



INSTITUTE FOR DEFENSE ANALYSES

Tutorial: Parametric Reliability Models

Yevgeniya K. Pinelis, Project Leader

William R. Whitley

September 2018

Approved for public release.
Distribution is unlimited.

IDA Non-Standard Document
NS D-9171

Log: H 2018-000282



The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Task BD-9-2299(90), "Test Science Applications," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of Yevgeniya K. Pinelis, Keyla Pagan-Rivera, and Jason Sheldon from the Operational Evaluation Division.

For more information:

Yevgeniya K. Pinelis, Project Leader
ypinelis@ida.org • (703) 845-6899

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

Copyright Notice

© 2018 Institute for Defense Analyses
4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (a)(16) [Jun 2013].

INSTITUTE FOR DEFENSE ANALYSES

IDA Non-Standard Document NS D-9171

**Tutorial:
Parametric Reliability Models**

Yevgeniya K. Pinelis, Project Leader

William R. Whitley

Executive Summary

Reliability is the probability that a mechanical, electrical, or software system will not fail after a certain amount of use. In operational testing, analysts frequently evaluate the suitability of a system based on its reliability. Operational tests generally produce failure data consisting of time or distance intervals until the system or one of its components fails. An analyst must then select a statistical model that fits this failure data to project the future reliability of the tested system. The exponential, two-parameter Weibull, and lognormal distributions are three commonly used parametric models of reliability, because their shape and relative simplicity make them suitable to model skewed positive failure data.

This paper is a tutorial that demonstrates how to plot reliability functions parametrically in R using the output of any reliability modeling software. It provides R code and sample plots of reliability and failure rate functions with confidence intervals for the three aforementioned probability distributions, and it provides the mathematical background of the models and recommendations for when to use them. It does not discuss every aspect of these distributions or the full derivations and parameter estimation techniques, although it does present a mathematical method to derive confidence bands for reliability and failure rate. The goal is to enable the reader to create and customize presentation-quality reliability charts for use in reports and briefings more easily using the scripting interface of R, rather than through the menu-based interfaces of other software packages.¹

Selecting a suitable model for a set of data is seldom a trivial decision, but the exponential, Weibull, and lognormal models provide a lot of flexibility for skewed positive data. The exponential model is the simplest and is often adequate to model the reliability of systems or components with constant failure rates. The more general two-parameter Weibull distribution can model reliability for systems or components with increasing or decreasing failure rates over time, and the lognormal distribution has a failure rate function that initially increases and then slowly decreases over time.² Model selection depends on both the anticipated behavior of the system's failure rate, which likely depends on engineering data, and on comparison of statistical criteria. In some cases, multiple models may fit the data sufficiently well that the analyst can choose the simpler option, provided it makes physical sense for the system under test. Although no model can predict reliability perfectly, the methods and code presented in this paper are useful to present reliability analysis for a wide range of hardware and software systems.

¹ The author has developed a web application to generate downloadable high-resolution graphs for the three models discussed in this paper. It is available at <https://test-science.shinyapps.io/ParametricReliabilityModels/>.

² Although less commonly used to model reliability than the first two, the lognormal distribution is useful for some reliability data sets and is also commonly used to model repair time.

Tutorial: Parametric Reliability Models

William R. Whitledge

August 2018

Abstract

This tutorial demonstrates how to plot reliability functions parametrically in R using the output from any reliability modeling software. It provides code and sample plots of reliability and failure rate functions with confidence intervals for three different skewed probability distributions: the exponential, the two-parameter Weibull, and the lognormal. These three distributions are the most common parametric models for reliability or survival analysis. This paper also provides mathematical background for the models and recommendations for when to use them.

Introduction

Reliability is the probability that a mechanical, electrical, or software system will not fail after a certain amount of use. In operational testing, analysts frequently evaluate the suitability of a system based on its reliability. System *use* is typically measured in units of time or distance. An operational test generally produces failure data consisting of time or distance intervals until the system or one of its components fails. An analyst must then select one of several statistical models that fits this failure data to project the future reliability of the tested system. The failure data collected during the operational test is treated as a sample of a failure population consisting of *all* future failures during system operation.

Since intervals between failures must be positive numbers, the group of statistical models appropriate for modeling reliability is restricted to distributions defined on non-negative domains. Skewed distributions also tend to be more appropriate than symmetric ones, because the data sets generally contain a cluster of points near zero with a few points scattered farther from zero in a long tail to the right. Consequently, the three most common parametric models of reliability (exponential, Weibull, and lognormal) share those characteristics. Most failure data fit one or more of these three models reasonably well, and the models are generally well-understood with closed form expressions, few parameters, and physically sensible failure rate functions.

The goal of this paper is to allow the reader to create and customize presentation-quality reliability charts more easily using the scripting interface of R, rather than through the graphical or menu-based interfaces of other software packages. Reliability charts, which plot the probability of no failure occurring through time t or distance x , are often useful in reports, papers, briefings, and analysis products.¹ This paper gives analysts, authors, and briefers a tool that enables customized plotting of reliability and the ability to answer technical questions about their reliability models. I have implemented these three models in a web-based application for reliability analysis and plotting at <https://test-science.shinyapps.io/ParametricReliabilityModels/>, which allows a user to create and download high-resolution images of the model graphs without writing or copying code.

¹In some cases, only the distribution mean is of interest to a particular audience or sponsor. Mean time between failure is a common metric of reliability in Department of Defense equipment acquisition programs, and the graph of reliability as a probability over time is not always requested. However, it is still helpful to plot the reliability curve and confidence band over time to better understand the model used to estimate the mean time between failure.

This paper includes mathematical overviews of the exponential, two-parameter Weibull, and lognormal reliability models and how they are implemented in the R code, but it does not cover every aspect of the probability distributions or the full derivations and parameter estimation techniques. It assumes the reader is familiar with statistics and is comfortable using software to fit different probability distributions to data sets, and it assumes the reader can use R for plotting. However, this paper assumes the reader is not an expert on the three distributions discussed or the maximum likelihood methods of estimating model parameters.²

This paper presents the mathematical background for each model followed by example plots of reliability and failure rate functions, which describe different aspects of failure behavior over time. The tutorial gives instructions and R code for plotting these functions and their confidence intervals in R using the `ggplot2` package. It assumes that the user has basic knowledge of reliability analysis and has already fit a reliability model³ to a set of continuous failure data using another software tool, such as the JMP Life Distribution tool by SAS, and has estimates of the model parameters and covariance matrices available.⁴ The R scripts provided in the appendices of this paper use those parameters to plot reliability curves over any independent variable domain and annotate the plots with the appropriate parametric reliability formulas.⁵

Exponential Model

In reliability analysis, researchers are typically interested in the probability of a system failing (or not failing) after a certain length of time running or distance traveled. They can model the probability of failure as a parametric probability distribution of random variable T , where T could be operating time, distance driven, engine cycles, hard disk drive rotations, or any continuous variable measured between system start and system failure. The probability of failure after t hours is called $F(t)$ and is the cumulative distribution function (CDF) of the model. The reliability $R(t)$ of the system is estimated by $1 - F(t)$.

The frequency of failure after operating for t hours is also of interest. Instantaneous failure rate, also called hazard rate, is described by the failure rate or hazard function $\lambda(t)$. This function is defined as the ratio of the probability density function (PDF) to the reliability function. The failure rate $\lambda(t)$ is always greater than or equal to zero, but it can increase, decrease, or remain constant over time depending on the probability distribution of T .

The exponential model for reliability assumes a *constant* failure rate over time. This means the probability of the system failing after much use is assumed to be identical to the probability of the

²See Casella pp. 315-323 or SAS pp. 1313-1315 for a description of maximum likelihood estimation. Also see ReliaSoft chapters 7, 8, and 10 (pp. 71, 103, and 187) for a more complete description of the three distributions.

³Reliability analysis is mathematically similar to survival analysis. The difference is that the reliability function estimates the probability of not failing after t hours, while the survival function estimates the probability of not dying after t hours.

⁴The models that are graphed in this paper were fit using the JMP Life Distribution tool. This tool can quickly fit and compare multiple probability distributions to a set of failure data, including the exponential, Weibull, and lognormal distributions. Other software packages, such as R and ReliaSoft's Weibull++, also contain reliability modeling tools. See <https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/nls> and <http://www.reliasoft.com/Weibull/index.htm>. Some of the theoretical model discussions below are tailored to the terminology found in JMP output, but the R code provided in the appendices is indifferent to the software used to estimate model parameters.

⁵The independent variable in reliability analysis is generally time (to failure or between failure), but it could also be, for instance, distance traveled or engine cycles (to failure or between failure).

system failing soon after starting to operate. The R code provided in the exponential appendix sets up the parameter $\alpha = 1/\lambda$, where λ is a constant failure rate and α is the mean of the exponential distribution. The code creates data columns to plot the exponential reliability function and its confidence band as a function of time (or any continuous variable). The exponential distribution is a special case of the Weibull distribution in which the Weibull shape parameter β is equal to one; the Weibull distribution with $\beta \neq 1$ is discussed in the next section.

The exponential distribution (ReliaSoft 2015, 71) has a PDF of

$$f(t) = \lambda e^{-\lambda(t-\gamma)} \quad (1)$$

where $f(t) \geq 0$

$\lambda > 0$

$t \geq \gamma$.

The term γ is called the location parameter and is often called the *failure free life* in reliability analysis, because it shifts the distribution to the right to signify that failures can only occur γ hours after operations begin. γ is assumed to be 0 for this paper, because physical systems generally do not operate for periods where the probability of failure is zero. Therefore, the exponential PDF reduces to $f(t) = \lambda e^{-\lambda t}$, giving an exponential CDF (ReliaSoft 2015, 72) of

$$F(t) = 1 - e^{-\lambda t} \quad (2)$$

The CDF is the failure probability function. One minus failure probability gives the probability of not failing, or reliability. More specifically, the reliability curve $R(t)$ is defined as the probability of no failure having occurred after t hours of operation. It does not say anything about when a failure is more or less likely to occur during the interval $[0, t]$; that information is given by the failure rate. In the exponential case, where $\lambda(t) = \lambda$, a failure is exactly as likely to occur after 100 hours of operation as it is after one hour of operation. One minus $F(t)$ gives an exponential reliability function of

$$R(t) = 1 - F(t) = e^{-\lambda t} \quad (3)$$

Taking the ratio of the PDF to reliability gives the failure rate function, or hazard function.

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (4)$$

In the exponential case

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda = \frac{1}{\alpha} \quad (5)$$

Figure 1 contains plots of the probability density, cumulative distribution, reliability, and failure rate for different values of α . Notice how reliability approaches zero more quickly as α decreases and its reciprocal failure rate λ increases.

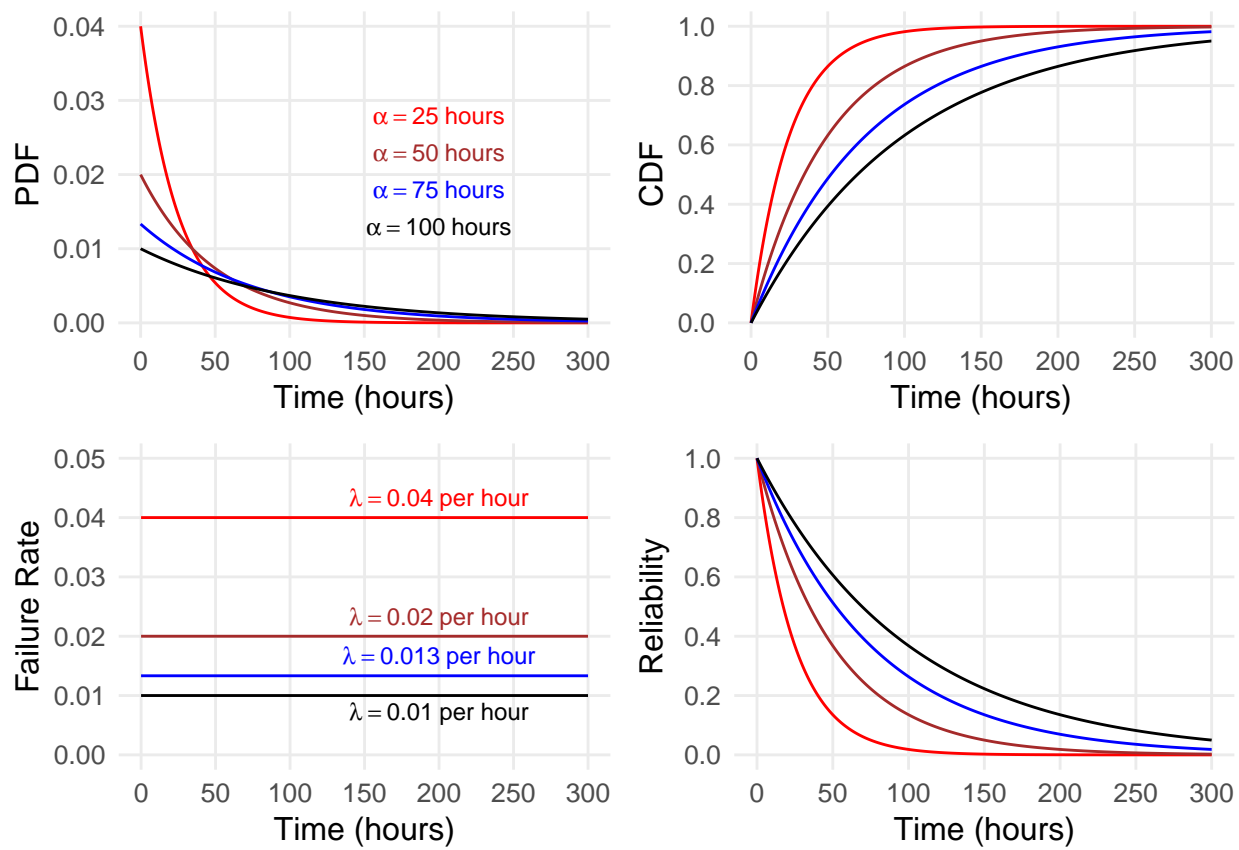


Figure 1: Exponential model plots

As described above, the mean of the exponential distribution (ReliaSoft 2015, 72) is

$$\bar{T} = \alpha = 1/\lambda \quad (6)$$

which is generally the mean time to failure or the mean time between failure in reliability analysis. It could also be the mean distance to failure. Confusion often arises surrounding discussions of the mean time to failure, because reliability analysis usually generates two values that include the word *mean*. One is the distribution mean, or $\bar{T} = \alpha$ for the exponential case; the other is the sample mean of the data points used to model $R(t)$. The sample mean is independent of the distribution used to model the data and is always defined as the familiar arithmetic average of the sample data:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i \quad (7)$$

where t_i is a single time to failure measurement and n is the sample size.⁶ With the exponential reliability model, $\hat{\bar{T}}$ (the estimate of \bar{T}) is equal to $\hat{\alpha}$. And, for the exponential distribution, this estimate is also equal to the sample mean, i.e. $\bar{t} = \hat{\alpha}$. But in general, $\bar{t} \neq \hat{\bar{T}}$. This inequality is true for both the Weibull and lognormal distributions discussed in the next two sections.⁷

Exponential Reliability Function

The exponential distribution and its reliability function are defined by only one parameter. Therefore, plotting the reliability function and its confidence interval is as straightforward as plotting three functions of the form e^{-qt} , where $q = \hat{\lambda}$, λ_L , or λ_U for the parameter estimate and its lower and upper confidence bounds (ReliaSoft 2015, 89). The JMP Life Distribution analysis tool produces an exponential fit that estimates α , or what it calls the *scale* parameter, and its confidence intervals. These are equivalent to λ for plotting the exponential reliability function.⁸

The exponential model code creates a data frame with time t varying from 0 to 600 hours. $\hat{\alpha}$, α_L , and α_U are set to 366, 137.4, and 975.2 hours, respectively, based on an exponential fit of real time between failure data in the JMP Life Distribution tool using a confidence level of 95 percent. This means α_L and α_U are the lower and upper 95 percent confidence bounds, respectively, for the estimate of α . JMP uses a maximum likelihood estimation (MLE) algorithm to estimate α and its confidence levels (Casella and Berger 2002, 315–23). For each point t , the R script calculates $\hat{R}(t)$, $R_L(t)$, and $R_U(t)$.

$$\hat{R}(t) = e^{-t/\hat{\alpha}} = e^{-t/366} \quad (8)$$

⁶This paper does not discuss data censoring, but the R scripts provided in the appendices are indifferent to the use of censoring in estimating the model parameters used to plot reliability.

⁷It is good practice to report not only the sample mean of a set of data, but also the distribution mean based on the estimated distribution parameters. By fitting a distribution to a set of data and presenting the reliability curve for that distribution, a researcher implicitly accepts the hypothesis that the sample data came from a population described by that distribution. Therefore, he should also report the estimate of the distribution mean.

⁸Different software packages may use different methods for fitting models, and they may report the same parameter under a different name or as an inverse of another parameter: α vs. λ , etc. It is important to be sure which parameter or parameter name the reliability modeling software uses.

Lower confidence bound:

$$R_L(t) = e^{-t/\alpha_L} = e^{-t/137.4} \quad (9)$$

Upper confidence bound:

$$R_U(t) = e^{-t/\alpha_U} = e^{-t/975.2} \quad (10)$$

Maximum likelihood theory provides estimates of the upper and lower confidence bounds on the estimate of α . In general, if h is a function of a parameter α (for instance $h = \alpha$ or $h = \lambda = 1/\alpha$), then the upper and lower confidence limits on \hat{h} can be calculated by $h_U = \hat{h} + K_\delta \sqrt{Var(\hat{h})}$ and $h_L = \hat{h} - K_\delta \sqrt{Var(\hat{h})}$, where \hat{h} is assumed to be normally distributed and K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level. If h is constrained to positive values, as in the exponential model where $h = \alpha$, then the natural logarithm, $\ln \hat{h}$, is assumed to be normally distributed instead of \hat{h} , and the upper and lower confidence limits are given instead by $h_U = \hat{h} e^{K_\delta \sqrt{Var(\hat{h})}/\hat{h}}$ and $h_L = \hat{h} e^{-K_\delta \sqrt{Var(\hat{h})}/\hat{h}}$. Therefore, if $\hat{h} = \hat{\alpha}$, then $\alpha_U = \hat{\alpha} e^{K_\delta \sqrt{Var(\hat{\alpha})}/\hat{\alpha}}$ and $\alpha_L = \hat{\alpha} e^{-K_\delta \sqrt{Var(\hat{\alpha})}/\hat{\alpha}}$. Likewise, if $\hat{h} = \hat{\lambda}$, then $\lambda_U = \hat{\lambda} e^{K_\delta \sqrt{Var(\hat{\lambda})}/\hat{\lambda}}$ and $\lambda_L = \hat{\lambda} e^{-K_\delta \sqrt{Var(\hat{\lambda})}/\hat{\lambda}}$ (Nelson 1990, 361–63).

When h is a continuous function of a single parameter α , the variance of \hat{h} is related to the variance of the parameter by

$$Var(\hat{h}) = \left(\frac{\partial h}{\partial \alpha} \bigg|_{\hat{\alpha}} \right)^2 Var(\hat{\alpha}) \quad (11)$$

The coefficient of $Var(\hat{\alpha})$ is the square of the partial derivative of h with respect to α evaluated at the maximum likelihood estimate of α , which is $\hat{\alpha}$. Equation 11 is important for calculating the more complicated reliability confidence intervals for the Weibull and lognormal models. For the exponential model it is sufficient to know that R, JMP, and other model fitting software provide numerical estimates of λ or α and their respective confidence limits and variances (The JMP Life Distribution tool estimates α , while the R `fitdistrplus` package estimates λ).⁹

Additional R code in the exponential appendix plots the exponential reliability function and 95 percent confidence band versus time using `ggplot2`. It also annotates the plot with the parametric reliability equation $\hat{R}(t) = e^{-t/\hat{\alpha}}$, replacing $\hat{\alpha}$ dynamically with its numeric value. Figure 2 is the plot of the exponential reliability model. I generally recommend plotting only a single curve and confidence band on one set of axes to avoid clutter or confusion. If you want to show two or more reliability curves together (perhaps for comparable components or systems or multiple failure types), consider making two or four small graphs of the same size. Replicate the R code two or four times, or for however many plots are needed: `plot1`, `plot2`, `plot3`, Then call the R function `grid.arrange(plot1, plot2, plot3, plot4, ncol=2, nrow=2)` for a 2×2 grid.¹⁰ If you must

⁹Note that the R code is indifferent to which software (JMP, R, WEIBULL++, MATLAB, etc.) was used to estimate the model parameters, but the code is designed to accommodate the standard output of the JMP Life Distribution tool, which provides estimates in terms of the mean time to failure, α . Take care to paste the correct values into the R code parameter variables in the appendix if you use another software program to fit a distribution.

¹⁰This function is available in the R library ‘gridExtra’. See <https://cran.r-project.org/web/packages/gridExtra/vignettes/arrangeGrob.html>.

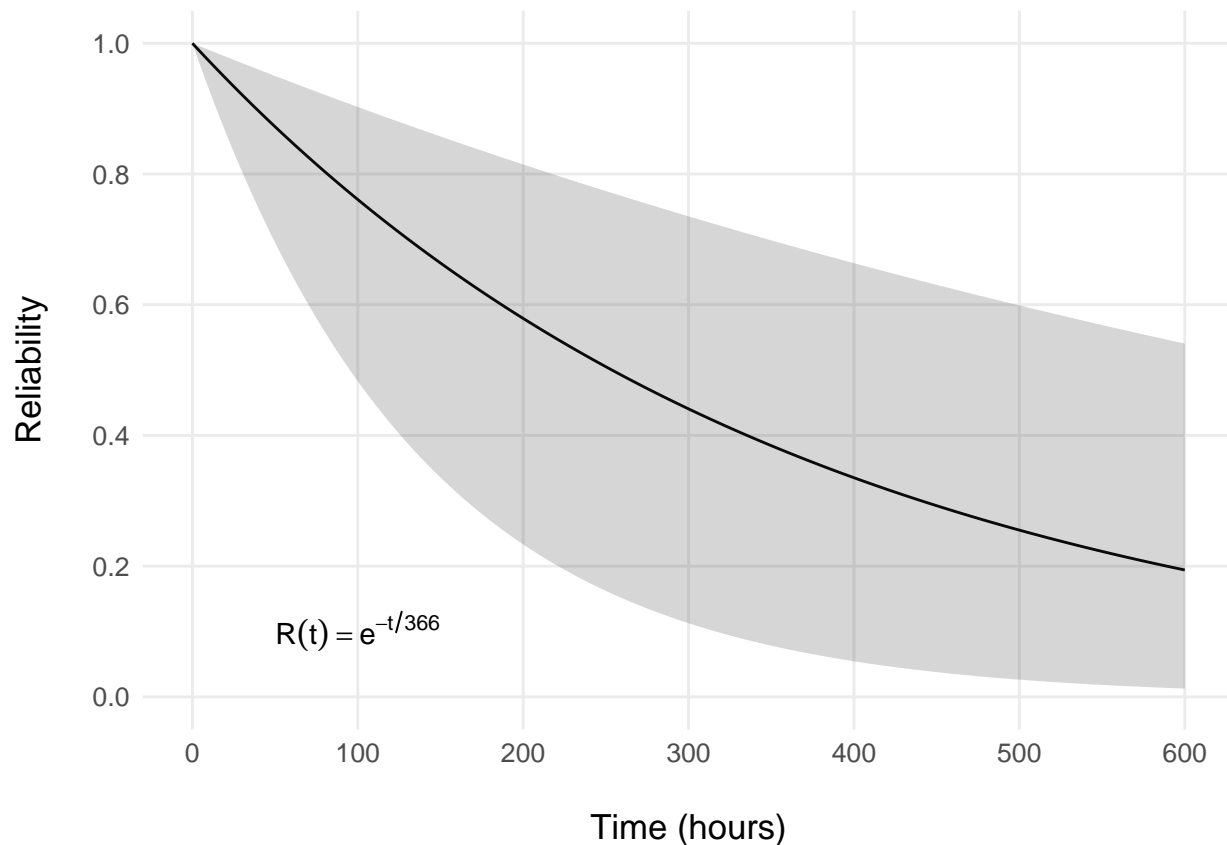


Figure 2: Exponential reliability model

show multiple curves on a single set of axes, consider dropping the confidence bands if the graph becomes too cluttered with overlapping bands.

If you only want to plot the first 100 hours of operating time, reduce the limits in the plotting function: `limits = c(0, 600)` becomes `limits = c(0, 100)`. Figure 3 shows the reliability model zoomed to these limits. Viewing the entire reliability curve ranging from $\hat{R} = 1$ to near $\hat{R} = 0$ is informative in understanding the whole model shape, but if users are expected to use the system under test for only a few hours, then consider only graphing $\hat{R}(t)$ over a shorter domain interval: $0 < t < 100$ hours, for instance. Be careful not to mislead your audience by showing a domain that is too narrow without providing the context of performance over a wider domain. Also, resist the temptation to plot $\hat{R}(t)$ over a domain that greatly surpasses the largest data point, $MAX(t_i)$, because it is difficult to estimate $\hat{R}(t)$ far away from the actual data. Parametric plotting is a powerful tool, but it is easy to overuse it when a script enables rapid arbitrary changes to the axis limits and rapid re-plotting.

Exponential Hazard Function

An analyst can also plot the estimated failure rate for the fitted exponential model along with its 95 percent confidence band. Recall from equation 5 that the failure rate function for the exponential model is just $\lambda(t) = \lambda = 1/\alpha$, which is constant over all t . The lower and upper confidence levels for

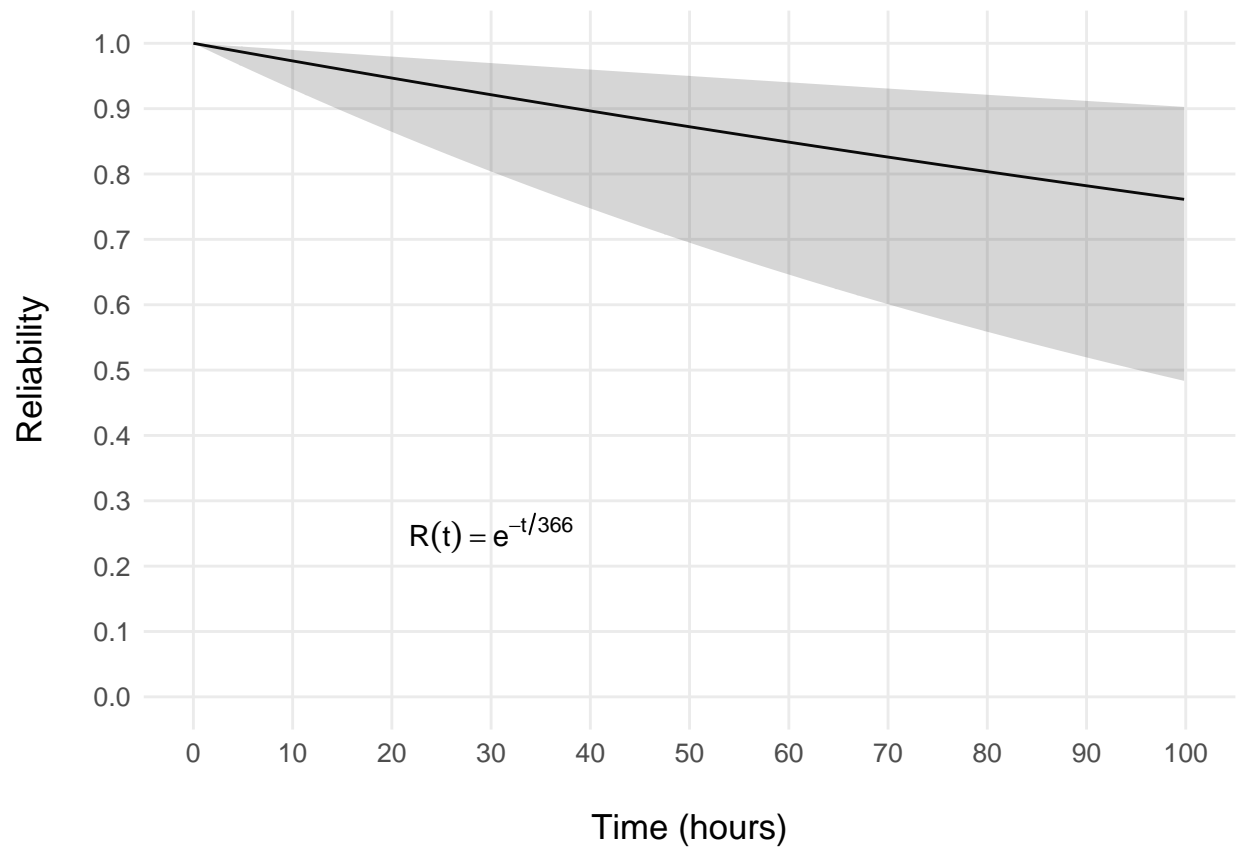


Figure 3: Exponential reliability model (zoomed)

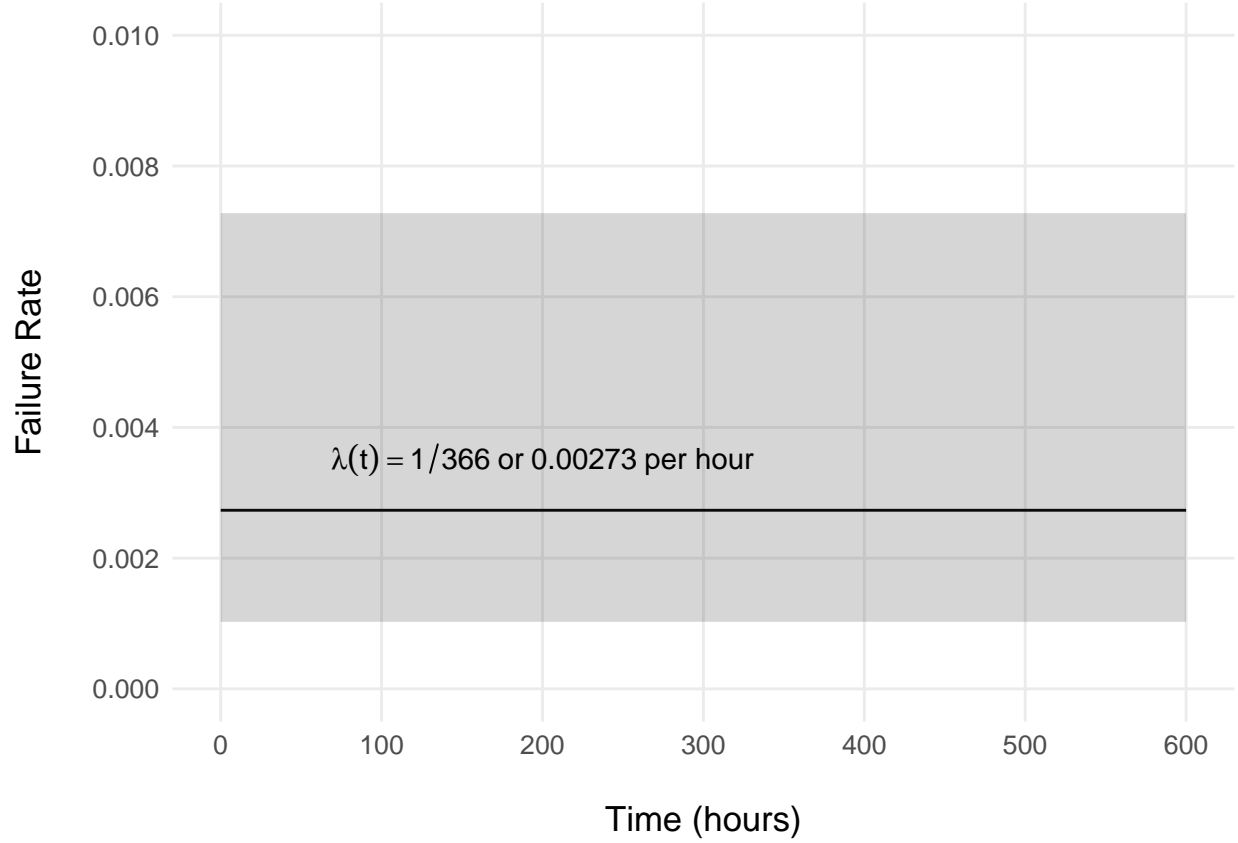


Figure 4: Exponential failure rate model

the exponential failure rate are also constants (Nelson 1990, 321).

$$\hat{\lambda} = 1/\hat{\alpha} \quad (12)$$

Lower confidence bound:

$$\lambda_L = 1/\alpha_U \quad (13)$$

Upper confidence bound:

$$\lambda_U = 1/\alpha_L \quad (14)$$

The final block of R code in the exponential model appendix plots the failure rate and its confidence band over 600 hours and annotates the graph with the estimated hazard function $\hat{\lambda} = 1/\hat{\alpha}$, replacing $\hat{\alpha}$ dynamically with its numeric value. Figure 4 shows the failure rate for the exponential model. The axes limits are adjustable like those of the reliability graph.

Weibull Model

Unlike the exponential model, the Weibull model for reliability does not assume a constant failure rate over time. Depending on the parameter values, the model may show the system under test becoming more or less likely to fail as it is used repeatedly. The R code in the Weibull model appendix sets up the parameters of the two-parameter Weibull distribution, and creates data columns to plot the Weibull reliability function and its confidence band as a function of time (or any continuous variable T).

The general Weibull distribution has three parameters, α , β and γ , which are called the *scale*, *shape* (or *slope*), and *location* (or *failure free life*). Again, γ is assumed to be 0, because physical systems generally do not operate for periods where the probability of failure is zero. Note that in the case of the exponential distribution, I assume the Weibull shape parameter $\beta = 1$. I do not assume that for the general Weibull case, and the model fitting software fits both $\alpha > 0$ and $\beta > 0$ without further constraint. The probability density function (ReliaSoft 2015, 103) for the full three-parameter Weibull distribution is

$$f(t) = \left(\frac{\beta}{\alpha}\right) \left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} e^{-(\frac{t-\gamma}{\alpha})^\beta} \quad (15)$$

where $f(t) \geq 0$, $t > \gamma$

$$\alpha > 0$$

$$\beta > 0$$

$$-\infty < \gamma < \infty.$$

Since I assume $\gamma = 0$, the PDF reduces to the two-parameter Weibull distribution

$$f(t) = \left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(\frac{t}{\alpha})^\beta} \quad (16)$$

which is more commonly used. The two-parameter Weibull CDF (ReliaSoft 2015, 105) is

$$F(t) = 1 - e^{-(\frac{t}{\alpha})^\beta} \quad (17)$$

and the resulting reliability function is

$$R(t) = 1 - F(t) = e^{-(\frac{t}{\alpha})^\beta} \quad (18)$$

The Weibull hazard function (ReliaSoft 2015, 105) is defined as

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \quad (19)$$

which, in the exponential case of $\beta = 1$, reduces to $\lambda(t) = \frac{1}{\alpha} = \lambda = \text{constant}$. The change in failure rate with respect to time is also interesting for the Weibull model. Taking the derivative of λ with respect to t gives

$$\frac{d\lambda}{dt} = (\beta - 1) \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \frac{1}{t} = \frac{\beta - 1}{t} \lambda(t) \quad (20)$$

where α , β , t , and $\lambda(t)$ are all positive. If $\beta = 1$ (exponential case), then $\frac{d\lambda}{dt} = 0$, meaning the failure rate is constant. If $\beta < 1$, then $\frac{d\lambda}{dt} < 0$, meaning the failure rate is decreasing with t . If $\beta > 1$, then $\frac{d\lambda}{dt} > 0$, meaning the failure rate is increasing with t .

When deciding whether to use the exponential model or the two-parameter Weibull model of reliability, you must weigh the exponential model's simplicity against the Weibull model's ability to account for a changing failure rate. Software and network-based information systems that operate mainly on servers and circuits, which tend to wear out slowly, may have a constant or nearly constant failure rate. The failure rate may decrease or increase with code modifications, hardware upgrades, or virtualization of servers. Systems like vehicles or aircraft with many mechanical parts that are prone to physical wear may have failure rates that increase with time, distance, or use. Other systems may exhibit decreasing failure rates over time where increased use actually improves the performance of the system. And for some systems, it may be appropriate to divide the lifespan of the system into segments where each segment has its own model of reliability with different failure rate functions.¹¹ Choosing a reliability model with a physically logical failure rate function will increase the usefulness of the model.

Figure 5 contains plots of the probability density, cumulative distribution, reliability, and failure rate for $\alpha = 100$ hours and different values of β , which is unitless.

The mean of the Weibull distribution (ReliaSoft 2015, 104) is defined as

$$\bar{T} = \gamma + \alpha \Gamma\left(\frac{1}{\beta} + 1\right) \quad (21)$$

where $\Gamma(\frac{1}{\beta} + 1)$ is the gamma function of $(\frac{1}{\beta} + 1)$. The gamma function is defined as

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx \quad (22)$$

Setting $\gamma = 0$ for this paper gives the two-parameter Weibull distribution mean of $\bar{T} = \alpha \Gamma(\frac{1}{\beta} + 1)$. In the exponential case where $\beta = 1$, $\Gamma(\frac{1}{\beta} + 1)$ evaluates to $\Gamma(2) = 1$ and $\bar{T} = \alpha$ as discussed above. The estimated Weibull distribution mean that comes from fitting a Weibull distribution to a set of data, represented by $\hat{\bar{T}} = \hat{\alpha} \Gamma(\frac{1}{\hat{\beta}} + 1)$, will probably not be equal to the sample mean $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$.

The Weibull fit in the JMP Life Distribution tool assumes $\beta \neq 1$ and provides estimates and confidence intervals of five values: *location*, *scale*, *Weibull α* , *Weibull β* , and *Mean*. Confusion can arise here, because the relation of these numbers to each other is not obvious. The JMP location

¹¹This approach can sometimes lead to a compound u-shaped failure rate curve with regions of decreasing, constant, and increasing rate of failure. This shape is sometimes called the bathtub curve.

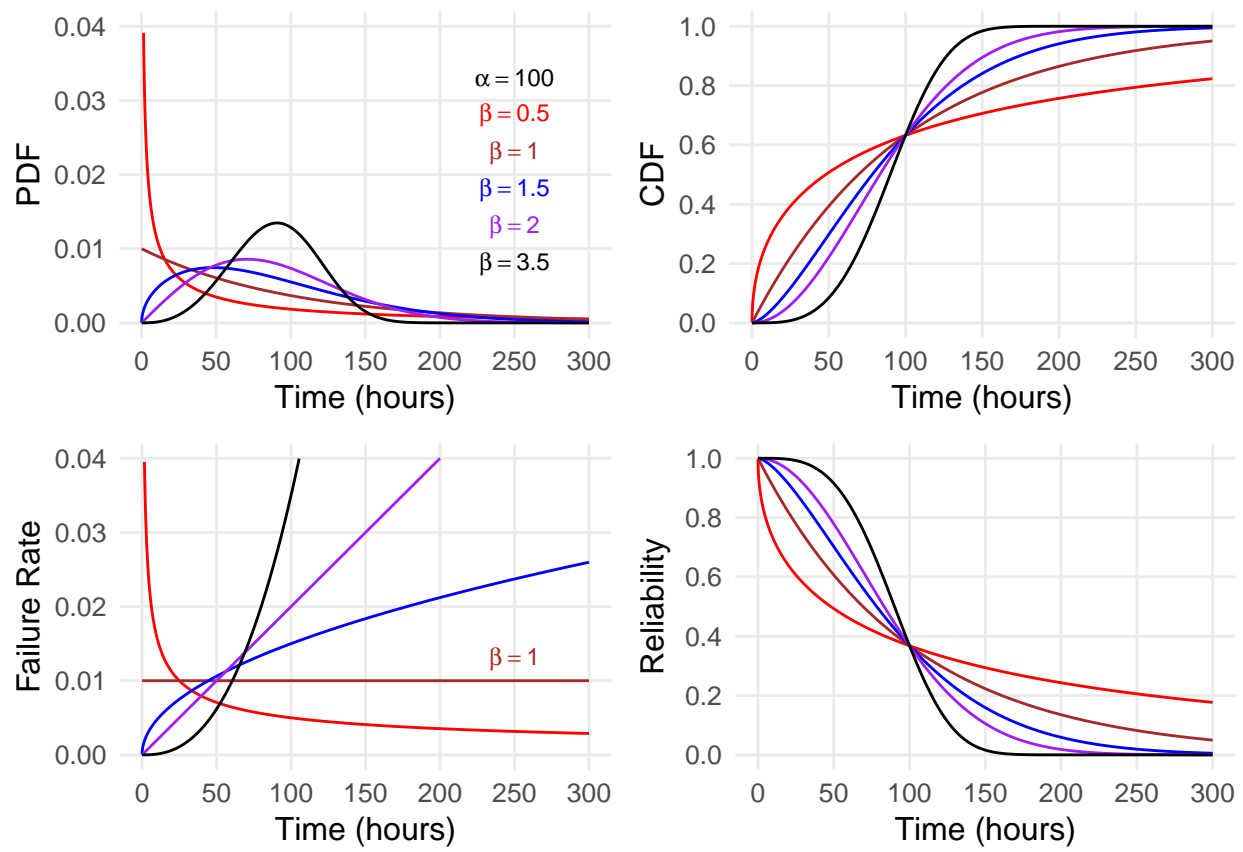


Figure 5: Weibull model plots

parameter is not the same as the Weibull location parameter γ described above. In JMP, *Weibull* $\alpha = e^{location}$ is the estimate of the Weibull scale parameter α . And JMP *Weibull* $\beta = \frac{1}{scale}$ is the estimate of the Weibull shape parameter β . These parameters that the JMP Life Distribution tool calls location and scale are important, because they are necessary to calculate the confidence interval on reliability.¹²

Weibull Reliability Function

As shown in equation 18, the Weibull reliability function is $R(t) = e^{-(\frac{t}{\alpha})^\beta}$, and the goal is to construct reliability boundaries $R_U(t)$ and $R_L(t)$ (ReliaSoft 2015, 122–26). By defining $u = \beta(\ln t - \ln \alpha)$ I can rewrite reliability as $R(t) = e^{-e^{\ln(\frac{t}{\alpha})^\beta}} = e^{-e^{\beta(\ln t - \ln \alpha)}} = e^{-e^u}$, and the reliability confidence levels can be written as

Upper bound:

$$R_U = e^{-e^{u_U}} \quad (23)$$

Lower bound:

$$R_L = e^{-e^{u_L}} \quad (24)$$

where u_U and u_L are the upper and lower confidence bounds on u . These bounds are estimated by:

$$\text{Upper bound: } u_U = \hat{u} + K_\delta \sqrt{Var(\hat{u})}$$

$$\text{Lower bound: } u_L = \hat{u} - K_\delta \sqrt{Var(\hat{u})}$$

where $\hat{u} = \hat{\beta}(\ln t - \ln \hat{\alpha})$ is the estimate of u , and K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level. I assume that \hat{u} is normally distributed with respect to α and β (Nelson 1990, 374). For a 95 percent confidence interval, choose $\delta = 0.95$, which gives $K_\delta = -1.96$.¹³

The variance of \hat{u} is estimated by

$$Var(\hat{u}) = \frac{\hat{\beta}^2}{\hat{\alpha}^2} Var(\hat{\alpha}) + \frac{\hat{u}^2}{\hat{\beta}^2} Var(\hat{\beta}) - \frac{2\hat{u}}{\hat{\alpha}} Cov(\hat{\alpha}, \hat{\beta}) \quad (25)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of α and β provided by JMP and $\hat{u} = \hat{\beta}(\ln t - \ln \hat{\alpha})$ (ReliaSoft 2015, 122–26). This equation is derived from a two-parameter version¹⁴ of equation 11, which says that if \hat{u} is a function of two parameters α and β , then

¹²JMP can cause additional confusion, because the JMP Parametric Survival Fit tool calls these same *location* and *scale* parameters *Intercept* and δ , respectively. The Parametric Survival Fit tool is not needed to plot reliability curves if the Life Distribution tool is used.

¹³For an 80 percent confidence band, choose $\delta = 0.80$ and $K_\delta = -1.28$ as found in a standard normal distribution table.

¹⁴In general, one can estimate a continuous function $h = h(\mu, \sigma)$ as $\hat{h} = h(\hat{\mu}, \hat{\sigma})$, where \hat{h} is assumed to be close to normally distributed with respect to its parameters. This is called the Invariance Property of Maximum Likelihood estimators. The variance of \hat{h} can then be estimated by $Var(\hat{h}) = (\frac{\partial h}{\partial \mu}|_{\hat{\mu}, \hat{\sigma}})^2 Var(\hat{\mu}) + (\frac{\partial h}{\partial \sigma}|_{\hat{\mu}, \hat{\sigma}})^2 Var(\hat{\sigma}) + 2(\frac{\partial h}{\partial \mu}|_{\hat{\mu}, \hat{\sigma}})(\frac{\partial h}{\partial \sigma}|_{\hat{\mu}, \hat{\sigma}})Cov(\hat{\mu}, \hat{\sigma})$ assuming the partial derivatives are continuous functions near $\hat{\mu}$ and $\hat{\sigma}$.

$$Var(\hat{u}) = \left(\frac{\partial u}{\partial \alpha} \Big|_{\hat{\alpha}, \hat{\beta}} \right)^2 Var(\hat{\alpha}) + \left(\frac{\partial u}{\partial \beta} \Big|_{\hat{\alpha}, \hat{\beta}} \right)^2 Var(\hat{\beta}) + 2 \left(\frac{\partial u}{\partial \alpha} \Big|_{\hat{\alpha}, \hat{\beta}} \right) \left(\frac{\partial u}{\partial \beta} \Big|_{\hat{\alpha}, \hat{\beta}} \right) Cov(\hat{\alpha}, \hat{\beta}) \quad (26)$$

The variance and covariance values for the Weibull parameters are also required to calculate $Var(\hat{u})$, but JMP does not provide these directly. The Life Distribution tool instead provides the covariance matrix of the estimates of the JMP *location* and *scale* parameters and *not* the Weibull location and scale parameters (γ and α):

$$\begin{bmatrix} Var(location) & Cov(location, scale) \\ Cov(location, scale) & Var(scale) \end{bmatrix} = \begin{bmatrix} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) \\ Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma}) \end{bmatrix} \quad (27)$$

where the I call the JMP *location* and *scale* parameters μ and σ . To obtain the covariance matrix for the Weibull α and β parameters, the following conversion formulas apply (SAS 2011, 1319):

$$Var(\hat{\alpha}) = [\exp(\hat{\mu})]^2 Var(\hat{\mu})$$

$$Var(\hat{\beta}) = \frac{Var(\hat{\sigma})}{\hat{\sigma}^4}$$

$$Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\exp(\hat{\mu})}{\hat{\sigma}^2} Cov(\hat{\mu}, \hat{\sigma})$$

Now the covariance matrix elements for the Weibull parameters α and β can be written as

$$\begin{bmatrix} Var(\hat{\alpha}) & Cov(\hat{\alpha}, \hat{\beta}) \\ Cov(\hat{\alpha}, \hat{\beta}) & Var(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} [\exp(\hat{\mu})]^2 Var(\hat{\mu}) & -\frac{\exp(\hat{\mu})}{\hat{\sigma}^2} Cov(\hat{\mu}, \hat{\sigma}) \\ -\frac{\exp(\hat{\mu})}{\hat{\sigma}^2} Cov(\hat{\mu}, \hat{\sigma}) & \frac{Var(\hat{\sigma})}{\hat{\sigma}^4} \end{bmatrix} \quad (28)$$

This covariance matrix (either for the JMP *location* and *scale* or *Weibull* α and β parameters) is the inverse of the *observed Fisher information matrix* (SAS 2011, 1313–23). It is defined as the inverse of the negative matrix of partial second derivatives of the log likelihood function of the Weibull parameters, evaluated at the final parameter estimates that maximize likelihood. Maximum likelihood estimation (MLE) is a common iterative technique for estimating statistical parameters, but it is beyond the scope of this paper to explain fully.¹⁵ It is sufficient to know that it produces a covariance matrix of the parameters that allows the analyst to calculate $Var(\hat{u})$ and the upper and lower confidence bounds on u .

Now an analyst can take the parameters provided by the JMP Life Distribution (Weibull distribution) and parametrically plot reliability and its confidence band as a function of t . The R code in the Weibull model appendix generates data columns for t , $\hat{R}(t)$, $R_L(t)$, and $R_U(t)$. Again, the code is indifferent to which software (JMP, R, WEIBULL++, MATLAB, etc.) was used to estimate the model parameters, but the code is designed to accommodate the standard output of the JMP Life Distribution tool.

Figure 6 contains the Weibull reliability model and 95 percent confidence band versus time. The graph is annotated with the parametric reliability equation $\hat{R}(t) = e^{-(\frac{t}{\alpha})^\beta}$. As discussed, I generally recommend plotting only a single curve and confidence band on one set of axes unless the two curves are spaced far enough apart to avoid looking cluttered. The code for plotting Figure 6 is given in the Weibull model appendix.

¹⁵See Casella pp. 315-323 or SAS pp. 1313-1315.

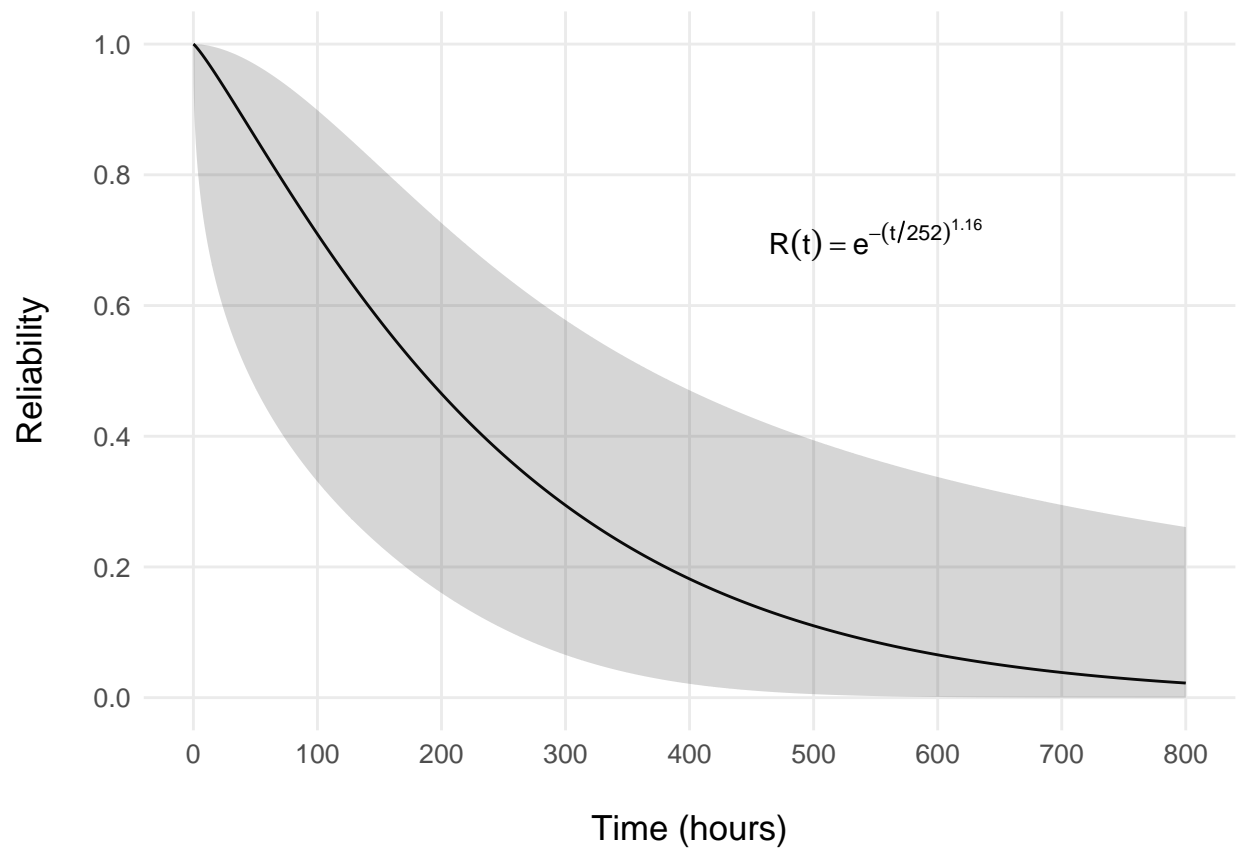


Figure 6: Weibull reliability model

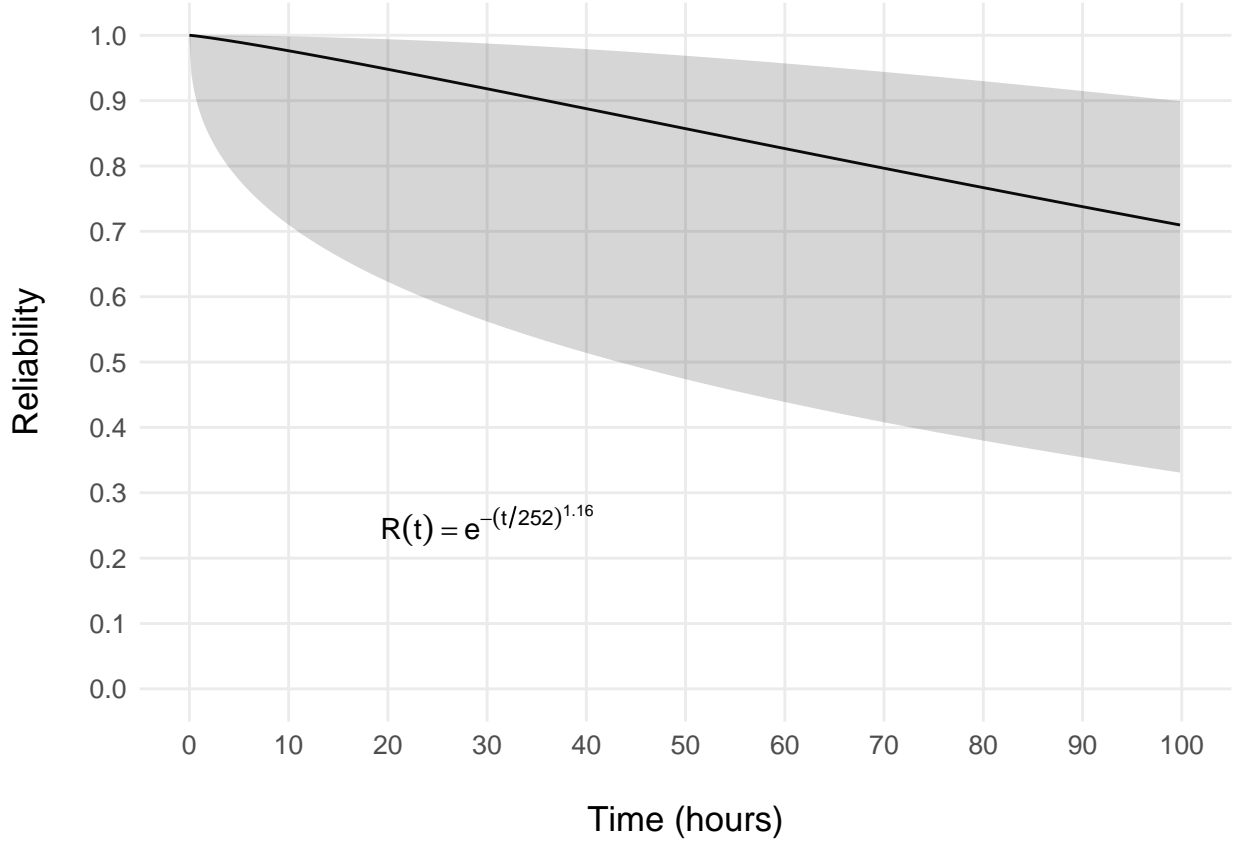


Figure 7: Weibull reliability model (zoomed)

If you only want to plot reliability over the first 100 hours, reduce the limits in the plotting function: `limits = c(0, 800)` becomes `limits = c(0, 100)`. Figure 7 shows the plot of reliability over this interval. As discussed under the exponential case, you may want to do this for presentation purposes if a typical user only uses the system under test for a mission that is a few hours in duration. In that case, you may not want to present $\hat{R}(t)$ for t much greater than the typical mission time. Remember, reliability is defined as the probability of no failure having occurred after t hours of operation, so quoting a reliability percentage after 50 hours of use might not be useful for a system that is only used for 30 minutes at a time. Also recall that plotting $\hat{R}(t)$ over a domain that greatly exceeds $MAX(t_i)$ is risky, because inference beyond the range of data is not reliable.

Weibull Hazard Function

An analyst can also plot the failure rate for the fitted Weibull model along with its 95 percent confidence band. To generate the confidence band for $\lambda(t)$, I take a similar approach to the method used above to generate the upper and lower confidence bounds on $R(t)$ (Nelson 1990, 356–75). Recall from equation 19 that

$$\lambda(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} \quad (29)$$

Taking the natural logarithm of λ gives

$$v(t) = \ln \lambda(t) = \ln \left[\frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} \right] \quad (30)$$

where $\lambda(t) = e^{v(t)}$. Rewriting this equation using the rules of logarithms gives

$$v(t) = (\beta - 1) \ln t + \ln \beta - \beta \ln \alpha \quad (31)$$

Now the upper and lower confidence levels of the hazard function become

Upper bound:

$$\lambda_U = e^{v_U} \quad (32)$$

Lower bound:

$$\lambda_L = e^{v_L} \quad (33)$$

where v_U and v_L are the upper and lower confidence bounds on v . These bounds can be estimated by the relations:

$$\text{Upper bound: } v_U = \hat{v} - K_\delta \sqrt{\text{Var}(\hat{v})}$$

$$\text{Lower bound: } v_L = \hat{v} + K_\delta \sqrt{\text{Var}(\hat{v})}$$

where $\hat{v} = (\hat{\beta} - 1) \ln t + \ln \hat{\beta} - \hat{\beta} \ln \hat{\alpha}$ is the estimate of v and K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level. Choose $\delta = 0.95$ for a 95 percent confidence interval, which gives $K_\delta = -1.96$.¹⁶ The variance of \hat{v} is found by applying equation 26 (Nelson 1990, 374):

$$\text{Var}(\hat{v}) = \left(\frac{\partial v}{\partial \alpha} \Big|_{\hat{\alpha}, \hat{\beta}} \right)^2 \text{Var}(\hat{\alpha}) + \left(\frac{\partial v}{\partial \beta} \Big|_{\hat{\alpha}, \hat{\beta}} \right)^2 \text{Var}(\hat{\beta}) + 2 \left(\frac{\partial v}{\partial \alpha} \Big|_{\hat{\alpha}, \hat{\beta}} \right) \left(\frac{\partial v}{\partial \beta} \Big|_{\hat{\alpha}, \hat{\beta}} \right) \text{Cov}(\hat{\alpha}, \hat{\beta}) \quad (34)$$

or, after taking the first partial derivatives and squaring

$$\text{Var}(\hat{v}) = \frac{\hat{\beta}^2}{\hat{\alpha}^2} \text{Var}(\hat{\alpha}) + \left(\ln t + \frac{1}{\hat{\beta}} - \ln \hat{\alpha} \right)^2 \text{Var} \hat{\beta} - \frac{2\hat{\beta}}{\hat{\alpha}} \left(\ln t + \frac{1}{\hat{\beta}} - \ln \hat{\alpha} \right) \text{Cov}(\hat{\alpha}, \hat{\beta}) \quad (35)$$

Equation 35 is sufficient to calculate confidence bounds on the failure rate. The R code for the Weibull failure rate plots the failure rate and its confidence band over 800 hours and over 100 hours in Figures 8 and 9, respectively, and it annotates each graph with the estimated hazard function $\lambda(t)$.

Observe that since the estimate of $\hat{\beta}$ is slightly greater than one for this Weibull model, the failure rate increases gradually over time t . Also note that the confidence band for the failure rate is

¹⁶I assume the function \hat{v} is approximately normally distributed.

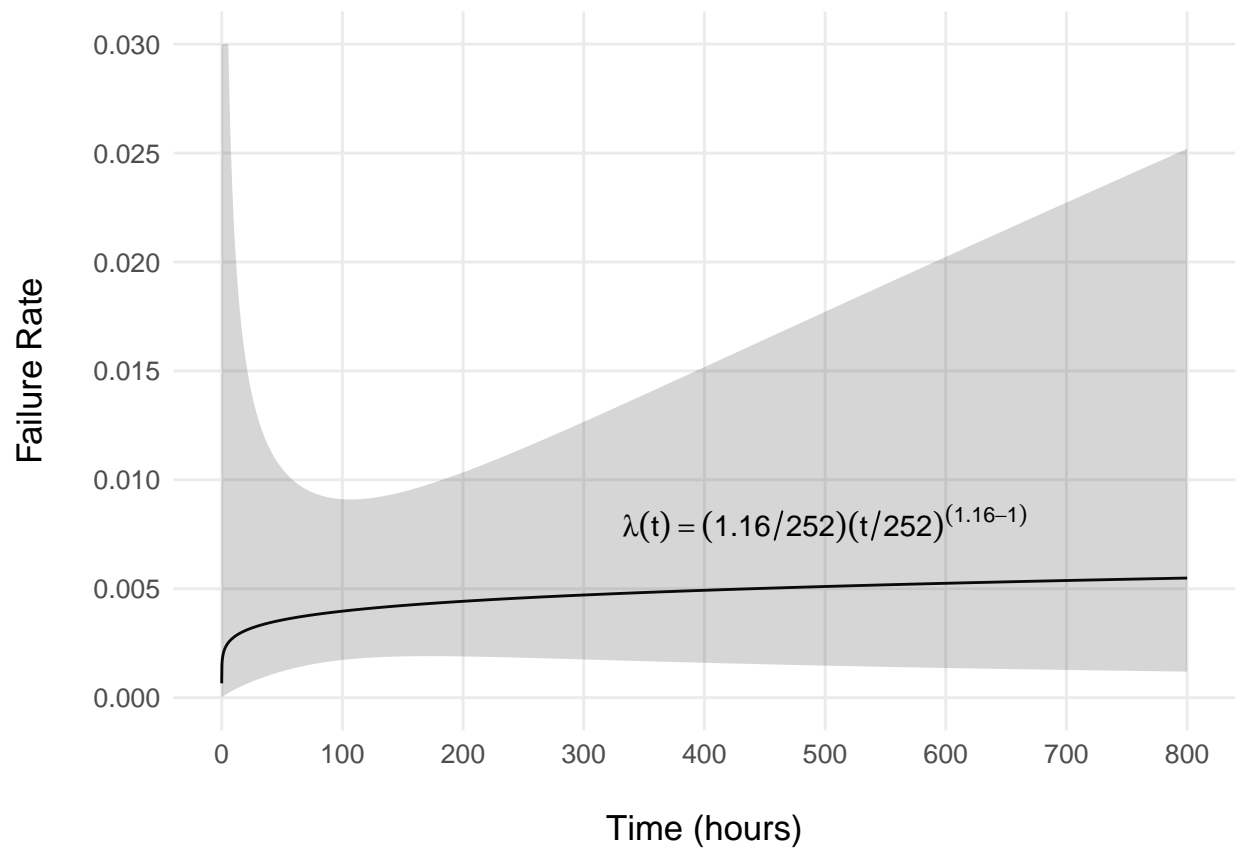


Figure 8: Weibull failure rate model

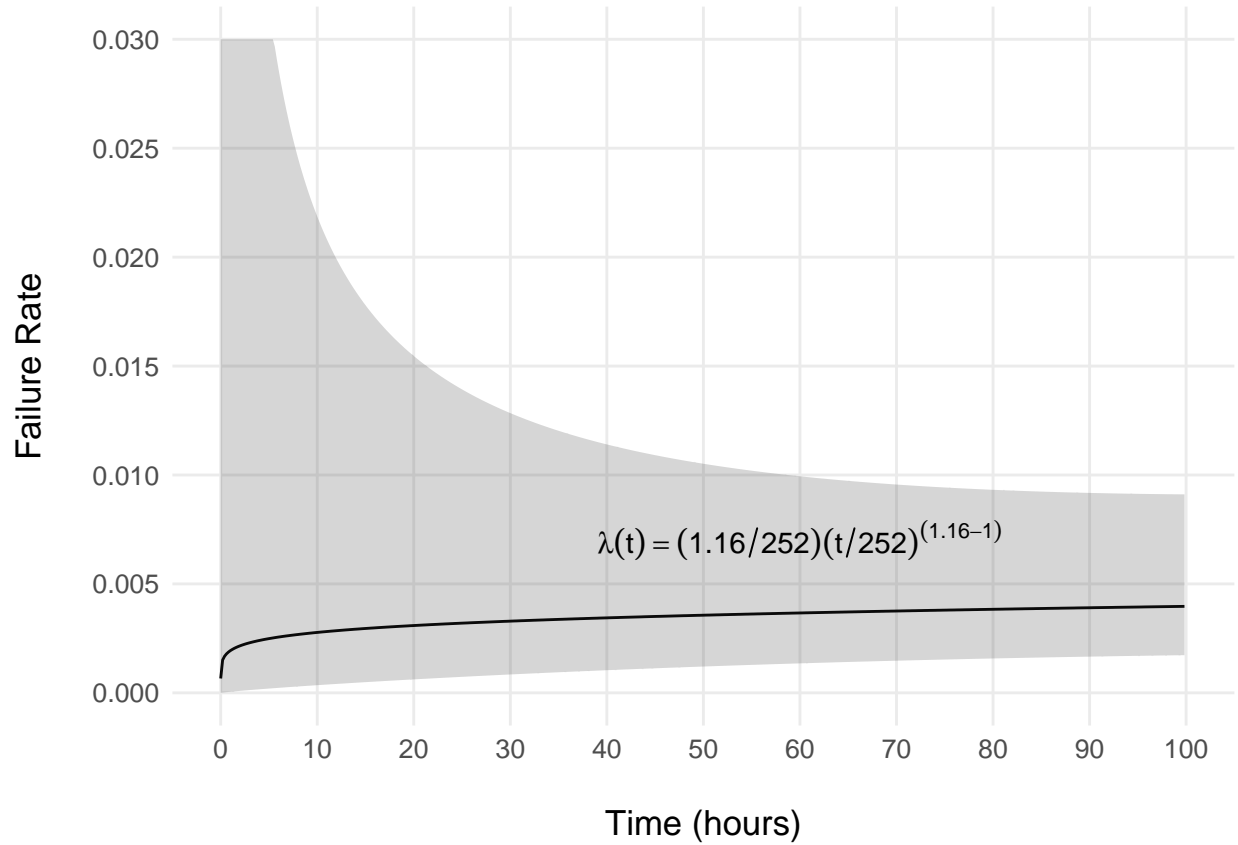


Figure 9: Weibull failure rate model (zoomed)

relatively wide near $t = 0$, meaning the uncertainty of the failure rate estimate a short time after system startup is relatively high. This makes sense for a system that is unlikely to fail within a short time after startup, because there is generally little failure data to support a precise estimate of the failure rate in the region of t near zero. To plot the failure rate over a shorter domain, reduce the limits in the plotting function: `limits = c(0, 800)` becomes `limits = c(0, 100)`.

Lognormal Model

The lognormal model for reliability assumes that the natural logarithm of the random variable T , usually time, is normally distributed. The R code provided in the lognormal model appendix sets up the parameters of the lognormal distribution and creates data columns to plot the lognormal reliability function and its confidence band as a function of t .

The lognormal distribution has a probability density function (ReliaSoft 2015, 187)

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2} \quad (36)$$

where $f(t) \geq 0$, $t > 0$

$-\infty < \mu < \infty$

$\sigma > 0$.

ReliaSoft 2015 calls the term μ the *scale* parameter and the term σ the *shape* parameter. JMP calls them *location* and *scale*, respectively. The lognormal cumulative distribution function (ReliaSoft 2015, 188) does not have a closed-form solution but is represented as an integral of the form

$$F(t) = 1 - \int_{\ln t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (37)$$

Equation 37 means one minus the probability of x being in the interval $\ln t \leq x < \infty$, where $x = \ln t'$ is the natural logarithm of time t' and is normally distributed with mean μ and standard deviation σ . Note that the lower limit $\ln t$ is the natural logarithm of the specific input time t to the function $F(t)$, whereas the time t' term in the variable $x = \ln t'$ is the variable of integration. The reliability function (ReliaSoft 2015, 188) is simply one minus the cumulative distribution function:

$$R(t) = \int_{\ln t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (38)$$

Note that swapping the limits of integration causes the cumulative distribution function and the reliability function to become $F(t) = \int_{-\infty}^{\ln t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ and $R(t) = 1 - \int_{-\infty}^{\ln t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$. Doing this is sometimes helpful for looking up values in standard normal tables, which may be organized to give the area under the standard normal curve from $-\infty$ through z . Recall that the hazard function is defined as the probability density function divided by the reliability function:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}}{\int_{\ln t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx} \quad (39)$$

Both the reliability function and the hazard function must be evaluated using standard normal tables, or an R code script can evaluate them instead! Finally, the mean of the lognormal distribution (ReliaSoft 2015, 188) is

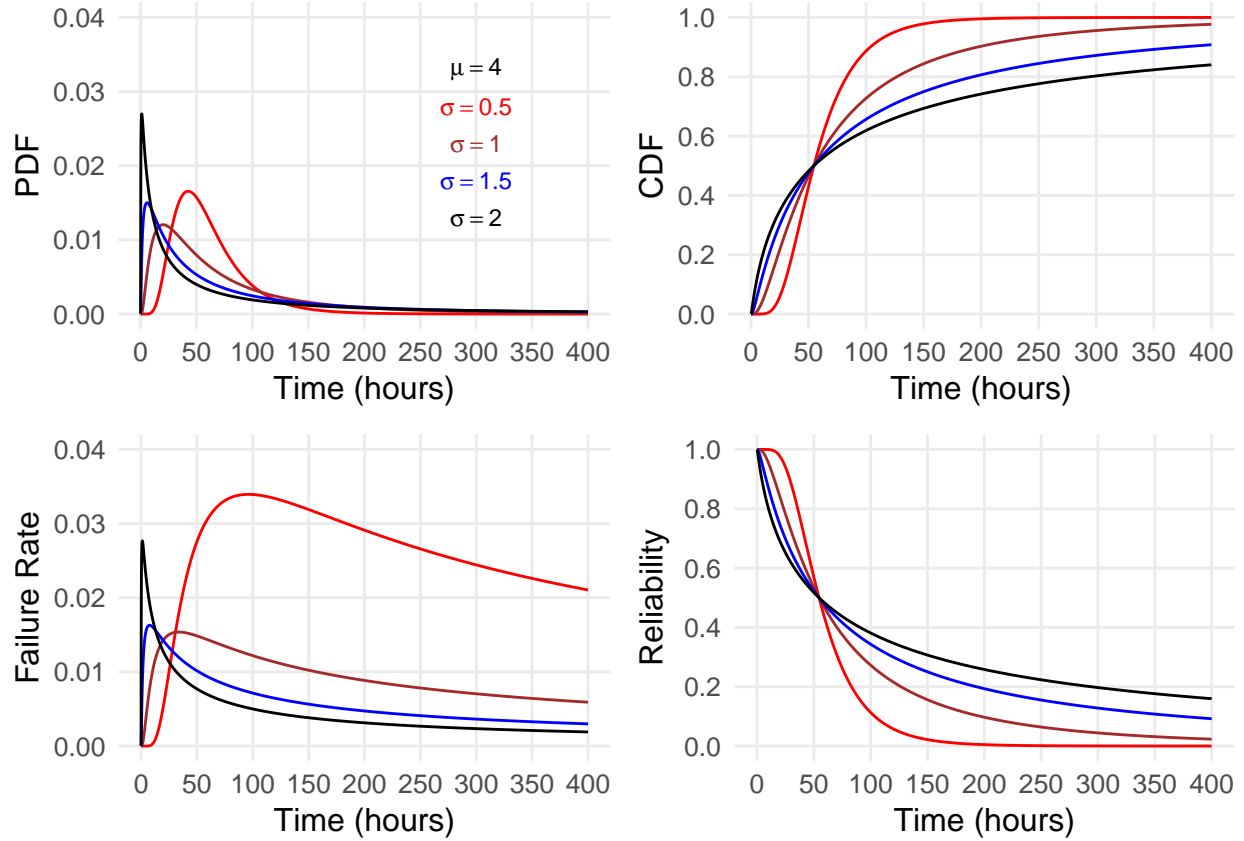


Figure 10: Lognormal model plots

$$\bar{T} = e^{\mu + \frac{1}{2}\sigma^2} \quad (40)$$

where the distribution mean estimate $\hat{\bar{T}} = e^{\hat{\mu} + \frac{1}{2}\hat{\sigma}^2}$ and the sample mean $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$ will generally not be equal.

Figure 10 contains plots of the probability density, cumulative distribution, reliability, and failure rate for different values of μ and σ . Note how the skewness increases as σ increases. Also note how, for a given set of parameters, the failure rate initially increases rapidly with time and then slowly decreases; this contrasts with the Weibull distribution in which failure rate is always decreasing, increasing, or constant over all time for a given set of parameters. The lognormal failure rate asymptotically approaches zero as time approaches infinity.¹⁷

The JMP Life Distribution tool provides estimates of μ , σ , and \bar{T} in its lognormal parametric fit

¹⁷To prove this, let $\mu = 0$ and $\sigma = \frac{1}{\sqrt{2}}$ for simplicity, and recall that the failure rate function can be written as $\lambda(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1-F(t)} = \frac{\frac{1}{t\sigma\sqrt{2\pi}} e^{-\left(\frac{\ln t - \mu}{\sigma}\right)^2}}{\frac{1}{2} \operatorname{erfc}\left(\frac{\ln t - \mu}{\sigma}\right)}$ where $\operatorname{erfc}(w)$ is the complementary error function of w . Substituting the parameter values $\mu = 0$ and $\sigma = \frac{1}{\sqrt{2}}$ and simplifying gives $\lambda(t) = \frac{2}{t\sqrt{\pi}} \frac{e^{-(\ln t)^2}}{\operatorname{erfc}(\ln t)}$. For very large values of w , the complementary error function can be approximated by $\operatorname{erfc}(w) \approx \frac{e^{-w^2}}{w\sqrt{\pi}}$. Therefore, as t approaches infinity, the failure rate function can be approximated by $\lambda(t) \approx \frac{2e^{-(\ln t)^2}}{t\sqrt{\pi} \frac{e^{-(\ln t)^2}}{\sqrt{\pi} \ln t}} = \frac{2 \ln t}{t}$, which slowly approaches zero.

and calls them *location*, *scale*, and *Mean*, respectively. These terms should be familiar from reading the Weibull distribution section above, but for the lognormal distribution these parameters are the ones actually needed to plot reliability. No conversion to another pair of parameters is necessary, and JMP provides a location-scale covariance matrix below its table of parameter estimates. Again, this is the inverse of the observed Fisher information matrix for the location and scale parameters, and is the inverse of the negative matrix of partial second derivatives of the log likelihood function of the lognormal parameters, evaluated at the final parameter estimates that maximize likelihood (ReliaSoft 2015, 200–201).

$$\begin{bmatrix} Var(location) & Cov(location, scale) \\ Cov(location, scale) & Var(scale) \end{bmatrix} = \begin{bmatrix} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) \\ Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma}) \end{bmatrix} \quad (41)$$

Lognormal Reliability Function

To compute the confidence intervals for the reliability function, let $\hat{z}'(x) = \frac{x - \hat{\mu}}{\hat{\sigma}}$ where $\hat{\mu}$ and $\hat{\sigma}$ are the JMP [maximum likelihood] estimates of μ and σ and $x = \ln t'$. The term t' is the variable of integration, and t is the lower limit of integration. Therefore, let $\hat{z} = \frac{\ln t - \hat{\mu}}{\hat{\sigma}}$. Substituting $\hat{z}(x)$ into the reliability function in equation 38 gives

$$\hat{R} = R(\hat{z}) = \int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \quad (42)$$

This is the reliability estimate as a function of \hat{z} which is a function of the parameters and the natural logarithm of time. The goal is to construct a confidence interval for reliability with upper and lower bounds of $R_U(z_U)$ and $R_L(z_L)$, where z_U and z_L are the upper and lower confidence bounds on z . These are estimated by

Upper bound: $z_U = \hat{z} + K_\delta \sqrt{Var(\hat{z})}$

Lower bound: $z_L = \hat{z} - K_\delta \sqrt{Var(\hat{z})}$

where K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level.¹⁸ For a 95 percent confidence interval, choose $\delta = 0.95$, which gives $K_\delta = -1.96$. The variance of \hat{z} is given by (ReliaSoft 2015, 201):

$$Var(\hat{z}) = \frac{1}{\hat{\sigma}^2} \left(Var(\hat{\mu}) + \hat{z}^2 Var(\hat{\sigma}) + 2\hat{z} Cov(\hat{\mu}, \hat{\sigma}) \right) \quad (43)$$

The upper and lower confidence bounds on reliability are thus

Upper bound:

$$R_U = \int_{z_U}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \quad (44)$$

¹⁸I assume that \hat{z} is approximately normally distributed with respect to the parameters μ and σ , and I use the technique described in Nelson, p. 374.

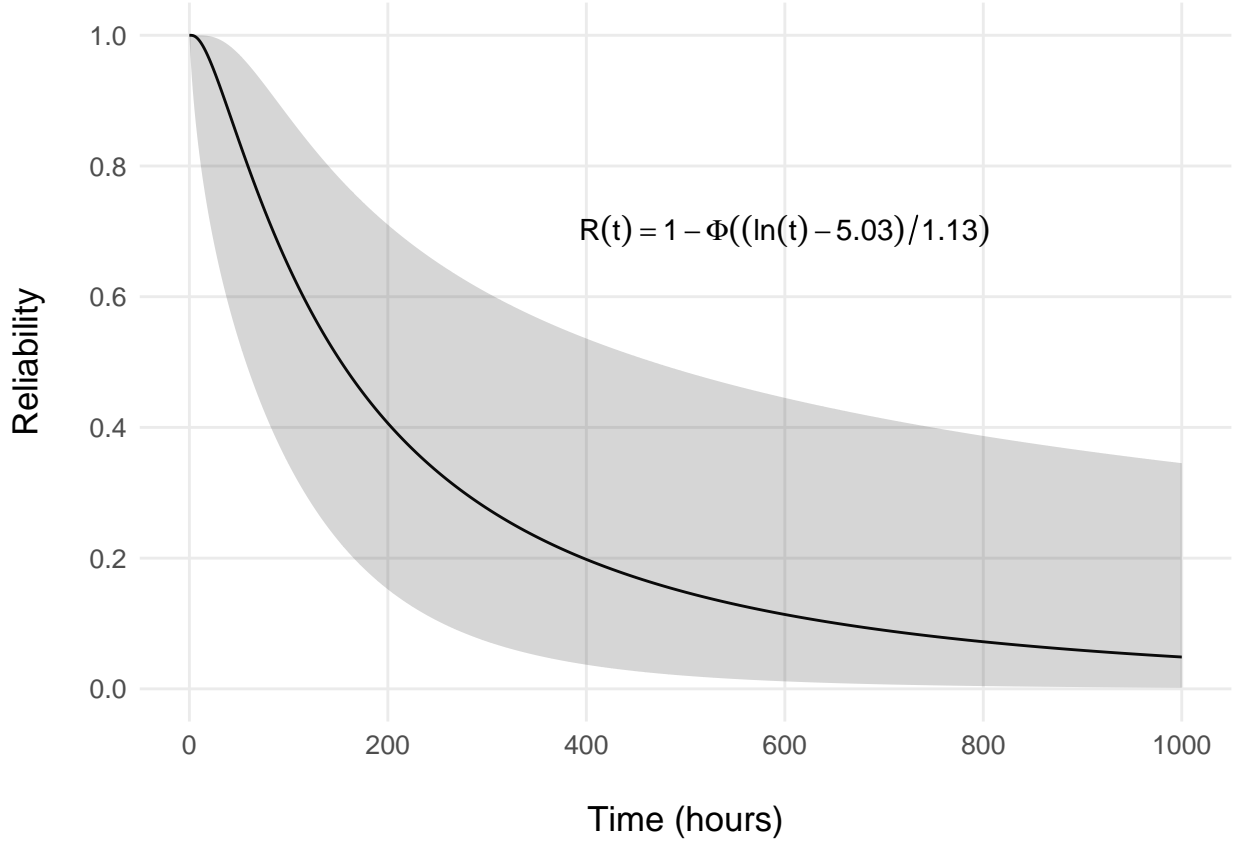


Figure 11: Lognormal reliability model

Lower bound:

$$R_L = \int_{z_L}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \quad (45)$$

Since JMP provides estimates of the location, scale, variance, and covariance values, you can now compute reliability and its upper and lower confidence bounds as functions of time t . The R code in the lognormal model appendix does this. Figures 11 and 12 show the estimated lognormal reliability function plotted with its 95 percent confidence band versus time. Figure 12 is zoomed to $0 < t < 100$ hours. Both graphs are annotated with the reliability equation $\hat{R}(t) = 1 - \Phi\left(\frac{\ln t - \hat{\mu}}{\hat{\sigma}}\right)$ where $\Phi(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw$ is the normal cumulative distribution function. To change the time interval of the reliability plot, reduce the axis limits in the plotting function: `limits = c(0, 1000)` becomes `limits = c(0, 100)`.

Lognormal Hazard Function

Finally, the failure rate can also be plotted parametrically with its confidence band. The derivation of the lognormal failure rate confidence band is more complicated than for the Weibull model, but

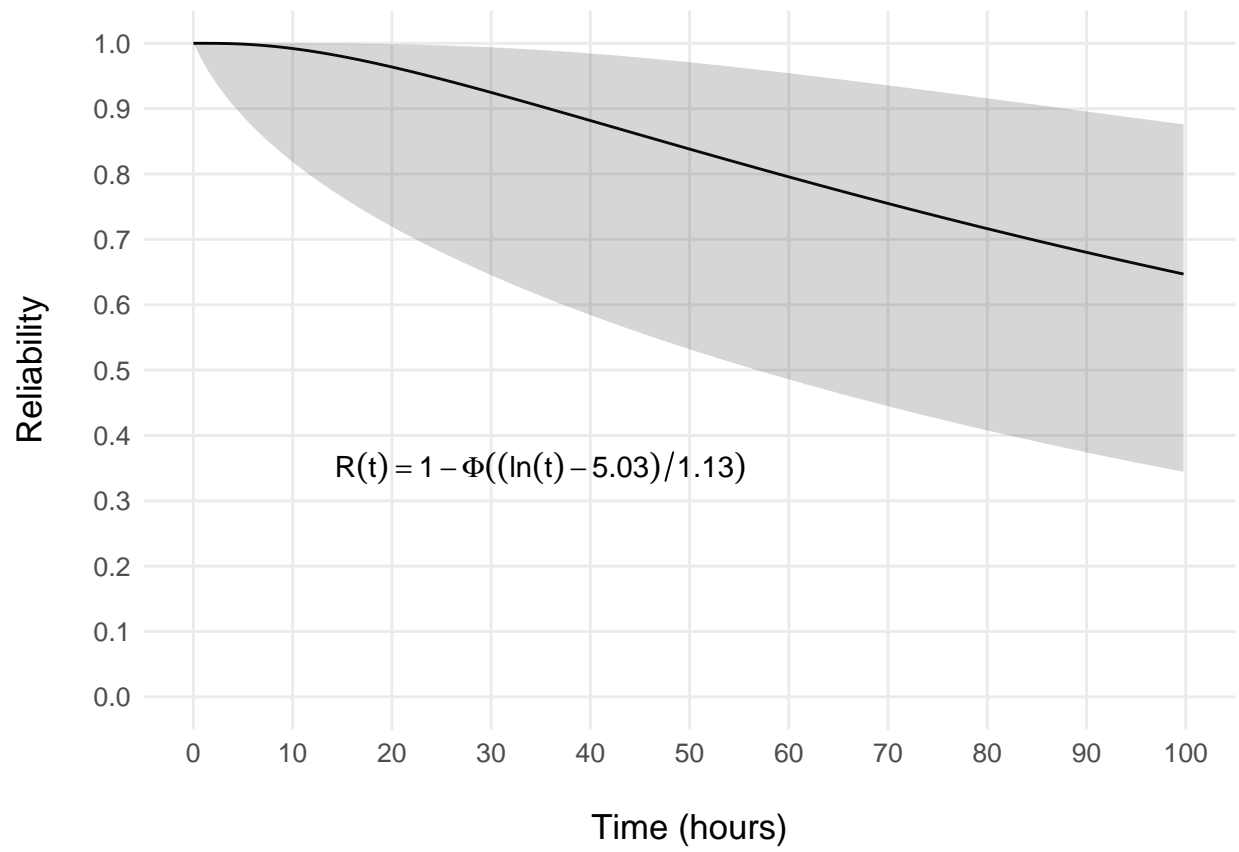


Figure 12: Lognormal reliability model (zoomed)

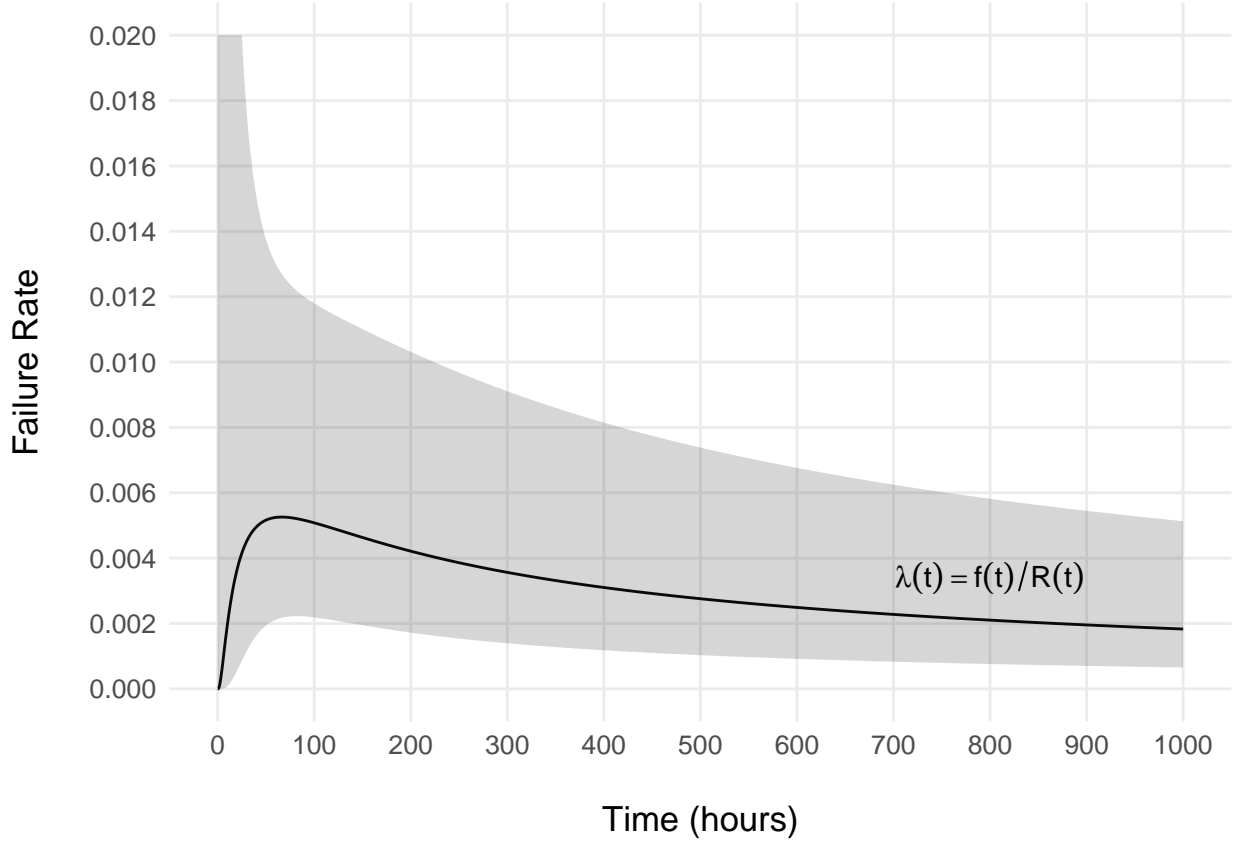


Figure 13: Lognormal failure rate model

the approach is the same. The lognormal appendix contains the complete derivation, but the key steps and results are presented here. Recall from equation 39 that the lognormal hazard function is given by

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}}{\int_{\ln t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} dx} \quad (46)$$

where x is the natural logarithm of time. Substituting $z' = \frac{\ln t' - \mu}{\sigma} = \frac{x - \mu}{\sigma}$ and $z = \frac{\ln t - \mu}{\sigma}$ into the hazard function and using the laws of logarithms and some algebra produces

$$\hat{v} = -\ln \hat{\sigma}\sqrt{2\pi} - \frac{1}{2}\hat{z}^2 - \hat{\sigma}\hat{z} - \hat{\mu} - \ln \left[\int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\hat{z}'^2} d\hat{z}' \right] \quad (47)$$

Since v is defined so that $\lambda = e^v$, the next step is to find upper and lower confidence bounds on v such that the upper and lower confidence bounds of λ are given by

Upper bound:

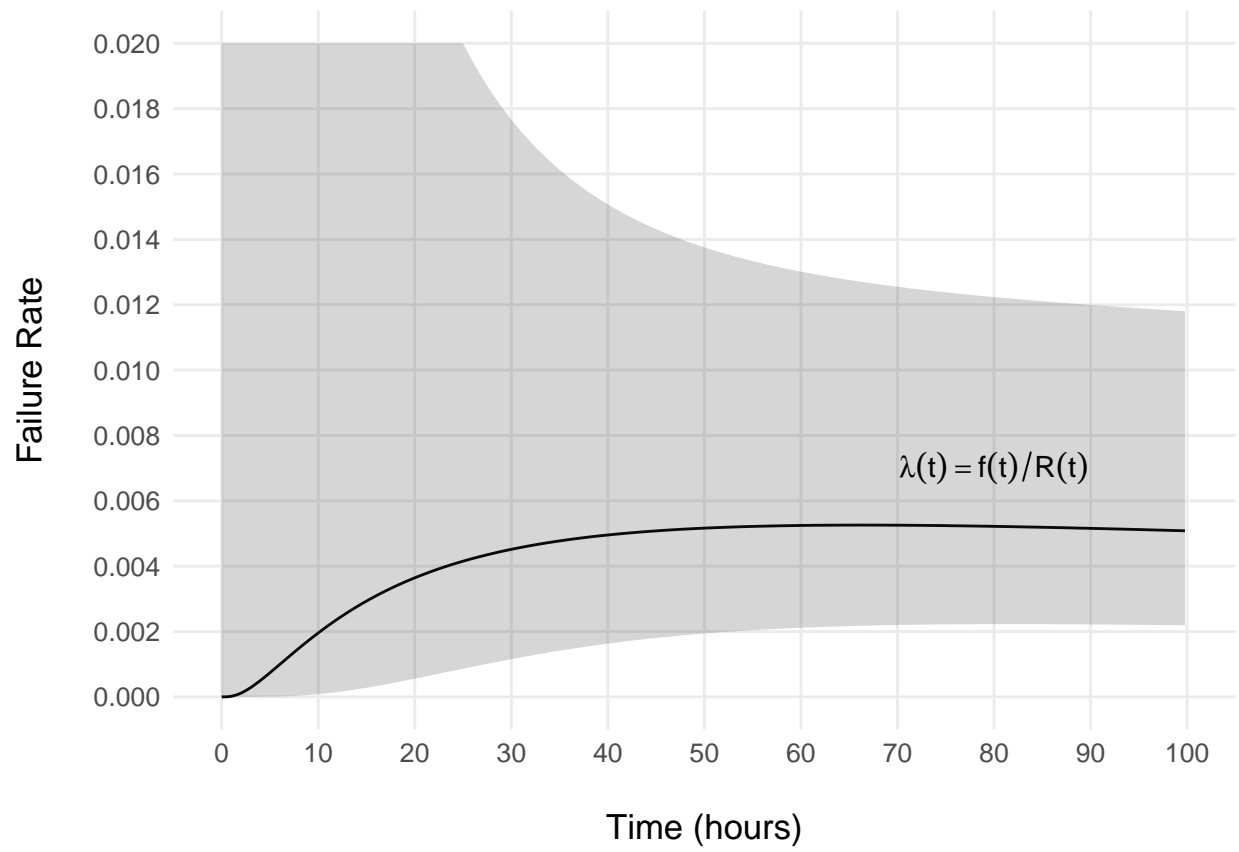


Figure 14: Lognormal failure rate model (zoomed)

$$\lambda_U = e^{v_U} \quad (48)$$

Lower bound:

$$\lambda_L = e^{v_L} \quad (49)$$

and the upper and lower confidence bounds on v are

$$\text{Upper bound: } v_U = \hat{v} - K_\delta \sqrt{\text{Var}(\hat{v})}$$

$$\text{Lower bound: } v_L = \hat{v} + K_\delta \sqrt{\text{Var}(\hat{v})}$$

K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level, and \hat{v} is assumed to be approximately normally distributed. For a 95 percent confidence interval, choose $\delta = 0.95$ to give $K_\delta = -1.96$. The variance of \hat{v} is estimated by rewriting equation 26 for the lognormal model (Nelson 1990, 374):

$$\text{Var}(\hat{v}) = \left(\frac{\partial v}{\partial \mu} \Big|_{\hat{\mu}, \hat{\sigma}} \right)^2 \text{Var}(\hat{\mu}) + \left(\frac{\partial v}{\partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right)^2 \text{Var}(\hat{\sigma}) + 2 \left(\frac{\partial v}{\partial \mu} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \left(\frac{\partial v}{\partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) \quad (50)$$

Evaluating these derivatives for the lognormal model, which includes taking derivatives of integrals, is not trivial. The appendix provides complete derivations, but the final results are

$$\frac{\partial v}{\partial \mu} \Big|_{\hat{\mu}, \hat{\sigma}} = \frac{1}{\hat{\sigma}} \left[\hat{z} - \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\hat{z}^2}}{\int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \right] \quad (51)$$

and

$$\frac{\partial v}{\partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} = \frac{1}{\hat{\sigma}} \left[\hat{z}^2 - 1 - \frac{\frac{1}{\sqrt{2\pi}} \hat{z} e^{-\frac{1}{2}\hat{z}^2}}{\int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \right] \quad (52)$$

Although these expressions appear formidable, remember that the integral in the denominators of both expressions is merely the lognormal reliability function and can be easily evaluated by using a standard normal table and the reliability function $\hat{R}(t) = 1 - \Phi\left(\frac{\ln t - \hat{\mu}}{\hat{\sigma}}\right)$. An analyst can now evaluate $\text{Var}(\hat{v})$. The full expression for $\text{Var}(\hat{v})$ is long, but the R code calculates all the derivatives and squares of derivatives numerically to generate the variance of \hat{v} . From that, you can calculate and plot $\hat{\lambda}$, λ_L and λ_U .

Figures 13 and 14 contain plots of the lognormal failure rate and its 95 percent confidence band over time. Figure 14 is zoomed to the interval $0 < t < 100$ hours. The failure rate has a very wide confidence band near $t = 0$, meaning there is a lot more uncertainty in the failure rate estimate shortly after system startup. This makes sense, because there is little failure data to support a precise estimate in this region of t near zero.

Conclusions

This tutorial provides some of the mathematical theory behind the exponential, Weibull, and lognormal probability distributions, and it provides methodology for plotting estimated reliability curves and confidence limits for each distribution parametrically. The supporting R code is in the appendices following this conclusion. The code and methodology presented assume the reader has already selected and fitted one of these distributions to a set of reliability data in another software package (JMP, R, WEIBULL++, MATLAB, etc.). JMP has a large and versatile reliability analysis capability, so this paper explains how JMP terminology relates to statistical terminology encountered in other sources. However, an analyst can obtain parameter and covariance estimates from any software package and plug them into the R scripts provided in the appendices. JMP is also capable of plotting reliability curves, and its profiler and survival curve features are useful. However, I have found the command line and scripting interfaces of R to be more capable of rapidly producing and tailoring the presentation-quality graphs of parametric functions and confidence bands that are needed for reports, papers, and briefings.¹⁹

The exponential, Weibull, and lognormal distributions are not the only skewed distributions available, but they are the most common in reliability analysis. The exponential distribution is the simplest to plot and interpret, and it is sometimes sufficient to describe reliability data. At other times the exponential fit is relatively poor, and the two-parameter Weibull or the lognormal distribution is a better fit for the data or makes more sense because of its variable failure rate. The lognormal distribution is also commonly used to model repair time.²⁰ The normal distribution is symmetric and not often used in reliability analysis, because reliability data are typically skewed and positive. Other skewed distributions are more difficult to interpret and do not always have closed-form definitions like the three described in this paper.

Each data set is unique, and a researcher should carefully select the distribution that fits the data best. When fitting a distribution, consider the expected failure rate behavior of the system under test: constant, increasing, decreasing, or some combination. Also, compare means for similarity and compare the goodness of fit criteria, such as the AIC and BIC,²¹ between the different distributions. Visual inspection of how the reliability curve fits the data points²² and how wide the confidence interval grows with increasing t are also important to consider when selecting a reliability model. If multiple distributions fit the data similarly with similar means, choose a simpler one if it makes sense from a failure rate perspective. Finally, when it comes to using these R scripts to plot final reliability curves over the desired t domain, be careful not to plot the curves far beyond the maximum value of t_i where no data exist.

¹⁹Visit <https://test-science.shinyapps.io/ParametricReliabilityModels/> to use the associated web application to generate high-resolution graphs for the three models discussed in this paper without writing or copying code.

²⁰Repair time can be used as a metric for system maintainability in operational testing, and it generally exhibits similar characteristics to reliability data in terms of skewness.

²¹The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are both estimators of relative quality of statistical models. The JMP Life Distribution tool will fit several distributions to a data set simultaneously and rank them by quality of fit using one of several estimators like the AIC and BIC.

²²This is sometimes called the fat pencil test.

Appendices

Exponential Model Code

This section contains the R code scripts used to generate the exponential reliability model estimates and plots.

```
# Exponential Reliability Model

# Enter model parameters based on your fitted exponential model.
# The values listed here come from an exponential model fit to actual
# time between failure data used to analyze a system's reliability.
alpha_est <- 366.0 # Estimate of the mean of the exponential distribution
alpha_UCL <- 975.2 # 95% upper confidence limit on parameter estimate
alpha_LCL <- 137.4 # 95% lower confidence limit on parameter estimate

# This code creates a data frame with six columns to plot versus time:
# 1-3) Estimated reliability and the associated confidence limits
# 4-6) Estimated failure rate and the associated confidence limits
# Additional columns are generated in the process of computing the
# upper and lower confidence limits on reliability and the failure
# rate. Adjust the boundaries for time (0, 600) to suit your
# model. The boundaries below are chosen to show the reliability
# curve decrease from 1 to 0.2.
ExpData <- data.frame(Time = seq(from = 0, to = 600, length.out = 3000))

# Reliability columns
ExpData$Rel <- exp(-1 * ExpData$Time/alpha_est)
ExpData$Rel_UCL <- exp(-1 * ExpData$Time/alpha_UCL)
ExpData$Rel_LCL <- exp(-1 * ExpData$Time/alpha_LCL)

# Hazard columns
ExpData$Haz <- 1/alpha_est
ExpData$Haz_UCL <- 1/alpha_LCL
ExpData$Haz_LCL <- 1/alpha_UCL
```

Exponential Reliability Plot

```
# Libraries:
library(ggplot2) # A library to make beautiful customizable plots.

# Create the equation text to annotate the chart.
equal <- sprintf("R(t) == e^{-t / %0.2f}", alpha_est)

# This code uses ggplot to graph the data frame containing
```

```

# the reliability function and confidence band.
gexp <- ggplot(data = ExpData, aes(x = Time, y = Rel)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = Rel_LCL, ymax = Rel_UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 1),
                    breaks = c(0, 0.2, 0.4, 0.6, 0.8, 1)) +
  scale_x_continuous(limits = c(0, 600),
                    breaks = c(0, 100, 200, 300, 400, 500, 600)) +
  labs(x = "\nTime (hours)", y = "Reliability\n") +
  annotate("text", x = 100, y = 0.1, label = as.character(equa1), size = 4,
         colour = "black", parse = TRUE)
gexp

```

Exponential Reliability Plot - zoomed

```

# This code uses ggplot to graph the data frame containing the reliability
# function and confidence band over a shorter interval: (0, 100).
gexp_short <- ggplot(data = ExpData, aes(x = Time, y = Rel)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = Rel_LCL, ymax = Rel_UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 1),
                    breaks = c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1)) +
  scale_x_continuous(limits = c(0, 100),
                    breaks = c(0,10,20,30,40,50,60,70,80,90,100)) +
  labs(x = "\nTime (hours)", y = "Reliability\n") +
  annotate("text", x = 30, y = 0.25, label = as.character(equa1), size = 4,
         colour = "black", parse = TRUE)
gexp_short

```

Exponential Failure Rate Plot

```

# Create the equation text to annotate the chart.
equa2 <- sprintf("lambda(t) == 1 / %0.2f ~ \"or\" ~ %0.5f ~ \"per hour\"",
                alpha_est, 1/alpha_est)

# This code uses ggplot to graph the failure rate and confidence band.

```

```

gexp_haz <- ggplot(data = ExpData, aes(x = Time, y = Haz)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = Haz_LCL, ymax = Haz_UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 0.01),
                    breaks = c(0, 0.002, 0.004, 0.006, 0.008, 0.01)) +
  scale_x_continuous(limits = c(0, 600),
                    breaks = c(0, 100, 200, 300, 400, 500, 600)) +
  labs(x = "\nTime (hours)", y = "Failure Rate\n") +
  annotate("text", x = 200, y = 0.0035, label = as.character(equa2), size = 4,
         colour = "black", parse = TRUE)
gexp_haz

```

Weibull Model Code

This section contains the R code scripts used to generate the Weibull reliability model estimates and plots.

```

# Weibull Reliability Model

# Enter model parameters based on your fitted Weibull model.
# The values listed here come from a Weibull model fit to actual
# time between failure data used to analyze a system's reliability.
mu_location <- 5.52959 # JMP 'location' parameter estimate
sigma_scale <- 0.86514 # JMP 'scale' parameter estimate
Alpha <- 252.0401 # JMP 'Weibull alpha' parameter estimate
Beta <- 1.15588 # JMP 'Weibull beta' parameter estimate
K_delta <- -1.96 # Use -1.96 for 95% confidence bands

# Enter the Covariance values from the Covariance Matrix
# in the JMP Life Distribution output.
Var_mu <- 0.13098 # Variance of JMP 'location' parameter
Var_sigma <- 0.077757 # Variance of JMP 'scale' parameter
Cov_mu_sigma <- -0.02202 # Covariance of JMP 'location' and 'scale' parameters

# This code estimates the variances of the Weibull alpha and beta parameters.
Var_Alpha <- (exp(mu_location)^2)*Var_mu
Var_Beta <- Var_sigma/(sigma_scale^4)
Cov_alpha_beta <- -(exp(mu_location)*Cov_mu_sigma)/(sigma_scale^2)

# This code creates a data frame with six columns to plot versus time:
# 1-3) Estimated reliability and the associated confidence limits
# 4-6) Estimated failure rate and the associated confidence limits

```

```

# Additional columns are generated in the process of computing
# the upper and lower confidence limits on reliability and the failure
# rate. These columns will not be plotted. Adjust the boundaries
# for time (0, 800) to suit your model. The boundaries are chosen to
# show the reliability curve decrease from 1 to about 0.05.
WeibData <- data.frame(Time = seq(from = 0.001, to = 800, length.out = 4000))

# Reliability columns
WeibData$Rel <- exp(-((WeibData$Time/Alpha)^Beta))
WeibData$u_t <- Beta*(log(WeibData$Time) - log(Alpha))
WeibData$Var_u_t <- (Var_Beta*(WeibData$u_t/Beta)^2) +
  (Var_Alpha*(Beta/Alpha)^2) -
  (Cov_alpha_beta*(2*WeibData$u_t/Alpha))
WeibData$u_U <- WeibData$u_t + K_delta*sqrt(WeibData$Var_u_t)
WeibData$u_L <- WeibData$u_t - K_delta*sqrt(WeibData$Var_u_t)
WeibData$R_U <- exp(-exp(WeibData$u_U))
WeibData$R_L <- exp(-exp(WeibData$u_L))

# Hazard columns
WeibData$Haz <- (Beta/Alpha)*(WeibData$Time/Alpha)^(Beta - 1)
WeibData$v_t <- log(Beta) - Beta*log(Alpha) + (Beta - 1)*log(WeibData$Time)
WeibData$var_v_t <- ((log(WeibData$Time) + 1/Beta - log(Alpha))^2)*Var_Beta +
  ((Beta/Alpha)^2)*Var_Alpha -
  (2*Beta/Alpha)*(log(WeibData$Time) + 1/Beta - log(Alpha))*Cov_alpha_beta
WeibData$v_U <- WeibData$v_t - K_delta*sqrt(WeibData$var_v_t)
WeibData$v_L <- WeibData$v_t + K_delta*sqrt(WeibData$var_v_t)
WeibData$Haz_U <- exp(WeibData$v_U)
WeibData$Haz_L <- exp(WeibData$v_L)

# Data frame for plotting
Time <- WeibData$Time
Reliability <- WeibData$Rel
UCL <- WeibData$R_U
LCL <- WeibData$R_L
Hazard <- WeibData$Haz
UCL_Haz <- WeibData$Haz_U
LCL_Haz <- WeibData$Haz_L
WeibDataGraph <- data.frame(Time, Reliability, UCL, LCL, Hazard, UCL_Haz, LCL_Haz)

```

Weibull Reliability Plot

```

# Libraries:
library(ggplot2) # A library to make beautiful customizable plots.

# Create the equation text to annotate the charts
equa3 <- sprintf("R(t) == e^{- (t / %0.0f)^%0.2f}", Alpha, Beta)

```

```

# This code uses ggplot to graph the data frame containing
# the reliability function and confidence band.
gweib <- ggplot(data = WeibDataGraph, aes(x = Time, y = Reliability)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL, ymax = UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 1),
                    breaks = c(0, 0.2, 0.4, 0.6, 0.8, 1)) +
  scale_x_continuous(limits = c(0, 800),
                    breaks = c(0, 100, 200, 300, 400, 500, 600, 700, 800)) +
  labs(x = "\nTime (hours)", y = "Reliability\n") +
  annotate("text", x = 550, y = 0.7, label = as.character(equa3), size = 4,
         colour = "black", parse = TRUE)
gweib

```

Weibull Reliability Plot - zoomed

```

# This code uses ggplot to graph the data frame containing the reliability
# function and confidence band over a shorter interval: (0, 100).
gweib_short <- ggplot(data = WeibDataGraph, aes(x = Time, y = Reliability)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL, ymax = UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 1),
                    breaks = c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1)) +
  scale_x_continuous(limits = c(0, 100),
                    breaks = c(0,10,20,30,40,50,60,70,80,90,100)) +
  labs(x = "\nTime (hours)", y = "Reliability\n") +
  annotate("text", x = 30, y = 0.25, label = as.character(equa3), size = 4,
         colour = "black", parse = TRUE)
gweib_short

```

Weibull Failure Rate Plot

```

# Create the equation text to annotate the charts
equa4 <- sprintf("lambda(t) == (%0.2f / %0.0f)*(t / %0.0f)^(%0.2f - 1)",
                Beta, Alpha, Alpha, Beta)

```

```

# Truncate the UCL to fit the y-axis limits.
# This step may not be necessary in general, but in this case the upper
# confidence level becomes very large near t = 0, making the graph difficult
# to read. Therefore, I truncate the UCL above 0.03 to allow R to plot the
# confidence band from t = 0. This does not detract from the usefulness of
# the graph.
WeibDataGraph$UCL_Haz <- ifelse(WeibDataGraph$UCL_Haz > 0.03,
                                0.03, WeibDataGraph$UCL_Haz)

# This code uses ggplot to graph the data frame containing
# the failure rate and confidence band.
gweib_haz <- ggplot(data = WeibDataGraph, aes(x = Time, y = Hazard)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL_Haz, ymax = UCL_Haz), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, .03),
                    breaks = c(0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03)) +
  scale_x_continuous(limits = c(0, 800),
                    breaks = c(0, 100, 200, 300, 400, 500, 600, 700, 800)) +
  labs(x = "\nTime (hours)", y = "Failure Rate\n") +
  annotate("text", x = 500, y = 0.008, label = as.character(equa4), size = 4,
         colour = "black", parse = TRUE)
gweib_haz

```

Weibull Failure Rate Plot - zoomed

```

# This code uses ggplot to graph the data frame containing the failure
# rate and confidence band over a shorter interval: (0, 100).
gweib_haz_short <- ggplot(data = WeibDataGraph, aes(x = Time, y = Hazard)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL_Haz, ymax = UCL_Haz), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, .03),
                    breaks = c(0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03)) +
  scale_x_continuous(limits = c(0, 100),
                    breaks = c(0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)) +
  labs(x = "\nTime (hours)", y = "Failure Rate\n") +
  annotate("text", x = 60, y = 0.007, label = as.character(equa4), size = 4,
         colour = "black", parse = TRUE)
gweib_haz_short

```


Lognormal Model Code

This section contains the R code scripts used to generate the lognormal reliability model estimates and plots.

```
# Lognormal Reliability Model

# Enter model parameters based on your fitted lognormal model.
# The values listed here come from a lognormal model fit to actual
# time between failure data used to analyze a system's reliability.
mu_location <- 5.03007 # JMP 'location' parameter estimate
sigma_scale <- 1.13275 # JMP 'scale' parameter estimate
K_delta <- -1.96 # Use -1.96 for 95% confidence bands

Var_mu <- 0.19666097 # Variance of JMP 'location' parameter
Var_sigma <- 0.10593702 # Variance of JMP 'scale' parameter
Cov_mu_sigma <- 0.01272291 # Covariance of JMP 'location' and 'scale' parameters

# This code creates a data frame with six columns to plot versus time:
# 1-3) Estimated reliability and the associated confidence limits
# 4-6) Estimated failure rate and the associated confidence limits
# Additional columns are generated in the process of computing
# the upper and lower confidence limits on reliability. These
# columns will not be plotted. Adjust the boundaries for time
# (0, 1000) to suit your model. The boundaries are chosen to
# show the reliability curve decrease from 1 to about 0.05.
LNData <- data.frame(Time = seq(from = 0.001, to = 1000, length.out = 4000))

# Reliability columns
LNData$z_hat <- (log(LNData$Time) - mu_location)/(sigma_scale)
LNData$CDF <- pnorm(LNData$z_hat)
LNData$Rel <- 1 - LNData$CDF
LNData$Var_z_hat <- ((1/sigma_scale)^2)*
  (Var_mu + Var_sigma*(LNData$z_hat)^2 + 2*LNData$z_hat*Cov_mu_sigma)
LNData$z_U <- LNData$z_hat + K_delta*sqrt(LNData$Var_z_hat)
LNData$z_L <- LNData$z_hat - K_delta*sqrt(LNData$Var_z_hat)
LNData$R_U <- 1 - pnorm(LNData$z_U)
LNData$R_L <- 1 - pnorm(LNData$z_L)

# Hazard columns
LNData$PDF <- (1/(LNData$Time*sigma_scale*sqrt(2*pi)))*
  exp(-(1/2)*((log(LNData$Time) - mu_location)/sigma_scale)^2)
LNData$Haz <- LNData$PDF/LNData$Rel
LNData$v_hat <- log(LNData$Haz)
LNData$dmu <- (1/sigma_scale)*
  (LNData$z_hat - (1/sqrt(2*pi))*
    exp(-(1/2)*LNData$z_hat^2)/(1 - pnorm(LNData$z_hat)))
LNData$dsigma <- (1/sigma_scale)*
```

```

(LNData$z_hat^2 - 1 - ((1/sqrt(2*pi))*LNData$z_hat*exp(-(1/2)*LNData$z_hat^2))/
  (1 - pnorm(LNData$z_hat)))
LNData$Var_v_hat <- Var_mu*(LNData$dmu^2) +
  Var_sigma*(LNData$dsigma^2) +
  2*Cov_mu_sigma*LNData$dmu*LNData$dsigma

LNData$v_U <- LNData$v_hat - K_delta*sqrt(LNData$Var_v_hat)
LNData$v_L <- LNData$v_hat + K_delta*sqrt(LNData$Var_v_hat)
LNData$Haz_UCL <- exp(LNData$v_U)
LNData$Haz_LCL <- exp(LNData$v_L)

# Data frame for plotting
Time <- LNData$Time
Reliability <- LNData$Rel
UCL <- LNData$R_U
LCL <- LNData$R_L
Hazard <- LNData$Haz
UCL_Haz <- LNData$Haz_UCL
LCL_Haz <- LNData$Haz_LCL
LNDataGraph <- data.frame(Time, Reliability, UCL, LCL, Hazard, UCL_Haz, LCL_Haz)

```

Lognormal Reliability Plot

```

# Libraries:
library(ggplot2) # A library to make beautiful customizable plots.

# Create the equation text to annotate the charts.
equa5 <- sprintf("R(t) == 1 - Phi((ln(t) - %0.2f)/%0.2f)",
  mu_location, sigma_scale)

# This code uses ggplot to graph the data frame containing
# the reliability function and confidence band.
glognorm <- ggplot(data = LNDataGraph, aes(x = Time, y = Reliability)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL, ymax = UCL), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
    panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 1),
    breaks = c(0, 0.2, 0.4, 0.6, 0.8, 1)) +
  scale_x_continuous(limits = c(0, 1000),
    breaks = c(0, 200, 400, 600, 800, 1000)) +
  labs(x = "\nTime (hours)", y = "Reliability\n") +
  annotate("text", x = 600, y = 0.7, label = as.character(equa5), size = 4,
    colour = "black", parse = TRUE)

```

```
glognorm
```

Lognormal Reliability Plot - zoomed

```
# This code uses ggplot to graph the data frame containing the reliability  
# function and confidence band over a shorter interval: (0, 100).  
glognorm_short <- ggplot(data = LNDataGraph, aes(x = Time, y = Reliability)) +  
  geom_line(lwd = 0.5, colour = "black") +  
  geom_ribbon(aes(ymin = LCL, ymax = UCL), alpha = 0.2) +  
  theme_bw() +  
  theme(panel.grid.minor = element_blank(),  
        panel.border = element_blank(), axis.ticks = element_blank()) +  
  theme(text = element_text(size = 13)) +  
  scale_y_continuous(limits = c(0, 1),  
                    breaks = c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1)) +  
  scale_x_continuous(limits = c(0, 100),  
                    breaks = c(0,10,20,30,40,50,60,70,80,90,100)) +  
  labs(x = "\nTime (hours)", y = "Reliability\n") +  
  annotate("text", x = 35, y = 0.35, label = as.character(equa5), size = 4,  
         colour = "black", parse = TRUE)  
glognorm_short
```

Lognormal Failure Rate Plot

```
# Create the equation text to annotate the charts.  
equa6 <- sprintf("lambda(t) == f(t) / R(t)")  
  
# Truncate the UCL to fit the y-axis limits.  
# This step may not be necessary in general, but in this case the upper  
# confidence level becomes very large near t = 0, making the graph difficult  
# to read. Therefore, I truncate the UCL above 0.02 to allow R to plot the  
# confidence band from t = 0. This does not detract from the usefulness of  
# the graph.  
LNDataGraph$UCL_Haz <- ifelse(LNDataGraph$UCL_Haz > 0.02,  
                             0.02, LNDataGraph$UCL_Haz)  
  
# This code uses ggplot to graph the data frame containing  
# the failure rate and confidence band.  
glognorm_haz <- ggplot(data = LNDataGraph, aes(x = Time, y = Hazard)) +  
  geom_line(lwd = 0.5, colour = "black") +  
  geom_ribbon(aes(ymin = LCL_Haz, ymax = UCL_Haz), alpha = 0.2) +  
  theme_bw() +  
  theme(panel.grid.minor = element_blank(),  
        panel.border = element_blank(), axis.ticks = element_blank()) +
```

```

theme(text = element_text(size = 13)) +
scale_y_continuous(limits = c(0, 0.02),
                    breaks = c(0,0.002,0.004,0.006,0.008,0.01,0.012,
                                0.014,0.016,0.018,0.02)) +
scale_x_continuous(limits = c(0, 1000),
                    breaks = c(0,100,200,300,400,500,600,700,800,900,1000)) +
labs(x = "\nTime (hours)", y = "Failure Rate\n") +
annotate("text", x = 800, y = 0.0034, label = as.character(equa6), size = 4,
         colour = "black", parse = TRUE)
glognorm_haz

```

Lognormal Failure Rate Plot - zoomed

```

# This code uses ggplot to graph the data frame containing the failure
# rate and confidence band over a shorter interval: (0, 100).
glognorm_haz_short <- ggplot(data = LNDDataGraph, aes(x = Time, y = Hazard)) +
  geom_line(lwd = 0.5, colour = "black") +
  geom_ribbon(aes(ymin = LCL_Haz, ymax = UCL_Haz), alpha = 0.2) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_blank(), axis.ticks = element_blank()) +
  theme(text = element_text(size = 13)) +
  scale_y_continuous(limits = c(0, 0.02),
                    breaks = c(0,0.002,0.004,0.006,0.008,0.01,0.012,
                                0.014,0.016,0.018,0.02)) +
  scale_x_continuous(limits = c(0, 100),
                    breaks = c(0,10,20,30,40,50,60,70,80,90,100)) +
  labs(x = "\nTime (hours)", y = "Failure Rate\n") +
  annotate("text", x = 80, y = 0.007, label = as.character(equa6), size = 4,
         colour = "black", parse = TRUE)
glognorm_haz_short

```

Derivation of the Lognormal Failure Rate Confidence Band

This section contains the full derivation of the confidence band for the lognormal model failure rate. The hazard function is given by

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}}{\int_{\ln t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} dx} \quad (53)$$

where x is the natural logarithm of time. For the following derivation let $x = \ln t'$, where t' is also time, because for every value t in $\lambda(t)$, the reliability function must integrate from the lower limit $\ln t$ to infinity. Using $x = \ln t'$ distinguishes the variable of integration from the integral lower limit. Substituting $z' = \frac{\ln t' - \mu}{\sigma} = \frac{x - \mu}{\sigma}$ and $z = \frac{\ln t - \mu}{\sigma}$ into the hazard function gives

$$\lambda(z) = \frac{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}z^2 - \sigma z - \mu}}{\int_z^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z'^2} dz'} \quad (54)$$

The function z' is estimated by the parameter estimates $\hat{\mu}$ and $\hat{\sigma}$, making $\hat{z}' = \frac{x - \hat{\mu}}{\hat{\sigma}}$. The estimated hazard function $\lambda(\hat{z})$, graphed in Figures 13 and 14, is therefore:

$$\hat{\lambda} = \lambda(\hat{z}) = \frac{\frac{1}{\hat{\sigma}\sqrt{2\pi}}e^{-\frac{1}{2}\hat{z}^2 - \hat{\sigma}\hat{z} - \hat{\mu}}}{\int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z'^2} dz'} \quad (55)$$

The goal is to construct confidence intervals for the failure rate estimate in a manner similar to the exponential and Weibull models. Taking the natural logarithm of λ gives

$$v = \ln \lambda = \ln \left[\frac{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}z^2 - \sigma z - \mu}}{\int_z^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z'^2} dz'} \right] \quad (56)$$

which can be rewritten using the rules of logarithms as

$$v = \ln \left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}z^2 - \sigma z - \mu} \right] - \ln \left[\int_z^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z'^2} dz' \right] \quad (57)$$

and finally as

$$v = -\ln \sigma\sqrt{2\pi} - \frac{1}{2}z^2 - \sigma z - \mu - \ln \left[\int_z^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z'^2} dz' \right] \quad (58)$$

The estimate of v based on the estimates of the parameters μ and σ is therefore

$$\hat{v} = -\ln \hat{\sigma} \sqrt{2\pi} - \frac{1}{2} \hat{z}^2 - \hat{\sigma} \hat{z} - \hat{\mu} - \ln \left[\int_{\hat{z}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \hat{z}'^2} d\hat{z}' \right] \quad (59)$$

Since v is defined so that $\lambda = e^v$, the next step is to find upper and lower confidence bounds on v such that the upper and lower confidence bounds of λ are given by

Upper bound:

$$\lambda_U = e^{v_U} \quad (60)$$

Lower bound:

$$\lambda_L = e^{v_L} \quad (61)$$

and the upper and lower confidence bounds on v are

$$\text{Upper bound: } v_U = \hat{v} - K_\delta \sqrt{\text{Var}(\hat{v})}$$

$$\text{Lower bound: } v_L = \hat{v} + K_\delta \sqrt{\text{Var}(\hat{v})}$$

Once again, K_δ is the $(1 - \delta)/2 \times 100\%$ percentile of the standard normal distribution with δ being the confidence level, and I assume that \hat{v} is approximately normally distributed with respect to μ and σ . For a 95 percent confidence interval, choose $\delta = 0.95$, which gives $K_\delta = -1.96$. The variance of \hat{v} is estimated by rewriting equation 26 for the lognormal model (Nelson 1990, 374):

$$\text{Var}(\hat{v}) = \left(\frac{\partial v}{\partial \mu} \Big|_{\hat{\mu}, \hat{\sigma}} \right)^2 \text{Var}(\hat{\mu}) + \left(\frac{\partial v}{\partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right)^2 \text{Var}(\hat{\sigma}) + 2 \left(\frac{\partial v}{\partial \mu} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \left(\frac{\partial v}{\partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) \quad (62)$$

Evaluating these derivatives for the lognormal model means taking derivatives of integrals. Starting with the partial derivative of v with respect to μ and applying the chain rule of calculus for derivatives²³ gives

$$\frac{\partial v}{\partial \mu} = -z \frac{\partial z}{\partial \mu} - \sigma \frac{\partial z}{\partial \mu} - 1 - \frac{1}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z'^2} dz'} \frac{\partial}{\partial \mu} \left[\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z'^2} dz' \right]$$

Then using the derivative $\frac{\partial z}{\partial \mu} = -1/\sigma$ and applying the Leibniz integral rule²⁴ gives

$$\frac{\partial v}{\partial \mu} = \frac{z}{\sigma} - \frac{1}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z'^2} dz'} \int_z^\infty \frac{\partial}{\partial \mu} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z'^2} dz' \right)$$

Evaluating the partial derivative with respect to μ and then integrating in the numerator gives

²³If $F(x) = f(g(x))$ for all x , then $\frac{dF}{dx} = \frac{df}{dg} \frac{dg}{dx}$.

²⁴If a and b are constants with respect to a variable x in the function $f(x, t)$, then the derivative with respect to x of an integral of f over t from a to b can be rewritten as the integral of the derivative with respect to x of the expression inside the integral: $\frac{d}{dx} \left(\int_a^b f(x, t) dt \right) = \int_a^b \frac{\partial}{\partial x} (f(x, t) dt)$.

$$\frac{\partial v}{\partial \mu} = \frac{z}{\sigma} - \frac{\int_z^\infty \frac{z'}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} = \frac{z}{\sigma} - \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'}$$

and therefore, evaluating at the parameter estimates gives

$$\left. \frac{\partial v}{\partial \mu} \right|_{\hat{\mu}, \hat{\sigma}} = \frac{1}{\hat{\sigma}} \left[\hat{z} - \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\hat{z}^2}}{\int_{\hat{z}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \right] \quad (63)$$

Although this expression still appears formidable, remember that the integral in the denominator is just the lognormal reliability function and can be easily evaluated by $\hat{R}(t) = 1 - \Phi\left(\frac{\ln t - \hat{\mu}}{\hat{\sigma}}\right)$. Turning now to the partial derivative of v with respect to σ and again applying the chain rule gives

$$\frac{\partial v}{\partial \sigma} = -\frac{\sqrt{2\pi}}{\sigma^2} - z \frac{\partial z}{\partial \sigma} - z - \sigma \frac{\partial z}{\partial \sigma} - \frac{1}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \frac{\partial}{\partial \sigma} \left[\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right]$$

and because $\frac{\partial z}{\partial \sigma} = -z/\sigma$, the expression above simplifies to

$$\frac{\partial v}{\partial \sigma} = \frac{z^2 - 1}{\sigma} - \frac{1}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \frac{\partial}{\partial \sigma} \left[\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right]$$

Applying the Leibniz integral rule again to the partial derivative in the second term gives

$$\frac{\partial}{\partial \sigma} \left[\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right] = \int_z^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right) \quad (64)$$

Now comes a potential pitfall. The term dz' is actually a function of σ , where $dz' = dx/\sigma$. This comes from differentiating z' with respect to x and substituting dz' into the reliability integral. Thus, care must be taken to apply the chain rule of calculus correctly here. Temporarily rewriting the integral in equation 64 in terms of x and dx can help.²⁵

$$\int_z^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right) = \int_{\ln t}^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right)$$

Evaluating the derivative with respect to σ using the product rule gives

$$\int_{\ln t}^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) = \int_{\ln t}^\infty \left[\frac{-1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \frac{1}{\sigma\sqrt{2\pi}} \left(\frac{-x+\mu}{\sigma} \right) \left(\frac{-x+\mu}{\sigma^2} \right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right] dx$$

Substituting $z' = \frac{x-\mu}{\sigma}$ back into the equation and simplifying gives

$$\int_z^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right) = \frac{1}{\sigma} \int_z^\infty \frac{z'^2}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' - \frac{1}{\sigma} \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'$$

²⁵Note that dz' is constant with respect to μ so there was no need to do this when calculating $\frac{\partial v}{\partial \mu}$ in equation 63.

which is written as the difference of two integrals. The second integral is familiar, but to solve the first integral with the z'^2 term I must integrate by parts.²⁶ First rearrange the terms in this integral so that

$$\int_z^\infty \frac{z'^2}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' = \frac{-1}{\sqrt{2\pi}} \int_z^\infty z'(-z'e^{-\frac{1}{2}z'^2}) dz'$$

Then, by letting $u = z'$, $v = e^{-\frac{1}{2}z'^2}$, $\frac{du}{dz'} = 1$, and $\frac{dv}{dz'} = -z'e^{-\frac{1}{2}z'^2}$, I can integrate by parts

$$\frac{-1}{\sqrt{2\pi}} \left[\int_z^\infty z'(-z'e^{-\frac{1}{2}z'^2}) dz' \right] = \frac{1}{\sqrt{2\pi}} \left[ze^{-\frac{1}{2}z^2} + \int_z^\infty e^{-\frac{1}{2}z'^2} dz' \right]$$

Substituting this result back into equation 64 and subtracting gives

$$\int_z^\infty \frac{\partial}{\partial \sigma} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right) = \frac{1}{\sigma \sqrt{2\pi}} ze^{-\frac{1}{2}z^2} \quad (65)$$

Finally, putting the pieces together gives

$$\frac{\partial v}{\partial \sigma} = \frac{z^2-1}{\sigma} - \frac{1}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \frac{\partial}{\partial \sigma} \left[\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz' \right] = \frac{z^2-1}{\sigma} - \frac{\frac{1}{\sigma \sqrt{2\pi}} ze^{-\frac{1}{2}z^2}}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'}$$

or more simply

$$\frac{\partial v}{\partial \sigma} = \frac{1}{\sigma} \left[z^2 - 1 - \frac{\frac{1}{\sqrt{2\pi}} ze^{-\frac{1}{2}z^2}}{\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z'^2} dz'} \right]$$

Thus, evaluating at the parameter estimates gives

$$\left. \frac{\partial v}{\partial \sigma} \right|_{\hat{\mu}, \hat{\sigma}} = \frac{1}{\hat{\sigma}} \left[\hat{z}^2 - 1 - \frac{\frac{1}{\sqrt{2\pi}} \hat{z} e^{-\frac{1}{2}\hat{z}^2}}{\int_{\hat{z}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\hat{z}'^2} d\hat{z}'} \right] \quad (66)$$

An analyst can now evaluate $Var(\hat{v})$. The full expression for $Var(\hat{v})$ is long, but the R code calculates all the derivatives and squares of derivatives numerically to generate the variance of \hat{v} . From that, you can calculate and plot $\hat{\lambda}$, λ_L and λ_U over time as shown in Figures 13 and 14.

²⁶Integration by parts is derived from the product rule in calculus. It says that if u and v are both functions of x , then $u(x)v(x) = \int \frac{du}{dx} v dx + \int \frac{dv}{dx} u dx$ or $\int \frac{dv}{dx} u dx = u(x)v(x) - \int \frac{du}{dx} v dx$.

References

Casella, George, and Roger L. Berger. 2002. *Statistical Inference*. Pacific Grove, CA, United States: Duxbury Thomson Learning.

Nelson, Wayne. 1990. *Applied Life Data Analysis*. Wiley Series in Probability and Statistics (Book 257). New York: Wiley-Interscience.

ReliaSoft. 2015. *Life Data Analysis Reference*. Tucson, AZ, United States: ReliaSoft Corporation. http://www.synthesisplatform.net/references/Life_Data_Analysis_Reference.pdf.

SAS. 2011. *SAS/Qc 9.3 User's Guide*. Cary, NC, United States: SAS Institute Inc. http://support.sas.com/documentation/cdl/en/qcug/63964/HTML/default/viewer.htm#qcug_reliability_sect046.htm.