Revista EIA

Statistical model for analizing negative variables with application to compression test on concrete

Revista EIA
ISSN 1794-1237
e-ISSN 2463-0950
Año XIX/ Volumen 19/ Edición N.38
Junio-Diciembre de 2022
Reia3806 pp. 1-19

Publicación científica semestral
Universidad EIA, Envigado, Colombia

Para citar este artículo / To reference this article /
Urrea Montoya, A.; Hernández Barajas, F.; Patiño Rodríguez, C.; Usuga Manco, O. (2022).
Statistical model for analizing negative variables with application to compression test on concrete.
Revista EIA, 19(38), Reia3806. pp. 1-19.
https://doi.org/10.24050/reia.v19i38.1526

Amylkar Urrea Montoya [1]
Freddy Hernández Barajas [2]
Carmen Patiño Rodríguez [1]
Olga Usuga Manco [1]

1. Universidad de Antioquia
2. Universidad Nacional de Colombia, Sede Medellín

Autor de correspondencia:

Usuga Manco, O. (Olga):
Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad de Antioquia.
Correo electrónico:
olga.usuga@udea.edu.co

Recibido: 29-04-2021
Aceptado: 28-7-2021
Disponible online: 01-06-2022

Abstract

In some areas of knowledge, we can find phenomena represented by negative variables ($\mathbb{R}^-$); having a statistical model is crucial to describe the phenomenon and explain it using other variables. This paper proposes a regression model to analyze negative random variables using the reflected Weibull distribution. This paper reports the RelDists package created in the R programming language to implement the proposed model. A Monte Carlo simulation study was conducted to explore the performance of the estimation procedure. The simulation study encompasses two cases: without covariates and with covariables. In the first case, we only have the response variable to estimate the distribution parameters. In the second case, we have the response variable and two explanatory variables to estimate the model parameters. Additionally, censored and uncensored data were considered in the simulation study. From the simulation study, we found that the estimation procedure achieves accurate estimations of the parameters as the sample size increases and the percentage of censoring decreases. In the paper, we present an application of the proposed model using experimental data from a compression test with concrete specimens. In the application, a model was fitted to explain the shrinkage strain using the variable time. The regression model for negative variables and the RelDists package can be used by academic, scientific, and business communities to perform reliability analysis.

Key Words: reliability, censored data, parameter estimation, GAMLSS, programming language R, maximum likelihood, regression model, compression test on concrete, negative random variable, reflected Weibull.

Modelo estadístico para el análisis de variables negativas con aplicación a pruebas de contracción en concreto

Resumen

En algunas áreas de conocimiento se pueden presentar fenómenos que son representados por variables aleatorias negativas ($\mathbb{R}^-$); contar con un modelo estadístico es crucial para representar esos fenómenos y explicarlos en función de otras variables auxiliares. En este trabajo se propone un modelo de regresión para el análisis de variables aleatorias negativas tomando como distribución para la variable respuesta la distribución Weibull reflejada. En este artículo reportamos el paquete RelDists creado en el lenguaje de programación R para facilitar el uso del modelo de regresión propuesto. Por medio de un estudio de simulación Monte Carlo se exploró el desempeño del proceso de estimación de parámetros. En el estudio de simulación se consideraron dos casos: sin covariables y con covariables. El primer caso se refiere a la situación en la cual sólo se tiene la variable respuesta y con ella se deben estimar los parámetros de la distribución. En el segundo caso se tiene la variable respuesta y variables explicativas que en conjunto se usan para estimar los parámetros del modelo de regresión. Adicionalmente, en el estudio de simulación se consideraron datos censurados y no censurados. Del estudio se encontró que el proceso de estimación logra estimar bien los parámetros del modelo a medida que el tamaño de la muestra aumenta y que el porcentaje de censura disminuye. En el artículo se muestra una aplicación del modelo propuesto usando datos experimentales provenientes de una prueba de contracción con probetas de concreto. En la aplicación se construyó un modelo para explicar la contracción de las probetas en función del tiempo. El modelo de regresión para variables aleatorias negativa y el paquete RelDists pueden ser usados por comunidades académicas, científicas y de negocios para el desarrollo de análisis de confiabilidad.

Palabras clave: confiabilidad, datos censurados, estimación de parámetros, GAMLSS, lenguaje de programación R, máxima verosimilitud, modelo de regresión, prueba de contracción en concreto, variable aleatoria negativa, Weibull reflejada.

1. Introduction

The worldwide market is more competitive and demanding than ever; therefore, companies are required to focus on achieving customer satisfaction through new initiatives and customer loyalty strategies. Hence, the systems and machines used to manufacture their products have to be reliable in order to work according to production plans. Reliability concepts can be applied to minimize the possibility of failures in machines, tools, and materials and allow early detection and correction of design deficiencies. Reliability can be studied by modeling the behavior of random variables involved in the process through various probability distributions.

Reliability, as defined by Meeker and Escobar (2014), is the probability that a system will perform its intended function under operating conditions for a specified period of time; it is a definition used in a variety of industrial settings. To quantify the reliability of an item, we use a probabilistic metric, which treats reliability as a probability of the successful achievement of an item's intended function (Modarres, Kaminskiy & Krivtsov , 2016). The metric can be obtained by using probability distributions that model the behavior of the data. Hence, if the true behavior is modeled and measured, then good estimations can be made to make decisions related to the prevention or the reduction of frequent failures.

Distribution fitting is a procedure of choosing a probability distribution model and finding parameter estimates for that distribution. This procedure requires judgment and expertise, and generally needs an iterative process of distribution choice, parameter estimation, and quality of fit assessment. Lifetime data commonly fit distributions such as gamma, Weibull, exponential or lognormal. Even though these distributions have been shown to be very flexible in modeling various types of lifetime distributions, in practice, they cannot be used to model all three phases of a bathtub curve at the same time. Some authors such as Xie and Lai (1996), Barreto-Souza, Santos, and Cordeiro (2010), Lee, Famoye, and Olumolade (2007), Zhang and Xie (2007) and Stacy (1962) have proposed some flexible probability distributions for effective models.

In this paper, we study the reflected Weibull (RW) distribution based on reflecting the Weibull distribution. It is a model appropriate for reliability testing the mechanical properties determined by the standardized fatigue test performance of ductile materials and the mechanics of the fatigue of bearings (Lai, 2014, Balakrishnan & Kocherlakota, 1985, Ali & Woo, 2006). The advantage of using this distribution, instead of the traditional distributions in reliability such as Weibull, lognormal or gamma, is that the estimations and interpretations are made on the natural scale of the variable, avoiding transformation of the original variable.

In the analysis of uncensored data, Wei et al. (2016) assembled some techniques within a framework in order to evaluate the reliability of vehicle components. One of the techniques used in this framework is called variable transformation, which can be obtained either using the Weibull or the RW distribution. Furthermore, the RW distribution has been used by Orjubin (2007) to model electromagnetic compatibility tests, applying the generalized extreme value distribution, which was obtained in terms of the RW distribution. In the analysis of censored data, Cohen (1975) used the RW, Pearson type III, IV, V, and lognormal distributions to estimate the parameters of an interval of censored data. The RW distribution had the best performance in this application. In addition, Cohen (1975) applied the three-parameter Weibull distribution to type II multicensored data, obtained in three different stages, by removing data once a certain number of failures had occurred.

In addition to the application of the RW distribution in various settings, the distribution has been useful for creating new distributions. Authors such as Caron et al. (2018) have proposed a Weibull link (skewed) model for categorical response data arising from binomial, as well as multinomial, models, where the link function is the RW distribution. Additionally, Al Abbasi, Risan, and Resen (2018) proposed the Kumaraswamy-reflected Weibull distribution as an extension to the RW distribution.

Some authors have developed estimation methods for the parameters of the RW distribution from the frequentist and Bayesian points of view. In the case of frequentist estimation methods, Nagatsuka, Kamakura, and Balakrishnan (2013) proposed a method called the location and scale parameters free maximum likelihood estimator. This method demonstrated an improved performance using large samples for the bias and root mean square error (RMSE) metrics in comparison with other parameter estimation methods such as the weighted maximum likelihood. Cohen and Whitten (1982) modified the maximum likelihood estimation (MLE) method for application in the three-parameter Weibull distribution. This modification can be adapted to two-parameter distributions such as the RW distribution. Regal and Larntz (1978) demonstrated parametric likelihood methods for analyzing censored data; these methods could be modified for other families of distributions. Cohen (2016) gave an illustrative example using censored data from a Weibull population. In the case of bayesian estimation methods, authors as Kalsoom, Nasir, and Syed (2019a, 2019b) have estimated the scale parameter of the RW Distribution under bayesian analysis by using non-informative priors and informative priors.

Although many authors have used the RW distribution, it is not easy to estimate its parameters and analyze its reliability because this distribution is not yet implemented in any software. For this reason, it is essential to build a library or package in any computational program that contains these kinds of nontraditional distributions to make these distributions accessible to users and researchers worldwide, so that they apply the RW distribution in different areas.

This paper aims to present a statistical model for negative variables associated with real phenomena using the RW distribution. Additionally, we implemented the RW distribution on the RelDists package in the R programming language (R Core Team, 2021) to encourage the use of this distribution in many applications. The RelDists package contains functions useful to obtain the density, quantiles, probability, hazard, and random samples for the RW distribution. With the RelDists package, it is possible to obtain estimations for the RW distribution or the RW regression model. We also studied the asymptotic behavior of RW estimators through a Monte Carlo simulation following the ADEMP (Aims, Data-generating mechanisms, Estimands, Methods, and Performance measures) methodology proposed by Morris, White, and Crowther (2019).

The paper is structured as follows. Section 2 introduces the RW distribution. Section 3 describes the parameter estimation method, and section 4 presents the RW regression model. In section 5, we introduce the RelDists package. Section 6 presents the Monte Carlo simulation study, and in section 7, we can find the results. Finally, an application with real data is shown in Section 8.

2. Materials and methods

2.1. Reflected Weibull distribution

By reflecting the Weibull distribution over the vertical axis $Y = -X$, Cohen (1975) introduced the three-parameter reflected Weibull distribution, and later, Almalki and Nadarajah (2014) reparametrized this distribution and obtained the two-parameter reflected Weibull (RW) distribution, which has a location parameter μ and scale parameter σ.

Probability density function: The probability density function (pdf) of the RW distribution is given by the following expression:

$$f(y;\mu,\sigma) = \mu\sigma(-y)^{\sigma-1}e^{-\mu(-y)^{\sigma}}, \tag{1}$$

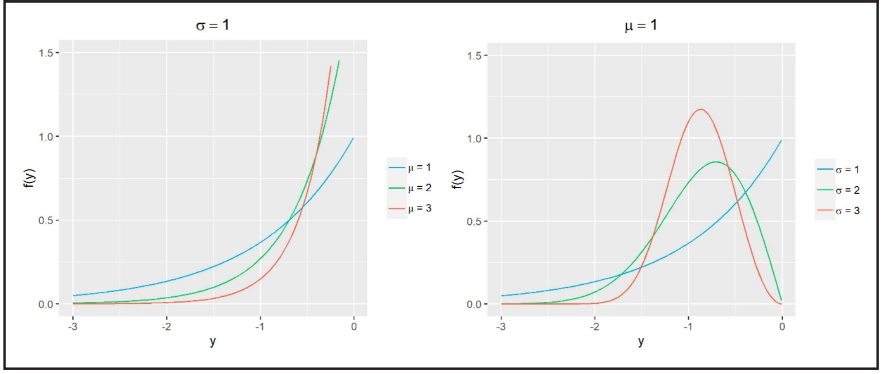
where $y < 0$ and $\mu,\sigma > 0$. If a random variable Y follows a RW distribution, we can denote as $Y \sim RW(\mu,\sigma)$. The pdf is useful to find the probability that the random variable Y is in a specific interval.

Figure 1 shows the pdf of the RW distribution for different values of μ and σ. When $\sigma = 1$ we obtain the exponential distribution. The left panel of Figure 1 shows the pdf shapes for the location parameter $\mu = 1,2,3$ and a common scale parameter $\sigma = 1$. The figure shows that if the scale parameter is fixed and the value of the shape parameter increases, the distribution of Y is J-shaped. The right panel of Figure 1

https://doi.org/10.24050/reia.v19i38.1526

shows the pdf shapes for the location parameter $\mu = 1$ and the scale parameter $\sigma = 1,2,3$. The figure shows that if the shape parameter is fixed and the value of the scale parameter increases, the distribution of Y becomes bell-shaped.

Figure 1. Probability density function of the RW distribution for different values of μ and σ. The curves in the left panel have common $\sigma = 1$ and the right panel shows curves with $\mu = 1$.



Cumulative density function: The cumulative density function (cdf) of the RW distribution is given by the following:

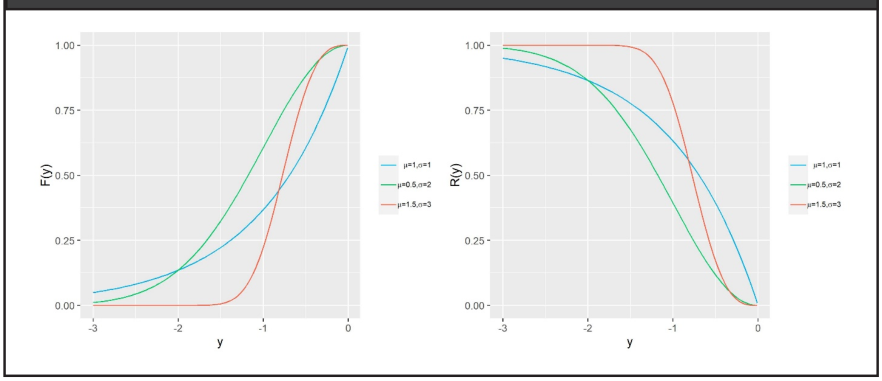
$$F(y;\mu,\sigma) = e^{-\mu(-y)^{\sigma}}, \tag{2}$$

where $y < 0$ and $\mu,\sigma > 0$. With the cdf, we obtain the probability that the random variable Y is less or equal to the value y. In addition, from the cdf, the reliability function (Rf) can be obtained from the following expression:

$$R(y;\mu,\sigma) = 1 - e^{-\mu(-y)^{\sigma}}, \tag{3}$$

where $y < 0$ with $\mu,\sigma > 0$. The Rf calculates the probability of an item operating for a certain amount of time without failure. Figure 2 shows the cdf (left) and the Rf (right) of the RW distribution for different values of μ and σ.

Figure 2. Cumulative density function and reliability function of the RW distribution for different values of μ and σ.



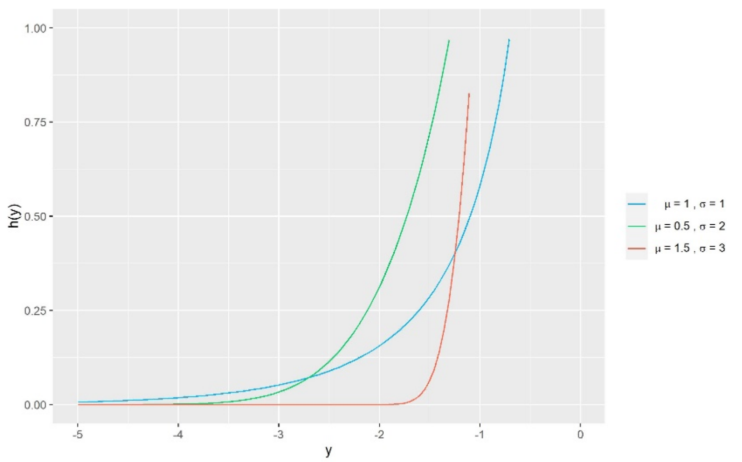
Hazard function: The expression for the hazard function (hf) of the RW distribution is given as follows:

$$h(y;\mu,\sigma) = \mu\sigma(-y)^{\sigma-1}, \tag{4}$$

where $y < 0$ and $\mu,\sigma > 0$. With the hf function, we can measure the probability that the product has not failed in a specific time.

Figure 3 shows the hazard function of the RW distribution for different values of μ and σ. These shapes shown in the figure are common in systems in the wear-out phase, such as the mechanical phase, in accumulation systems, such as pipes, and in the wearing of tools.

Figure 3. Hazard function of the RW distribution for different values of μ and σ.



Quantile function: The quantile function (qf) of the RW distribution is given by the following:

$$q(p;\mu,\sigma) = -\left(-\frac{1}{\mu}\log(p)\right)^{\frac{1}{\sigma}}, \tag{5}$$

where $0 < p < 1$ and $\mu,\sigma > 0$. The qf is obtained by getting the inverse function of the cdf. The qf calculates the time at which a fraction or proportion p of the units is expected to fail.

Random function: The random function (rf) of the RW distribution was built from the Inverse Transform Algorithm (Ross, 2012) and is given by the following:

$$r(p;\mu,\sigma) = -\left(-\frac{1}{\mu}\log(p)\right)^{\frac{1}{\sigma}}, \tag{6}$$

where p is a random number from the uniform distribution in the interval [0,1] and $\mu,\sigma > 0$. The rf is used as the basis for generating deviations from the RW distribution.

Revista EIA | https://doi.org/10.24050/reia.v19i38.1526

2.2. Parameters estimation

This section presents the estimation method for the RW distribution based on the maximum likelihood method. Let be $y_1, y_2, \ldots, y_n$ a random sample from (1) and $\boldsymbol{\theta} = (\mu, \sigma)^T$ be the vector of unknown parameters. The log-likelihood function for $\boldsymbol{\theta}$ is given by the following:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\log(\mu) + \log(\sigma) + (\sigma - 1) \log(-y_i) - \mu((-y_i)^{\sigma}) \right) \tag{7}$$

By taking the partial derivates of $l(\boldsymbol{\theta})$ with respect to μ and σ, and letting them be equal to zero, we have the following two equations:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu} = \sum_{i=1}^{n} \left\{ \frac{1}{\mu} - (-y_i)^{\sigma} \right\} = 0 \tag{8}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = \sum_{i=1}^{n} \left\{ \frac{1}{\sigma} + \log(-y_i) - \mu(-y_i)^{\sigma} \log(-y_i) \right\} = 0 \tag{9}$$

The maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ are obtained by solving the likelihood equations (8) and (9). One contribution of this paper is that we created a procedure in the language programming R (2020) to take advantage of the gamlss package to solve the expressions (8) and (9) to obtain the maximum likelihood estimates of $\boldsymbol{\theta}$. The gamlss package has three algorithms to obtain the maximum likelihood estimates: the Rigby and Stasinopoulos (RS) algorithm, the Cole and Green (CG) algorithm, and the mixed algorithm.

2.3. Reflected Weibull regression model

In this section, we present the RW regression model using the maximum likelihood method in the framework of the generalized additive model for location, scale and shape (GAMLSS) (Rigby & Stasinopoulos, 2005). These models have some advantages over linear models because they are flexible, and they have no limitations with the Gaussian assumption (Hernández et al., 2021). The RW regression model assumes that the observations y_i for $i = 1, \ldots, n$ are independent with a probability density function $f(y; \mu, \sigma)$ given in expression (1). The structure of the RW regression model is as follows:

$$g_1(\mu) = \eta_1 = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \boldsymbol{Z}_{j1} \varsigma_{j1}, \tag{10}$$

$$g_2(\sigma) = \eta_2 = \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \boldsymbol{Z}_{j2} \boldsymbol{\varsigma}_{j2}, \tag{11}$$

where $g_k(.)$ is a known monotonic link function for $k = 1,2$ to map the linear predictor into the parameter domain, by default, these functions are log ($\square$); $\boldsymbol{X}_k$ are known design matrices of order $n \times J_k'$ associated with the fixed effects $\boldsymbol{\beta}_k$ of $J_k' \times 1$ and $\boldsymbol{Z}_{jk}$ are known design matrices of order $n \times q_{jk}$ associated with the random effects $\boldsymbol{\varsigma}_{jk}$ of $q_{jk} \times 1$ with multivariate normal distribution. The quantity J_k' represents the number of covariates used in the fixed effects of η_k, while J_k represents the number of random effects in η_k.

The diagnostics for the RW regression model is based on the normalized quantile residuals (Dunn & Smyth, 1996). The main advantage of the normalized quantile residuals is that the true residuals always have a standard normal distribution when the assumed model is correct. Since the checking of model assumptions via the normality of residuals is well established within the statistical literature, the normalized quantile residuals provide us with a familiar way to check the adequacy of the fitted model (Stasinopoulos et al., 2017).

2.4. RelDists package

We developed the RelDists package in the language programming R (2020) to implement new distributions proposed in the reliability field, and one of these distributions is the RW distribution. With the functions in the RelDists package, the user could estimate the RW distribution parameters, estimate effects for the RW regression model, and obtain the pdf, cdf, hf, qf and rf for the RW distribution. The package also provides useful tools commonly used in the reliability field, such as parameter estimation, graphic analysis, and regression analysis. The online documentation of the package can be consulted in the url https://ousuga.github.io/RelDists.

To install the RelDists package the user could copy and paste the next instructions in R console:

```
if (!require("devtools"))
install.packages("devtools")
devtools::install_github("ousuga/RelDists",force=TRUE)
library(RelDists)
```

The functions implemented in RelDists package for the RW distribution are described in Table 1.

Table 1. Functions and arguments for the RW distribution in the RelDists package.

Function name	Function
Probability density function	`dRW(y, mu, sigma)`
Cumulative density function	`pRW(q, mu, sigma)`
Hazard function	`hRW(y, mu, sigma)`
Quantile function	`qRW(p, mu, sigma)`
Random function	`rRW(n, mu, sigma)`
Reflected Weibull family	`RW(mu.link="log", sigma.link="log")`

The probability density function dRW(y, mu, sigma)allows the user to draw the probability density curve and identify the area in which the random variable Y is most likely to take values. The red curve of the right panel in Figure 1 can be obtained with:

```
curve(dRW(x, mu=1, sigma=3),from=-3, to=-0.01)
```

In practice, the cumulative density function pRW(q, mu, sigma)allows the user to obtain the probability that the random variable is less than or equal to a specific value. From the red curve on the left panel in Figure 2, it is observed that the probability that the random variable Y takes a value lower than -1.0 is small and is exactly equal to 0.22, which is obtained from:

```
pRW(q=-1, mu=1.5, sigma=3)
```

The hazard function for an item with Y=-2, assuming μ=0.5 and σ=2, can be obtained with the instruction:

```
hRW(y=-2, mu=0.5, sigma=2)
```

The quantile function qRW(p, mu, sigma)is useful for identifying specific quantiles, for example, the 0.25, 0.50, and 0.75 quartiles. In the case of $Y \sim RW(\mu = 1, \sigma = 3)$, the median $\tilde{Y} = -0.88$ is obtained as follow:

```
qRW(p=0.5, mu=1, sigma=3)
```

In practice, the random function rRW(n, mu, sigma) can be used to generate random samples from the RW distribution for creating simulation scenarios to validate something of interest in RW the distribution. For example, for the $Y \sim RW(\mu = 1, \sigma = 3)$ we can generate ten random numbers as follows:

```
rRW(n=10, mu=1, sigma=3)
```

Finally, the family function RW(mu.link="log", sigma.link="log") is used in the fitting procedure to estimate the parameters for the RW distribution and the regression coefficients for the RW model. For example, if a user has a random sample stored in a vector called ransam, the code to estimate the parameters for the RW distribution is as follows:

```
fit <- gamlss(ransam ~ 1, family=RW)
coef(fit)
```

The RelDists package uses three methods for parameter estimation provided by GAMLSS models (Stasinopoulos et al., 2017). The method is specified in the argument called method, with default method=RS(). The user may specify method=CG(), or a combination of both algorithms with method=mixed(). In the case of the RelDists distributions, it is recommended to use the CG method because the cross derivates are calculated manually, offering less computation time to fit the models.

3. Simulation study

In this section, we present a Monte Carlo simulation study to analyze the performance of the estimation procedure considering two cases, namely, "without covariates" and "with covariates". In both cases, we simulated left censored responses and noncensored data. For this simulation study we use the ADEMP methodology shown in Figure 4 proposed by Morris, White, and Crowther (2019), which was slightly modified for adaption to this study. The methodology involves defining aims, data-generation mechanisms, estimands, methods, and performance measures.

Figure 4. ADEMP methodology used in the simulation study.

The measures used to evaluate the parameter estimation procedure were the mean value and the mean squared error (*MSE*) of the estimated components of the parameter vector $\boldsymbol{\theta}$ in each case. The mean value was chosen to see how similar the estimation was compared with the true value. The *MSE* was selected because it contains the bias-variance trade off, which is important to see how good a model is fitted. The mean value for each estimated parameter $\hat{\theta}_k$ is given by the following:

$$\bar{\theta}_k = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_{ki}, \tag{12}$$

where m is the number of the simulation. The *MSE* for the *k* estimated parameter $\hat{\theta}_k$ is given by the following:

$$MSE_k = Var(\hat{\theta}_k) + \left(Bias(\hat{\theta}_k, \theta_k)\right)^2 \tag{13}$$

The methodology for the cases analyzed in the simulation study is below.

Case 1: simulation without covariates

In this case, we analyzed the asymptotic behavior of the estimated parameters, $\hat{\mu}$ and $\hat{\sigma}$, of the RW distribution when there are not covariates, that is, when μ and σ are fixed quantities. We consider different sample sizes n=30,50,...,970,1000 and different percentages of censored data *pcd* = 0%,10%,30%,50%. The datasets were generated from a model assuming μ = 0.8 and σ =1.2 as given below.

$$\begin{gathered} Y_i \sim RW(\mu, \sigma), \\ \mu = 0.8, \\ \sigma = 1.2 \end{gathered} \tag{14}$$

For the combinations of sample size *n* and percentage of censored data *pcd*, we followed the next steps to simulate the data and to estimate the parameters:

1. Generate a random sample of size n from the population *RW* (μ = 0.8, σ = 1.2).
2. Modify the simulated random sample to ensure a *pcd* of left censored observations.
3. Obtain and store the estimations $\hat{\mu}$ and $\hat{\sigma}$ using the proposed procedure.
4. Repeat m=100000 times the steps 1 to 3.

Case 2: simulation with covariates

In this case, we analyzed the asymptotic behavior of the estimated parameters of the RW regression model along different sample sizes n = 30,50,...,970,1000 and different percentages of censored data pcd = 0%,10%,30%,50%. The datasets were generated from the next model:

$$\begin{aligned} Y_i &\sim RW(\mu_i, \sigma_i), \\ \log(\mu_i) &= \beta_0 + \beta_1 X_{i1}, \\ \log(\sigma_i) &= \gamma_0 + \gamma_1 X_{i2}, \end{aligned} \tag{15}$$

where the variables X_1 and X_2 were generated from $X_1 \sim U(0.4,0.6)$ and $X_2 \sim U(0.4,0.6)$.

The parameter vector was fixed as $\boldsymbol{\theta} = (\beta_0 = 1.5, \beta_1 = -1.5, \gamma_0 = 2, \gamma_1 = -2)^T$. For the combinations of sample size n and percentage of censored data pcd, we followed the next steps to simulate the data and to estimate the parameters:

1. Generate a random sample of n values from a RW distribution given in the model.
2. Modify the simulated random sample to ensure the pcd of left censored observations.
3. Obtain and store the estimations $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_0$ and $\hat{\gamma}_1$ using the proposed procedure.
4. Repeat m = 100000 times the steps 1 to 3.

4. Results

In this section, we present the results of the simulation study. For each case, we show a figure for the mean and the MSE to explore the evolution of the estimated parameters as n and pcd increase.

Case 1: simulation without covariates

Figure 5 describes the mean of the estimated parameters $\hat{\mu}$ and $\hat{\sigma}$ for different censored percentages and sample sizes n. We can observe that as the sample size increases, the mean value of $\hat{\mu}$ and $\hat{\sigma}$ tend to the true value of the parameters represented by the red lines. Additionally, we can observe from the figure that as the percentage of censored data decreases, there is a lower bias in the estimations of $\hat{\mu}$ and $\hat{\sigma}$.

Figure 5. Mean of the estimated parameters $\hat{\mu}$ *and* $\hat{\sigma}$ versus *n* for different percentage of censored data. Red lines correspond to the true values $\mu = 0.8$ and $\sigma = 1.2$.

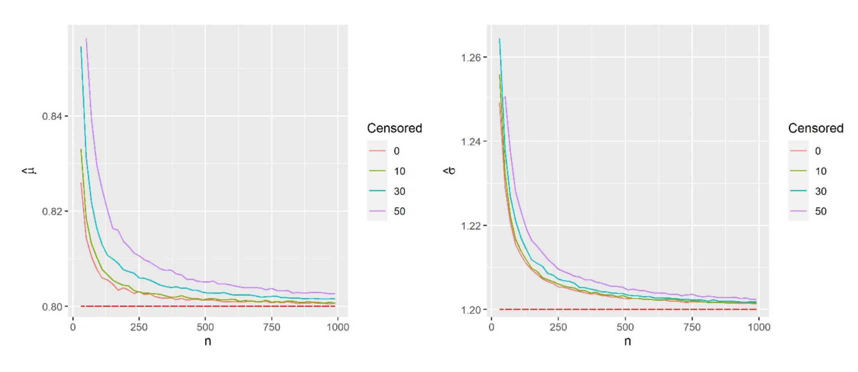
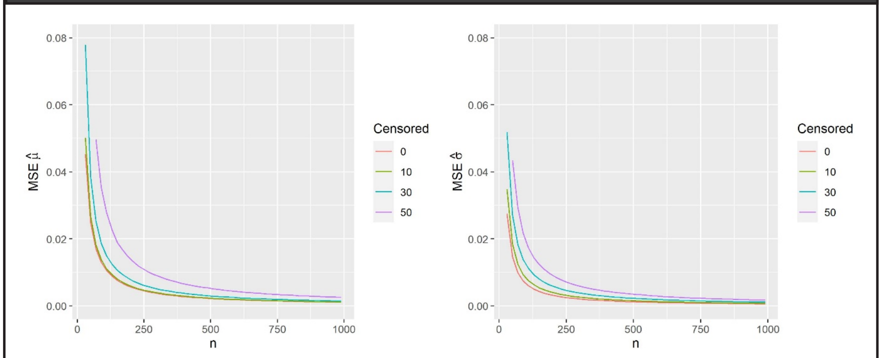


Figure 6 shows the MSE trend for the estimated parameters $\hat{\mu}$ and $\hat{\sigma}$ for each value of *n*. As the sample size increases, the value of the MSE decreases, bringing MSE near zero. Additionally, as the percentage of censored data increases, the MSE tends to increases.

Figure 6. Mean squared error for the estimated parameters $\hat{\mu}$ *and* $\hat{\sigma}$ versus *n* for different percentage of censored data.



For both measures, mean and MSE, it can be observed that the estimation is different for every censored data. For sample sizes with $n < 500$, we observed that the impact of the censure is significant, being worse for 50% and 30% of censored data; for sample sizes with $n > 500$, the MSE decreases, and the lines begin to overlap.

Finally, from Figure 5 and Figure 6 we observe, as expected, as the *n* increases and/or the percentage of censoring *pcd* decreases, the estimation of μ and σ tends to be closer to the true parameters.

Case 2: simulation with covariates

Figure 7 shows the mean of the estimated parameters $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ versus *n* for different percentages of censored data. From this figure, we observe a general pattern, as the sample size increases, the estimations tend to the true values.

Revista EIA https://doi.org/10.24050/reia.v19i38.1526

The censure seems to have a slight effect on the estimations, except for the curves associated with β_0 and β_1 that move away from the true values when the percentage of censored data is 50%.The estimations for γ_0 and γ_1 seem not to be affected by the percentages of censure. Additionally, the oscillations for each curve tend to decrease, even for the estimations of γ_1. Last, the percentage of censoring seems only to affect the estimations of β_0 and β_1.

Figure 7. Mean of the estimated parameters $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ versus *n* for different percentages of censored data. Red line corresponds to the objective value.

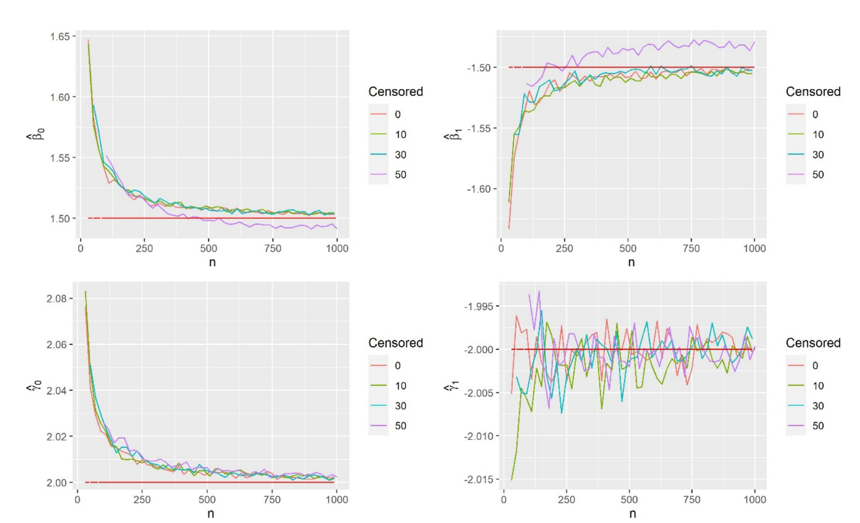
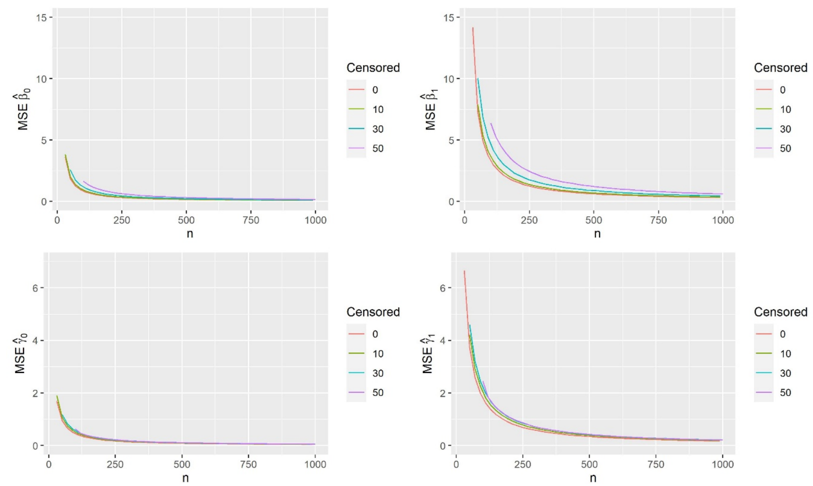


Figure 8 shows the MSE trend of the estimated parameters β_0, β_1, γ_0 and γ_1 of the regression model for each value of *n*. As the sample size increases in the panels, the value of the MSE decreases, bringing it closer to zero. The error between the real values of the parameters and the estimated values is increasingly less. The MSE for β_0 decreases quickly because it has less variance than β_1, which has much variance, especially in the percentages 30% and 50%. Something similar happens with γ_0 that has less variance than γ_1, but for both of them the bias is small causing overlapped curves.

Figure 8. Mean squared error for the estimated parameters $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ versus n for different percentages of censored data.

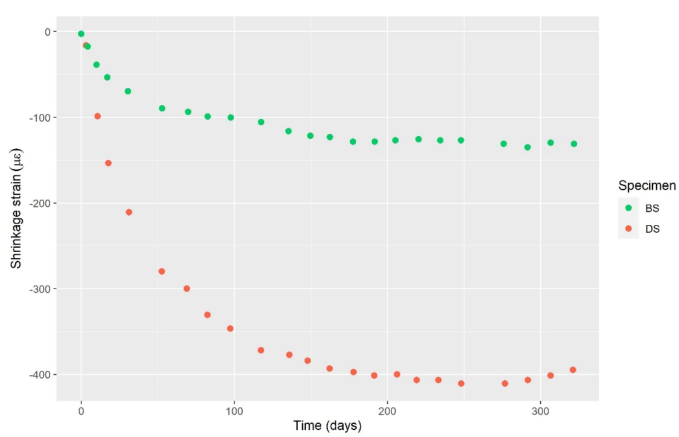


In both cases, the behavior of the mean and the MSE are consistent with similar simulations using the Weibull distribution made by authors such as Gibbons and Vance (1983), Guure and Ibrahim (2012), Guure and Ibrahim (2013), Kim and Yum (2008), Kim et al. (2011), Phadnis et al. (2020) and Odell, Anderson, and D'Agostino (1992).

5. Application

The study developed by Huang et al. (2019) analyzed the creep behavior of small scale concrete specimens under eccentric compression loads. To examine the effect of load eccentricity and drying of concrete over time, four 100×100×450 mm concrete specimens (prisms) were tested. Specimens DS were used to measure drying shrinkage strains, whereas specimens BS were epoxy coated on all faces to examine basic shrinkage without concrete drying effects. Figure 9 shows the measured drying shrinkage strains (Y) of specimens BS and DS.

Figure 9. Measured drying shrinkage strains of specimens BS and DS.



Revista EIA https://doi.org/10.24050/reia.v19i38.1526

For both datasets (BS and DS) were fitted models to explain the variable shrinkage strain (Y) in specimens as a function of the time. For each model, four distributions were assumed for the response variable: reflected Weibull, log-normal, Generalized Inverse Gaussian, and Gamma.

To compare the models we used the generalized Akaike information criterion ($GAIC$) (Akaike, 1983) given by the following:

$$GAIC(k) = -2\log(\hat{l}) + k \times df,$$

where $\hat{l}$ corresponds to the likelihood of the current fitted model, df denotes the total degrees of freedom (the effective number of parameters) of the model, and k is the penalty for each degree of freedom. Hence, $GAIC$ (k = 2) gives the Akaike information criterion (AIC) (Akaike, 1983) and $GAIC$ (k = log (n)) gives the Schwarz Bayesian criterion (SBC) or the Bayesian information criterion (BIC) (Schwarz, 1978). The model with the lowest value of $GAIC$ (k) for a chosen value of k is selected as the "best" model (Stasinopoulos et al., 2017).

Additionally, we compare the models using the pseudo correlation coefficient R^2_{Pseudo} proposed by Nagelkerke (1991) and the correlation coefficient $\rho(Y,\hat{Y})$ between the observed values Y and the estimated mean values $\hat{Y}$. Table 2 shows the measures obtain for each model in the two datasets. From this table, we can observe that the model assuming the reflected Weibull distribution was the model with the best measures (bold numbers).

Table 2. Measures to compare the models fitted to the BS and DS datasets. Bold numbers represent the best values in each column.

	BS dataset			DS dataset		
Distribution for Y	$GAIC$	R^2_{Pseudo}	$\rho(Y,\hat{Y})$	$GAIC$	R^2_{Pseudo}	$\rho(Y,\hat{Y})$
Reflected Weibull	**217.62**	**0.75**	**0.96**	**249.03**	**0.80**	**0.97**
log-normal	240.31	**0.75**	0.90	272.83	0.79	0.92
Generalized Inverse Gaussian	235.61	0.66	0.93	269.92	0.71	0.94
Gamma	233.61	0.66	0.93	267.92	0.71	0.94

Detailed results for the reflected Weibull models fitted to specimens BS and DS are shown in Table 3 and Table 4, respectively. For both models, we found that a 3-degree polynomial of the time (T) is appropriate to explain the parameter μ whereas that for the parameter σ we found that no terms are needed.

Table 3. Estimated effects for the model of specimens BS.

Estimated effects for $\log(\mu)$				
	Estimate	Standard error	t-value	p-value
Intercept	-23.038	4.072	-5.658	0.000
T	-9.149	1.755	-5.212	0.000
T^2	5.657	1.441	3.926	0.000
T^3	-3.319	1.264	-2.625	0.017
Estimated effects for $\log(\sigma)$				
	Estimate	Standard error	t-value	p-value
Intercept	1.621	0.175	9.266	0.000

Table 4. Estimated effects for the model of specimens DS.

Estimated effects for $\log(\mu)$				
	Estimate	Standard error	t-value	p-value
Intercept	-34.478	0.552	-66.043	0.000
T	-9.551	1.518	-6.292	0.000
T^2	6.385	1.739	3.672	0.000
T^3	-4.077	1.534	-2.658	0.016
Estimated effects for $\log(\sigma)$				
	Estimate	Standard error	t-value	p-value
Intercept	1.789	0.013	135.100	0.000

With the information in Table 3, it is possible to write the mathematical expressions for the estimated parameters $\hat{\mu}$ and $\hat{\sigma}$ of the RW model of specimen BS. In the RW regression model, both parameters μ and σ, must be positive, for this reason, the log link function is used in the following expressions:

$$\log(\hat{\mu}) = -23.038 - 9.149\, T + 5.657\, T^2 - 3.319\, T^3$$

$$\log(\hat{\sigma}) = 1.621$$

Similarly, using the information in Table 4, it is possible to write the mathematical expressions for the estimated parameters $\hat{\mu}$ and $\hat{\sigma}$ of the RW model of specimen DS:

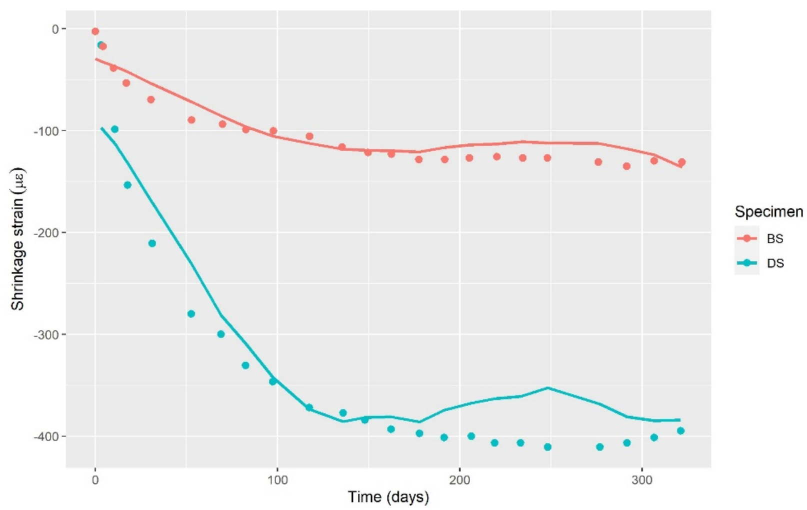
$$\log(\hat{\mu}) = -34.478 - 9.551\, T + 6.385\, T^2 - 4.077\, T^3$$

$$\log(\hat{\sigma}) = 1.789$$

Using the fitted models above, we can obtain the expected value $E(Y|T = t)$ for the response variable. Figure 10 depicts the observed data (points) and the expected values of Y as a function of time for both BS and DS specimens. From this figure, we observe that curves follow the experimental data. This fact confirms that both fitted reflected Weibull models explain the phenomena.

Revista EIA https://doi.org/10.24050/reia.v19i38.1526

Figure 10. Fitted drying shrinkage strains for specimens BS and DS.



6. Conclusions

The RelDists package is an useful tool for the academic, scientific, and business communities, because it has functions to achieve probabilities, percentiles, densities, and random variates for the RW distribution. Additionally, the RelDists package has helpful functions to perform parameter estimations and regression models. These characteristics allow users to do reliability studies that provide support for decision-makers.

From the simulation study, we found that the estimations tend to go to the true values as the sample size increases, and the percentage of censoring decreases. Additionally, we found that for large sample sizes, the percentage of censoring seems to have less impact over the maximum likelihood estimators. This behavior points out that the regression model offers accurate predictions as there is more information available.

References

Akaike, H. (1983). Information measures and model selection. Proceedings of the 44th Session of the International Statistical Institute. Budapest: Akademiai Kiado. pp. 277–291.

Al Abbasi, J. N.; Risan, H. K.; Resen, I. A. (2018). Application of Kumaraswamy Extreme Values Distributions to Earthquake Magnitudes in Iraq and Conterminous Regions. International Journal of Applied Engineering Research, 13(11), pp. 8971–8980.

Ali, M. M.; Woo, J. (2006). Skew-symmetric reflected distributions. Soochow Journal of Mathematics, 32(2), pp. 233–240. https://doi.org/10.1080/01966324.2008.10737716

Almalki, S. J.; Nadarajah, S. (2014). Modifications of the Weibull distribution: A review. Reliability Engineering & System Safety, 124(4), pp. 32–55. https://doi.org/https://doi.org/10.1016/j.ress.2013.11.010

Balakrishnan, N.; Kocherlakota, S. (1985). On the double Weibull distribution: order statistics and estimation. Sankhya: The Indian Journal of Statistics, Series B, 47(2), pp. 161–178.

Barreto-Souza, W.; Santos, A. H. S.; Cordeiro, G. M. (2010). The beta generalized exponential distribution. Journal of Statistical Computation and Simulation, 80(2), pp. 159–172. https://doi.org/10.1080/00949650802552402

Caron, R.; Sinha, D.; Dey, D.; Polpo, A. (2017). Categorical data analysis using a skewed Weibull regression model. Entropy. 20(3), pp. 176–193. Multidisciplinary Digital Publishing Institute. https://doi.org/https://doi.org/10.3390/e20030176

Cohen, A. C. (1975). Multi-censored sampling in the three parameter Weibull distribution. Technometrics, 17(3), pp. 347–351. https://doi.org/https://doi.org/10.2307/1268072

Cohen, A. C. (2016). Truncated and censored samples: theory and applications. Boca Raton. CRC press. pp 85-172.

Cohen, C. A.; Whitten, B. (1982). Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. Communications in Statistics-Theory and Methods, 11(23), pp. 2631–2656. https://doi.org/https://doi.org/10.1080/03610928208828412

Dunn, P. K.; Smyth, G. K. (1996). Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5(3), pp. 236–244. https://doi.org/https://doi.org/10.2307/1390802

Gibbons, D. I.; Vance, L. C. (1983). Estimators for the 2-parameter Weibull distribution with progressively censored samples. IEEE Transactions on Reliability, 32(1), pp. 95–99. https://doi.org/10.1109/TR.1983.5221484

Guure, C. B.; Ibrahim, N. A. (2012). Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data. Mathematical Problems in Engineering, 2012. https://doi.org/https://doi.org/10.1155/2012/329489

Guure, C. B.; Ibrahim, N. A. (2013). Methods for estimating the 2-parameter Weibull distribution with type-I censored data. Research Journal of Applied Sciences, Engineering and Technology, 5(3), pp. 689–694. https://doi.org/10.19026/rjaset.5.5010

Hernández, B. F.; Cano, B. U.; Caicedo, E. A. C. (2021). Modelos GAMLSS para analizar el grado secado de calcio dihidratado. Revista EIA, 18(35), pp. 1–13. https://doi.org/https://doi.org/10.24050/reia.v18i35.1439

Huang, H.; Garcia, R.,;Huang, S.S.; Guadagnini, M.; Pilakoutas, K. (2019). A practical creep model for concrete elements under eccentric compression. Materials and Structures, 52(6), pp. 1–18. https://doi.org/10.1617/s11527-019-1432-z

Kalsoom, U.; Nasir, W.; Syed, A. (2019). On estimation of reflected Weibull distribution using bayesian analysis under informative prior. 15th Islamic Countries Conference on Statistical Sciences (ICCS-15), 49. Pakistan. pp. 93-106. http://www.isoss.net/downloads/ProcICCS15.pdf.

Kim, C.; Jung, J.; Chung, Y. (2011). Bayesian estimation for the exponentiated Weibull model under Type II progressive censoring. Statistical Papers, 52(1), pp. 53–70. https://doi.org/https://doi.org/10.1007/s00362-009-0203-2

Kim, J. S.; Yum, B.-J. (2008). Selection between Weibull and lognormal distributions: A comparative simulation study. Computational Statistics & Data Analysis, 53(2), pp. 477–485. https://doi.org/https://doi.org/10.1016/j.csda.2008.08.012

Lai, C.-D. (2014). Generalized weibull distributions. Springer.

Lee, C.; Famoye, F.; Olumolade, O. (2007). Beta-Weibull distribution: some properties and applications to censored data. Journal of Modern Applied Statistical Methods, 6(1), pp. 173–186. https://doi.org/DOI: 10.22237/jmasm/1177992960

Meeker, W. Q.; Escobar, L. A. (2014). Statistical methods for reliability data. John Wiley & Sons.

Modarres, M.; Kaminskiy, M. P.; Krivtsov, V. (2016). Reliability engineering and risk analysis: a practical guide. Boca Raton. CRC press. pp. 1-13.

Morris, T. P.; White, I. R.; Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in Medicine, 38(11), pp. 2074–2102. https://doi.org/https://doi.org/10.1002/sim.8086

Nagatsuka, H.; Kamakura, T.; Balakrishnan, N. (2013). A consistent method of estimation for the three-parameter Weibull distribution. Computational Statistics & Data Analysis, 58(1), pp. 210–226. https://doi.org/https://doi.org/10.1016/j.csda.2012.09.005

Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. Biometrika, 78(3), pp. 691–692. https://doi.org/https://doi.org/10.1093/biomet/78.3.691

Odell, P. M.; Anderson, K. M.; D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. Biometrics, 48(3), pp. 951–959. https://doi.org/https://doi.org/10.2307/2532360

Orjubin, G. (2007). Maximum field inside a reverberation chamber modeled by the generalized extreme value distribution. IEEE Transactions on Electromagnetic Compatibility, 49(1), pp. 104–113. https://doi.org/10.1109/TEMC.2006.888172

Phadnis, M. A.; Sharma, P.; Thewarapperuma, N.; Chalise, P. (2020). Assessing accuracy of Weibull shape parameter estimate from historical studies for subsequent sample size calculation in clinical trials with time-to-event outcome. Contemporary Clinical Trials Communications, 17(1). https://doi.org/10.1016/j.conctc.2020.100548

R Core Team. (2021). R: A Language and Environment for Statistical Computing. https://www.r-project.org/

Regal, R. R.; Larntz, K. (1978). Likelihood methods for testing group problem solving models with censored data. Psychometrika, 43(3), pp. 353–366. https://doi.org/https://doi.org/10.1007/BF02293645

Rigby, R. A.; Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), pp. 507–554. https://doi.org/https://doi.org/10.1111/j.1467-9876.2005.00510.x

Ross, M. S. (2012). Simulation. San Diego. Elsevier.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), pp. 461–464.

Stacy, E. W. (1962). A generalization of the gamma distribution. The Annals of Mathematical Statistics, 33(3), pp. 1187–1192.

Stasinopoulos, M. D.; Rigby, R. A.; Heller, G. Z.; Voudouris, V.; Bastiani, F. (2017). Flexible Regression and Smoothing Using GAMLSS in R. Boca Raton. CRC Press.

Wei, Z.; Start, M.; Hamilton, J.; Luo, L. (2016). A unified framework for representing product validation testing methods and conducting reliability analysis. SAE International Journal of Materials and Manufacturing, 9(2), pp. 303–314. https://doi.org/https://doi.org/10.4271/2016-01-0269

Xie, M.; Lai, C. D. (1996). Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. Reliability Engineering & System Safety, 52(1), pp. 87–93. https://doi.org/https://doi.org/10.1016/0951-8320(95)00149-2

Zhang, T.; Xie, M. (2007). Failure data analysis with extended Weibull distribution. Communications in Statistics—Simulation and Computation, 36(3), pp. 579–592. https://doi.org/https://doi.org/10.1080/03610910701236081