STATISTICS IN TRANSITION new series, June 2026
Vol. 27, No. 2, pp. xx–XX, https://doi.org/10.59139/stattrans-2026-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Mean absolute deviations for the Weibull distribution: applications in survival analysis and insurance claims

Weiqi Zhang[1], **Zibo Wang**[2], **Eugene Pinsky**[3]

Abstract

The Weibull distribution is widely applied in fields such as survival analysis, reliability engineering, failure analysis, and extreme value theory. Traditionally, Maximum Likelihood Estimation (MLE) has been commonly used to estimate the parameters of this distribution. In this paper, we derive a new formula for the mean absolute deviation (MAD) about the median. We use this formula to derive a MAD-based parameter-estimation method that is computationally simpler than MLE. We apply our results to analyze the survival times of breast cancer patients and insurance claim amounts, providing evidence from biomedical and actuarial domains. We estimate parameters using MLE, MAD, and quantile-based approaches. The results show that the proposed MAD-based approach is superior to the other two methods. It demonstrates the practical application of MAD methods in survival analysis and financial risk modeling of insurance claims, where accurate modeling is crucial for understanding extreme outcomes.

Key words: Weibull.

1. Introduction

In various fields, from healthcare to manufacturing, accurate statistical models are essential for evaluating risks and understanding the reliability and survival of systems or patients (Carroll, 2003). These models play a crucial role in quantifying the probabilities of events occurring over time. By capturing the temporal dynamics of such events, they enable a deeper understanding of risk factors, reliability patterns (Almeida, 1999), and longevity predictions across various domains, including engineering (Kang 2018), finance (Chen, 2011; Gebizlioglu 2011), and healthcare (Quiroz, 2024), to name just a few. The Weibull distribution is widely utilized in survival analysis and reliability engineering due to its flexibility in modeling various failure rates, making it a versatile tool for analyzing time-to-event data and estimating the probability of survival or failure under different conditions (Fogliatto, 2019), (Yoosefi, 2018).

For the Weibull distribution, the Maximum Likelihood Estimation (MLE) is the main method for parameter estimation (Cohen, 1965; Balakrishnan, 2008). However, MLE has some limitations, and the MLE method is subject to errors when the data distribution is

---

[1] Department of Computer Science, Metropolitan College, Boston University, United States. ORCID: https://orcid.org/0009-0009-2536-9529.

[2] Department of Computer Science, Metropolitan College, Boston University, United States. ORCID: https://orcid.org/0009-0008-6243-3131.

[3] Department of Computer Science, Metropolitan College, Boston University, United States. E-mail: epinsky@bu.edu. ORCID: https://orcid.org/0000-0002-3836-1851.

© W. Zhang, Z. Wang, E. Pinsky. Article available under the CC BY-SA 4.0 licence

skewed, or outliers are present (Jacquelin, 1993). In such cases, researchers are exploring alternative methods to provide more accurate and robust estimates (Cohen,1982; Teimouri, 2011; Pobockova, 2014). Some researchers use the Quantiles method as an alternative to maximum likelihood estimation (Jokiel, 2024). The use of Quantiles offers an insightful approach for analyzing the variability in the estimated values of α, revealing important aspects of the data's structure and concentration patterns (Pinsky, 2024). This method, which estimates parameters based on specific percentiles, captures the distribution's characteristics in a simplified manner but still fails to eliminate the effects of outliers.

Research has shown that MAD exhibits greater resistance to extreme values and delivers more reliable estimates, even in the presence of data irregularities, thus reducing the impact of outliers (Hortobagyi, 2022). In healthcare, building accurate and robust statistical models of patient survival can inform treatment strategies, guide resource allocation, and improve patient care by accurately estimating survival probabilities (Cancho, 2020). Accurate survival estimates are critical for predicting outcomes and assessing risk probabilities.

In this paper, we derive formulas for calculating MAD around the median of the Weibull distribution. The median is particularly important in cancer survival studies because it provides a robust measure of central tendency that is less susceptible to extreme values, making it a reliable reference point for survival outcomes. In addition, we derive three methods for calculating parameters using MLE, quantiles, and MAD. Finally, we used all three methods to predict survival in breast cancer patients and model insurance claim amounts, demonstrating the method's versatility across biomedical and actuarial contexts. It was found that MAD provided more accurate predictions with less error than MLE and quantiles. The aim of this study is to provide a method for calculating Weibull distribution parameters using MAD to improve the validity of statistical analysis by enabling more reliable estimation techniques and modeling methods in survival analysis and financial risk assessment.

2. Mean absolute deviations for the Weibull distribution

The Weibull distribution is defined by the following probability density function (PDF) and cumulative distribution function (CDF):

$$f(x) = \frac{k}{\alpha}\left(\frac{x}{\alpha}\right)^{k-1} e^{-(x/\alpha)^k} \quad \text{and} \quad F(x) = 1 - e^{-(x/\alpha)^k}, \quad x \geq 0 \tag{1}$$

The quantile function $Q(p)$ is $Q(p) = \alpha\left(-\log(1-p)\right)^{1/k}$. The quartiles for Weibull distributions are $Q_1 = \alpha(\log {}^{4}\!/_{3})^{1/k}$, $M = \alpha(\log 2)^{1/k}$ and $Q_3 = \alpha(2\log 2)^{1/k}$. The mean μ and variance σ^2 are given by:

$$\mu = \alpha\Gamma\left(1+\frac{1}{k}\right), \quad \sigma^2 = \alpha^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right] \tag{2}$$

where $\Gamma(\cdot)$ denotes the Gamma function (Olver, 2010):

$$\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}\,dt \tag{3}$$

To compute mean absolute deviations, we will find it convenient to introduce the auxiliary integral:

$$I(z) = \int_{x \leq z} x\,dF(x) \tag{4}$$

We can consider $I(z)$ as the partial mean of X computed over all $x \leq z$. The mean absolute deviation (MAD) of X from a constant c is defined as:

$$H(X,c) = \int_{x \leq c} (c - x)\,dF(x) + \int_{x > c} (x - c)\,dF(x) \tag{5}$$

This can be rewritten as:

$$H(X,c) = c\,(2F(c) - 1) + \mu - 2I(c) \tag{6}$$

We can express $I(c)$ as:

$$I(c) = \int_{x \leq c} x\,dF(x) = cF(c) - \int_{x \leq c} F(x)\,dx \tag{7}$$

Substituting this into equation (6) gives:

$$\begin{aligned} H(X,\,c) &= (\mu - c) + 2\alpha \int_0^{c/\alpha} \left(1 - e^{-z^k}\right) dz \\ &= (\mu + c) - 2\alpha \int_0^{c/\alpha} e^{-z^k}\,dz \end{aligned} \tag{8}$$

From the definition of the upper incomplete Gamma function (Olver, 2010)

$$\Gamma(s,\,x) = \int_x^{\infty} t^{s-1} e^{-t}\,dt \tag{9}$$

we obtain from equation (8)

$$H(X,\,c) = (\mu + c) + \frac{2\alpha}{k}\Gamma\left(\frac{1}{k},\,\left(\frac{c}{\alpha}\right)^k\right) - \frac{2\alpha}{k}\Gamma\left(\frac{1}{k},\,0\right) \tag{10}$$

For the Weibull distribution, the mean $\mu = \alpha\Gamma(1 + {}^1\!/_k) = ({}^\alpha\!/_k)\Gamma({}^1\!/_k)$. Thus, the mean absolute deviation around the mean is:

$$H(X,\,\mu) = \frac{2\alpha}{k}\Gamma\left(\frac{1}{k},\,\Gamma^k\left(1 + \frac{1}{k}\right)\right) \tag{11}$$

The median M of the Weibull distribution is $M = \alpha\,(\log 2)^{1/k}$. Therefore, the mean absolute deviation around the median is:

$$H(X,\,M) = \alpha\left[(\log 2)^{1/k} - \Gamma\left(1 + \frac{1}{k}\right) + \frac{2}{k}\Gamma\left(\frac{1}{k},\,\log 2\right)\right] \tag{12}$$

Example 1: For $k = 1$, we have the exponential distribution with rate $1/\alpha$. The mean $\mu = \alpha$

and median $M = \alpha \log 2$. The corresponding mean absolute deviations are: $H(X, \mu) = (2\alpha/e)$ and $H(X, M) = \alpha \log 2$

Example 2: For $k = 2$, we have the Rayleigh distribution with scale $\sigma = (\alpha/\sqrt{2})$. The mean is $\mu = \alpha(\sqrt{\pi}/2)$. Using the identities $\Gamma(3/2) = (\sqrt{\pi}/2)$ and $\Gamma(1/2,x) = \sqrt{\pi}\,\mathrm{erfc}(\sqrt{x})$, we compute:

$$H(X,\mu) = \alpha\sqrt{\pi}\,\mathrm{erfc}\left(\frac{\pi}{2}\right) \tag{13}$$

The median is $M = \alpha\sqrt{\log 2}$. Therefore, the MAD around the median for the Rayleigh distribution is:

$$H(X,M) = \alpha\left[\sqrt{\log 2} - \frac{\sqrt{\pi}}{2} + \sqrt{\pi}\,\mathrm{erfc}\left(\sqrt{\log 2}\right)\right] \tag{14}$$

Example 3: Consider the Frechet (inverse Weibull) distribution with shape $k > 0$, scale $\alpha > 0$ and location 0 [**?**]. Its density function (PDF) $fx)$ and cumulative distribution function (CDF) $F(x)$ given by:

$$f(x) = \frac{k}{\alpha}\left(\frac{x}{\alpha}\right)^{-k-1} e^{-(x/\alpha)^{-k}} \quad \text{and} \quad F(x) = 1 - e^{-(x/\alpha)^{-k}}, \quad x > 0 \tag{15}$$

The quantile function $Q(p)$ is $Q(p) = \alpha\,(-\log p)^{-1/k}$. The quartiles for this distribution are $Q_1 = \alpha\,(\log 4)^{-1/k}$, $M = \alpha\,(\log 2)^{-1/k}$, and $Q_3 = \alpha(\log 4/3)^{-1/k}$. The mean and variance are defined only for $k > 1$ and $k > 2$ respectively and are:

$$\mu = \alpha\Gamma\left(1 - \frac{1}{k}\right), \qquad \sigma^2 = \alpha^2\Gamma\left(1 - \frac{2}{k}\right) - \Gamma^2\left(1 - \frac{1}{k}\right) \tag{16}$$

For $k > 1$, the mean absolute deviations exist, and we can derive mean absolute deviations in a manner similar to Weibull in equations (6), (7), (8). For example, for $H(X, M)$ we obtain

$$H(X,M) = \alpha\left[(\log 2)^{-1/k} - \Gamma\left(1 - \frac{1}{k}\right) + \frac{2}{k}\Gamma\left(\frac{1}{k},\, (\log 2)^{-1}\right)\right] \tag{17}$$

in complete analogy with the Weibull distribution.

3. Confidence intervals and tail estimation

Let us examine how the MAD deviations can be used to estimate tail probabilities. The most general bound for any distribution (with finite variance) is Chebyshev's inequality (Feller, 1956):

$$P\left(|X - \mu| \geq b\sigma\right) \leq \frac{1}{b^2} \tag{18}$$

This inequality is useful for $b \geq 1$. This inequality follows from the so-called Pearson inequality (Feller, 1956) with $r = 2$:

$$P\left(|X - \mu| \geq bV_r^{1/r}\right) \leq \frac{1}{b^2}, \qquad \text{where} \qquad V_r = \mathrm{E}\left(|X - \mu|^r\right) \tag{19}$$

For $r = 1$, a much less-known inequality exists for bounds in terms of mean absolute deviation $H(X,\mu)$ from the mean, namely

$$P\Big(|X-\mu| \geq bH(X,\mu)\Big) \leq \frac{1}{b} \tag{20}$$

Similarly, there is an inequality in terms of the mean absolute deviation from the median $H(X,M)$ given by (PhamGia, 2001)

$$P\Big(|X-M| \geq bH(X,M)\Big) \leq \frac{1}{b} \tag{21}$$

Let us compute bounds based on the MAD deviation. First, define $\delta = H(X,\mu)/\sigma$. Note that for all distributions we have $H(X,M) \leq H(X,\mu) \leq \sigma$ and therefore $\delta \leq 1$. We can rewrite the MAD-based inequality for $H(X,\mu)$ in equation (20) in terms of σ as follows:

$$P(|X-\mu| \geq b\sigma) = P\left(|X-\mu| \geq \frac{b\sigma}{H(\mu)} \cdot H(\mu)\right) \leq \frac{\delta}{b} \tag{22}$$

Also, we can consider the Peek inequality.

$$P(|X-\mu| \geq b\sigma) \leq \frac{1-\delta^2}{b^2-2b\delta+1} \tag{23}$$

Comparing equations (18) and (22) we find that MAD-based upper bound for $H(X,\mu)$ is lower than Chebyshev's upper bound for $1 \leq b \leq 1/\delta$. Comparing equations (22) and (23) we find that the Peek inequality is lower than both MAD and Chebyshev for $b > 1/\delta$. This is shown in Figure 1.

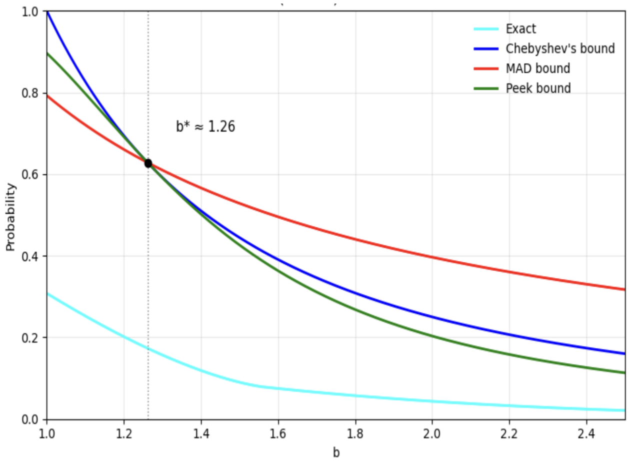

Figure 1. Comparison of Tail Probabilities for the Standard Symmetric Case

In Figure 1, we compare the exact tail probabilities with three classical bounds: the Chebyshev bound (equation (18)), the MAD-based bound (equation (22)), and the Peek bound (equation (23)). For the Weibull distribution with $k = 1.6$ and $\alpha = 1.0$, we obtain $\mu = 0.90$, $\sigma = 0.57$, $H(\mu) = 0.45$, and $\delta = 0.79$, which yields the intersection point $b^* = 1/\delta \approx 1.26$. For $1 \leq b < 1.26$, the MAD-based inequality provides a tighter upper bound than either Chebyshev or Peek. For $b > 1.26$, both Chebyshev and Peek improve upon the MAD bound, with the Peek inequality offering the sharpest estimate. Nevertheless, none of these inequalities approximate the exact tail probabilities well, particularly for larger values of b.

4. Parameter estimation

The most commonly used method for estimating the Weibull distribution parameters α (scale) and k (shape) is Maximum Likelihood Estimation (MLE). MLE efficiently estimates parameters by maximizing the likelihood function, making it statistically robust and precise with large samples. However, MLE can be computationally intensive and sensitive to small sample sizes or outliers. For this reason, we compare MLE with alternative methods, such as the Quantiles and Mean Absolute Deviation (MAD) about the median approaches, which offer simpler computations and may provide more robust estimates in certain cases.

4.1. Parameter estimation using quantiles

Another simple method for estimating the Weibull distribution's parameters is to use quantiles. This approach is based on the relationship between the median and the third quartile. The procedure is as follows:

1. for the Weibull distribution, the median M and the third quartile Q_3 are related via quantiles have the following formulas:

$$\begin{cases} M = \alpha(\log 2)^{1/k} \\ Q_3 = \alpha(2\log 2)^{1/k} \end{cases} \quad \Longrightarrow \quad Q_3 = 2^{1/k} M \tag{24}$$

2. Calculate the sample median M, and find the third quartile Q_3, the point where 75% of the data falls below.
3. Using the formula between the median and third quartile, estimate the shape parameter k as follows:

$$2^{1/k} = \frac{Q_3}{M} \quad \Longrightarrow \quad k = \frac{\log 2}{\log Q_3 - \log M} \tag{25}$$

4. Once k is estimated, compute the scale parameter α using the following equation:

$$\alpha = \frac{M}{(\log 2)^{1/k}} = \frac{M}{(\log 2)^{\frac{\log(Q_3)-\log(M)}{\log 2}}} \tag{26}$$

4.2. Parameter Estimation Using MAD (around Median)

An alternative to Maximum Likelihood Estimation is to use the Mean Absolute Deviation (MAD) around the sample median, which can be computationally simpler. The procedure to estimate the parameters α and k is as follows:

1. Compute the sample median $M = \text{median}\{x_1, \ldots, x_n\}$, which represents the central tendency of the data.
2. Compute the Mean Absolute Deviation H about the sample median M:

$$H = \frac{1}{n} \sum_{i=1}^{n} |x_i - M| \tag{27}$$

   where H measures the spread of the data around the median.

3. From the Weibull distribution properties, we use the following relationship:

$$\frac{H}{M} = 1 - \frac{1}{k(\log 2)^{1/k}} \left[\Gamma\left(\frac{1}{k}\right) + 2\Gamma\left(\frac{1}{k}, \log 2\right) \right] \tag{28}$$

   where Γ is the Gamma function, and $\Gamma(a,b)$ is the incomplete Gamma function. This equation provides an estimate of k, the shape parameter.

4. Once k is estimated, compute α, the scale parameter, from the following equation:

$$\alpha = \frac{M}{(\log 2)^{1/k}} \tag{29}$$

5. Estimation results

To assess the accuracy of the proposed parameter estimation, we will generate Weibull distributions and compare our method with maximum likelihood and quantile methods.

To generate Weibull distribution, we generate n random values $p_i \in (0,1)$ and compute $x_i = Q(p_i)$ from the quantile function. We repeat this procedure 100 times and report the estimation results (averages and standard deviation SD for parameters k and α) for the Weibull distribution in Table 1. As shown in the table, the results for k are more accurate, with a much lower SD, than those obtained with the quantile method.

Figure 2 presents the relative percentage error for estimating the shape k and scale α parameters.

As can be seen from Figure 2, for estimating the shape k, the Mean Absolute Deviation from Median method (MAD-M) is very close to Maximum Likelihood, and both methods give better results than the quantile method. On the other hand, for estimating the scale α, the MAD-M method is worse than the quantile and MLE for smaller sample sizes n (up to 200), but for large n, the proposed method is similar to MLE and superior to the quantile method. One advantage of the MAD-M method is its reduced sensitivity to outliers in the estimated parameters.

Table 1. Estimation Results for the Weibull Distribution ($k = 1$, $\alpha = 2$)

Method	$n = 100$				$n = 200$			
	k		α		k		α	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MLE	1.0132	0.0830	2.0036	0.2055	1.0123	0.0564	2.0021	0.1520
MAD$-M$	1.0162	0.1013	2.0147	0.2514	1.0132	0.0698	2.0049	0.1827
Quantile	1.0468	0.1962	2.0079	0.2318	1.0287	0.1347	2.0016	0.1697
	$n = 400$				$n = 800$			
MLE	1.0031	0.0402	2.0021	0.1091	1.0015	0.0277	1.9996	0.0751
MAD$-M$	1.0021	0.0483	2.0012	0.1257	1.0011	0.0345	2.0005	0.0901
Quantile	1.0113	0.0923	1.9986	0.1220	1.0066	0.0666	1.9986	0.0852

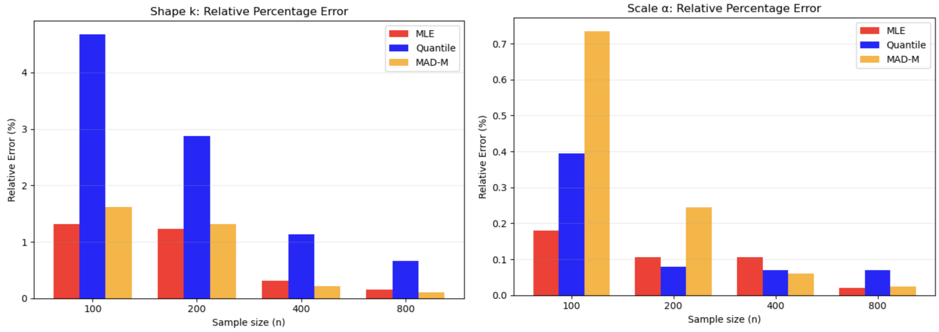

Figure 2. Relative Percentage Error Plots for the Weibull Distribution ($k = 1$ and $\alpha = 2$)

6. Case Study I: Analysis of Breast Cancer Survival Time Using MLE and MAD Methods

Data for this analysis were sourced from the cBioPortal platform, specifically the Breast Cancer (METABRIC) dataset, which contains clinical data for thousands of breast cancer patients. In this study, we analyze patient data from the METABRIC dataset, focusing on key clinical attributes related to breast cancer survival. To streamline the presentation, we employ several common abbreviations in our tabular data. Type of Surgery is abbreviated as mastectomy (Mast) and breast-conserving surgery (BCS). For Cancer Type Detailed, we use Invasive Ductal Carcinoma (IDC), Mixed Ductal and Lobular Carcinoma (MDLC), and Invasive Lobular Carcinoma (ILC). The sample is presented in Table 2.

These abbreviations are widely used in clinical oncology literature and facilitate a more concise representation of patient data, particularly when discussing multiple cases in survival analysis. The selected columns, including age at diagnosis, type of surgery, cancer subtype, chemotherapy status, and overall survival (measured in months), are crucial for understanding patient outcomes and the factors influencing survival rates.

Table 2. Sample of Breast Cancer Patients from the METABRIC Dataset
(Mast. = Mastectomy, BCS = Breast-Conserving Surgery, IDC = Invasive Ductal Carcinoma, MDLC = Mixed Ductal and Lobular Carcinoma, ILC = Invasive Lobular Carcinoma).

Patient ID	Age at Diagnosis	Type of Surgery	Cancer Type	Chemotherapy	Survival (Months)
MB-0000	75.65	Mast.	IDC	No	140.5
MB-0002	43.19	BCS	IDC	No	84.63
MB-0005	48.87	Mast.	IDC	Yes	163.70
MB-0006	47.68	Mast.	MDLC	Yes	164.93
MB-0008	76.97	Mast.	MDLC	Yes	41.37
MB-0010	78.77	Mast.	IDC	No	7.80
MB-0014	56.45	BCS	IDC	Yes	164.33
MB-0020	70.00	Mast.	ILC	Yes	22.40

In this study, we used three statistical methods—Maximum Likelihood Estimation (MLE), Quantiles, and Mean Absolute Deviation (MAD)—to estimate the parameters of the Weibull distribution for breast cancer survival data. The Weibull distribution is popular in survival analysis because it can model different survival time patterns and risk profiles (Muhammad, 2024). Our main aim was to see how well each method predicted the median survival time and scale parameter α (Hirst, 2021).

In this analysis, we use survival time data from breast cancer patients and randomly split the dataset into equal training and test sets (50%/50%). To ensure robust results, the training-testing process is repeated 1,000 times. In each iteration, considering the importance of median survival time in handling skewed data and outliers in medical and survival studies, we apply the Maximum Likelihood Estimation (MLE) and the Quantiles and Median Absolute Deviation (MAD) methods to estimate the scale parameter (α) of the Weibull distribution and predict the median survival rate of patients, using the test data for error analysis.

It is critical to analyze the scale parameter α and the mean error of the median prediction because these metrics directly affect the accuracy and reliability of survival predictions in the clinical setting. The scale parameter α determines the distribution of survival times; inaccurate estimates can lead to significant bias in predicting survival probabilities. Similarly, the predicted median survival time is a key metric in medical statistics, indicating how long half of the patients are expected to survive, and it serves as an important benchmark in survival studies. Minimizing the error in these two metrics is therefore critical to generating more accurate estimates and improving the overall reliability of survival models.

The results of the analysis are summarized in Table 3, showing the differences in performance between the Maximum Likelihood Estimation MLE, Quantiles, and MAD methods.

The median survival time predicted using the MLE method was 104.58 months with a mean prediction error of 12.15 months. The Quantiles method improved in prediction accuracy, predicting a median of 116.68 months with a mean error of 4.22 months. The MAD method, on the other hand, provided the closest prediction to the actual median, with a median prediction of 116.58 months and a significantly lower mean error of 3.79 months.

Regarding the Weibull distribution parameters, the MAD method also showed higher accuracy than the other methods. The shape parameter k was estimated to be 1.56 by the

Table 3. MLE, Quantile, and MAD results for Breast Cancer Survival Data.

Metric	MLE	Quantiles	MAD
Shape parameter k	1.56	1.51	1.51
Median Survival Time from prediction method (months)	104.58	116.68	116.58
Error in Predicted Median (months)	12.15	4.22	3.79
Error in Predicted Median (%)	10.75	3.70	2.46
Scale parameter α	132.18	148.83	148.73
Error in Scale parameter α	7.50	3.83	3.39
Error in Scale parameter α (%)	5.47	2.46	2.33

MLE method and 1.51 by both the Quantiles and MAD methods. For the scale parameter α, which determines the distribution of the survival time, the estimates were 132.18 by the MLE method, 148.83 by the Quantiles method, and 148.73 by the MAD method. For the mean error of α, the MAD method had the lowest error of 3.39, the Quantiles method had 3.83, and the MLE method had the highest error of 7.50. These results suggest that the MAD method is more robust than the MLE and Quantiles methods for predicting median survival time and scale parameters in cancer research and, therefore, can provide more accurate survival predictions in the clinical setting.

This finding has important implications for practical research. Accurate prediction of median survival time is critical in clinical survival analysis as it is a key metric for assessing treatment efficacy and patient prognosis (Hazim, 2025). The MAD method's ability to efficiently handle outliers ensures that predictions are not overly influenced by extreme survival times, making it a more robust tool for predicting patient outcomes. In contrast, despite its widespread use, the MLE method may not be as robust as the MAD method in biased datasets, leading to less accurate predictions in some cases.

7. Case Study II: Analysis of Insurance Claims Using MLE and MAD Methods

Data for this analysis come from an insurance claims dataset that captures demographic and clinical attributes, as well as claim amounts. The columns include patient age, gender, BMI (Body Mass Index), blood pressure, diabetic status, number of children, smoking status, region, and total claim amount. These factors are commonly used in actuarial modeling to understand claim severity and to identify risk drivers across demographics and geographic areas. For brevity, we employ several abbreviations in our tabular presentation: regions are abbreviated as Southeast (SE), Northwest (NW), and Southwest (SW); diabetic status and smoking status are recorded as Yes/No (Y/N). In this sample slice, *children* are all 0 and *smoker* is No for all rows; other values may appear in the full dataset. The sample is presented in Table 4. The selected columns are pertinent to claim severity modeling and facilitate comparisons across sex, metabolic factors (e.g., BMI, diabetic status), and region.

The results of the analysis are summarized in Table 5, showing the differences in performance between the Maximum Likelihood Estimation (MLE), Quantiles, and MAD methods.

Table 4. Sample of Insurance Claims (Selected Records)

Patient ID	Age	Gender	BMI	Blood Pressure	Diabetic	Children	Smoker	Region	Claim
1	39.0	Male	23.2	91	Yes	0	No	SE	1,121.87
2	24.0	Male	30.1	87	No	0	No	SE	1,131.51
8	19.0	Male	41.1	100	No	0	No	NW	1,146.80
19	49.0	Male	35.4	97	Yes	0	No	SW	1,263.25
25	50.0	Female	20.8	85	Yes	0	No	SE	1,607.51
29	58.0	Female	31.1	87	No	0	No	SE	1,621.88
31	29.0	Male	20.4	80	Yes	0	No	NW	1,625.43
40	49.0	Female	39.8	100	Yes	0	No	SE	1,633.96

Table 5. MLE, Quantile, and MAD results for Insurance Claims Data.

Metric	MLE	Quantiles	MAD
Shape parameter (k)	1.15	1.23	1.15
Median Claim from prediction method (USD)	9,641.87	9,341.35	9370.01
Absolute Error in Predicted Median (USD)	483.84	450.84	427.56
Relative Error in Predicted Median (%)	5.16	4.81	4.56
Scale parameter (α)	13,235.47	12,655.27	11,356.77
Absolute Error in Scale parameter (α) (USD)	812.25	873.21	519.36
Relative Error in Scale parameter (α) (%)	6.13	6.90	4.57

In this study, we applied Maximum Likelihood Estimation (MLE), Quantiles, and Mean Absolute Deviation (MAD) to fit a Weibull distribution to insurance claim amounts. The Weibull distribution is widely used in actuarial modeling because it accommodates skewed and heavy-tailed data often observed in claim severity (Hamza, 2023). We focused on two key metrics: the median claim amount, which provides a robust measure of central tendency, and the scale parameter α, which governs overall claim dispersion and impacts pricing and capital planning.

The dataset was split into equal training and test subsets, and the estimation process was repeated multiple times to ensure stable averages. As shown in Table 5, the MAD method yielded the smallest relative errors for both the predicted median (4.56%) and scale parameter α (4.57%), outperforming MLE and Quantiles. These results highlight MAD's robustness in handling variability and outliers, making it a practical alternative for improving the reliability of insurance risk models and premium estimation.

8. Conclusion

In this paper, we presented the formula for the Mean Absolute Deviation (MAD) about the median of the Weibull distribution and provided the calculations for the scale and shape parameters using three methods: Maximum Likelihood Estimation (MLE), the Quantiles method, and the Mean Absolute Deviation (MAD) around the median method.

The application of these methods to real-world data underscores the importance of selecting an appropriate statistical approach based on the dataset's characteristics. In survival analysis, where outcomes can vary widely, methods such as MAD that reduce the influence of outliers yield more reliable and clinically relevant predictions. Similarly, in actuarial science, modeling insurance claim amounts with MAD improves parameter estimation robustness, particularly for heavy-tailed or skewed distributions where extreme values are common. While MLE and Quantiles remain valuable techniques, they are more sensitive to outliers and may produce less stable estimates. Therefore, for future research involving survival data or financial risk modeling of insurance claims, particularly when the data follow a Weibull distribution, the MAD method based on the median is recommended as the most accurate and robust approach for parameter estimation and prediction.

Acknowledgements

Conflict of Interest: We declare that there are no conflicts of interest regarding the publication of this paper.

Author Contributions: All the authors contributed equally to the effort.

Funding: This research was conducted without any external funding. All aspects of the study, including design, data collection, analysis, and interpretation, were carried out using the resources available within the authors' institution.

Data Availability (including Appendices): All the relevant data, Python code for analysis, detailed annual tables, and graphs are available via: https://github.com/vickyzhang7/Mean-Absolute-Deviation-for-Weibull-Distribution

References

Almeida, J. B., (1999). Application of Weibull Statistics to the Failure of Coatings. *Journal of Materials Processing Technology*. https://doi.org/10.1016/S0924-0136(99)00177-6.

Balakrishnan, N., Kateri, M., (2008). On the Maximum Likelihood Estimation of Parameters of Weibull Distribution Based on Complete and Censored Data. textitStatistics & Probability Letters. https://doi.org/10.1016/j.spl.2008.05.019.

Bloomfield, P., Steiger, W. L., (1984). *Least Absolute Deviations*, https://doi.org/10.1007/978-1-4684-8574-5.

Cancho, V. G., Macera, M. A. C., Suzuki, A. K., Louzada, F., and Zavaleta, K. E. C., (2020). A New Long-term Survival Model with Dispersion Induced by Discrete Frailty. *Lifetime Data Analysis*, 26(2). https://doi.org/10.1007/s10985-019-09472-2.

Carroll, K .J., (2003). On the Use and Utility of the Weibull Model in the Analysis of Survival Data. *Controlled Clinical Trials*. https://doi.org/10.1016/S0197-2456(03)00072-2.

Chen, Q., Gerlach, R., (2011). *The Two-sided Weibull Distribution and Forecasting Financial Tail Risk*, University of Sydney Business School, Discipline of Business Analytics.

Cohen, A. C., (1965). Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Samples. *Technometrics*. https://doi.org/10.1080/00401706.1965.10490300.

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*. https://doi.org/10.1080/03610928208828412.

Feller, J., (1956). *Probability Theory and Applications*. https://doi.org/10.1016/S0895-7177(01)00109-1.

Fogliatto, M. S. S., Santos, T. M. O., Bessani, M., and Maciel, C., (2019). *Survival Analysis of Electrical Power Distribution Systems Using Weibull Regression*, Proceedings of the 2019 SBAI. https://doi.org/10.17648/sbai-2019-111513.

Gebizlioglu, O. L., Senoglu, B., and Kantar, Y. M., (2011). Comparison of Certain Value-at-Risk Estimation Methods for the Two-Parameter Weibull Loss Distribution. *Journal of Computational and Applied Mathematics*. https://doi.org/10.1016/j.cam.2011.01.044.

Hamza, A., (2023). A Bayesian Approach to Weibull Distribution with Application to Insurance Claims Data. textitJournal of Reliability and Statistical Studies. https://doi.org/10.13052/jrss0974-8024.1611.

Hazim, K., Abdul Sada, M. T., (2025). The Survival Power Weibull Distribution With Application, *Statistics, Optimization & Information Computing*. https://doi.org/10.19139/soic-2310-5070-2318.

Hirst, T. C., Sena, E. S., and Macleod, M. R., (2021). Using Median Survival in Meta-analysis of Experimental Time-to-Event Data. *Systematic Reviews*, 10. https://doi.org/10.1186/s13643-021-01824-0.

Hortobagyi, G. N., Stemmer, S. M., Burris, H. A., Yap, Y.-S., and Sonke, G. S., (2022). Overall Survival with Ribociclib plus Letrozole in Advanced Breast Cancer. *New England Journal of Medicine*, 386(10). https://doi.org/10.1056/nejmoa2114663.

Imran, M., Alsadat, N., (2024). The Development of an Extended Weibull Model with Applications to Medicine, Industry and Actuarial Sciences. *Scientific Reports*. https://doi.org/10.1038/s41598-024-61308-8.

Jacquelin, J., (1993). Generalization of the Method of Maximum Likelihood (Insulation Testing). *IEEE Transactions on Electrical Insulation*. https://doi.org/10.1109/14.192241.

Jokiel-Rokita, A., Piatek, S., (2024). Estimation of Parameters and Quantiles of the Weibull Distribution. *Statistical Papers*. https://doi.org/10.1007/s00362-022-01379-9.

Kang, D., Ko, K. and Huh, J., (2018). Comparative Study of Different Methods for Estimating Weibull Parameters: A Case Study on Jeju Island, South Korea. *Energies*. https://doi.org/10.3390/en11020356.

Olver, F. W., Lozier, D. W., Boisvert, R. F., and Clark, C. W., (2010). NIST Handbook of Mathematical Functions, *Cambridge University Press*, New York.

Pham-Gia, T., Hung, T. L., (2001). *The Mean and Median Absolute Deviations*. https://doi.org/10.1090/S0002-9947-1956-0090927-3.

Pinsky, E., Zhang, W. and Wang, Z., (2024). Pareto Distribution of the Forbes Billionaires. *Computational Economics*. https://doi.org/10.1007/s10614-024-10730-1.

Pobockova, I., Sedliackova, Z., (2014). Comparison of Four Methods for Estimating the Weibull Distribution Parameters. *Applied Mathematical Sciences*. https://doi.org/10.12988/ams.2014.45389.

Quiroz Flores, A., (2024). Machine Learning for Survival Analysis. *In Oxford Handbook of Engaged Methodological Pluralism in Political Science*, Vol. 1. https://doi.org/10.1093/oxfordhb/9780192868282.013.48.

Teimouri, M., Hoseini, S. M. and Nadarajah, S., (2011). Comparison of Estimation Methods for the Weibull Distribution. textitStatistics. https://doi.org/10.1080/02331888.2011.559657.

Welch, B. L., Johnson, N. L. and Kotz, S., (1972). Distributions in Statistics: Continuous Univariate Distributions. *Journal of the Royal Statistical Society. Series A (General)*, 135(3). https://doi.org/10.2307/2344623.

Yoosefi, M., Baghestani, A. R. and Khadembashi, N., (2018). Survival Analysis of Colorectal Cancer Patients Using Exponentiated Weibull Distribution. *International Journal of Cancer Management*, 11(3). https://doi.org/10.5812/ijcm.8686.