Al-Rafidain Journal of Computer Sciences and Mathematics (RJCM)

www.csmj.mosuljournals.com

Enhancing Reliability Engineering Through Weibull Distribution in R

Sameera Abdulsalam Othman

Department of Mathematics, College of Basic Education, University of Dohuk, Iraq
Email: Sameera.othman@uod.ac

Article information

Article history:
Received :1/12/2023
Accepted :15/1/2024
Available online: 25/6/2024

Abstract

Accelerated life testing is a fundamental practice in reliability engineering, making the evaluation of component or device performance over extended lifetimes impractical to encounter during design. This study delves into the application of the Weibull distribution to model lifetime data, showcasing its versatility in real-world scenarios. The evaluation includes critical metrics such as Akaike's information criterion (AIC), Bayesian information criterion (BIC), coefficient of determination, and standard error for distribution comparison. Utilizing Maximum Likelihood Estimation (MLE) for parameter estimation, a simulation study is conducted with varying sample sizes, and the R programming language is employed for in-depth analysis. Real data analysis involves Weibull using goodness-of-fit criteria. Maximum Likelihood Estimates (MLEs) are obtained, and the likelihood ratio test demonstrates the Weibull model's superior alignment with the data. The study concludes with the simplicity of producing Quick Fit plots for analysis using R software. The presented approach provides a comprehensive understanding of reliability characteristics, combining theoretical insights with practical applications and numerical analyses. The estimated parameters (β=0.973725, η=14167.5) and statistical measures (K-Smirov, AIC, BIC, Anderson-Darling, Cramer-von Misses) underscore the thoroughness of the evaluation process. The likelihood ratio test further substantiates the Weibull distribution's closer alignment with the input data compared to the standard 2-parameter Weibull distribution. These findings offer a significant methodology for accelerated life testing and model selection, providing essential practical insights into reliability engineering.

Keywords:
Probability distributions, R code, data modeling, failure times, Reliability.

Correspondence:
Author: Sameera Abdulsalam Othman
Email: Sameera.othman@uod.ac

1. Introduction

A key component of reliability engineering is accelerated life testing. To evaluate the performance of a component or device over lifetimes that would be impractical to encounter under design conditions at the time of product introduction, it is a means to shorten the time to failure. Identification of stress factors that can be changed in a controlled manner during testing to hasten the degradation of component materials is the key to this testing. The study of dependability benefits from the modeling of failure times. Therefore, probability distributions that link a given value of the examined variable with the chance of occurrence must be used to statistically model the objects under study [1].

Exponential, Gamma, Lognormal, and Weibull distributions are those that are most frequently employed to represent failure times, according to [2]. Choosing the distribution that most closely matches the failure times constitutes the analysis [3]. Software, particularly R, is widely utilized for identifying the distributions that best-fit failure times, supporting both numerical and analytical approaches [4] [5]. Techniques, including graphical methods like the probability paper, are employed to establish or suggest accurate failure time data models [6]. Nonparametric functional regression techniques, such as the Kernel Model and KNN Model, offer alternative approaches for scalar Y and functional x [7]. Algorithmic applications for transforming positive original responses have been proposed to produce a family of distributions [8].

Proposing a study exploring the application of the transmuted Weibull distribution for modeling lifetime data, [9] highlights its versatility in real-world scenarios. In this study, the emphasize is on demonstrating how the Weibull distribution can effectively describe +lifetimes, utilizing examples from actual data. The two-parameter Weibull distribution is specifically utilized to model datasets, and the comparison of distributions involves metrics such as Akaike's information criterion (AIC), Bayesian information criterion (BIC), coefficient of determination, and standard error. The estimation of distribution parameters is carried out using the maximum likelihood method.

2. Methodology

First, the approach to working with the Weibull distribution is presented in this section. When a random variable X has the following probability density function (pdf), it is said to have a Weibull distribution with parameters $\beta > 0$ and $\eta = 1$

$$g\,(x) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^{\beta}}, \quad x > 0 \qquad (1)$$

The probability density function (PDF) of the Weibull distribution is depicted in **Fig. 1** for different values of the shape parameter (β) while keeping the scale parameter (η) fixed at 1.

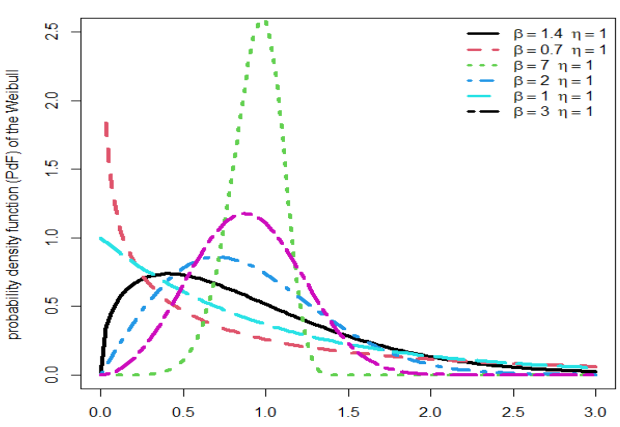

Fig. 1. Probability density function (PDF) of the Weibull distribution for different values of (β) and (η) =1.

The cumulative distribution function (cdf) of the Weibull Distribution can be expressed as follows.

$$G(x) = 1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}} \qquad (2)$$

The cumulative distribution function (CDF) of the Weibull distribution is illustrated in **Fig. 2** for varying values of the shape parameter (β), with the scale parameter (η) held constant at 1.

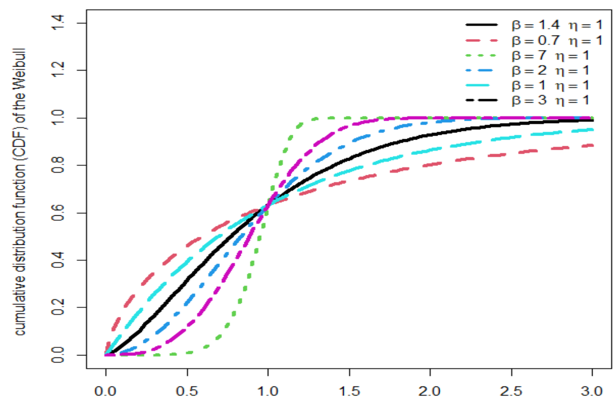

Fig. 2. Cumulative distribution function (CDF) of the Weibull distribution for different values of β and η =1.

Reliability is the complement of the Cumulative Distribution Function (CDF), representing the probability that failure will not happen until time (t), as given by [10]. Reliability of the Weibull distribution is illustrated in **Fig. 3** for varying values of the shape parameter (β), with the scale parameter (η) held constant at 1.

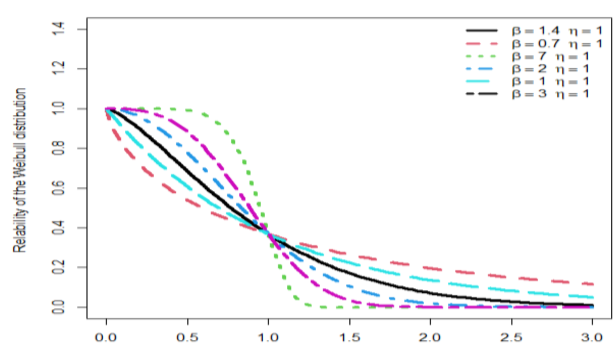

Fig. 3. Reliability of the Weibull distribution for different values of β and η =1.

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}} \qquad (3)$$

where η corresponds to the mean time to failure ($mttf$) specifically when the slope,β is set to one. The relationship between β and mttf is established through a gamma function of β, as demonstrated in the subsequent equation [11]:

$$mttf = \eta\Gamma[1 + \frac{1}{\beta}] \quad (4)$$

When $\beta = 1.0$, mttf = η , the Exponential distribution.
When $\beta > 1.0$, mttf is less than η.
When $\beta < 1.0$, mttf is greater than η.
When $\beta = \frac{1}{2}$, mttf= 2η .

It is essential to differentiate between $mtbf$ (Mean Time Between Failures) and $mttf$ (Mean Time To Failure), as they represent distinct concepts. $mtbf$ denotes the average time interval between occurrences of failures and is computed by dividing the cumulative operational time of all units by the total count of observed failures. These two parameters possess dissimilar characteristics, although they equate when instances of system suspensions are absent. However, under scenarios involving suspensions, substantial discrepancies may emerge. mtbf finds relevance primarily in systems that are capable of being repaired. Moreover, the Weibull hazard function, denoted as h(t), plays a critical role in depicting the instantaneous rate of failures and is mathematically expressed as follows:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{\beta}{\eta}(\frac{x}{\eta})^{\beta-1} e^{-(\frac{x}{\eta})^{\beta}}}{e^{-(\frac{x}{\eta})^{\beta}}}$$

undergoes compensation and simplification, resulting in

$$h(x) = \frac{\beta}{\eta}(\frac{x}{\eta})^{\beta-1} \quad (5)$$

The Weibull hazard function, of the Weibull distribution is illustrated in **Fig. 4** for varying values of the shape parameter (β), with the scale parameter (η) held constant at 1

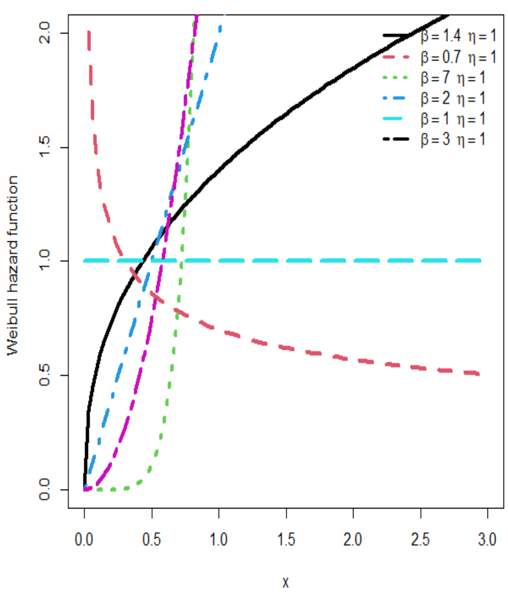

Fig. 4. Weibull hazard function for different values of β and η =1.

The mean residual life MRL at a given time x measures the expected remaining lifetime of an individual of age x. It is given by [12,13]

$$m(x) = E(X - x|X \geq x)$$

$$= \frac{1}{R(x)}\int_0^{\infty} R(u)\, du \quad (6)$$

It's noteworthy that $m(0)$ represents the mean time to failure. The Mean Residual Life (MRL) can be expressed in terms of the cumulative hazard rate function as demonstrated by the integral equation: Here is how you can write your equations: For the substitution, you can write: $u = (\frac{t}{\eta})^{\beta}$. Then $du = \frac{\beta}{\eta}(\frac{t}{\eta})^{\beta-1}dt$.

Since $\frac{t}{\eta} = u^{1/\beta}$, this gives us $du = \frac{\beta}{\eta}(u^{1/\beta})^{\beta-1}dt = \frac{\beta}{\eta}u^{1-1/\beta}\, dt$, so that

$$dt = \frac{\eta}{\beta}u^{1/\beta-1}du.$$

$$m(x) = e^{(\frac{x}{\eta})^{\beta}} \int_{(\frac{x}{\eta})^{\beta}}^{\infty} \frac{\eta}{\beta} u^{1/\beta-1} e^{-u}\, du$$

$$= \frac{\beta}{\eta} e^{(\frac{x}{\eta})^{\beta}}\, \Gamma\left(\frac{1}{\beta}, \left[\frac{x}{\eta}\right]^{\beta}\right),$$

$$m(x) = \eta e^{(\frac{x}{\eta})^{\beta}}\, \Gamma\left(1 + \frac{1}{\beta}, \left[\frac{x}{\eta}\right]^{\beta}\right) - \left[\frac{x}{\eta}\right]^{\beta} e^{(\frac{x}{\eta})^{\beta}}$$

$$= \eta e^{(\frac{x}{\eta})^{\beta}}\, \Gamma\left(1 + \frac{1}{\beta}, \left[\frac{x}{\eta}\right]^{\beta}\right) - x$$

(7)

Here $\Gamma(\eta, x) = \int_x^{\infty} e^{-1} z^{\eta-1}\, dz$ represents the upper incomplete Gamma function.

3. Maximum Likelihood Estimators

The likelihood function for a sample of n independent and identically distributed observations $x_1, x_2, \ldots, x_n$, from the Weibull distribution is given by[12,13,14]:

$$L(\eta, \beta) = \prod_{i=1}^{n} \frac{\beta}{\eta}(\frac{x_i}{\eta})^{\beta-1} e^{-(\frac{x_i}{\eta})^{\beta}} \quad (8)$$

To simplify computations, take the natural logarithm of the likelihood function:

$$\ln L(\eta, \beta) = \sum_{i=1}^{n} \left[ln\frac{\beta}{\eta} + (\beta - 1)ln\left(\frac{x_i}{\sigma}\right) - \left(\frac{x_i}{\sigma}\right)^{\beta}\right]$$

Differentiate the log-likelihood function for the parameters β and η, and set the derivatives equal to zero to find the values that maximize the likelihood.

Derivative for β:

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{1}{\beta} + ln\left(\frac{x_i}{\sigma}\right) - \left(\frac{x_i}{\sigma}\right)^{\beta} ln\left(\frac{x_i}{\sigma}\right)\right] = 0$$

Derivative for η:

$$\frac{\partial l}{\partial \eta} = \sum_{i=1}^{n} \left[-\frac{\beta}{\eta^2} + \left(\frac{\beta - 1}{\eta}\right) - \beta \left(\frac{x_i}{\sigma}\right)^{\beta-1} \right] = 0$$

The MLE is obtained by maximizing this log-likelihood function. To simplify the math, the log transformation is used. The optimization is performed iteratively by negating the log-likelihood function and finding its minimum.

This iterative optimization is achieved using an optimization function in the nlm function in R.

4. Application

4.1 Simulation Study

Simulation study, as a method, involves the representation or emulation of real-world phenomena using specific models. Given the intricacies of complex operations encountered in reality, which may be challenging to comprehend and analyze directly, models resembling real-world scenarios become invaluable. Simulation serves as a tool to enhance understanding and analysis by providing insights into the underlying processes or real-world situations.

In this section, we present a simulation study wherein data is generated using the inverse transformation method of the cumulative distribution function.

$$x = -\eta(-\ln(1 - F(x)))^{\frac{1}{\beta}}$$

The primary objective is to assess the performance of estimators, specifically Maximum Likelihood Estimators (MLEs). The evaluation is based on the comparison of their estimates and Mean Squared Errors (MSEs). The simulation tests are carried out with varying sample sizes (n = 25, 50, 75, 100, 150) for the Weibull distribution. The implementation utilizes the R programming language, adjusting values for the two parameters ($\boldsymbol{\beta}$, η). The experiment is iterated 1000 times for each combination of sample size and shape parameter values. **Tables 1, 2, and 3** present the estimated parameters and MSEs for the estimations of ($\boldsymbol{\beta}$, η) in three distinct cases. Case (1) is outlined in **Table. 1**, Case (2) in **Table. 2**, and Case (3) in **Table. 3**.

These simulation results offer practical applications by improving device design and guiding maintenance strategies. Understanding the performance of MLEs under different conditions provides reliability engineers with valuable insights, enabling better decision-making in designing components that meet or exceed expected lifetimes. Maintenance strategies can be optimized based on a deeper understanding of failure patterns, enhancing efficiency and cost-effectiveness.

This study's practical significance extends to the engineering community, providing a foundation for robust practices in accelerated life testing and reliability analysis. The results contribute to informed decision-making, fostering advancements in device reliability and maintenance strategies.

These tables illustrate the Mean Squared Error (MSE), a metric that gauges the average squared difference between the estimated and true parameter values. Lower MSE values signify greater accuracy in parameter estimation. Notably, the MSE values exhibit a downward trend with increasing sample size, aligning with the expectation that larger sample sizes contribute to more precise parameter estimates. Additionally, the selection of initial values for β and η appears to influence the performance of the estimation methods, as evidenced by the corresponding MSE values. The tables include Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, likely employed for model selection. Lower AIC and BIC values suggest a more favorable fit for the model. In essence, these tables offer a comprehensive overview of estimation performance across diverse conditions, facilitating the evaluation and comparison of different estimation methods.

4.2 Real Data

In this section, we showcase instances where the Weibull distribution (Wd) model is juxtaposed with other related models. To ensure a balanced comparison, we employ various goodness of fit criteria. The R software is used to conduct numerical analyses to determine the distribution that best fits each data set. The Maximum Likelihood Estimates (MLEs) of the parameters of the distributions are displayed in the subsequent tables. The models are selected using the Akaike Information Criterion (AIC), also known as the Bayesian Information Criterion (BIC). The data used in this context is purely for illustrative purposes. All crucial numerical computations have been executed using the R software. Our first dataset pertains to the analysis of gear data, obtained from the smithdat folder within the SuperSMITH installation, as shown in **Table. 4**. These data points, representing subjects, have been fitted using the Weibull distribution, and the estimated parameters are outlined in the table below. It's noteworthy that the subject data have been modeled using both the Weibull and the transmuted Weibull distributions. **Table. 5** presents the Maximum Likelihood Estimates (MLEs) and maximal log-likelihood values for the Weibull distributions. The likelihood ratio test demonstrates that the Weibull distribution better fits the input data than the standard 2-parameter Weibull distribution.

Where
A = Anderson-Darling statistic.
W = Cramer-von Misses statistic;

Quick Fit functions have fit characteristics included in the function name, and reasonable defaults are used, making it straightforward to get a full analysis. **Fig. 5** simply demonstrates the simplicity of producing a Quick Fit plot.

Table 1. MSE of the parameter estimations of estimation at the sample sizes (25,50,100,150) for the initial value set (β=1.4, $\boldsymbol{\eta}$ =1).

Methods	Sample size	Parameters	Estimate				MSE
			Value	MSE	AIC	BIC	
wd	25	β	1.17624	0.050067	37.1835	39.6213	0.8243463
		$\boldsymbol{\eta}$	0.77094	0.05246599			
wd	50	β	1.392868	5.085831e-05	80.79324	84.61728	0.7721594
		$\boldsymbol{\eta}$	0.9338179	0.004380075			
wd	100	β	1.428079	0.0007884332	-155.737	-147.922	0.770635
		$\boldsymbol{\eta}$	0.9994493	3.033195e-07			
wd	150	β	1.417022	0.0002897649	244.5604	250.5817	0.773764
		$\boldsymbol{\eta}$	0.9813339	0.0003484228			

Table 2. MSE of the parameter estimations of estimation at the sample sizes (25,50,100,150) For the initial value set (β=0.7, $\boldsymbol{\eta}$ =1).

Methods	Sample size	Parameters	Estimate				MSE
			Value	MSE	AIC	BIC	
wd	25	β	0.7417809	0.00174564	56.86037	59.29812	9.343934
		$\boldsymbol{\eta}$	0.9683636	0.001000863			
wd	50	β	0.6964223	1.279985e-05	100.5557	104.3798	6.985404
		$\boldsymbol{\eta}$	0.872006	0.01638245			
wd	100	β	0.7140396	0.0001971093	225.752	230.9624	6.380177
		$\boldsymbol{\eta}$	0.9988995	1.211052e-06			
wd	150	β	0.7020947	4.387814e-06	352.3819	358.4032	6.27817
		$\boldsymbol{\eta}$	1.057965	0.003359905			

Table 3. MSE of the parameter estimations of estimation at the sample sizes (25,50,100,150) For the initial value set (β=7, $\boldsymbol{\eta}$ =1).

Methods	Sample size	Parameters	Estimate				MSE
			Value	MSE	AIC	BIC	
wd	25	β	5.881204	1.251704	-13.3299	-10.8921	0.07438244
		$\boldsymbol{\eta}$	0.9493018	0.002570306			
wd	50	β	7.363925	0.1324413	-48.0245	-44.2004	0.0662862
		$\boldsymbol{\eta}$	1.013052	0.0001703603			
wd	100	β	7.837957	0.70192	-129.096	-123.885	0.0580619
		$\boldsymbol{\eta}$	0.9835058	0.000272057			
Wd	150	β	7.044783	0.002005491	-126.376	-120.355	0.0542712
		$\boldsymbol{\eta}$	1.001298	1.683608e-06			

Table 4. The gear data obtained from the smithdat folder on the SuperSMITH installation.

4325.816	6089.124	6281.571	7329.370	7586.772
8361.412	9136.757	9794.200	10939.03	10942.62
11090.46	11635.25	12160.14	13057.69	14307.81

Table 5. The normality tests of the original and transformed datasets.

Datasets	MLE	Std. Dev.	K-Smirov		Statistics			
			Stat.	p-value	AIC	BIC	W	A
wd	β =0.973725	0.138803	0.11596	0.9227	379.47	381.46	0.03914	3.134
	$\boldsymbol{\eta}$=14167.5	3467.59	Min(-log(Likelihood)) =-211.479					

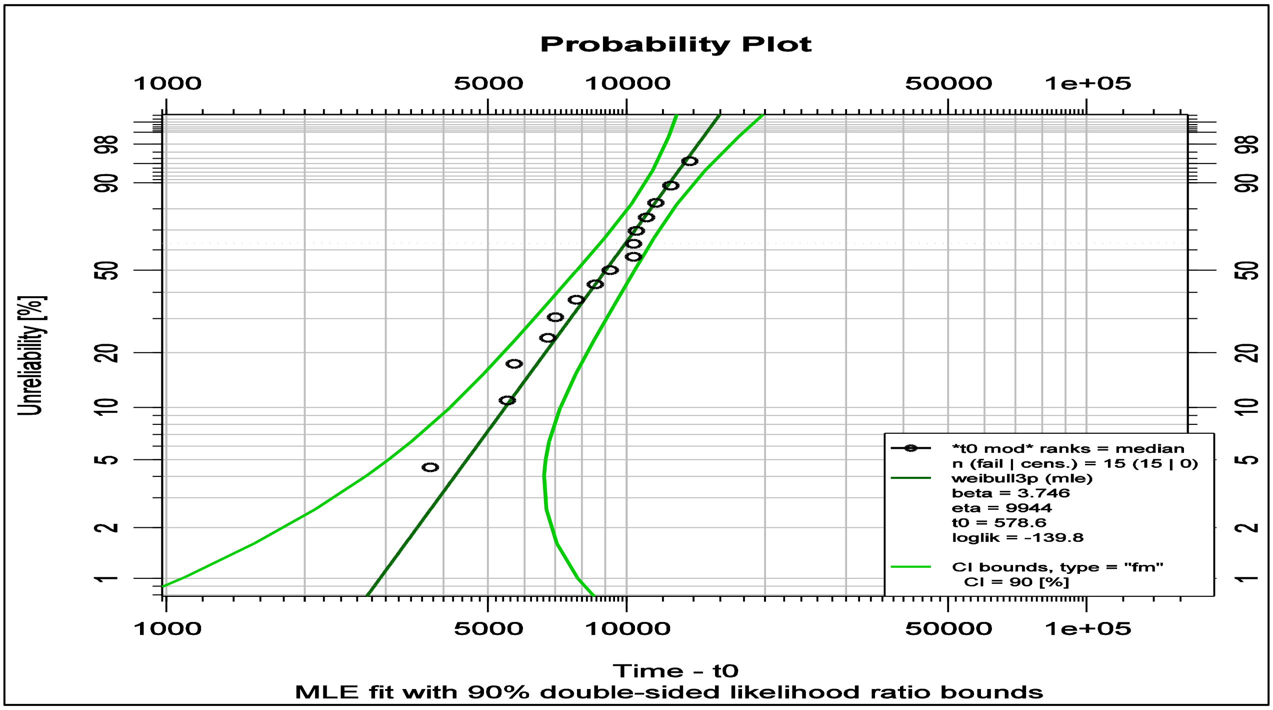

Fig 5. A Probability Plot with Quick Fit.

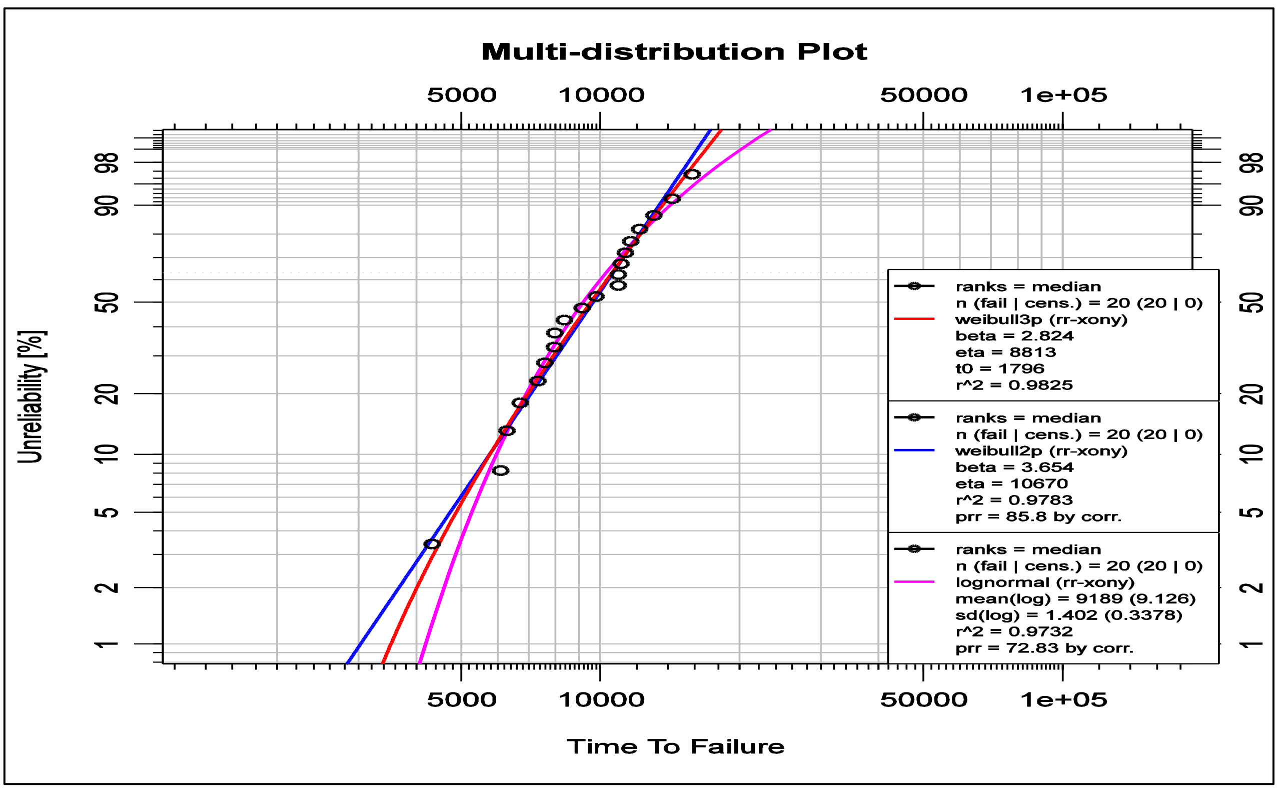

Fig. 6. Multi-distribution with Weibull two, three parameters, and lognormal.

Fig. 6 shows a multi-distribution, and the best distribution is the Weibull with a 2-parameter. When comparing the R^2 in the three cases Weibull with 2 -parameter= 0.9783 and lognormal = 0.9732) it is concluded that the best value is transformed Weibull with 2 parameters.

The precursor computation for establishing bounds on the likelihood ratio involves creating a likelihood contour at a specified confidence level for a given model. These contours represent horizontal sections through the peaked likelihood mound centered around the maximum likelihood estimate. The contour slices are generated at ratio values determined by the following relationship[15]:

$$\text{ratio test} = \text{mle} - \frac{qchisq(CL, def)}{2}$$

where mle is the maximum log-likelihood estimate, CL is the confidence limit, and def represents the degrees of freedom. The degrees of freedom are set to 1 when comparing the model fit itself and 2 when making comparisons against other data. we can show that in **Fig. 7**.

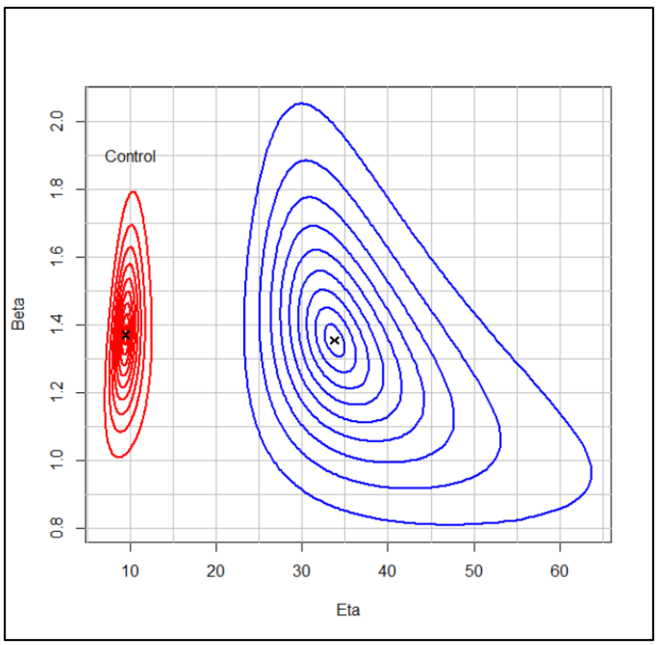

Fig. 7. Comparison of datasets by likelihood contour based on a submitted data set with 3 failure points and approximately 30,000 right-censored suspensions, values.

The points on a specific confidence level contour are used to define confidence interval bounds. **Fig. 8** shows how the extreme Beta value points form asymptotes for the bouns on a 2-parameter mode.

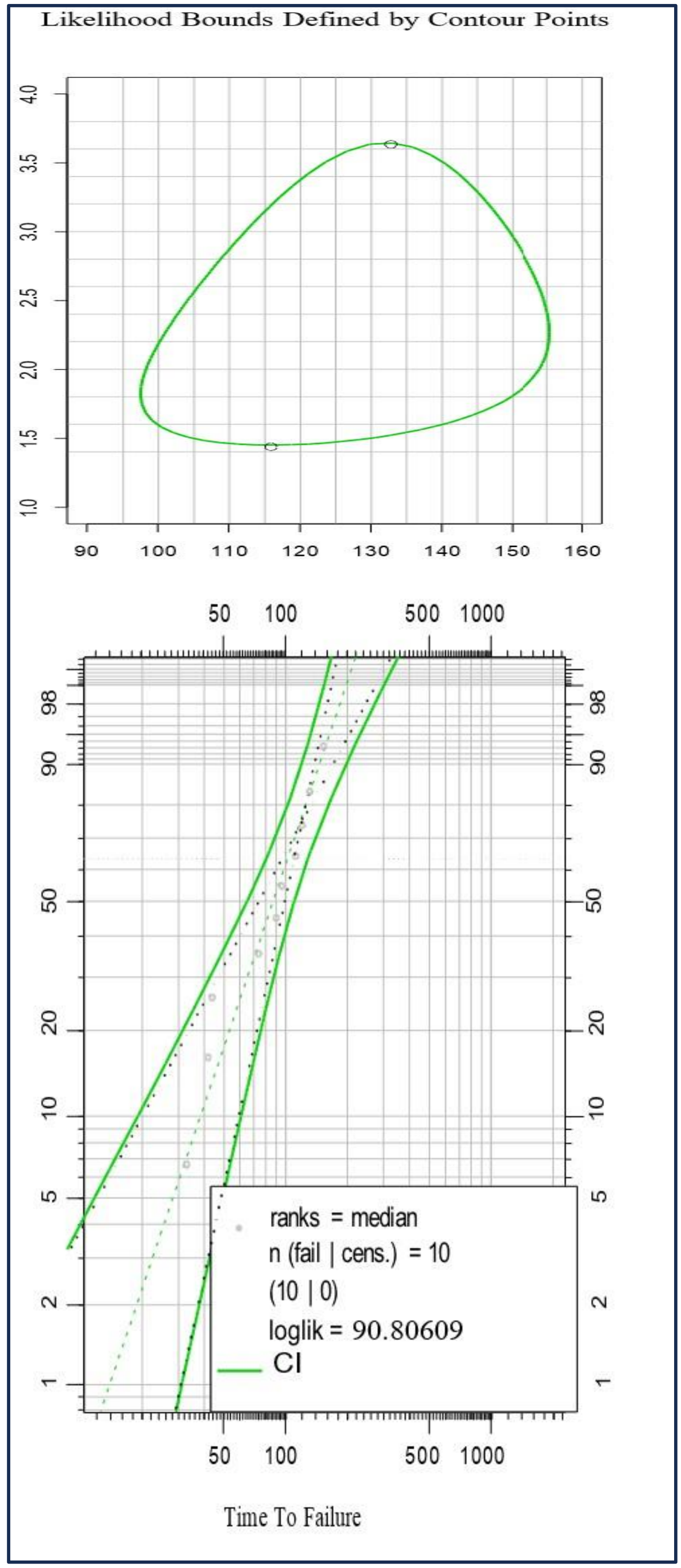

Fig. 8. Likelihood ratio bounds formed by confidence level contour.

The Fisher Matrix bounds including uncertainty in the third parameter. The data used for **Fig. 6** have been applied to form these bounds as bold purple lines in **Fig. 9**.

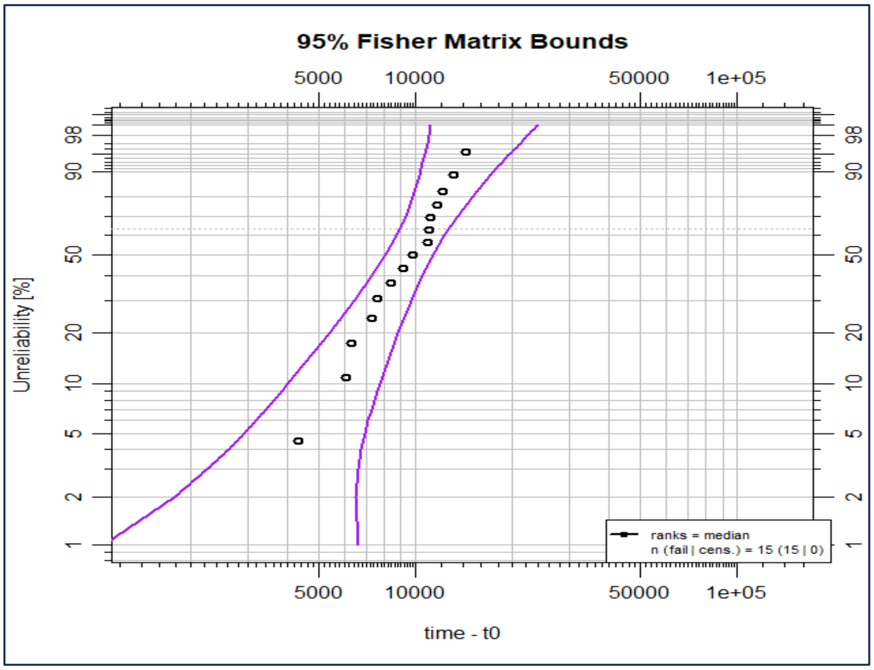

Fig. 9. Unusually formed Fisher Matrix bounds on a 2-parameter model.

5. Conclusion

In conclusion, this research showcases the Weibull distribution's efficacy in accelerated life testing, emphasizing its versatility in real-world scenarios. Employing Maximum Likelihood Estimation (MLE) and the R programming language, a simulation study with varied sample sizes demonstrates the Weibull distribution's robustness. Real data analysis validates the approach using rigorous criteria like AIC and BIC, particularly in gear data modeling. The likelihood ratio tests the Weibull model's superior fit. The study provides a practical understanding of reliability characteristics, blending theoretical insights with numerical analyses. Quick Fit plots in R demonstrate the simplicity of our proposed approach. Overall, this research contributes valuable insights to reliability engineering, offering a comprehensive methodology for accelerated life testing and model selection concisely and practically. Additionally, the research introduces novel insights into likelihood contours, confidence intervals, and Fisher Matrix bounds, enriching the understanding of uncertainty in parameter estimation.

References

[1] CRESPO, A.A. Estatística Fácil. 17ª Edição. LTC Editora, 2005.

[2] HAVIARAS, G. Metodologia para análise de confiabilidade de pneus radiais em frota de caminhões de longa distância. Dissertação (Mestrado em Engenharia Automotiva), Escola Politécnica da Universidade de São Paulo, 2005, p.124.

[3] WUTTKE, R.A.; SELLITTO, M.A. Cálculo da disponibilidade e da posição na curva da banheira de uma válvula de processo petroquímico. Revista Produção On Line, Santa Catarina, v. 8, n. 4, 2008, dez.

[4] FOGLIATO, F.S.; RIBEIRO, J.L.D. Confiabilidade e Manutenção Industrial. Elsevier, 2009, Rio de Janeiro.

[5] R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2020. URL: https://www.Rproject.org/.

[6] FALCETTA, E.F. Análise da Confiabilidade de Produtos baseada em Dados de utilização da Garantia. Dissertação (Mestrado profissionalizante em Engenharia), Universidade Federal do Rio Grande do Sul, 2000, Porto Alegre.

[7] [7] Omar, K.M.T., Othman, S.A. Comparative Analysis of Predictive Performance in Nonparametric Functional Regression: A Case Study of Spectrometric Fat Content Prediction. International Journal of Statistics in Medical Research, 2023, (12), pp. 179-184.

[8] Othman, S.A., Ali, H.T.M. Improvement of the nonparametric estimation of functional stationary time series using Yeo-Johnson transformation with application to temperature curves. Advances in Mathematical Physics, 2021, pp.1-6.

[9] Othman, S.A., Omar, K.M.T., Abdulazeez, S.T. Reliability Analysis and Statistical Fitting for the Transmuted Weibull Model in R. Mathematical Statistician and Engineering Applications, 2023, 72(2), pp. 161-177.

[10] Hartzell, A.L., Da Silva, M.G., Shea, H.R. MEMS reliability. Springer Science & Business Media, 2010.

[11] Espinel-Blanco, E., Flórez-Solano, E., Arévalo-Rueda, J. Study of the modified Weibull function to analyze reliability in engineering components that fail and are repairable. Journal of Physics: Conference Series, 2021, July, 1981(1), p. 012011, IOP Publishing.

[12] Othman, S.A., Qadir, S.T. and Abdulazeez, S.T.,. Comparative Analysis of Scaling Parameter Estimation for the Lindley Distribution in Wait Time Analysis: Simulated and Real Data Applications. Tuijin Jishu/Journal of Propulsion Technology,2023, 44(2).

[13] Dibal, N.P., Bakari, H.R., Yahaya, A.M. Estimating the parameters in the two-parameter Weibull model using simulation study and real-life data. IOSR Journal of Mathematics, 2016, 12, pp.38-42.

[14] Othman SA, Othman AR. QUANTILE MOMENT ESTIMATION OF NEW ALPHA POWER WEIBULL FOR BANK ASSISTANCE WAITING TIME. Journal of Research Administration. 2023 Dec 11;5(2):8426-42.

[15] Chen, G., & Balakrishnan, N. 1995. A general purpose approximate goodness-of-fit test.,Journal of Quality Technology, 27(2), 154-16.