mechanism or failure mode if it does exist.

2. **A Part or System Exposure Rate per unit time -** The expected average exposure or called usage per month or usage per year are required for modeling the exposure rate. A reasonable estimate is all that is needed for modeling, for example knowing the average use could be improved by knowing the distribution of usage across the population. The better the use estimate, the better the model and the reliability results will become. Weibull-Ease [24] is a program that allows one to model this very well.

3. **Introduction Rate of New Units into the Field -** This number represents the known periodic introduction of new units into the field population. It is usually the monthly ship number, but may include an estimate of delays if shipments go to a distributor or to storage before release into the market. The better the estimate, then the better the model and results may become. Note, this might require a roll-out failure mode in order to get the data analysis correct [25].

4. **The Age of Replacement Parts -** Since most parts do not carry a time clock, we cannot easily tell how old a part is just by observing it in a system. Therefore, knowing its age at the time a part is introduced into the field is important to describing the future history of the field population. This number may reflect the screening in-house or the time in storage before customer use began. A good estimate of the population distribution is required.

In the simplest possible example, all N systems in the field ***are of the same age***, with no replacement components or failures having occurred since the start of all N systems. We are then ignoring any installation problems identified or other additional reliability information that may have been produced before the systems began operation at the customer site. This additional information can include inconsistencies in installation quality, problems with interface to customer equipment, any incomplete repairs or work-around solutions that may have been used previously. These problems will have a variety of effects that should be included in any warranty analysis or warranty costing models. Initially, we will ignore these issues and treat them as if they have less impact then the planned and scheduled maintenance actions or the unplanned repair actions. Installation repairs and all customer set-up quality issues would ordinarily be treated as part of the field failures problems. Often, companies look at an installation metric, as well as the failures occurring in the first 7 days or the first 30 days of operation. These metrics typically measure early life failures or infant mortality. For many companies, failures occurring after 30 days of successful operation are treated as system level failure that could be contained or prevented by maintenance actions or other proactive activities.

Example 4.2 - A voltage regulator system already in the field with prior history exists has 2780 total systems in the field spread over 29 years prior years of operation. Comparison of the new updated design versus any older design would be valid and relevant analysis. The history of the older design as well as the new design is summarized in Table 19. Only some data is presented from the original example which was taken from the Abernethy [1] and was modified for clarity of the example. Fifty-three field failures were noted over the 29 years and 101 suspensions (removals from the field for no-failure reason) occurred as summarized in Table 20. A two parameter Weibull analysis (Fig 16) yielded a β = 9.04 and η = 31.21 years based upon MLE from the complete data shown in the two tables. Analysis is discussed below.

Table 19 - Voltage Regulator System Example

Age in years	Fails in year	Suspensions in year	Systems	Approximate Reliability at given age
1 yrs.	0	0	160	1.000000
2	0	0	160	1.000000
3	0	0	160	1.000000
4	0	0	140	0.999999
5	0	0	140	0.999998
6	0	0	140	0.999992
7	0	0	120	0.999978
8	0	0	120	0.999948
9	0	2	120	0.999889
10	0	0	100	0.999780
11	1	0	100	0.999592
12	0	2	100	0.999282
13	0	2	100	0.998794
14	1	4	100	0.998049
15	0	4	100	0.996949
16	0	4	100	0.995366
17	2	5	100	0.993139
18	4	5	100	0.990074
19	2	6	100	0.985931
20	3	6	100	0.980428
21	5	6	80	0.973235
22	6	7	80	0.963974
23	3	8	60	0.952218
24	2	8	60	0.937501
25	4	9	40	0.919326
26	2	9	40	0.897182
27	15	6	20	0.870569
28	2	4	20	0.839032
29	1	4	20	0.802202
	----	----	----	
Total	53	101	2780	

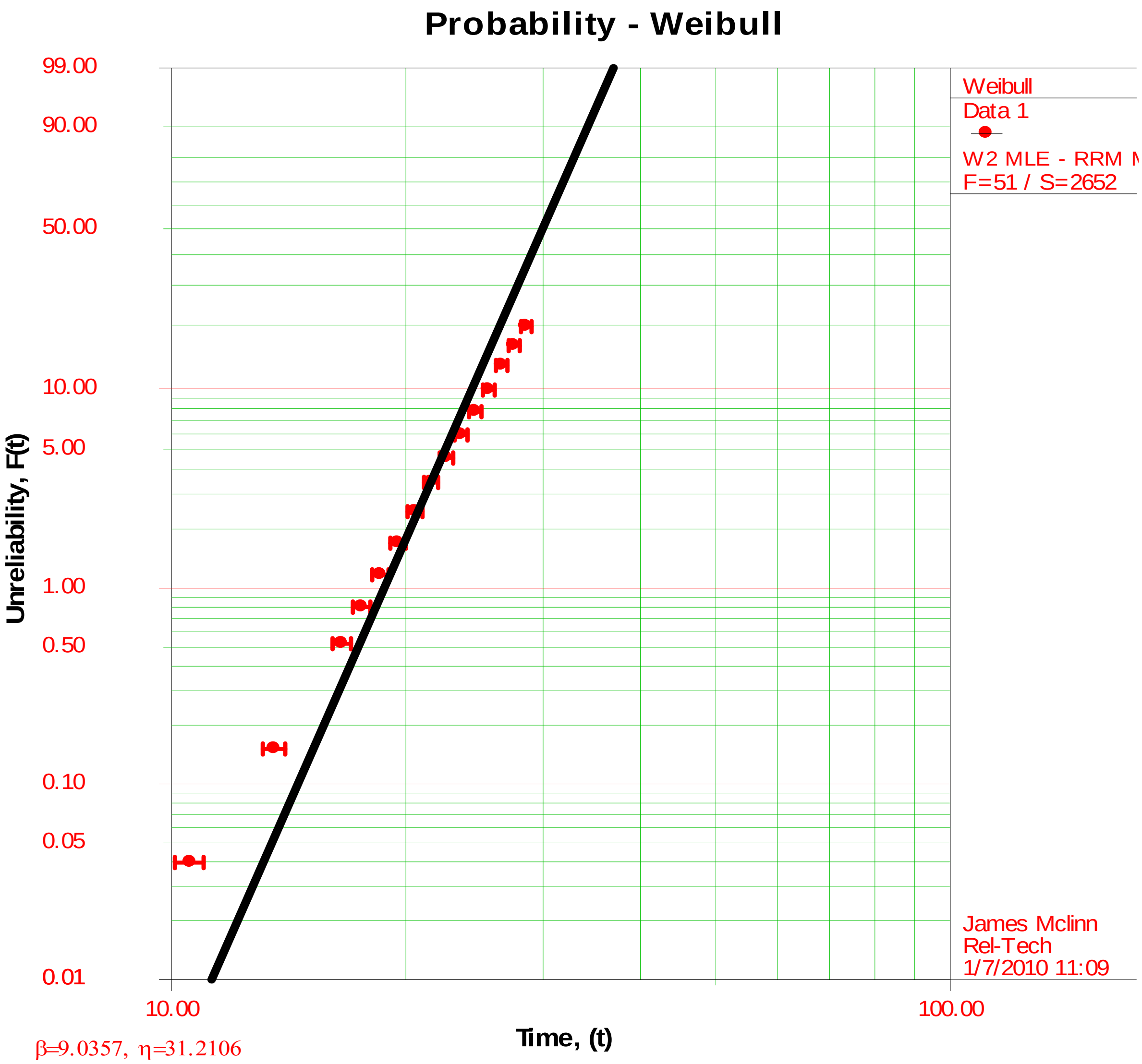


Fig. 16 – Voltage Regulator Field Data

In addition to the failure information shown in Table 19, which is information about the 2780 systems operating in the field, there were 53 field failures. Table 20 reflects detailed data about 101 known suspensions from the data set of Table 19, but this is incomplete suspension information. We will treat this additional data as a histogram of suspensions spread over time. Table 20 information will be added to the Weibull reliability data of Table 19 for the creation of Figure 16. It is not unusual to have units removed from operation by the customer and it not be known to the manufacturer. The suspension data is treated as "interval type" data for purposes of analysis. This simplifies the data entry task, and spreads the suspensions over the time interval in question, since it is not known exactly when the suspensions actually occurred. Units that fail in the field are repaired or replaced quickly, so this need not be considered.

Table 20 - Suspensions of Volt Regulator Field Data	
Number of units suspended	Age passed without failure before suspension
40 units	In first 20 years, these were removed from operation
61 units	21 to 29 years were removed without failure

Figure 16 shows the interval data. That is, the failures are spread over the year, since the exact time of failure is unknown. This interval is shown by the wings on each data point. The best doesn't go through most of the interval data points because the vast majority of points represent unfailed units. Only the failed units are plotted with the MLE, but the line is influenced by the unfailed units.

The Addition of Cost to the Model

Now that there is a model for time to failure, one can include costs of failure and perform the trade-off with planned maintenance. It is possible to identify an optimum policy for planned maintenance (or renewal or refurbishment as your own situations suggests) that minimizes the total cost. Table 21 reflects some of the cost issues associated with the field failure model generated from the data shown in Figure 16. A cost of **$1500** has been established for each field failure and a cost of **$250** for each planned replacement. These costs are low and reflect local repair and replacement facilities such as would occur with a product widely distributed through-out the United States. If there were a single manufacturing facility in the U.S. with only a few repair depots, these failure costs might increase to about $2500 and planned repair costs increase to $750 when based out of a single repair. Table 21 shows annual estimates the cost for failure combined with repair in every year out to 29 years. The annualized (planned) replacement cost is the total cost to replace all remaining operating units if initiated during the year reflected. The cost is annualized over the number of years since the last replacement. It reflects a planned approach to reduce the replacement cost to such a low level.

The projected annualized replacement cost for the 11 th. year entry reflects the following information. If all of the good remaining systems were replaced that year, the cost reflected as an annualized replacement cost would be:

$$\text{Annualized replacement cost} = \frac{(2777.8)(\$250)}{11 years} = \$\ 63131.82 \text{ per year}$$

This Weibull approach allows us to estimate an optimum replacement scheme and balance it against the planned repair, fixed customer costs and maintenance costs. Based upon this approach, we select the 18 th year as the minimum total cost, here $42973.02, to combine repair, fixed costs and replacement costs of the field units. The reliability at this point is 0.993117, which is usually very acceptable. We would expect only about 19 failures to have occurred at this point in time, out of 2780 systems.

Table 21 - Cost Calculation to Optimize Replacements

Year	Expected Fails in time	Fail Cost	Planned Repair Cost	Annualized Total Cost
0 to 5 Years	0.00018	$0.16	$139,000	$139,000.16
6 years	0.00093	$0.70	$115,833.33	$115,834.03
7 "	0.00376	$2.42	$99285.71	$99288.13
8 "	0.01258	$7.07	$86875.00	$86882.07
9 "	0.03647	$18.24	$77222.22	$77240.46
10 "	0.09447	$42.51	$69450.00	$69492.51
11 "	0.22361	$91.48	$63136.36	$63227.84
12 "	0.49082	$184.06	$57854.17	$58038.22
13 "	1.01115	$350.01	$53365.38	$53715.40
14 "	1.97414	$634.54	$49517.86	$50152.40
15 "	3.67559	$1102.68	$46133.33	$47236.01
16 "	6.57421	$1849.02	$43187.50	$45036.52
17 "	11.3464	$3003.45	$40588.24	$43591.69
18 "	18.9476	$4736.91	$38236.11	**$42973.02**
19 "	30.7226	$7276.41	$36105.26	$43381.67
20 "	48.5434	$10922.26	$34200.00	$45122.26
21 "	74.8348	$16036.03	$32464.29	$48500.31
22 "	112.680	$23048.17	$30863.64	$53911.81
23 "	165.871	$32453.05	$29380.43	$61833.48
24 "	239.178	$44845.87	$28041.67	$72877.54
25 "	337.696	$60785.29	$26820.00	$87605.29
26 "	465.908	$80637.86	$25663.46	$106301.32
27 "	628.621	$104770.12	$24611.11	$129381.23
28 "	824.402	$132493.13	$23544.64	$156037.77
29 "	1018.466	$158037.83	$21818.97	$179856.79

We can calculate the impact of selecting this optimum time for replacement versus allowing failures combined with any customer reliability requirements for the systems in the field. It is possible to add additional items in this cost calculation to reflect different costs at different companies. These would include the effects of a regular inflation or the present value of money to better identify the true costs of field failures. This balance point between high reliability and total cost often has a much higher reliability than most people believe. Tradition quality theories have suggested that the optimum field reliability is in the range of 0.9 to 0.95. In the 1990s, we have learned the optimum cost effective reliability is often greater than 0.95 [26]. This reflects the fact that direct costs and hidden costs are actually much higher then once believed. The hidden costs often appear as the costs of quality, but are more properly termed the "costs of poor quality" or the "costs of failure". In addition, studies about lost business have increased the costs to retain customers or find customers to replace those lost through low quality and reliability.

The biggest issue for Weibull Analysis is often the assumptions that must be made to cover the unknowns of the true field situation of the customer. These may strongly impact the outcome and influence the calculation of characteristic life and shape parameter. Had we changed the initial information and underlying model, **including the distribution of the suspensions**, then the results of Table 21 could significantly change. Just change the β value from 9.04 to another number such as 7.0 or 6.0 and the optimum time to replacement and total costs may become very different. These types of problems can be easily done with canned software [1] and/or the use of Excel for modeling.

4.2 - Selection of "the Best" Statistical Distribution

The selection of a distribution to model the time-to-failure distribution is less well understood because there are no clear ground rules to aid people in the many situations that may be encountered. The following will shed some light upon a proposed set of ground rules. The following examples may not cover all situations that you may encounter. Typically, three distributions are most often used when analyzing the time-to-failure (continuous) distribution; these are the **Weibull, Normal and Lognormal**. To a lesser extent the Extreme Value Distribution and Exponential distribution are sometimes employed in data analysis. The Exponential distribution can't adequately describe an improvement component or a wear dominated system. I will briefly describe each distribution below and suggest some areas and examples where they might naturally appear.

Weibull - This is a general 2 or 3 parameter distribution that often describes situations where there are multiple avenues or modes of failure. This is sometimes referred to as a "weakest link" in the chain model. In this situation, we recognize that the Weibull distribution is a type of smallest extreme value distribution. Any situation in which the failure modes can be described by the failure of a weak element can be described by the Weibull distribution. Look at any deteriorating systems, these can all be described by the Weibull distribution. All declining failure rate systems (improving reliability) can be described by the Weibull distribution. Examples of situations include ball bearings, most leaf and coil spring applications, many semiconductors, most discrete components and many complex systems.

Normal - This distribution appears in reliability in two ways. The first is through quality conformance inspection situations, such as direct measurements related to the central limit theorem. The distribution of a series of averages is always normally distributed even if the raw data points are not [Ref. 3, p36]. The data is said to be additive in nature. The other way is to look at the interactions of the contributions to failure of a part or system. When these interactions are *additive* in nature, the Normal distribution may describe the part or system well. Two types of common parts tend to fall into this category, the incandescent-type light bulbs and many batteries. A number of mechanical situations are often modeled with the Normal distribution, but each situation of time-to-failure should be checked individually by the engineer. Don't assume a mechanical distribution is Normal.

LogNormal - This distribution is often employed to describe nonlinear deteriorating parts or systems. Thus, we often see the LogNormal distribution appear in accelerated life tests. This distribution can be used to describe mechanical situations such as the growth of stress cracks in a metal or plastic. Another common use is to describe some semiconductor time to failure (life) situations. The LogNormal distribution is also said to describe a situation where a number of *multiplicative factors* interact to produce failure results. Most maintenance time-to-repair distributions are modeled by the LogNormal distribution. If the log of times to failure are normally distributed, then the times themselves are LogNormal. There are several quick tests that one may use to determine if data is approximately LogNormal. These include plotting the data on a LogNormal graph, or calculating the mean, standard deviation, skew and kurtosis. There is a relationship between the LogMean and the LogVariance which includes the mean and standard deviation of the Normal distribution. This is, if x_i = Ln(t_i) and the data set x_i are normally distributed, then the t_i are LogNormal. The data set x_i is described by a mean, μ and standard deviation σ for a Normal distribution. We can then write:

$$\textbf{LogMean} = e^{\mu+(0.5)(\sigma^2)} \quad \text{and } \textbf{LogVariance} = [e^{\sigma^2} - 1][e^{2\mu+\sigma^2}] \qquad \textbf{(26)}$$

Extreme Value Distribution – This distribution is used to describe events such as the failure of a weakest link. If there is a series connection of n elements under a stress forming a system, then the time to first failure of a set of such systems would form an Extreme Value Distribution [14, p320]. There are actually three different distributions. One formula for this Extreme distribution is:

$$\text{Extreme Value Reliability} = e^{-e^{(\frac{t-\gamma}{\eta})}}$$

The Weibull distribution represents one type of Extreme Value.

When both 2 and 3 parameter distributions are considered for both the Weibull, Normal or LogNormal a wide variety of distributions can be covered. The user often needs more information or some additional help when selecting the best distribution to model reliability data. Newer software allows us to test data with a variety of distributions in both rank regression (RR) and maximum likelihood models (MLE). This merely complicates the problem of selecting a "best-fit distribution" as it may be easy to show the two different distributions are about equally good fits to a data set. Help is needed in making a decision, so Section 4.3 lists a series of rules that may provide this help when selecting a distribution model. These rules can be used with the software results to help make the best decisions. Don't depend upon any one measure such as the goodness-to-fit in a canned software package. Always use all the information available, including past history, failure modes and design similarity. Later, additional rules will be provided for accelerated life testing situations.

4.3 - Selecting a Model of Time to Failure

The following list of rules aids in selecting the best distribution to employ when modeling data. Whenever failure mechanism and failure mode information exists, then this should also be included in the selection process.

1) Consider the **response of the part or system** to the stress involved with the test. These stresses produce very specific failure modes. A consideration of the "physics of failure" may be a big help here especially when a single failure or a small number of mechanisms may dominate. Consider all past test and any other information available. Is it a weakest link situation, competing failure modes, oxidation, temperature driven corrosion, stress driven corrosion or diffusion-dominated failure mode situation? What do you know about real causes of failure or the failure mechanisms present? Each of these situations might be described by a **simple equation** that relates results and stresses. For example, one relations would be

   $$\text{Diffusion of Metal} = A\,(\text{current density})^{-5}$$

2) **Select** either the rank regression **method** (RR) or the MLE. Confused over which to pick? If suspensions are present, select MLE. Look at the software package carefully, in a few cases other options are present.

3) Select a distribution and **try a preliminary mathematical fit** with the data using a

selected model for RR or MLE. Does all the data make sense and fit a simple line? Should some data points be selected or perhaps disregarded? Look at the correlation coefficient (for RR) or the Maximum Likelihood Estimator (usually $\mathcal{L}$). These two numbers measure how well the data and the best fit line coincide. Next, consider adding the third parameter for the distribution. Does this improve the data fit? Does this help reduce problems or does it create more questions? Does the time offset make sense with what is known about the parts and the physics of failure for the situation?

4) Consider the **engineering or physics** of the situation. Perhaps there is an equation that relates some of the variables and stresses for a dominant failure mode. Does it make sense with what is known to be true and any past history? Does the model reflect "wear out" when it is known it to be present? Does the model reflect an improving reliability when it is known it to be present in the design?

5) Consider the situation of an "**interval test**" where data is collected periodically rather than continuously. We have incomplete information, but still can measure and estimate reliability. The failures have occurred over some interval of time, but the exact time to failure is unknown. Figure 17 is an example of such data from Nelson, p 145 and is plotted with suspensions.

6) Consider the situation of an **incomplete data set**. How can we fill in any of the missing data? Monte Carlo, simple statistics or a simple model might do the job. This helps to smooth data and may remove some of the noise present in the data.

6) When all done, look at all of these factors to decide which **mathematical model** best fits the data. Weigh all the relevant factors when making the decision. Remember - there is no single rule to follow in all cases.

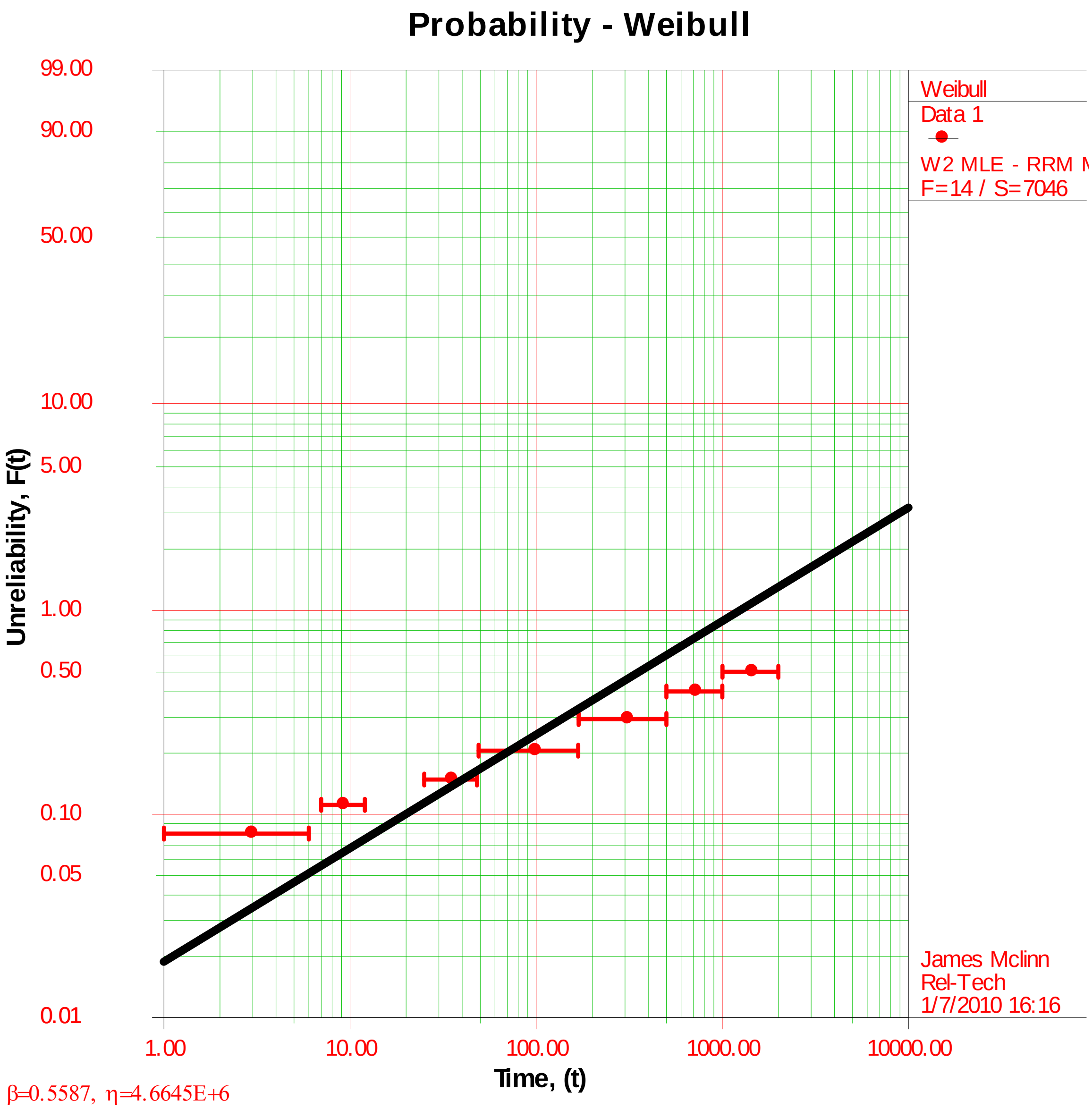


Figure 17 – Interval Test with Irregular Time Intervals

4.4 - Estimating Reliability in a Multi-Stress Level Test

Consider the following accelerated life test, ALT, in Table 22, showing data from Nelson (***Accelerated Testing,*** p. 157) for a rolling bearing system. The test consisted of 40 samples total with ten operated at four different stress levels. The entries each represent millions of rotations to failure and are complete, that is, all of the units in test were run to failure.

Table 22 - Life Test Data for Bearings

Stress level	@ 87%	@ 99%	@ 109%	@ 118%
Failure 1	1.67	0.80	0.012*	0.073
Failure 2	2.20	1.00	0.18	0.098
Failure 3	2.51	1.37	0.20	0.117
Failure 4	3.00	2.25	0.24	0.135
Failure 5	3.90	2.95	0.26	0.175
Failure 6	4.70	3.70	0.32	0.262
Failure 7	7.53	6.07	0.32	0.270
Failure 8	14.70	6.65	0.42	0.350
Failure 9	27.76	7.05	0.44	0.386
Failure 10	37.4	7.37	0.88	0.456

Two immediate distribution issues are present. First, what is the best time-to-failure distribution, recognizing that the starred entry at 109% stress is probably an outlier. Next, what is the acceleration factor that relates operating stress to ultimate bearing life? Remember, Palmgren's relations suggests that normal steel bearings might follow a power law with the power N = -3.0. This evaluation of this power should be consistent with any time-to-failure distribution. We don't have a situation where it makes a difference if we use rank regression or MLE since the data is complete. A best solution exists for this data set and it will be based upon first selecting the best distribution to describe all of the data at each stress level. Next, we look at the Power Law relationship limits which describes the relationship between the various stress levels. Then we employ the rest of the rules given above.

4.5 - Estimating Reliability when Suspensions Dominate

One difficult situation occurs when we have a large amount of field data with only a few failures. That is good economically, but makes it difficult to estimate reliability, which is high. Also difficult is the ability to determine if there is a potential systematic problem versus background noise. Systematic problems need attention, replacement or repair for the future. This is a common situation for a reliability engineer. They may be asked to estimate corporate risk with very sketchy information. The uncertainties on field data may include:

1) Time in the field such as total operating time, daily, weekly or monthly operating time.
2) The impact of running engineering changes has on the field performance.
3) The existence of customer use, misuse or abuse situations.
4) A good knowledge of when failures exactly occurred.

5) Delays in starting units in the field or delays in reporting failures.

6) Dead time in the field which includes how many units are not operating.

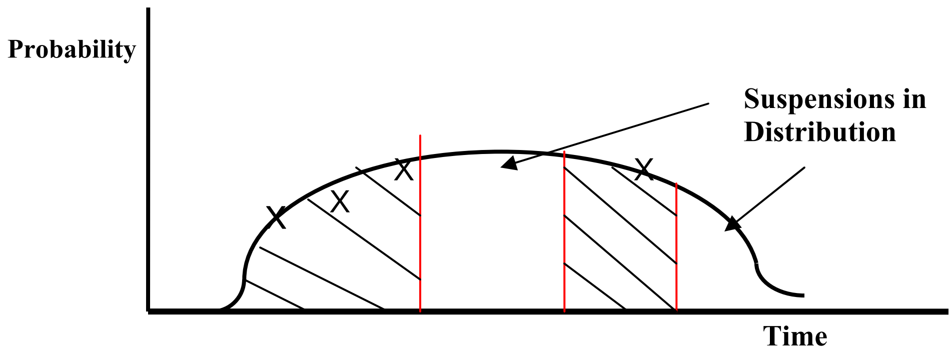


Figure 18 - Time to Failure with Suspensions

In general, when there are many suspensions, we select the MLE as the best analysis method, no matter the sample size or the number of failures. The prior examples in this book followed that general rule. Later examples will build upon this idea. Problems can arise in this situation for a line on the Weibull graph that represent the best fit to the whole data set may not pass through the few failures. This is because the suspensions (unfailed items) move the best fit line away from the early failures in the MLE method.

4.6 - Analyzing Field Return Data

The following example highlights this situation. Imagine we have a very reliable product, with more than 1,000 units having been in the field for 48 months. During this time only a small number have failed. The vast majority remain operating in the field. We know something about the failure modes of the few failures themselves. Table 23 is a summary of the last nine months of field data of this 4 year period. The assumptions about the field data include:

1) Shipped units begin collecting field operating time right away
2) Failures are not repaired or replaced

Solving the following two equations with the data from Table 23, we determine:

$$\mathbf{MTBF} = \eta\,\Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and} \quad \mathbf{R} = \mathbf{e}^{-\left(\frac{48}{\eta}\right)^{\beta}}$$

yielding η = 51583 months and β = 0.639

Cumulative reliability is the result of collecting the reliability each month in a spread sheet. Looking carefully at the columns we see that in the sixth month (June) of this table 56 new units were shipped and a total of 1228 units are operating with 19431 unit operating months exist. Only eighteen valid failures have occurred so far in this 45 th month since shipments of this product began. Table 23 is a snap shot of a much bigger data set, but presents enough data to work out some basic Weibull issues.

Table 23 – Sample Data for Field Returns

Month	Jan	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.
Units in Field	983	1026	1071	1123	1172	1228	1281	1341	1399
Failures Removed	23	23	23	24	24	25	25	26	26
Valid Removals	17	17	17	17	17	18	18	19	19
Total Unit Month in Field	13811	14837	15908	17031	18203	19431	20712	22053	23452
Cum Valid Defects per Exposure Month	$\frac{17}{13811}$	$\frac{17}{14837}$	$\frac{17}{15908}$	$\frac{17}{17031}$	$\frac{17}{18203}$	$\frac{18}{19431}$	$\frac{18}{20712}$	$\frac{19}{22053}$	$\frac{19}{23452}$
Field Reliability Each month	0.9988	0.9989	0.9989	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992

The reliability in June is calculated simply as:

$$R = 1 - \frac{18}{19431} = 0.999074 \sim 0.9991$$

The reliability is easily updated each month through the spread sheet as are each of the entries.

The dated can be plotted on a Weibull graph **manually**, as no canned program presently has the option to plot data of this nature. Plotting only every fourth or fifth data point is required to see the overall trend. Some ripples are expected in the data on the Weibull graph because it usually shows some variation from month to month. Using accumulated data of this type helps to reduce the noise, but variations are still present. The calculations shown here are based on the last year of data to arrive at the estimates for β and η. Once the data is in a simple spread sheet format, it is easy to update the data set monthly. When more details about field failures or failure modes become available, it may be possible to improve the calculation of reliability. The following is a short sample of additional detail that may impact the results in a big way.

Table 24 - Field Data: All Reasons for Removal

Month	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.
Customer Perception Non-valid Returns	6	6	6	7	7	7	7	7	7
Valid Reasons									
Loose Assembly	3	2	3	4	3	1	3	3	1
Damaged	2	3	2	3	4	3	2	1	2
Worn	10	9	11	9	8	11	11	11	12
Erratic Operation	1	2	0	1	2	1	1	3	1
Noisy	1	1	1	0	0	2	1	1	3
	----	----	----	----	----	----	----	----	----
Total Valid Returns (Failures) in Month	17	17	17	17	17	18	18	19	19

We can analyze this data for the valid failures of September to see what they tell us. Table 24 completes the necessary information to plot the data on a Weibull graph. We will look at the impact of the failures, but not separate any of the failure modes at this time. The "time to" information as well as the detailed failure modes may provide additional critical information for plotting.

The initial analysis of the data of Table 24 is detailed in Table 25 and then shown in Figure 19. This figure suggests that there may be more than one sub-population of time to failure. This is not surprising since Table 24 lists what appear to be five failure modes. The Weibull plot of the Table 25 data appears to show an S-Curve that is consistent with **two or more competing** failure modes. The calculation for the first eight failures as shown in Figure 20 has a characteristic life of 22.56 months with a slope of 1.815 based upon using the MLE. The abundance of suspensions suggests the MLE is the proper method, despite the difficulty posed by entering the results into a program. The remaining 11 data points seem to fit a different line, one with a characteristic life of 105.67 months and a slope of 3.56 as shown in Figure 21. Overall, the nineteen data points **do not** show as good a fit to a single line as the fit to a single line for each subgroup.

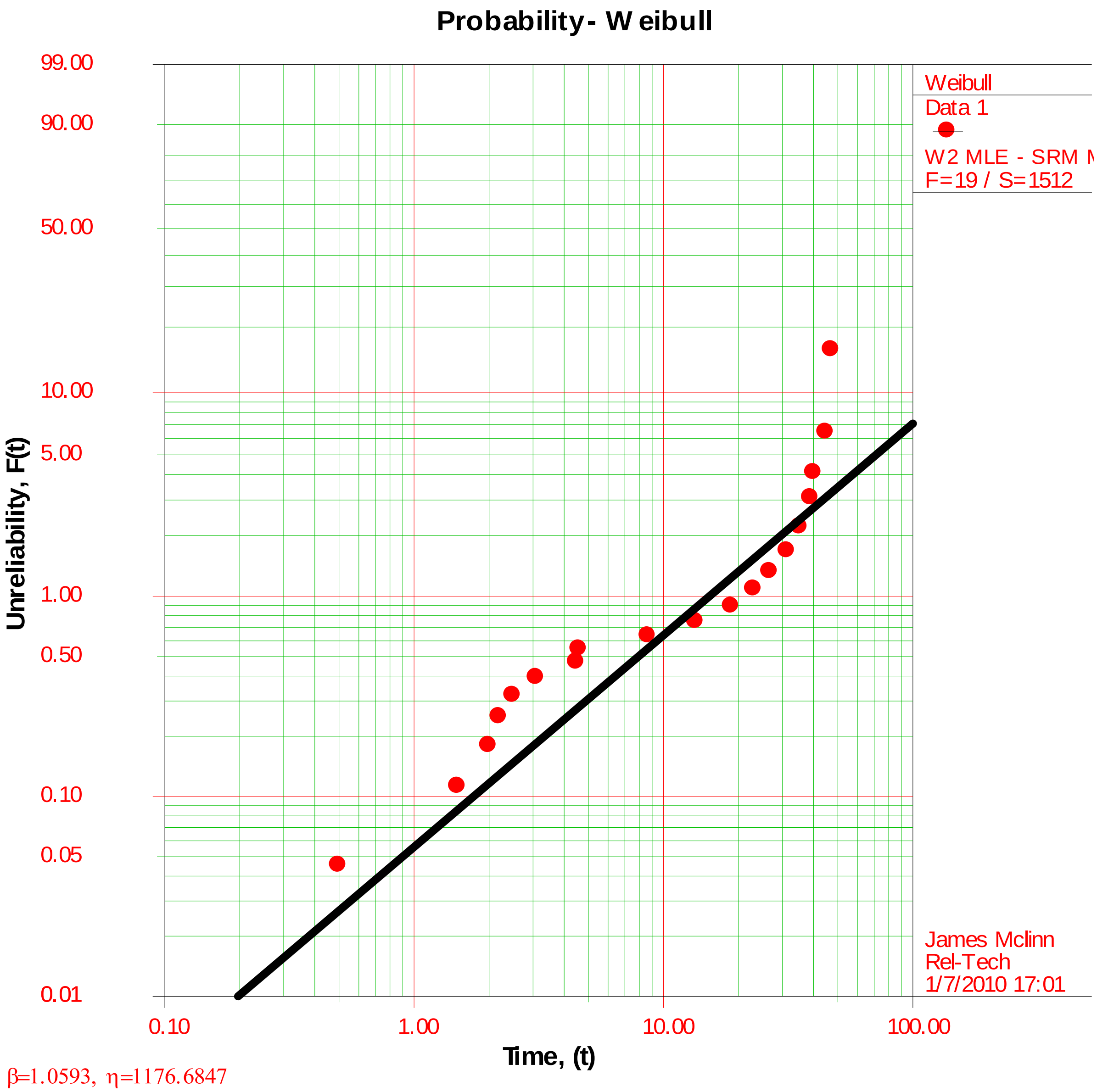


Figure 19 – Plot of All Field Data

Table 25 - Supplemental Field Failure Data

Failure Number	Time in Field	Failure Mode Description
1	0.5 months	Loose assembly Y
2	1.5 "	Noisy output
3	2.0 "	Worn component X
4	2.2 "	Noisy assembly Y
5	2.5 "	Erratic operation
6	3.1 "	Noisy assembly Y
7	4.5 "	Damaged component Z

8	4.6 "	Worn component X
9	8.7 "	Worn component X
10	13.5 "	Worn component Z
11	18.8 "	Worn component X
12	23.1 "	Worn component X
13	26.8 "	Worn component Z
14	31.4 "	Damaged component X
15	35.3 "	Worn component Z
16	39.1 "	Worn component X
17	40.2 "	Worn component X
18	45.0 "	Worn component Z
19	47.3 "	Worn component X

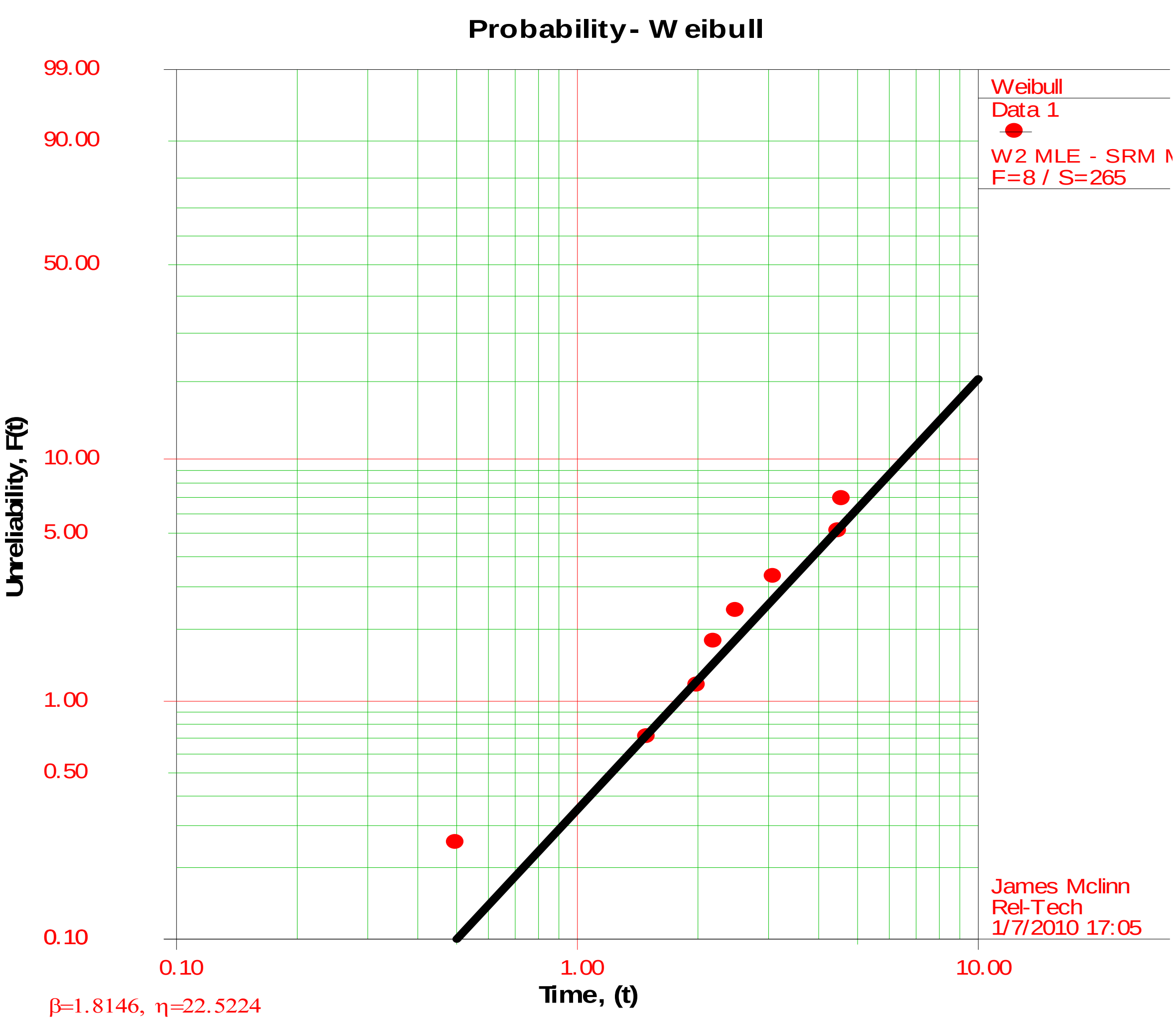


Figure 20 – The First Eight Data Points

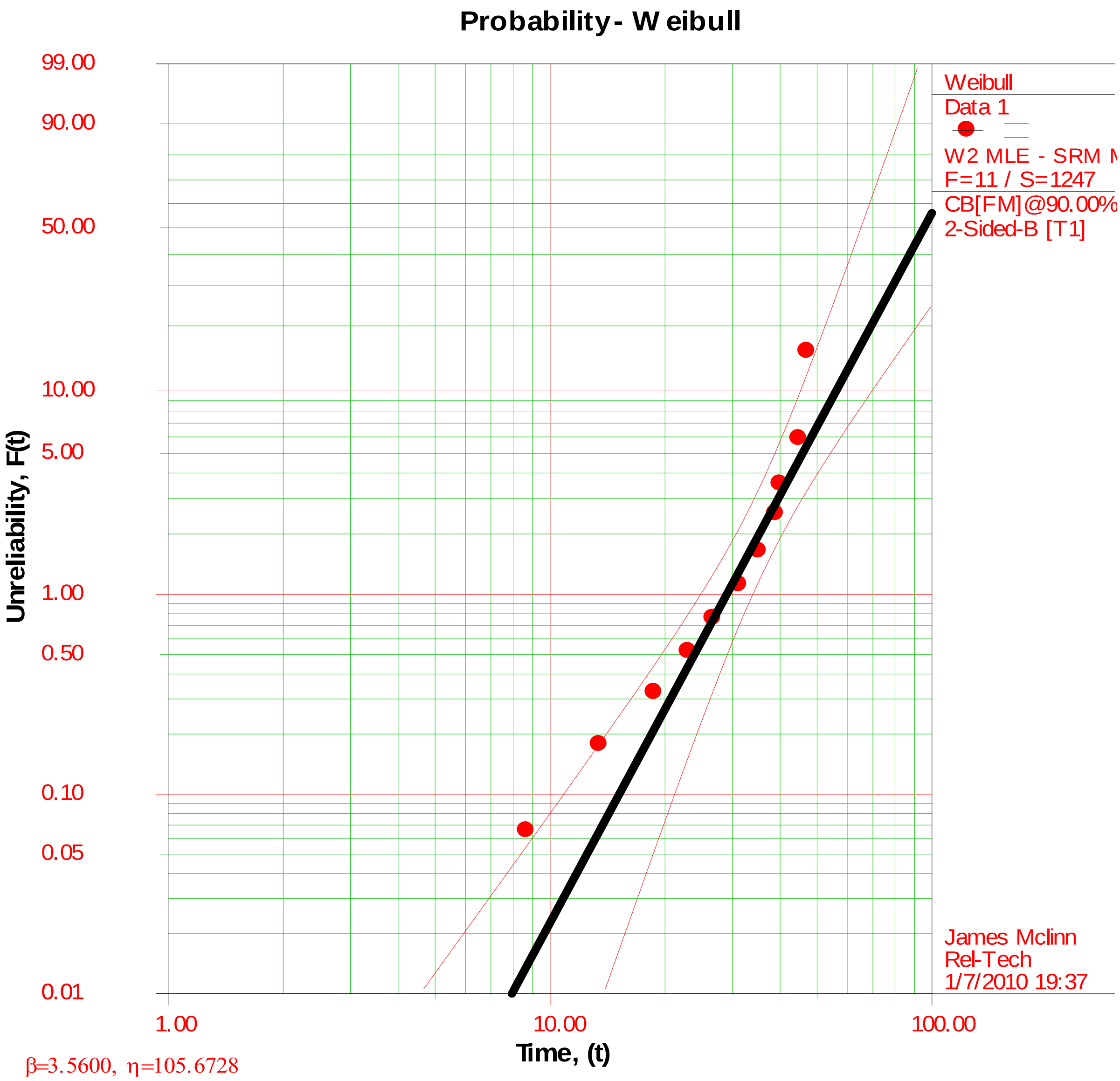


Figure 21 – The Last Eight Data Points

Compare the lines in Figures 20 and 21 to the whole data in Figure 19. The two figures show the first 8 data points only and the last 11 points, respectively. Knowing more about the failure modes would allow us to make a better analysis and perhaps identify what is actually occurring. Look at the first eight points. These seem to fit reasonably well with a slope of 1.82 and a characteristic life of 22.5 months. The data plots like one failure mode, but there are three loose, two worn and three other failure modes. The last 11 points have 10 worn components and one damaged. The first eight points with five different failure modes really show no clear trend based upon this data. The last eleven data points are mainly one failure mode, yet they show some curved line. The 90% confidence intervals around the best fit line show that two data points fall outside this range. These subtleties might be lost when dealing with so few failures and lots of suspensions. In general, the numerical data is presented in a small corner of the Weibull graph showing limited detail as in the lower left corner of Figure 21.

The upper right corner shows the analysis method, the sample size and the number of suspensions. The software program can be a big help to any thorough analysis. It is still up to the engineer to do the final analysis and interpret the results.

5.0 - Weibull Analysis and Weibull Confidence Issues

The Weibull function, in order to be very useful, needs to address confidence issues. Confidence allows us to estimate risk better than a variety of other approaches in many situations. Confidence, in a crude way, can be said to estimate how well "we think we know" something. (I am convinced there are an unlimited number of ways to estimate confidence. In one course, I present 17 different confidence calculation methods to the students.) Most of the confidence calculations are based upon one of the following fundamental approaches to estimating confidence. Many approaches create poor, that is very broad confidence estimates and so are easily discarded as non-viable choices. Most of the time we are able to limit the confidence calculation methods to a small number of valuable ones. Table 26 represents basic choices and so does not cover all of the possibilities that exist, nor does it exploit in detail some of the ones mentioned. Mechanical reliability confidence approaches are usually different from electronic ones. Reference 10 is a good text for more details on confidence.

Table 26 - Basic Methods to Estimate Confidence Intervals

1. β Binomial - This is the approach developed by Leonard Johnson of the auto industry in the late 1950s. We still use it in limited situations today, especially with purely graphical approaches. Tables of 5% and 95% limits are often based upon this simple method. This approach does not handle situations with suspensions well. Better and easier methods have replaced this one over time.

2. Fisher Matrix - This method has come into common use because of modern personal computers. It was developed by R.A. Fisher and is the basis of many of the "confidence interval calculations" associated with most standard Weibull analysis programs. Few reliability books show more than sketchy details because of the complexity of the matrix. Examples show matrix elements only.

3. Maximum Likelihood Estimates - This method is based upon the concept of the maximum likelihood function based upon the exponential distribution. It is a time-consuming approach that has some value and has been made easy through the use of computers. This approach counts suspensions equally as important as failures. In the case of large suspensions, the "best fit line" sometimes does not appear to fit the few failures very well. Reference 16 has a good summary.

4. Special distributions - These include Chi-Square and Normal distributions. These are often employed to estimate confidence in many situations. A few examples are in this book and in 10.

5. Kaplan-Meier - Reliability approaches to interval testing and field data are often best handled by this method. One example using the Greenwood Confidence calculation method follows in this chapter.

6. Calculation of Confidence Bounds - This involves the estimation of upper and lower bounds by one of several means. Examples include basic parallel versus serial estimates or use of the extreme value distribution. These approaches tend to be very conservative and are sometimes found in reliability texts.

Other methods to calculate confidence bounds do exist, so many of the other possibilities are not shown here. The problem of confidence interval calculation is often a matter of method as well as applicability. We may approach it from purely statistical thinking on one hand or work a problem as a simple engineering estimate. Each method leads us down very different thought processes with different calculation methods and sometimes very different results. We may use a sophisticated computer program for confidence or estimate the appropriate confidence limits by use of a Chi-Square table or similar simple approach. The first examples will use "canned Weibull" calculation programs. There are often three complex calculation approaches, either β Binomial, Fisher Matrix or Maximum Likelihood Estimator available.

We can look at the problem of confidence by finding a distribution that estimates the "uncertainty" of our knowledge of the separate reliability distribution. Often we invoke one of the Lemmas of **statistical physics**. Simply stated, it is "that a long-time average of a small number of samples of an event provides the same information as a short-time large group average". In reliability language, if we have a large sample size and watch it for a short time, we should see the same range of variability as watching a small sample for a much longer time. This is the process we use when establishing confidence calculations based upon a single small test. Even when we correct statistically for the "small sample size" present, the results may show bias. The Six Sigma methodology employs both **long-term** and short-term estimates of variability to correct for incomplete knowledge of process. The correction factor employed is typically 1.5σ to create long term standard deviations from short term data.

Two distributions are involved in most reliability situations. These are the reliability distribution and the confidence distribution. In some situations, we may wish to relate the two distributions; in most cases there is no clear connection between both. Often our goal is merely to estimate each distribution as best we can. Figure 22 displays both distributions on one Weibull graph. The reliability distribution is the heavy solid straight line that represents best fit line for the data points, marked by Xs. The two light lines that form the band of confidence around this best fit reliability line are the estimates of the limits of confidence to the data. This band is shown as the two light lines, one on each side of the solid line and "appearing to be parallel". At each end of the confidence lines, the confidence curves away from the data.

In most cases the confidence limits are not simply parallel lines and will actually "bend away from" the reliability line at low failure numbers and appear to approach the solid line as the number of failures increases. When many suspensions are present over a long time, the confidence limits will also diverge from the reliability line. Figures 9 and 21 show this behavior.

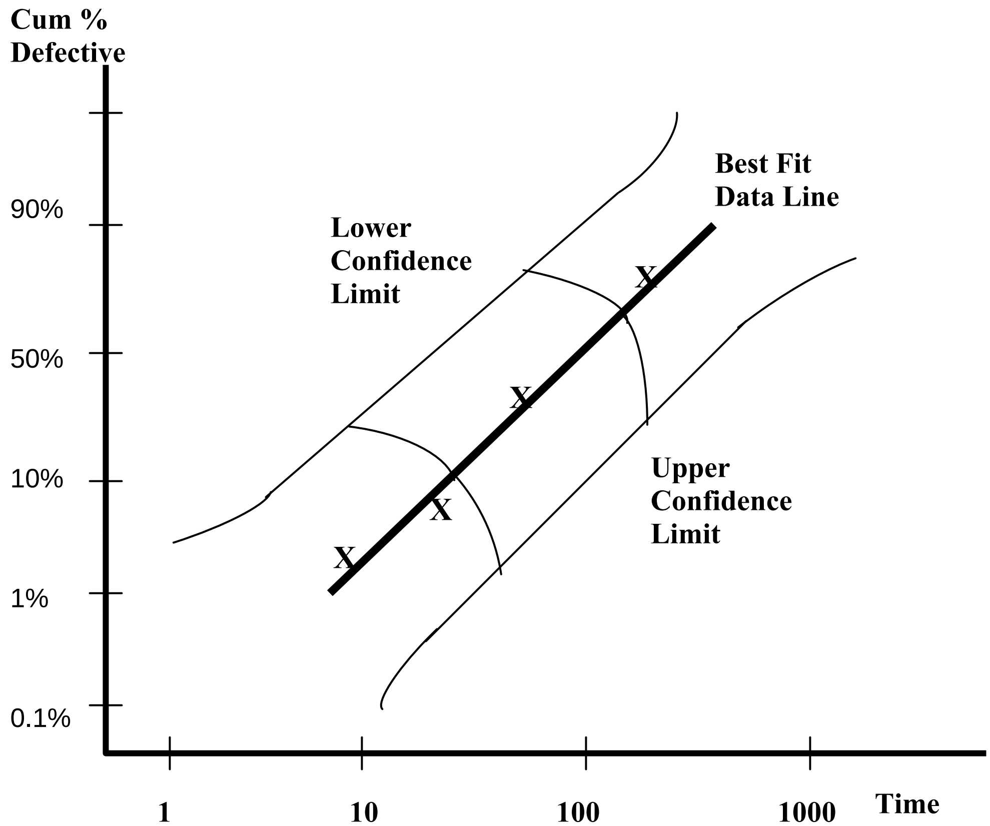


Figure 22 - Reliability and Confidence Distributions

5.1 – Beta Binomial Distribution

The β Binomial approach is based upon the two formulas (see references 10, 18 and 19). The first is the simple Beta binomial formula and the second is the improved version (shown as equations 27 and 28). The Median Rank may be obtained by solving the following equation for the j th failure in a sample of N. For example, solving for the 4 th failure in a sample of 10 would yield Z = 0.355

$$0.50 = \sum_{k=j}^{N} \binom{N}{j} Z^{j} (1-Z)^{N-j}$$

Confidence limits may be estimated by setting the probability at 95% and 5% limits. In this situation, the 95% rank would be 0.6966 and the 5% rank would be 0.1500. It assumes a good estimate of the

Weibull slope, β and the characteristic life, η. We calculate two rank limits for all the data points, at 5% and at 95% the graph would appear as in Figure 21. These limits are represented by a continuous distribution which is centered upon the "best fit" reliability curve estimate. We have selected the common 5% and 95% points, but this method can be calculated at any desired confidence level. We often see tables of these rank numbers in a reliability text appendix. See O'Connor's ***Practical Reliability***, Appendix 6, as an example of this type of non-parametric calculations. References [4, 10] have additional examples of the formula details. The following two formulas can be employed to estimate improved confidence limits. There exist a more detailed set that provides "tighter limits" in most cases.

$$t_{(i,5\%)} = \eta \left[\ln\left(\frac{1}{1-F_{(i,5\%)}}\right)\right]^{\frac{1}{\beta}} \qquad t_{(i,95\%)} = \eta \left[\ln\left(\frac{1}{1-F_{(i,95\%)}}\right)\right]^{\frac{1}{\beta}} \quad (27)$$

$$LCL = \frac{\frac{j}{n-j+1}}{F_{(1-\alpha,2(n-j+1),2j)} + \frac{j}{n-j+1}} \qquad UCL = \frac{\frac{j}{n-j+1} F_{(\alpha,2j,2(n-j+1))}}{F_{(\alpha,2j,2(n-j+1))} \frac{j}{n-j+1} + 1} \quad (28)$$

where **j** = failure number, **n** = total units on test, α measures confidence and **F** is the 3 parameter F distribution which can be found in Juran's ***Quality Control Handbook*** [26].

There are difficulties with the use of these two sets of formulas. Equation 27 formulas should be considered approximate, while Equation 28 formulas do not take into account the Weibull parameter $\tilde{\beta}$**.** The approximations for either formula set may not be very good for small sample sizes, such as less than 10 test points (failures plus suspensions). A second concern is that these four formulas do not do a good job of taking into account the effects of suspensions. Both sets of formulas also lead to a very conservative (i.e. very wide) estimate of bands. The value of these formulas is that quick and easy estimates may be produced, especially when calculating limits by hand. Thus, there is a tendency to use these among reliability engineers when calculating by hand or estimating graphically. An example of their use on a Weibull graph is shown in Figure 22. The following section explains how one type of confidence limit is placed upon a Weibull graph.

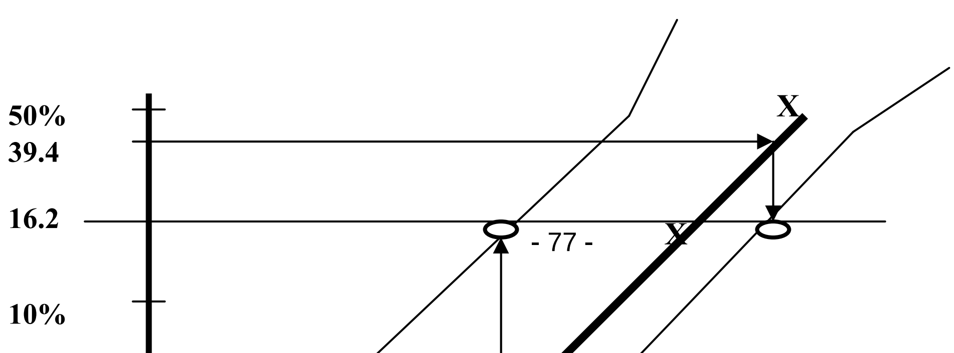


Figure 23 - Calculating Graphical Confidence Limits

There are 10 units on test and we wish to establish confidence limits estimates at 5% and 95% for several of the failures. We can do this very quickly and easily, but only roughly with the β Binomial tables previously cited. The method described below works for each ***plotted point*** of a Weibull graph. The *loci* of these points will provide a conservative estimate of the upper and lower confidence limits. Only the limit on the second failure point is demonstrated here. These confidence bounds need not be symmetrical around the best fit line or data points.

We start by plotting the reliability data on the Weibull graph and then drawing a best fit line to the data set. This could be an eyeball estimate straight line or one drawn by the RR or MLE via software. Go to the appropriate table for the 5% limits, typically labeled a median rank table for 5%. Find the entry for 10 samples and the second failure. This is a table entry of 3.68%. Next, we create a line at this percentage that is parallel to the time axis and extends to the solid "best fit" reliability line. Where this line intersects the reliability line, we erect a vertical line that extends up to the location, as measured in percentage, of the second failure in 10 or 16.3%. Mark this location, here marked with an "**o**". Next, go to the 95% limit tables and look up the same entry. For 10 samples and the second failure the entry is 39.4%. We create a horizontal line at this percentage that extends to the solid "best fit" reliability line. From here we go straight down to the level of 16.3%. Again, we mark the point with an "**o**". Repeat this exercise for *each failure* point. The band that is created is an estimate of the confidence limits for this data set. It would look like the confidence lines we see on Figure 23.

5.2 - A Fisher Matrix Approach

The Fisher Matrix [FM] is an approach that attempts to describe the confidence bounds based upon more detailed mathematical functions. This method has come into more common use because it is easily adapted to the personal computer [1]. Total details are beyond the scope of this short book, so only a summary will be presented. FM bounds provide tighter confidence estimates than Beta-Binomial for moderate sample sizes.

An example of the Fisher Matrix approach is the following simple formula using the log Likelihood function of two variables θ_1 and θ_2 to estimate parameters.

Ln|L| = Λ =

$$\sum_{i=1}^{R} \ln |\ f(T_i;\theta_1,\theta_2)\,| + \sum_{j=1}^{M} \ln |\,1 - F(S_j;\theta_1,\theta_2)\,| + \sum_{k=1}^{P} \ln\{F(I_i;\theta_1,\theta_2) - F(I_{i-1};\theta_1,\theta_2)\}$$

Where N units are on test, M are operating, S are suspended and P units have failed within a time interval. The simple Fisher Matrix can then be written as:

$$F = \begin{vmatrix} -\dfrac{\delta^2\Lambda}{\partial\theta_1^2} & -\dfrac{\delta^2\Lambda}{\partial\theta_1\,\partial\theta_2} \\ -\dfrac{\delta^2\Lambda}{\partial\theta_2\partial\theta_1} & -\dfrac{\delta^2\Lambda}{\partial\theta_2^2} \end{vmatrix}^{-1} = \begin{vmatrix} \text{Var}(\theta_1) & \text{CoVar}(\theta_1,\theta_2) \\ \text{CoVar}(\theta_2,\theta_1) & \text{Var}(\theta_2) \end{vmatrix} \qquad \textbf{(29)}$$

The symmetrical confidence bounds are evaluated by the following formula:

$$\text{Confidence bounds} = \text{Expected Value of G} \pm Z_\alpha \left[\sqrt{Var(G)}\right] \qquad \textbf{(30)}$$

With G being a function of the variables θ_1 and θ_2 and the variance of G being defined by

$$\text{VarG}(\theta_1,\theta_2) = \frac{\delta^2 G}{\partial\theta_1^2}\text{Var}(\theta_1) + \frac{\delta^2 G}{\partial\theta_1^2}\text{Var}(\theta_2) + 2\,\frac{\delta G}{\partial\theta_1}\,\frac{\delta G}{\partial\theta_2}\ \text{CoVar}(\theta_1,\theta_2) + \text{error term}$$

If G is a best fit line on the Weibull graph to a data set, we can calculate the upper and lower confidence bounds based upon that line. We treat these lines as confidence intervals to the raw data. Now if we let θ_1 be η and θ_2 be β, we can write the Log Likelihood function for the Weibull as:

$$\text{Log}(\beta,\eta) = \text{rLog}(\beta) - \text{r}\beta\text{log}(\eta) + (\beta\text{-}1)\sum Log(t_i) - (\eta)^{-\beta}\sum_{i=1}^{n}(t_i)^{\beta}$$

Table 27 presents an example of several different confidence calculations (from Abernethy). These represent some of the variety of ways in which confidence limits may be calculated. The limits that are generated by Weibull software programs may vary from program to program depending upon what model was selected. The data set shown in Table 27 shows some of the different values generated for β, and η as well as upper and lower confidence bounds. See, for example, WinSmith™ (by Abernethy), STATPAC™ (GE-Nelson), Weibull 6.0™ (ReliaSoft), WeibullEasy™ by Berner for additional details.

5.3 - Likelihood Ratio Bounds

These estimates of confidence bounds, based on the likelihood function that was described in detail by Jerald Lawless in ***Statistical Models and Methods for Lifetime Data*** published by Wiley in 1982. This method is more complex and detailed than the Fisher Matrix approach, so it is beyond the scope of this short book. Nelson (***Accelerated Testing***, p. 145) and Abernethy [1] provide examples of the application of this approach in their books. See also references 14 and 16 for more details and examples. Each author provides a derivation of the basic equations and examples of their use in their own software programs.

Table 27 - Comparison of Confidence Limits

For the data set of 10 points with failures at 403 hrs., 571, 659, 768, 845, 942, 1018, 1109, 1229 and 1348, we can estimate the limits as follows. (Abernethy, reference 1)

	β Value	**β Bounds**	**Characteristic Life η**	**η Bounds**
Fisher Matrix	2.999	1.95 - 4.60	1000	832 - 1201
MLE	3.47	2.20 - 5.05	992	831 - 1170
Monte Carlo From Simulation	2.925	1.78 - 4.88	998	818 - 1174

We can see that there is some variation between the different methods shown. There is some variation in the formulas employed to arrive at any of these numbers. The formulas represent estimates of the confidence bounds based upon assumptions. Thus, it is not surprising that a range of answers exist. Fortunately, modern computers have made a comparison of the various methods easy. We can quickly look at the results from any of a number of methods. What is the correct answer? The Abernethy data suggests that the range of data may not be large, but the estimate of β is key, as this parameter has a great influence upon the results.

5.4 - Other Distributions for Confidence

Normal Distribution

We can approximate upper and lower confidence limits through the Normal distribution. A number of ways exist based upon SPC techniques. The following is one set of equations for these limits from Nelson. The formulas of Equation 31 are approximate and can work with censored data when the number of unfailed units is greater than 10. Thus, if we had 27 units on test and stopped after 14 failures, we could approximate the upper and lower confidence limits with this method. If this sample size condition is not met, then another approximation such as the Poisson which employs the Chi-Square distribution, might be more appropriate.

$$F_{iL} = F_i - Z_a\sqrt{\frac{F_i\ (1-F_i\ )}{n}} \qquad F_{iU} = F_i + Z_a\sqrt{\frac{F_i\ (1-F_i\ )}{n}} \qquad \textbf{(31)}$$

The first four failures of a sample of 27 would have the following upper and lower 95% confidence limits based upon Equation 31. The median ranks at 95% are also shown in Table 28 for reference. These ranks were calculated as shown previously in section 5.0.

Table 28 - Confidence Limits at 95%

	F_{iL} - Lower	F_{iU} - Upper	MR_{low}	MR_{up}
1 st failure	-0.0227, not plotted	0.0968 = 9.68%	0.19%	10.50%
2 nd failure	-0.0088, not plotted	0.1570 = 15.7%	1.33%	16.40%
3 rd failure	0.0116 = 1.16%	0.2106 = 21.06%	3.10%	21.53%
4 th failure	0.0357 = 3.57%	0.2606 = 26.06%	5.22%	26.27%

Plotting this group of numbers for F_{iL} and F_{iU} would provide a graph similar to Figure 21.

5.5 - Kaplan-Meier (Survival) Method

This method is often employed with interval data, the type of data generated from periodic inspection of a life test in progress. Examples of periodic measurements include measuring for defects once a month or running long tests with irregular times to look for accumulated failures. The classical Kaplan-Meier approach was presented in ***Quality Progress***, August and October 1994 in the Statistics Corner. Here, I will treat each time interval as a separate, successive "independent sample" of the reliability, that is, as a serial sequence of events with successive reliability estimates. This thought process allows us to approach a number of reliability problems beyond Weibull applications. Table 29 contains a simplified example of some hypothetical field data. A more detailed example is in ***Statistical Analysis of Reliability Data***, [10]. Table 29 shows a snap-shot of the first four months in the field for a new product. It estimates the monthly reliability each month and then the cumulative reliability. The latter was calculated by multiplying the monthly reliability numbers. While we did not record the ship units per month, the data set suggests it is typically between 15 and 20 new units sent to the field in each of the first four months. For example, the fourth month shows a cumulative of 78 units-months in the field with a total of three failures in the fourth month itself. We have no information about how failures were repaired or replaced, only that we know the *accurate total* of unit months by our counting scheme.

Table 29 - Interval Field Data Example

Month	**Field exposure Cum. n_i**	**Failures in month -f_i**	**Reliability of month**	**Cum. Reliability**	**Tolerance $\pm$95%**
1	15	1	14/15 = 0.9333	0.9333	0.1262
2	27	2	0.9259	0.8642	0.1338
3	52	3	0.9423	0.8143	0.1299
4	78	3	0.9615	0.7830	0.1266

The tolerance will be estimated at 95% limits from Equation 32, using the columns above as the data.

$$\text{Tolerance} = \pm (Z_{a/2})(\text{Cum. Rel.})\left[\sqrt{\sum \frac{f_i}{n_i(n_i - f_i)}}\right] \qquad \textbf{(32)}$$

Note the wide tolerance range in Table 29 that exists at the fourth month. The data suggests the true reliability at this point in time must lie in the broad range of:

$$\mathbf{0.6564 \leq R_4 \leq 0.9096}$$

A variety of confidence methods could narrow down this range and then be employed to compare to that derived by the Kaplan-Meier Method. Reference 10 provides several examples of this method and Greenwood's method for calculating confidence limits. All detailed calculations are not shown here, however the following table shows some of the calculation chain needed to complete the tolerance estimates at 95% confidence.

Month	Cum Rel.	month f/n(n-f)	Cum f/n(n-f)	Sqrt.(Cum)
1	0.9333	1/(15)(14) = 0.00476	0.00476	0.06900
2	0.8642	1/(27)(25) = 0.00148	0.00624	0.07902
3	0.8143	1/(52)(49) = 0.00039	0.00664	0.08146
4	0.7830	1/(78)(75) = 0.00017	0.00681	0.08250

With Equation 32, the tolerance for month two = (1.96)(0.8642)(0.07902) = 0.1338. Tolerances at other confidence percentages are easily calculated.

5.6 - Analysis of an Accelerated Life Test Problem

The selection of a distribution to model the time-to-failure distribution is critical for estimating the results of any accelerated life test. Since there are no clear ground rules for many typical situations, a set of ground rules must be adopted. In the situation of accelerated life testing, these rules can be expanded to more fully deal with the many possibilities that may be faced. Table 31 provides a list of ground rules and these should be a big help when selecting the best distribution in the situation of multiple stresses. Follow the ground rules and use common sense as well as all the information at hand when making any decisions about distributions.

Table 30 shows an example of 40 units all operated to failure at four different stress levels as updated from the initial results shown in Table 22. The first failure at 109% stress was eliminated from the table as not plausibly related to the test results itself, hence the title “none” was entered. The rest of the data is the same. Now based upon this modified data set, the analysis of the data will be discussed.

Table 30 - Updated Life Test for Bearings

Stress level	**87%**	**99%**	**109%**	**118%**
Failure 1	1.67	0.80	**none**	0.073
Failure 2	2.20	1.00	0.18	0.098
Failure 3	2.51	1.37	0.20	0.117
Failure 4	3.00	2.25	0.24	0.135
Failure 5	3.90	2.95	0.26	0.175
Failure 6	4.70	3.70	0.32	0.262

Failure 7	7.53	6.07	0.32	0.270
Failure 8	14.70	6.65	0.44	0.350
Failure 9	27.76	7.05	0.44	0.386
Failure 10	37.4	7.37	0.88	0.456

The first problem is to look at possible time-to-failure distributions for each of the four stress conditions. Two possible distributions can be eliminated quickly, these are the Exponential and Normal distributions. These represent a poor fit for each time to failure at a given stress. This leaves the two common distributions LogNormal and Weibull as possible time-to-failure distributions. Other distributions such as Extreme Value or Logistic will not be considered. The following is a list of **seldom mentioned** criteria that help us select the best distribution to model the test results.

Table 31 - Reliability Ground Rules for Selecting Best Fits to data for ALT

1. The shape of the distribution, β, **may** be **dependent on the level of stress**. We expect some evidence of wear to exist with the product. Thus, Weibull slopes above 1.0 are expected unless the systems are strongly dependent upon infant failures. Remove any extraneous points from the analysis if they can be proven to be extraneous.

2. There should be a **well-defined relationship** between stress and life, even if is complex. We can look at ηinitially for such behavior. Be sure to look at other points such as the time to 10% failure as this may have a different relationship for stress. Next, consider the behavior of γ for stress related possibilities. The value of this offset should be greater than 5% of the time-to-first failure to be considered real. Otherwise it may reflect noise in the data set. Don’t assume a time offset exists merely because one can be calculated by a software program. To be real, a delay mechanism must exist.

3. "**The more you know** about the physical aspects of a test, the better you can do the analysis" according to Dr. Abernethy. Remember to consider the "physics of failures", the operating environment, and the history and possible effects of the test measurement method itself. Keki Bhote describes this as "talking to the parts".

4. Consider past history for the stress/life relationship as well as the time-to-failure relationship if it exists from the same or similar products. Remember also that we can look up models in many books such as those produced by the Reliability Analysis Center or common engineering texts. For bearings, we expect a power law relationship between stress and life. This would look like:

$$\textbf{Life} = \textbf{B(Stress)}^{-N}$$

The value of N is usually around three for ball and roller bearings.

5. Look at both the two and three parameter Weibull analysis. One common rule of thumb is to expect to improve the rank regression fit by at least half of the residual when we go to a three parameter analysis from a two parameter. That is, if we had an r^2 of 0.89 with a two parameter analysis, we expect to increase the r^2 to at least 0.95 with a 3 parameter analysis. If not, we probably do not have a good justification for the three parameter fit. See the additional comments in item 2. When using an MLE approach, the fit should improve (increase in size) by at least 10% (i.e. to go from -121 to at least -133.5) to be considered valid.

6. The **LogNormal or Normal Distribution** should be considered. Test each as you would the Weibull and see how well each distribution may perform within the four acceleration levels. Next look across each stress level and determine the fit. Consistency of shape is expected, don't compare a 3 parameter Normal to a 2 parameter LogNormal. Rank the top six distribution choices (Weibull, Normal and LogNormal for 2 and 3 parameter) and eliminate the bottom three or four. Test the remainder against each other.

7. Increase sample size if you can in order to get better data and reduce noise or impact of outliers. Remember, there may be a more complex function or situation that is being approximated by the distribution that you select. The distribution may look like a simple Weibull or LogNormal at small sample sizes, but could change as the sample size increases. Be sure to add all the suspension data when present.

8. Approximate the situation if the data is presented as interval. This may mean that once a month data is reported and the true time of failure during the month is unknown. For example during February, four failures are reported, but only at the end of the month. The exact date of failure is unknown. Spread the four failures evenly over the month.

Treat the operating time lost during the month as half a month average for each. This becomes 4(0.5) = 2.0 lost operating months. If shipments occur during the month and the exact time of start is unknown, likewise spread these out over the month in a uniform fashion. These approximations represent a reasonable estimate of the unknowns.

9. Never suspend common sense. Remember common engineering and statistics principles. This means statistical tests probably will not be used to make decisions, since most standard tests are unable to identify differences between increasing and decreasing failure rate situations because the sample sizes tend to be too small. Common sense and knowledge of the failure modes or mechanisms allow us to better estimate and model most of the time.

The following table simply summarizes many possibilities for β, η and γ for Weibull as well as Lnμ, Lnσ and γ for LogNormal distribution for this multi-level accelerated life test. Figure 24 shows the results of a simple independent analysis of the four levels of stress. All data is shown as a two parameter Weibull, though the three parameter option was also investigated. The data summary is presented in Table 32 for all the options to be discussed.

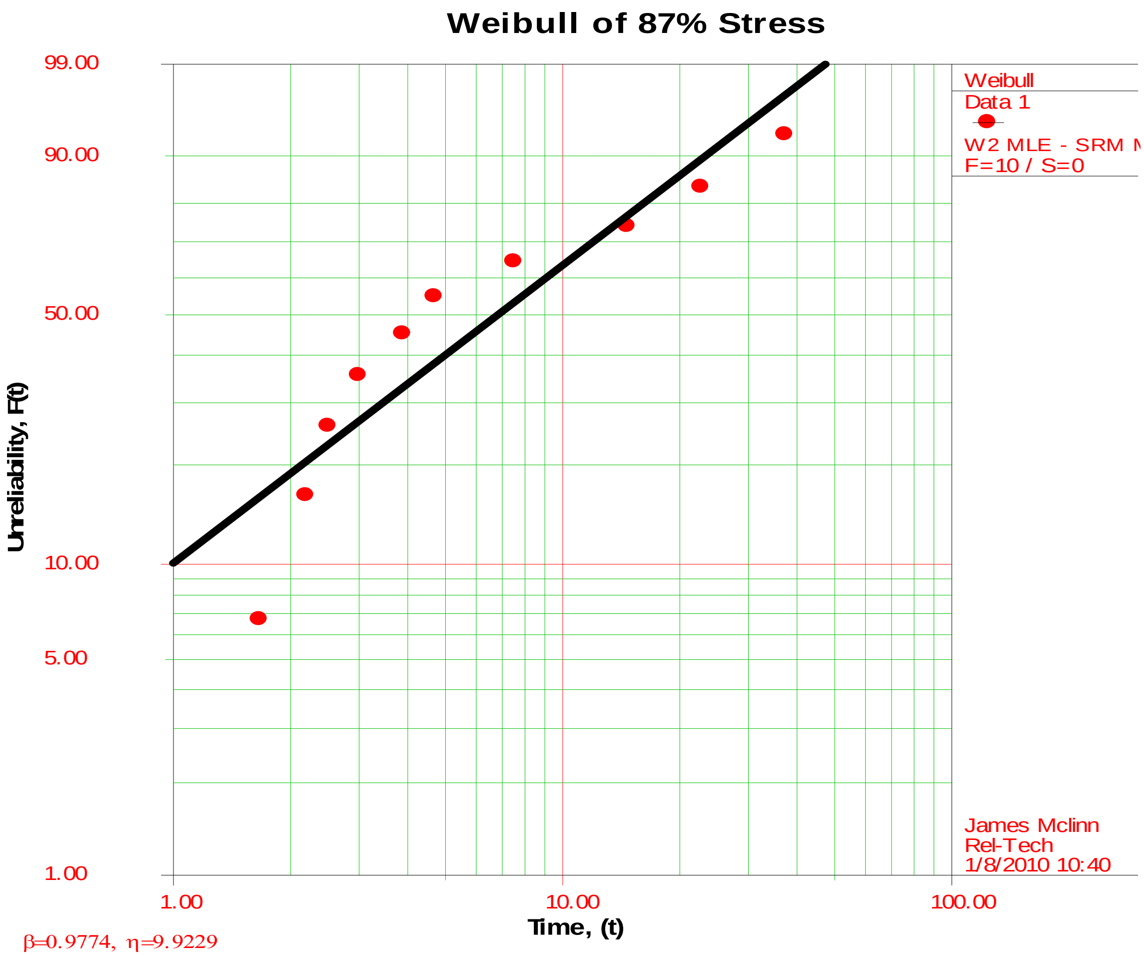


Figure 24 – Data from 87% Stress

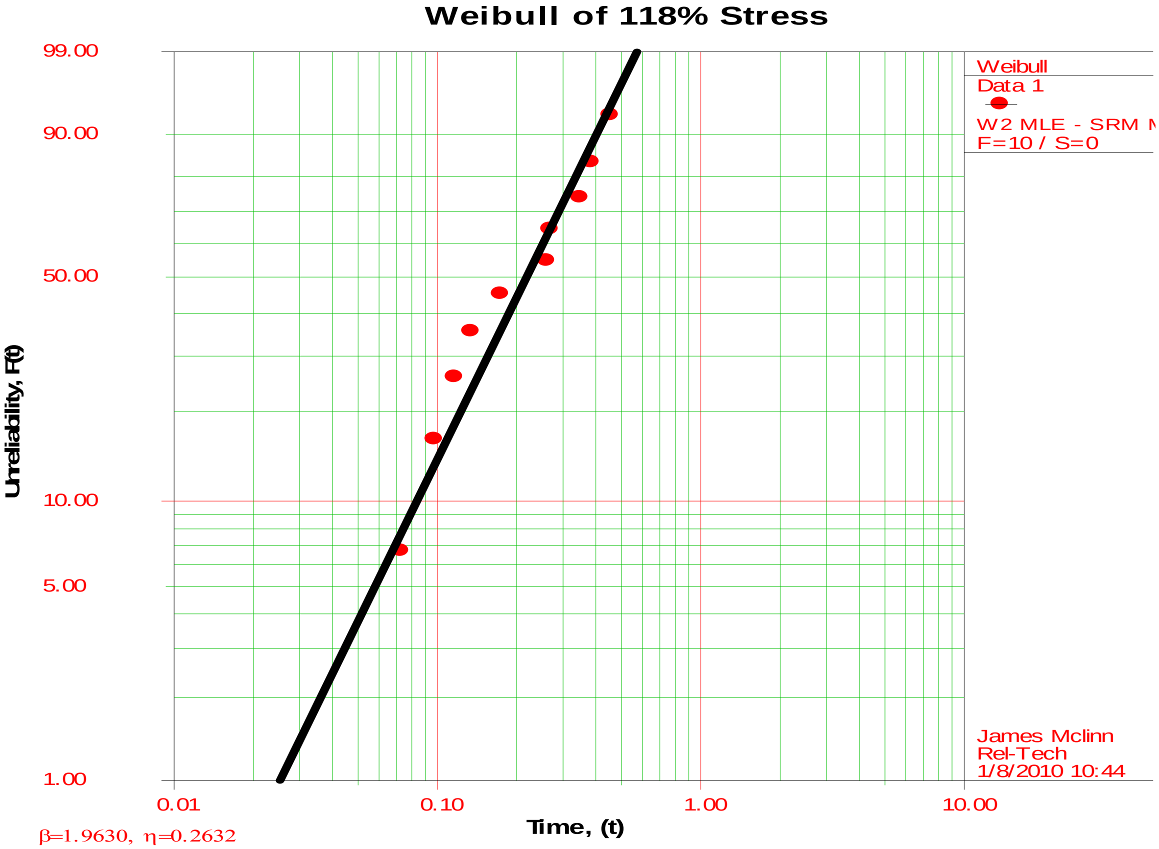


Figure 25 – Data from 118% Stress

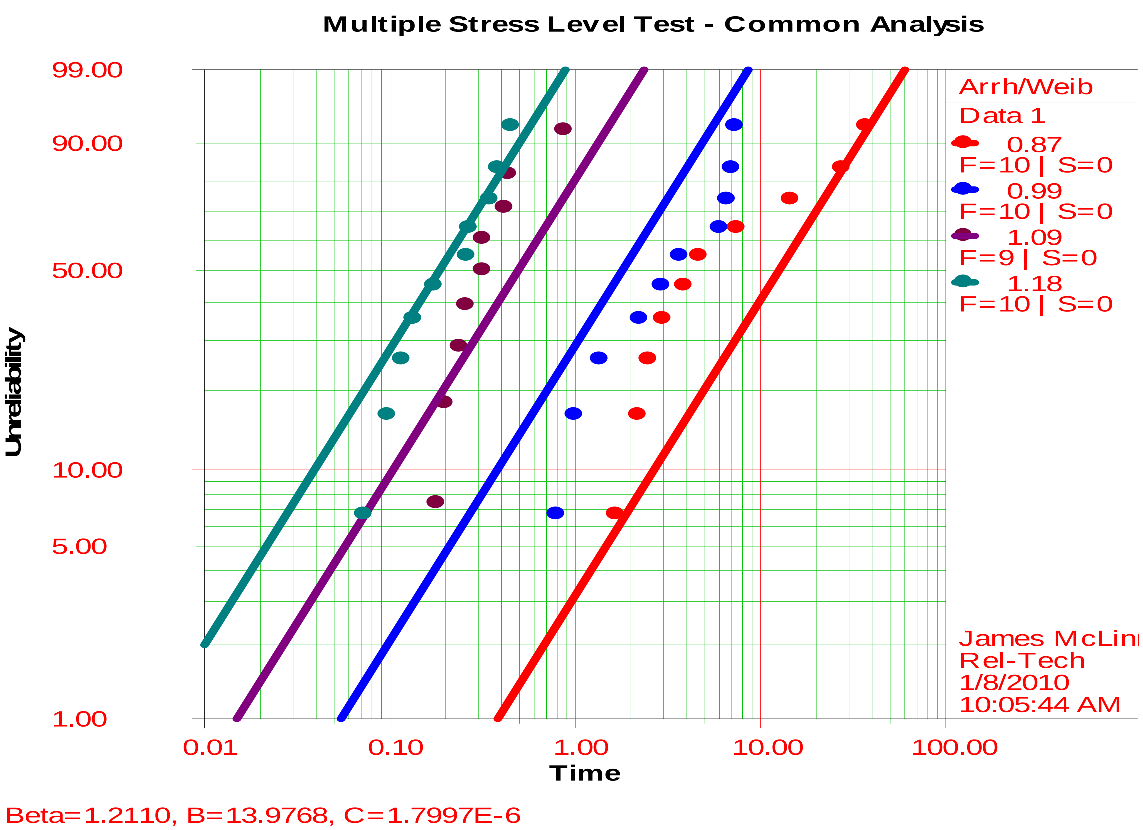


Figure 26 - Common Analysis of the Four Stresses

Table 32 - Bearing ALT Results

	Weibull - β, η, γ		LogNormal - Lnμ, lnσ, γ	
Stress Level	**2 parameter**	**3 parameter β and γ**	**2 parameter**	**3 parameter μ and γ**
87%	β = 0.977 η = 9.92	0.671, 1.59 η = 6.39	Ln μ = 1.77 Ln σ = 1.06	1.20, 0.372 Ln σ = 1.76
99%	β = 1.57 η = 4.37	1.39, 0.29 η = 3.98	Ln μ = 1.094 Ln σ = 0.841	1.42, -0.92 Ln σ = 0.642
109%*	β = 1.949 η = 0.412	0.99, 0.17 η = .192	Ln μ = -1.133 Ln σ = 0.484	0.145, 0.15 Ln σ = 1.033
118%	β = 1.96 η = 0.263	1.31, 0.06 η = 0.190	Ln μ = -1.627 Ln σ = 0.630	0.264, -0.02 Ln σ = 0.507

***** indicates the one peculiar point was removed from this set of ten, clearly it did not belong in that data set.

5.7 - A Short Analysis of ALT Data

The following is a short summary of the five commonly employed tests for selecting the best distribution Weibull versus LogNormal in this case of the bearings based upon the data in Table 28. Appendix A shows a statistical discussion of this choice of distributions.

1. The distribution of time to wear out. - The two parameter Weibull shows some expected bearing wear except at 87%. The β values at 99%, 109% and 118% were 1.57, 1.95 and 1.96. These numbers are reasonable. The β value at 87% stress was 0.98 which is close enough to 1.0 that it could be considered constant. This data suggests some slight stress dependence of β but is otherwise reasonable for β and η. A three parameter Weibull analysis suggests, while it fits the data slightly better for individual stress levels, it is even more inconsistent when rank regression is compared to MLE. This is possible, but must be considered somewhat suspicious since the two methods should be the same in this case of complete data. The value of β in the three parameter analysis jumps up and down as a function of stress and should be considered suspicious.

The LogNormal analysis suggests more consistency in the value of the Ln μ since this consistently decreases as stress increases for the two parameter analysis. Now Lnσ should not be a function of stress for two parameters, but shows a factor of two change as stress increases. The three parameter LogNormal does not fit as well as the two parameter. The two parameter **Weibull** and the two parameter **LogNormal** seem to **the best choices** here. **Call this even overall.**

2. The relationship between stress and the life models we generated - We actually have a good relationship between η and stress for the two parameter Weibull model and a less good fit for the three parameter Weibull. The two parameter LogNormal fit is good, but the three parameter is a very poor fit. We conclude when both time-to-failure distributions are considered across all of the tests, **two parameter Weibull** provides the best fit with the two parameter LogNormal next best.

3. Does a time offset make sense? - A test of time offset values suggests a serious problem for three parameter LogNormal model, so this must be rejected. The three parameter Weibull is not perfect, but is an acceptable fit here. Why all are a positive time offset exists is still a mystery. Since these are thought to be wear-dependent structures, we would have expected a negative number for all these offsets. This negative would indicate some wear occurred from operation of the system before the start of the life test. When all the offsets are positive, this indicates a “wear-in” situation where some operation must occur before the failure distribution begins. The LogNormal three parameter offsets are inconsistent, so this analysis is rejected. **Weibull** is the best choice based upon a three parameter analysis.

4. Stress versus life - Calculations of the expected power law coefficient, N, suggest Weibull provides a better model overall, since we have much better consistency across the four stress conditions. We do not get the expected relationship, but do get one that is not impossible where N = 13.98 for the two parameter Weibull model. Some of the LogNormal analysis leads to values of N that are negative numbers. These suggest illogical results when the LogNormal distribution is used. **Weibull two parameter** is a better fit here.

5. Look at the "goodness-to-fit", a traditional approach - A measure of the "goodness-to-fit" of the data suggests the best fit across all the data is with the Weibull distribution across all of the four possible analysis methods and test conditions. This is reinforced by a more detailed analysis. (see Figures 24, 25 and 26) The LogNormal distribution just cannot be made to work well with all the data. **Weibull two parameter** is the choice here.

6. Common Analysis – In this method the best fit analysis for all the data points has been made. The best fit slope for all data sets is a β of 1.21. This number represents a compromise between the lowest slope of 0.98 and the highest slope of 1.96. This common graph dramatically shows that the slope changes as a function of stress since the best fit line don't fit the data very well. The stress relationship between levels of stress may be expressed as the variable B in Figure 26. This number is 13.98 and represents the simple power law relationship between the stresses. It is higher than expected for this type of bearing system. A similar analysis with the LogNormal data is unable to calculate a meaningful number.

Conclusions - Based upon four of six measures , we conclude the Weibull is the best way to consistently describe or fit this bearing data, even though it is possible to reasonably model the data with a LogNormal under certain limited conditions. Does this prove the data is Weibull? **No!**

From the limited information we have considered, we can conclude **Weibull two parameter provides the best fit** across the criteria explored and considered. If we change the data, add information or try other time-to-failure models, a new one may appear as more probable than the Weibull fit. In the prior example, Weibull was a better choice in four of five cases. We need more information before we can say with engineering certainty that this is the case. This approach was supported by the data and Figures 24, 25 and 26 as presented.

5.8 - Dealing with an Outlier

The first point at 109% stress did not seem to fit the graph of the 109% line. It is possible to test this single point for inclusion in the data set at 109%. In their 1977 book, ***Reliability in Engineering Design***, Kapur and Lamberson identify an F test that is employed to determine if the first point of a series of data is typical of the rest of the data. In the situation, where the first data point is being tested for being abnormally short, the statistical test is described by the following inequality:

$$F_{(\alpha,\ 2n\text{-}2,2)} < \frac{\sum_{i=2}^{n} t_i}{(n-1)(t_1)} \qquad \textbf{(33)}$$

At **109% accelerated test conditions,** we have $\alpha = 10\%$, 2n-2 = 18 and n-1 = 9. Filling in the data from above we get:

$$9.43 < \frac{3.26}{(9)(0.012)} = 30.185$$

Thus, since the condition is true, the first data point is not representative of the rest of the data set.

Now applying the same statistical test to the first data point of the other 3 accelerated test conditions as a comparison we have:

118% accelerated test condition

$$9.43 < \frac{2.249}{(9)(0.073)} = 3.423$$

Since the condition is not true, the first data point is typical of the rest of the data at this stress.

99% accelerated test condition

$$9.43 < \frac{38.41}{(9)(0.80)} = 5.335$$

Since the condition is not true, the first data point is typical of the rest of the data at this stress.

87% accelerated test condition

$$9.43 < \frac{103.7}{(9)(1.67)} = 6.90$$

Since the condition is not true, the first data point is typical of the rest of the data at this stress.

The statistical test has shown that three of four first data points appear to belong with the remaining data at each stress level, while the remaining first point at 109% did not belong. Thus, the first data point at 109% was excluded from this analysis. A follow up with engineering or failure analysis would be appropriate to confirm or deny this statistical conclusion.

There are a variety of tests for a data set that measure consistency to some model. Ascher and Feingold present a number in chapter five of their book [27]. They use tests such as Laplace or Centroid for data trends, the Mann Reverse Arrangement Test and the Lewis-Robinson Test for renewal.

An Author's Note

This edition of the Monograph represents an increase in pages and examples as well as improvements in clarity for the user. The subject of Weibull Analysis continues to expand around the world as a way to quickly and effectively analyze data. Every reliability engineer should be very familiar with these topics.

The last example ends this book on the use of the Weibull distribution. Many people are especially interested in the applications of Weibull for the analysis of multilevel and multiple stress accelerated life tests. This special subject is covered in more detail in ***Practical Accelerated Life Testing***, James McLinn, published by the Reliability Division.

James A. McLinn CRE, CQE, CMQ/OE., Fellow ASQ
Updated January 2010 - JMREL2@Aol.com
Hanover, Minnesota

Appendix A – A Statistical Argument for Choice of Distribution.

A statistical test was developed to determine if the underling best fit distribution in a life test is the Weibull distribution of the LogNormal distribution [28]. The basis of this statistic is Pearson's correlation co-efficient (Rank Regression on Y will be used here) for each proposed distribution. The ratio of the coefficient for the Weibull distribution to the LogNormal distribution is approximately

Normally distributed. The expected value of this ratio is **greater than one** if the underlying distribution is Weibull and **less than one** if the underlying distribution is LogNormal. This test works best for 10 or more failures and more than half of the sample consumed. The data set for bearings meets these criteria. Each of the stresses will be tested for Weibull versus LogNormal and based upon the results a determination will be made.

87% Weibρ = 0.9208 LogNρ = 0.9681 Ratio = 0.9208/0.9601 = 0.9511

99% Weibρ = 0.9710 LogNρ = 0.9607 Ratio = 0.9710/0.9607 = 1.011

109% Weibρ = 0.9144 LogNρ = 0.9599 Ratio = 0.9144/0.9599 = 0.9526
the 9 data point model was employed for this estimate

119% Weibρ = 0.9785 LogNρ = 0.9830 Ratio = 0.9785/0.9830 = 0.9954

Overall Conclusion - Three of the four stress levels showed the ration to be less than 1.00. Based upon this statistic the best fit for the four bearing data sets would be the LogNormal distribution. When combined with the prior tests, four of seven tests of section 5.7 suggest that the best distribution is Weibull. In this example the difference is so small that it is **hard to distinguish** between either distribution as a best fit.

References

1. Abernethy, Dr. Robert, ***The New Weibull Handbook***, fourth edition, published by Dr Abernethy, 536 Oyster Road, North Palm Beach, Florida 33408-4328 and the SAE, 2000, ISBN 96-202898
2. Jensen, Finn, and Petersen, Niels, ***Burn-in: An Engineering Approach to the Design and Analysis of Burn-in Procedures***, Wiley and sons, New York, 1982
3. O'Connor, P.D.T., ***Practical Reliability Engineering***, fourth edition, Wiley, New York, 2002

4. Dodson, Brian, ***Weibull Analysis with Software***, second edition ASQ, Milwaukee, 2006

5. Weibull on the Web – see Wikipedia, Reliasoft, Relex etc. see below

6. Bain, Lee and Englehardt, Max, "Reliability test Plans for One-Shot Devices based upon repeated Samples", Journal of Quality Technology, Vol. 23, No. 4, pp.304-311, Oct. 1991

7. Halliman, Arthur, "A Review of the Weibull Distribution", Journal of Quality Technology, Vol. 25, No. 2, pp.85-93, April 1993

8. Dodson, Bryan, "Determining the Optimum Schedule for Preventative Maintenance", Quality Engineering, Volume 6, Number 4, pp.667-680, 1995

9. Web reference for Weibull data or information. See http://www.barringer1.com/wdbase.htm

10. Crowder, M.J., Kimber, A.C., Smith R.L. and Sweeting, T.J., ***Statistical Analysis of Reliability Data***, Chapman & Hall, London, 1991

11. Nelson, Wayne, ***Accelerated Testing***, Wiley and Sons, New York, 1990

12. King, James R., **Probability Plots for Decision Making, Technical and Engineering Aids for Management (TEAM)**, PO Box 25 , Tamworth, New Hampshire

13. McLinn, James A., ***Practical Weibull Analysis***, , published by the Reliability Division of ASQ, 2001. Contact the publisher of the Reliability Review for a copy.

14. Elsayed, Elsayed A., ***Reliability Engineering*,** Addison Wesley, New York, 1996

15. Abernethy, R., Breneman, Medlin and Reiman, ***Weibull Analysis Handbook*,** published by Pratt and Whitney Aircraft, November 1983, and issued as AFWAL-TR-83-2079, available from NTIS

16. **Reliasoft *Weibull Handbook*,** 1997, published by Reliasoft, 1450 S. Eastside Loop, Tucson, Arizona 85710-6703

17. Kececioglu, Dr Dimitri, Sun, Feng-Bin, ***Electronic Stress Screening***, Prentice Hall, Saddle River, 1995

18. Campean, I.F., *Exponential Age Sensitive Method for Failure Rank Estimation*, Quality and Reliability Engineering International, Wiley, 2000, Vol. 16, pp. 291-300

19. Wang, Wendai, *Improving the Quality of Field Data Reliability Analysis*, Proceedings of the 2001 RAMS, IEEE, Piscataway, pp. 285-289

20. Rigdon, S.E., and Basu, A.P., *The Power Law Process: A Model for the Reliability of Repairable Systems*, Journal of Quality Technology, Vol. 21, No. 4, pp. 251-260, October 1989

21. Gan and Koehler, *A Goodness-to-Fit Test Based on P-P Probability Plots*, Journal of Quality Technology, Vol. 24, No. 2, pp. 96-102, April 1992

22. Bain, Lee and Englehardt, Max, *Approximate Distributional Results Based on the Maximum Likelihood Estimators for the Weibull Distribution*, The Journal of Quality Technology, Vol. 18, No. 3, 1986, pp. 174-181

23. Leemis, Lawrence, ***Reliability:Probabilistic Models and Statistical Methods***, Prentice-Hall, Englewood Cliffs, 1995

24. Weibull-Ease™, a software program for Weibull analysis that includes a customer use distribution. Available from Applications Research, 4927 St. Croix Ave., Golden Valley, Minnesota 55422 and 763 521-0217

25. McLinn, James A. "*Field Failure Rate, more than you may think*", Proceedings of the 2010 RAMS, IEEE, Piscataway, pp. TBD

26. Juran, Joseph and Gryna, Frank, ***The Quality Control Handbook***, 4 th edition, p. 4.19, McGraw-Hill, New York, 1988

27. Ascher, Harold and Feingold, Harry, ***Repairable System Reliability***, Dekker, New York, 1984

28. Cain, Stephen R., Distinguishing Between LogNormal and Weibull Distributions, IEEE Transactions on Reliability, Vol. 51, #1, March 2002, pp. 32-38

The following **Weibull articles and books** contain examples on the use of a variety of Weibull techniques.

Allmen, C.R. and Lu, M.W., **Sample Size for Failure Detection and Reliability Demonstration**, Proceeding, 1993 Reliability And Maintainability Symposium, pp. 386-389

Bai, D.S., Chun, Y.R., Cha M.S., **Time-censored Ramp Tests with Stress Bound for Weibull Life Distributions**, IEEE Transactions on Reliability, March 1997, Volume 46, # 1, pp. 99-107

Bain, L. and Englehardt, M., **Approximate Distributional Results Based on the Maximum Likelihood Estimators for the Weibull Distribution,** The Journal of Quality Technology, Vol. 18, No. 3, 1986, pp. 174-181

Bain, L. and Englehardt, M., **Reliability Test Plans for One-Shot Devices Based on Repeated Samples,** Journal of Quality Technology, Vol. 23, No. 4, pp. 304-311, Oct. 1991

Campean, I.F., ***Exponential Age Sensitive Method for Failure Rank Estimation***, Quality and Reliability Engineering International, 2000, Vol. 16, Wiley, pp. 291-300

Chan, Victor and Meeker, William Q., **A Failure-Time Model for Infant-Mortality and Wearout Failure Modes**, IEEE Transactions on Reliability, Vol. 48, #4, pp. 377-387

Clark, W.B., **Analysis of Reliability Data for Mechanical Systems**, Proceedings of 1991 RAMS, pp. 438-441

Crow, L.H., **Evaluating the Reliability of Repairable Systems**, Proceedings of the 1990 Annual Reliability and Maintainability Symposium, RAMS, pp. 275-279

Dodson, B., **Determining the Optimum Schedule for Preventative Maintenance**, Quality Engineering Volume 6, Number 4, pp. 667- 680, 1995

Ellis, D.E., **Extended Service Coverage Cost Predictions**, Proceeding of 1990 RAMS, pp. 233-236

Elerath, J., **Analysis of Multiply Censored Run-In Data**, Proceeding of 1990 RAMS, pp. 501-506

Gan and Koehler, **A Goodness-to-Fit Test Based on P-P Probability Plots,** Journal of Quality Technology, Vol. 24, No. 2, pp. 96-102, April 1992

Giuntini, R.E. and Giuntini, M.E., **Simulating a Weibull Posterior Using Bayes Interference**, Proceedings of 1993 RAMS, pp. 48-55

Halliman, Arthur Jr., **A Review of the Weibull Distribution**, Journal of Quality Technology, Vol. 25, No. 2, pp. 85-93, April 1993

Kimura, M., **Oxide Breakdown Mechanism and Quantum Physical Chemistry for Time-Dependent Dielectric Breakdown**, Proceedings of 1997 IEEE International Reliability Physics Symposium, pp. 180-200

Klyatis, Lev, Teskin, Oleg and Fulton, James W., **Multi-Variate Weibull Model for Predicting System-Reliability from Testing Results of the Components**, Proceedings of the RAMS Conference, 2000, Los Angeles, pp. 144-149, Published by IEEE

Lu, C.J. and Meeker, W.Q., **Using Degradation Measures to Estimate a Time-to-Failure Distribution**, Technometrics 35(2), pp. 161-174 (1993), published by ASQ

Mackisak, M.S. and Stillman, R.H., **A Cautionary Tale about Weibull Analysis**, IEEE Transactions on Reliability, June 1996, V 45 #2, pp. 244-248

Rigdon, S.E. and Basu, A.P., **The Power Law Process: A Model for the Reliability of Repairable Systems,** Journal of Quality Technology, Vol. 21, No. 4, pp. 251-260, October 1989

Schneider and Weissfeld, **Interval Estimation Based upon Samples from the Weibull Distribution,** Journal of Quality Technology, Vol. 21, No. 3, pp. 179-186, July 1989

Wang, Wendai, ***Improving the Quality of Field Data Reliability Analysis***, Proceedings of the RAMS Conference, 2001, published by IEEE, pp. 285-289

Usher, J., **A Program to Estimate Repairable Systems Reliability**, Quality Engineering, Vol. 8, #2, Dec. 1995 pp. 323- 328, published by ASQ and Marcel Dekker

Usher, J., **A Computer Program for Component Life Estimation using Multiply Censored Life Data**, Computers in Industrial Engineering, 22(1), pp. 19-36 (1992)

Usher, J., **Case Study: Reliability Models and Misconceptions**, Quality Engineering, Vol. 6, #2, 1993-94 pp. 261- 271, published by ASQ and Marcel Dekker

Vannoy, E.H., **Improving "Mil-HDBK-217 Type" Models for Mechanical Reliability Prediction**, Proceedings of 1990 RAMS, pp. 341-345

Wasserman, G.S., **A Modeling Framework for Relating Robustness Measures with Reliability**, Quality Engineering, Vol. 8, #4, 1996, pp. 681- 692, published by ASQ and Marcel Dekker

Weibull Models, ISBN 978-0-471-36092-6 Wiley, published 2003, 408 pages

Software for Weibull

Weibull-Ease™ Available from Applications Research, 4927 St. Croix Ave., Golden Valley, Minnesota 55422 and 763 521-0217

WinSMITH™ and **BiWinSMITH ™** available from Dr. Abernethy 561 842-4082

and available from Innovative Software Designs, 540 Pellis Rd., Greensburg, Pennsylvania 15601, USA 724 836-8800. Windows compatible for IBM.

Weibull Analyst, Teque Inc., 11686 N. Daniels Dr. Germantown, Wisconsin 53022, 1 - 414 255-7210

Reliability and Maintenance Analysts, sold by Applied Reliability Concepts and Training, 3933 Highway 61 N. Boonville, Indiana 47601, 812 897-0271

Weibull 7.0 ++™ , available from Reliasoft, 1450 S. Eastside Loop. Tucson, Arizona 85710, 1-888-886-0410, offers a variety of software programs for reliability tools and tasks. Also Support@reliasoft.com

Relex Software, 540 Pellis Rd. Greensburg, Pa. 15601, 1 724-836-8800 offers a variety of software programs for reliability tools and tasks. Also info@relex.com

Advanced Logistics Development, ALD, Distributed through SOHAR Inc. 5731 W. Slauson Ave., Suite 175, Culver City, California, 90230 and 1 310 338-0990 info@sohar.com

Web sites for Reliability - Many have software which can be downloaded and tried

http://www.applicationsresearch.com/

http://en.wikipedia.org/wiki/Weibull_distribution

http://www.weibull.com/

http://www.it/.nist.gov/div898/handbook/eda/section3/eda3668.htm

http://www.mathpages.com/home/Kmath122/Kmath122.htm

http://www.qualitydigest.com/jan99/html/body_weibull.html

http://www.barringer1.com/wdbase.htm