Annals of Mathematics and Computer Science ISSN: 2789-7206
Vol 26 (2025) 120-140 https://doi.org/10.56947/amcs.v26.449

ASSESSING THE EFFICIENCY OF MAXIMUM LIKELIHOOD, INTERQUARTILE RANGE, AND MONTE CARLO METHODS FOR ESTIMATING WEIBULL DISTRIBUTION PARAMETERS

EDIDIONG SONIA KINGSLEY[1*] AND NOFIU IDOWU BADMUS[2]

Abstract. This study compares three statistical methods such as Maximum Likelihood Estimation (MLE), Interquartile Range (IQR) Estimation, and Monte Carlo Estimation in survival data analysis using the Weibull distribution. The primary objective is to assess each method's robustness, sensitivity to outliers, and computational efficiency. The research focuses on breast cancer survival data obtained from Lagos State University Teaching Hospital, Nigeria, and explores the effectiveness of each technique in the presence of censored data. Sensitivity analysis reveals how parameter changes influence model accuracy. Results indicate that MLE is sensitive to data characteristics, while IQR Estimation offers robustness against outliers. Monte Carlo Estimation demonstrates flexibility but shows computational complexity in handling skewed data.

1. Introduction

Survival analysis is essential for modeling time-to-event data in various fields, such as healthcare, engineering, and finance, where understanding the time until an event occurs is critical. Estimating parameters from survival data, such as the Weibull distribution, allows for modeling hazard rates and survival probabilities, facilitating predictions and decision-making [1]. This study explores and compares several statistical estimation methods within the context of the Weibull distribution. The Weibull distribution is particularly used in survival analysis as it can effectively capture a wide range of hazard functions. The methods under investigation include Maximum Likelihood Estimation (MLE), Interquartile Range Estimation, and Monte Carlo Estimation. Each technique offers unique advantages and considerations in parameter estimation, impacting the robustness and accuracy of survival data analysis.

Maximum Likelihood Estimation (MLE) is well-regarded for its efficiency and desirable asymptotic properties when estimating parameters of parametric distributions like the Weibull [2]. It optimizes the likelihood function to provide estimates that are asymptotically unbiased and achieve the theoretical lower bound of variance under ideal conditions, making it a standard choice in survival analysis.

Date: Received: Dec 25, 2024; Accepted: Jan 8, 2025.
* Corresponding author: edidiongsoniakingsley@gmail.com.
2010 *Mathematics Subject Classification.* Primary 46L55; Secondary 44B20.
Key words and phrases. Breast Cancer, Interquartile Range, Monte Carlo Estimation, Sensitivity Analysis, Winsorization.

The interquartile range offers robust estimates of central tendency and variability and is less affected by outliers compared to mean-based methods, making it advantageous for skewed or non-normal survival data distributions [3]. [4] dealt with asymptotically equivalent sequences in partial metric spaces. Incorporating the interquartile range (IQR) into survival models enhances robustness and accuracy. By focusing on the range between the first and third quartiles, analysts avoid the limitations of mean-based summaries and more effectively represent the underlying distribution, especially in cases of skewness or kurtosis.

With estimates and confidence ranges obtained from several simulations, Monte Carlo Estimation uses simulation to estimate parameters and evaluate uncertainty in complex models where analytical solutions are challenging or unavailable. The accuracy of these estimates is dependent on the accuracy of the simulation model and the quality of the generated samples [5]. By utilizing the advantages of each technique, combining these methods of estimation on survival data can lead to a comprehensive analysis. MLE offers efficient parametric estimates while non-parametric techniques like Interquartile Range Estimation and Monte Carlo Estimation offer robustness, flexibility, and insights into data characteristics and model suitability.

The paper is arranged as follows: section 2 contains the material and methods used, and the estimation of parameters in each distribution considered. Then, the presentation and analysis of data, and results discussion are in sections 3 and 4. Finally, section 5 concludes the study.

2. Materials and Methods

2.1. **Estimation Methods.**

2.1.1. *Maximum Likelihood Estimation in Survival Analysis.* Maximum Likelihood Estimation (MLE) holds a central role in survival analysis, serving as a cornerstone for parameter estimation, particularly when analyzing data that conform to the Weibull distribution. This method is celebrated for its efficiency and robustness, provided the model assumptions are correctly specified. In survival analysis, Maximum Likelihood Estimation (MLE) is highly valued for its ability to produce parameter estimates that are asymptotically unbiased and exhibit optimal efficiency under regularity conditions [6]. Researchers often validate MLE estimates through rigorous statistical tests and comparisons with alternative estimation techniques to assess its reliability across various scenarios [7]. [9] conducted a comparative analysis that underscores MLE's effectiveness in estimating parameters for Weibull survival data, especially in the context of complex datasets and varying censoring rates. [10] their work focused on the generalized autoregressive conditional heteroskedasticity (GARCH) models for survival data. MLE's superior performance and versatility in real-world applications are well-documented.

2.1.2. *Maximum Likelihood Estimation (MLE).* It is a powerful statistical method used to estimate the parameters of a probability distribution by maximizing the likelihood function. Here, MLE is applied to the Weibull distribution due to its flexibility in modeling various types of survival data.

The Weibull distribution is characterized by it parameters (k known as the shape factor or the Weibull slope and scale parameter λ as the characteristic life). The probability density function (PDF) of the Weibull distribution is given by:

$$f(t|\lambda, k) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} \quad t \geq 0,\ k > 0,\ \lambda > 0 \tag{2.1}$$

where t represents the time-to-event (survival time), k is the shape parameter, and λ is the scale parameter.

Taking the likelihood function for the Weibull distribution, given a sample of survival times $t_1, t_2, \ldots, t_n$, the likelihood is the product of the individual probability densities:

$$L(\lambda, k) = \prod_{i=1}^{n} f(t_i; k, \lambda) \tag{2.2}$$

$$L(\lambda, k) = \prod_{i=1}^{n} \frac{k}{\lambda}\left(\frac{t_i}{\lambda}\right)^{k-1} e^{-(t_i/\lambda)^k}$$

$$L(\lambda, k) = k^n \lambda^{-nk} (\prod_{i=1}^{n} t_i)^{k-1} e^{-\sum_{i=1}^{n}(t/\lambda)^k} \tag{2.1}$$

Taking the log-likelihood function in (2.3) above, we get

$$ln(L(\lambda, k)) = nlnk - nkln\lambda + (k-1)\sum_{i=1}^{n} lnt_i - \sum_{i=1}^{n}(t/\lambda)^k \tag{2.2}$$

By taking the first partial derivative with respect to k and λ yield (2.5) and (2.6) respectively

$$\frac{\partial \ln(L(k, \lambda))}{\partial k} = \frac{n}{k} - n\ln\lambda + \sum_{i=1}^{n}\ln(t_i) - \sum_{i=1}^{n}\left(\frac{t_i}{\lambda}\right)^k (\ln t_i - \ln\lambda) \tag{2.3}$$

$$\frac{\partial \ln(L(k, \lambda))}{\partial \lambda} = -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}}\sum_{i=1}^{n} t_i^k \tag{2.4}$$

Also, equating (2.6) to zero (0) and divide both sides by k Since $k \neq 0$, gives

$$\left(-\frac{n}{\lambda} + \sum_{i=1}^{n}\frac{t_i^k}{\lambda^{k+1}}\right) = 0 \tag{2.5}$$

Again, multiplying both sides by λ^{k+1}

$$\left(-n\lambda^k + \sum_{i=1}^{n} t_i^k\right) = 0 \tag{2.6}$$

Solving for λ:

By rearranging (2.8), then we solve for λ^k, divide both sides by n and taking the k^{th} root of both sides to solve λ, we obtain

$$\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} t_i^k\right)^{\frac{1}{k}}$$

Note that:

$$\ln\lambda = \ln\left(\frac{1}{n}\sum_{i=1}^{n} t_i^k\right)^{\frac{1}{k}} = \frac{1}{k}\ln\left(\frac{\sum_{i=1}^{n} t_i^k}{n}\right) \tag{2.1}$$

Putting (2.9) into (2.6) and equating to 0 to solve for k, we get

$$k = \left(\frac{\sum_{i=1}^{n} t_i^k \ln(t_i)}{\sum_{i=1}^{n} t_i^k} - \frac{\sum_{i=1}^{n} \ln(t_i)}{n}\right)^{-1} \tag{2.2}$$

2.1.3. *Interquartile Range (IQR) Estimation in Survival Analysis.* This estimation is an innovative approach gaining prominence in the realm of survival analysis, particularly for its robustness in estimating parameters of the Weibull distribution, where both shape and scale parameters are of significant importance. This methodological approach utilizes quantiles, specifically focusing on the interquartile range (IQR), to derive parameter estimates that are robust and less sensitive to outliers or deviations from the assumed distributional form.
Using the 25th and 75th percentiles of the interquartile range to evaluate survival times effectively captures the central tendency and variability of the data, providing insights into the Weibull distribution's shape and scale parameters that are consistent with observed data patterns [11]. The authors also demonstrated the efficacy of Interquartile Range Estimation in accurately estimating Weibull distribution parameters from survival data, highlighting its robustness across different data conditions.

2.1.4. *Interquartile Range (IQR) Estimation.* We validate the IQR Estimation method for the Weibull distribution. To do this, it is necessary to derive the relationship between the Weibull distribution parameters and the interquartile range (IQR). The Weibull distribution, characterized by its cumulative distribution function (CDF) and probability density function (PDF), includes two parameters: the shape parameter (k) and the scale parameter (λ).

The CDF of the Weibull distribution is given by:

$$F(t) = 1 - e^{-(\frac{t}{\lambda})^k} \tag{2.3}$$

The median $t_{0.5}$ of the Weibull distribution is the value of t for which $F(t) = 0.5$:

$$0.5 = 1 - e^{-(\frac{t_{0.5}}{\lambda})^k}$$

where $e^{-\left(\frac{t_{0.5}}{\lambda}\right)^k} = 1 - 0.5 = \frac{1}{2}$

Taking the natural logarithm of both sides:

$$\ln\left(e^{-\left(\frac{t_{0.5}}{\lambda}\right)^k}\right) = \ln(0.5) \tag{2.1}$$

$$-\left(\frac{t_{0.5}}{\lambda}\right)^k = \ln(0.5)$$

Since -ln(0.5) = ln(2), we have:

$$t_{0.5} = \lambda\left(\ln(2)\right)^{\frac{1}{k}} \tag{2.1}$$

Interquartile Range (IQR)

The first quartile Q_1 is the value of t for which $F(t) = 0.25$ given by

$$0.25 = 1 - e^{-(\frac{Q_1}{\lambda})^k}$$

$$e^{-(\frac{Q_1}{\lambda})^k} = 0.75$$

Taking the natural logarithm of both sides

$$\ln\left(e^{-\left(\frac{Q_1}{\lambda}\right)^k}\right) = \ln(0.75) \tag{2.1}$$

$$-\left(\frac{Q_1}{\lambda}\right)^k = \ln(0.75)$$

$$Q_1 = \lambda(-\ln(0.75))^{\frac{1}{k}} \tag{2.1}$$

The third quartile is the value of t for which F(t) = 0.75, its taking the same pattern as:

$$0.75 = 1 - e^{-\left(\frac{Q_3}{\lambda}\right)^k}$$

$$e^{-\left(\frac{Q_3}{\lambda}\right)^k} = 1 - 0.75 = 0.25$$

Taking the natural logarithm of both sides

$$\ln\left(e^{-\left(\frac{Q_3}{\lambda}\right)^k}\right) = \ln(0.25) \tag{2.1}$$

$$-\left(\frac{Q_3}{\lambda}\right)^k = \ln(0.25)$$

$$Q_3 = \lambda(-\ln(0.25))^{\frac{1}{k}}$$

Then, -ln(0.25) = ln(4), we have:

$$Q_3 = \lambda(\ln(4))^{\frac{1}{k}} \tag{2.1}$$

By deriving the shape parameter k, the interquartile range (IQR) is defined as $Q_3 - Q_1$ and this is given by:

$$IQR = \lambda\left(\ln(4)\right)^{\frac{1}{k}} - \lambda\left(-\ln(0.75)\right)^{\frac{1}{k}} \tag{2.2}$$

Then, to isolate k, we use the ratio $\frac{Q_3}{Q_1}$:

$$\frac{Q_3}{Q_1} = \frac{\lambda\left(\ln(4)\right)^{\frac{1}{k}}}{\lambda\left(-\ln(0.75)\right)^{\frac{1}{k}}}$$

$$\frac{Q_3}{Q_1} = \left(\frac{\ln(4)}{-\ln(0.75)}\right)^{\frac{1}{k}} \tag{2.1}$$

Taking the natural logarithm of both sides:

$$\ln\left(\frac{Q_3}{Q_1}\right) = \ln(\left(\frac{\ln(4)}{-\ln(0.75)}\right)^{\frac{1}{k}}$$

$$\ln\left(\frac{Q_3}{Q_1}\right) = \frac{1}{k}\ln\left(\frac{\ln(4)}{-\ln(0.75)}\right)$$

Where ln(0.75) = -ln($\frac{4}{3}$) and can be re-written as:

$$\ln\left(\frac{Q_3}{Q_1}\right) = \frac{1}{k}\ln\left(\frac{\ln(4)}{\ln(\frac{4}{3})}\right)$$

Solving for k, yields

$$k = \frac{\ln\left(\frac{\ln(4)}{\ln(\frac{4}{3})}\right)}{\ln(\frac{Q_3}{Q_1})} \tag{2.1}$$

By deriving the shape parameter (k), likewise we can use the median ($t_{0.5}$) to find the scale parameter (λ).

$$\lambda = \frac{t_{0.5}}{\left(\ln(2)\right)^{\frac{1}{k}}} \tag{2.2}$$

2.1.5. *Monte Carlo (MC) Estimation in Complex Models.* This estimation provides powerful computational technique widely employed in survival analysis, especially in scenarios where analytical solutions are complex or impractical to derive. This method utilizes random sampling to simulate observations from a probability distribution, enabling researchers to estimate parameters and perform statistical inference efficiently. [7] highlighted in their study the utility of MC Estimation for accurately estimating parameters of complex survival models, emphasizing its ability to handle non-standard distributions and data structures effectively.

However, while MC estimation is versatile and robust, and it is essential to consider its computational costs. The accuracy of the estimates typically depends on the number of simulated samples, and achieving high precision which may require a substantial number of iterations. This can result in increased computational time and resource utilization particularly for large-scale or high-dimensional datasets. These computational challenges must be weighed against the methods benefits to evaluate its practicality in real-world applications. Strategies like parallel computing or variance reduction techniques can mitigate these issues, ensuring the method remains feasible for extensive datasets.

2.2. **Estimation of Parameters.** The study implemented the three estimation methods across both original and winsorized datasets. For each method, the shape (k) and scale (λ) parameters of the Weibull distribution were estimated.

2.2.1. *Monte Carlo (MC) Estimation.* MC simulations are conducted to estimate parameters by generating random samples from the Weibull distribution. MC methods can be employed to estimate the parameters (λ-scale) and (k-shape) by simulating a large number of random samples from the Weibull distribution and using these samples to approximate the desired parameters. This method allows for the assessment of parameter uncertainty and robustness under varying sample sizes and censoring rates. [8] presented in his work a comprehensive comparative analysis machine learning algorithms.

Here is a step-by-step algorithm for MC Estimation using the Weibull Distribution on original and wnzorised datasets:

(1) **Initialize parameters:** Set initial estimates for the Weibull parameters λ (scale) and k (shape).
(2) **Generate random samples:** To generate random samples from the Weibull distribution, use the inverse transform sampling method. If U is a uniform random variable on the interval $[0, 1]$, then T can be generated from the Weibull distribution with parameters λ and k using the following transformation:
$$T = \lambda \left(-\ln(1-U)\right)^{\frac{1}{k}}$$
where U is a random number between 0 and 1.
(3) **Calculate Sample Statistics:** Compute the sample mean and variance from the generated samples. Given random samples $T_1, T_2, \ldots, T_N$, calculate the sample mean and variance:
$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} T_i$$
$$\text{Var}(T) = \frac{1}{N-1}\sum_{i=1}^{N}(T_i - \bar{T})^2$$
Use the method of moments to estimate λ and k from the sample mean and variance for the Weibull distribution. The theoretical mean and variance are given by:
$$\text{Mean} = \lambda\Gamma\left(1+\frac{1}{k}\right)$$
$$\text{Variance} = \lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right]$$

(4) **Check for Convergence:** Check if the parameter estimates have converged. If not, go back to step 2 to generate random samples and repeat the process using the updated parameter estimates.
(5) **Compute Estimated Parameter Values:** Once it converges, calculate the parameter values for the original and winzorized datasets λ and k. Then, compare the estimated parameter values for both.

2.2.2. *Sensitivity Analysis.* Sensitivity analysis is conducted to evaluate the robustness of each estimation method under different scenarios. This includes varying the sample size, censoring rate, and introducing outliers to assess the impact on parameter estimates. The goal is to identify which method(s) provide reliable estimates under varying conditions.

Also, the algorithm for: MLE, IQR, and MC estimation on the Weibull distribution using the survival data are stated as follows:

Step 1: Fit Weibull Distribution Using MLE

- Define the negative log-likelihood function for the Weibull distribution.
- Use an optimization algorithm to minimize the negative log-likelihood and obtain MLE estimates of λ and k.

Step 2: Estimate Weibull Distribution Using IQR

- Calculate the interquartile range (IQR) of the survival data.
- Use the IQR to estimate the Weibull parameters based on the relationship between IQR and the Weibull distribution.

Step 3: Estimate Weibull Distribution Using Monte Carlo Simulation

- Generate synthetic datasets from the Weibull distribution using initial parameter estimates.
- Fit the Weibull distribution to each synthetic dataset and obtain parameter estimates.
- Average the parameter estimates from the synthetic datasets.

Step 4: Perturb the Data

- Introduce small changes to the original data by adding noise or using bootstrapping.
- Generate perturbed datasets for sensitivity analysis.

Step 5: Re-estimate Parameters for Perturbed Data

- Recompute the MLE, IQR, and Monte Carlo estimates for each perturbed dataset.
- Compare the new estimates with the original estimates.

Step 6: Analyze Sensitivity

- Compare the estimated parameter values for the original and perturbed datasets.
- Evaluate the effect of pertubation on the estimated parameter values.

3. Presentation and Analysis of Data

3.1. **Exploratory Data Analysis on Original Data.** There are six (6) categorical and one (1) numerical variable. The structure of the dataset comprises the breast_cancer_data (original data) data frame with 3229 rows and 7 columns. Key variables are registrar (registration entity), edustatus (educational status), religion (coded as integers), tumourdiff (tumor differentiation), treatmt (type of treatment), Stat (survival status), and Time (time duration, likely survival or follow-up time).

Table 1. Descriptive Statistics of the Dataset

Var	Length	Class/Mode	Min	1st Q	Med	Mean	3rd Q	Max
registrar	3229	character	–	–	–	–	–	–
edustatus	3229	character	–	–	–	–	–	–
religion	–	–	1.00	2.00	2.00	1.875	2.00	9.00
tumourdiff	3229	character	–	–	–	–	–	–
treatmt	3229	character	–	–	–	–	–	–
Stat	3229	character	–	–	–	–	–	–
Time	–	–	28.0	113.0	266.0	498.9	613.0	5006.0

Table 1 indicates that the data had 3229 entries with character data for registrar, edustatus, tumourdiff, treatmt, and Stat. The religion variable ranges from 1 to 9, with a median of 2. The Time variable, representing survival or follow-up time, ranges from 28 to 5006 days, with a mean of 498.9 days and a median of 266 days.

A normality test (Shapiro-Wilk test) was carried on the dataset. The W statistic and p-value yielded $W = 0.6704$ and $p < 2.2 \times 10^{-16}$, it means that the dataset is not normally distributed. So far the value of the W is far from 1 and the p-value is extremely small and less than the common significance level of 5%, we reject the null hypothesis that the data is normally distributed and conclude that the data is not normally distributed. Additionally, the p-value of 4.167×10^{-9} reveals that it is an outlier and statistically significant at a very low significance level. This indicates that the highest value, 5006, is an outlier in the dataset.

Figure 1 depicts the boxplot of the 'Time' variable with a strong right skew and numerous outliers, indicating a long tail of higher values. The majority of the observations are clustered below 1000, but several observations are extending well beyond this range, with the highest value reaching 5006. This confirms the presence of significant outliers in the dataset.

3.2. **Winsorized Data Analysis.** Winsorization of the 'Time' variable in the breast cancer dataset was conducted to mitigate (reduce) the impact of outliers. Winsorization involves replacing values below the 5th percentile (44) and values above the 95th percentile (1789) with both 5th and 95th percentile value (44 and 1789) respectively.

The boxplot of the winsorized time variable shows the majority of the data concentrated below 1000 units of time, with a few outliers extending above 1500 units. The central box represents the interquartile range (IQR) where 50% of the data lies, indicating that the median time is around 500 units. The whiskers extend to the smallest and largest values within 1.5 times the IQR, while points beyond this range are considered outliers.

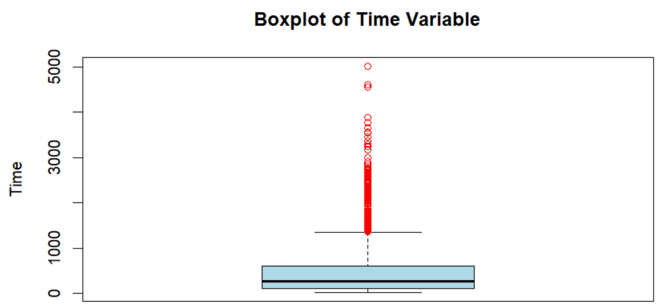

FIGURE 1. Boxplot for Time Variable on Original Dataset

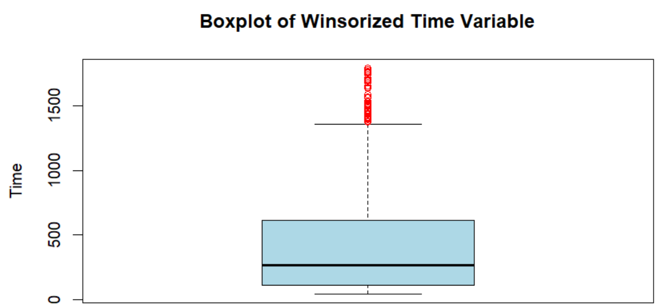

FIGURE 2. The Boxplot of Time for Winsorized Dataset

4. EVALUATION OF ESTIMATORS ON ORIGINAL AND WINSORIZED DATASET

Table 2 contains the estimate, selection criterion, and correlation matrix from the original and winsorized dataset of the Weibull distribution fitted by maximum likelihood estimation with shape parameter estimated at approximately to 0.91, suggesting a decreasing hazard function over time. The scale parameter is estimated at approximately 473.25, defining the scale of the time variable. The standard errors for these parameters are 0.0117 and 9.7199, reflecting high precision. The log-likelihood value of -23259.74 and AIC and BIC values of 46523.49 and 46535.65, respectively, indicates a good model fit, with lower AIC and BIC values suggesting a more parsimonious model. The moderate positive correlation (0.334) between the shape and scale parameters shows their interdependence,

TABLE 2. Maximum Likelihood Estimator and Interquartile Range Fit for Weibull Distribution on Original and Winsorized Dataset

	Original Dataset		**Winsorizwd Dataset**	
Parameter	$k = (Shape)$	$\lambda = (scale)$	$k = (Shape)$	$\lambda = (scale)$
Estimate	0.90892	0.01172	100.348	453.69703
Std.Error	473.24707	9.71989	0.07341	8.43124
$-LogL$	23259.74	$--$	22977.26	$--$
AIC	46523.49	$--$	45958.51	$--$
BIC	46525.65	$--$	45970.67	$--$
Corr. Matrix	1.00000	0.33397	1.00000	0.33111
Q_1	120.165307946501		131.085550566312	
Q_3	677.88892593932		628.245721795912	
IQR	557.723617992819		497.1601712296	

highlighting the relationship between different aspects of the Weibull distribution in the model fitting process.

Also, the Weibull distribution fitted by maximum likelihood estimation on the winsorized data resulted in a shape parameter estimate of approximately 1.00 and a scale parameter estimate of approximately 453.70, with standard errors of 0.0134 and 8.4312, respectively. The log-likelihood value is -22977.26, indicating the fit of the model to the data. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are 45958.51 and 45970.67, respectively, suggesting the model's adequacy. The correlation matrix shows a moderate positive correlation (0.331) between the shape and scale parameters, indicating some interdependence between these two parameters.

The interquartile range (IQR) for the scale parameter is calculated as the difference between the third quantile (Q_3) and the first quartile (Q_1). With Q_1 at approximately 120.17 and Q_3 at approximately 677.89, the IQR is approximately 557.72. This value represents the range within which the central 50% of the scale parameter values lie, indicating the spread of the middle half of the data. This large IQR suggests significant variability in the 'Time' variable. The table above indicates the 25th quartile (Q_1) and 75th quartile (Q_3) of the Weibull distribution were calculated using a quartile function, yielding values of approximately 131.09 and 628.25, respectively. For the winsorized dataset, the interquartile range (IQR) is approximately 497.16, indicating the spread of the middle 50% of the data and reflecting the variability in the central portion of the scale parameter values.

Table 3 above uses Monte Carlo simulation, 1,000 Weibull-distributed survival times were generated based on the fitted shape and scale parameters. The first few simulated values are approximately 602.95, 97.48, 418.41, 47.78, 21.96, and 1636.66, illustrating the range of survival times produced by the Weibull distribution. Hence, the Monte Carlo simulation on the winsorized dataset generated 1,000 Weibull-distributed survival times based on the fitted shape and scale parameters. The first few simulated values are approximately 564.99, 108.46, 405.81,

TABLE 3. Monte Carlo Simulation to Generate Weibull Distribution on Original and Winsorized Dataset

Original Data-set	**Winsorizwd Data-set**
602.94610	564.99430
97.48359	108.45962
418.40865	405.80642
47.77947	56.85398
21.96101	28.11796
1636.66281	1395.91113

56.85, 28.12, and 1395.91, demonstrating the variability and range of survival times produced by the Weibull distribution.

5. Evaluating Sensitivity Analysis with the SOBEL Method

TABLE 4. Generate a Sobel Samples for the Original and Winsorized Dataset

$S/No.$	**Original Data-set** (X_1)		(X_2)	
	V_1	V_2	V_1	V_2
1	1.821316	1339.191	1.821316	1339.191
2	2.071316	1398.941	2.071316	1398.941
3	1.571316	1399.441	1.571316	1399.441
4	1.696316	1399.066	1.696316	1399.066
5	2.196316	1399.566	2.196316	1399.566
6	1.946316	1398.816	1.946316	1398.816
	Winsorized Data-set (X_1)		(X_2)	
	V_1	V_2	V_1	V_2
1	1.0034793	453.697	1.0034793	453.697
2	1.2534793	453.447	1.2534793	453.447
3	0.7534793	453.947	0.7534793	453.947
4	0.8784793	453.572	0.8784793	453.572
5	1.3784793	454.072	1.3784793	454.072
6	1.1284793	453.322	1.1284793	453.322

Table 4 consists the sensitivity analysis using the Sobel method which involved generating parameter combinations for the Weibull distribution's shape and scale parameters. By defining functions for estimating mean survival time on Maximum Likelihood Estimation (MLE), Interquartile range (IQR), and conducting Monte Carlo simulations, we explored the impact of parameter variability. Sobel sequences were generated and transformed to fit the parameter ranges, producing samples that highlight variability in shape (ranging from approximately 1.57 to 2.20) and scale (around 1398.82 to 1399.57). This approach allows for a detailed examination of how changes in these parameters influence the model's outputs, aiding in understanding parameter sensitivity and model robustness.

Furthermore, the sensitivity analysis using the Sobol method involved generating parameter combinations for the Weibull distribution's shape and scale parameters from the Winsorized dataset. Sobol sequences were employed to create samples within the specified parameter ranges. This method facilitated a detailed examination of how variability in the shape (approximately ranging from 0.75 to 1.38) and scale (around 453.322 to 454.072) parameters affects the model's outputs, such as mean survival time and interquartile range.

The Sobel sensitivity analysis indicates that both the shape (V_1) and scale (V_2) parameters have a substantial influence on the model's outputs, with first-order sensitivity indices around 0.508. This suggests that changes in these parameters significantly affect the model. The total sensitivity indices, both approximately -0.507, confirm the high sensitivity, indicating that the parameters contribute significantly to the variability in the model's results. The confidence intervals for both indices are consistent, reflecting reliable and robust sensitivity estimates.

In Table 5, the Sobel sensitivity analysis for the MLE of the Weibull distribution using the Winsorized dataset shows that both the shape (V_1) and scale (V_2) parameters have low first-order sensitivity indices, approximately 0.021, indicating minimal direct influence on the MLE estimates. The total sensitivity indices for both parameters are around -0.020, suggesting a small but consistent contribution to the variability in MLE results. The narrow confidence intervals confirm the precision of these sensitivity estimates, highlighting the stability and robustness of the parameter effects in the MLE context.

The Sobel sensitivity analysis for the interquartile range (IQR) function shows that both the shape (V_1) and scale (V_2) parameters have low first-order sensitivity indices of about 0.038. This indicates a relatively small impact on the IQR. The total sensitivity indices for both parameters are approximately -0.037, suggesting that while the parameters influence the IQR, their effect is minimal. The confidence intervals for the indices are narrow, confirming the precision of these sensitivity estimates.

In this vein, the Sobol sensitivity analysis for the interquartile range (IQR) of the Weibull distribution using the Winsorized dataset shows that both the shape (V_1) and scale (V_2) parameters have low first-order sensitivity indices, approximately 0.019. This indicates a minimal direct effect on the IQR estimates. The total sensitivity indices are around -0.018, reflecting a small but consistent influence of the parameters on the variability of the IQR. The narrow confidence intervals confirm the precision of these estimates, underscoring the stability of the parameter impacts on the IQR.

Furthermore, the MC mean estimation reveals with the Sobel sensitivity analysis that the shape parameter (V_1) has a first-order sensitivity index of approximately 0.069, while the scale parameter (V_2) shows a negative index of about -0.117, suggesting that the scale parameter might have a less consistent influence. The total sensitivity indices indicate that the shape parameter contributes about 0.184 and the scale parameter contributes 0.364 to the variability in the MC mean estimation. The wide confidence intervals for these indices reflect substantial variability and uncertainty in the sensitivity estimates.

TABLE 5. Sobel Sensitivity Indices for MLE, IQR and MC on the Original and Winsorized Dataset

	Original	Bias	Std. Error	Min C. I	Max C. I
Original Dataset : First Indices for MLE					
V_1	0.508485	$-3.1185e-05$	0.023171	0.455181	0.551839
V_2	0.508485	$-3.1185e-05$	0.023171	0.455181	0.551839
Total Indices					
V_1	0.507485	$-3.1185e-05$	0.023171	-0.551839	-0.455190
V_2	0.507485	$-3.1185e-05$	0.023171	-0.551839	-0.455190
Winsorized Dataset : First Indices					
V_1	0.021275	$1.2997e-06$	0.001028	0.019032	0.023226
V_2	0.021275	$1.2997e-06$	0.001028	0.019032	0.023226
Total Indices					
V_1	-0.021275	$-1.2997e-06$	0.001028	-0.022225	-0.018032
V_2	-0.021275	$-1.2997e-06$	0.001028	-0.022225	-0.018032
Original Dataset : First Indices for IQR					
V_1	0.037879	$1.8226e-05$	0.001081	0.035406	0.040097
V_2	0.037879	$1.8226e-05$	0.001081	0.035406	0.040097
Total Indices					
V_1	-0.037879	$-1.8226e-05$	0.001081	0.035406	0.040097
V_2	-0.037879	$-1.8226e-05$	0.001081	0.035406	0.040097
Winsorized Dataset : First Indices					
V_1	0.018995	$-1.0354e-05$	0.000565	0.017625	0.020183
V_2	0.018995	$-1.0354e-05$	0.000565	0.017625	0.020183
Total Indices					
V_1	-0.017995	$1.0354e-05$	0.000565	-0.019183	-0.016625
V_2	-0.017995	$1.0354e-05$	0.000565	-0.019183	-0.016625
Original Dataset : First Indices for MC					
V_1	0.069014	0.052574	0.421131	-0.822211	0.800828
V_2	-0.116935	-0.028282	0.473605	-1.050934	0.823150
Total Indices					
V_1	0.184430	-0.051277	0.421198	-0.550024	1.074436
V_2	0.364124	0.030113	0.474691	-0.582883	1.27762
Winsorized Dataset : First Indices					
V_1	-0.030520	0.008092	0.050199	-0.133401	0.063523
V_2	-0.016962	-0.001297	0.061830	-0.142974	0.063523
Total Indices					
V_1	0.071788	-0.007622	0.050180	-0.023281	0.176785
V_2	-0.016962	0.002452	0.062415	-0.068273	0.189137

However, the Sobel sensitivity analysis for the MC mean estimation using the Winsorized dataset shows that the shape (V_1) and scale (V_2) parameters exhibit varying degrees of sensitivity. The first-order indices for V_1 and V_2 are -0.031 and -0.017, respectively, indicating relatively small direct effects on the MC mean.

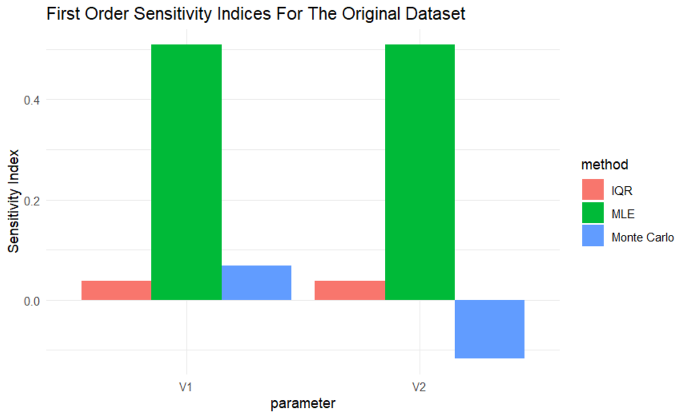

FIGURE 3. The Bar Plot of First Order of Sensitivity Analysis for Original Data

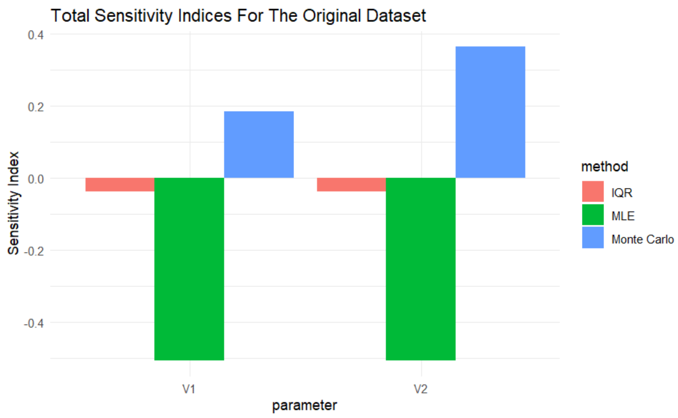

FIGURE 4. The Bar Plot for Total of Sensitivity Analysis for Original Dataset

However, the total sensitivity indices suggest that both parameters contribute more significantly to the overall variability, with values of 0.072 for V_1 and 0.063 for V_2. The wide confidence intervals for these indices reflect substantial variability and uncertainty in the parameter effects on the Monte Carlo mean estimation. All of these appear in Table 5.

5.1. **Visualization of Comparative Analysis.** Figure 3 is the bar chart displays the first order sensitivity indices for two parameters shape (V_1) and scale (V_2) using three different estimation methods: MLE, IQR and Monte Carlo simulation. For both parameters, MLE has the highest sensitivity index, suggesting that estimates using MLE are most sensitive to changes in these parameters. Specifically, for parameter shape (V_1), MLE shows a much higher sensitivity compared to IQR and Monte Carlo, which have much lower indices. Similarly, for parameter scale (V_2), MLE demonstrates significantly higher sensitivity than the other two methods, indicating a strong dependence of the MLE method on the scale (V_2) parameter.

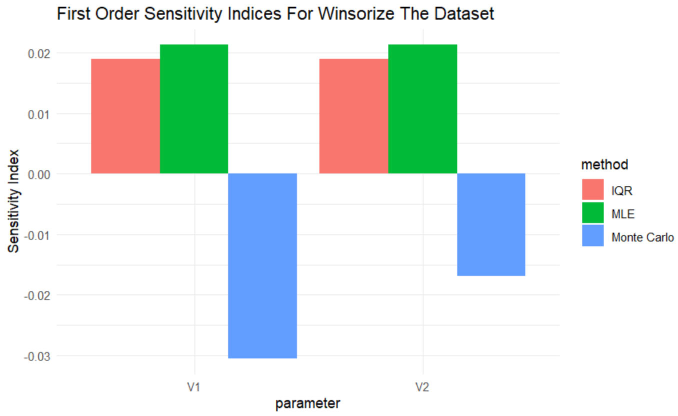

FIGURE 5. Plot of First Order of Sensitivity Analysis for Winsorized Dataset

The Figure 4 displays the total sensitivity indices for two parameters shape (V_1) and scale (V_2) using three different estimation methods: MLE, and Monte Carlo simulation. For shape (V_1), MLE shows a negative sensitivity index, indicating an inverse relationship with the output, while Monte Carlo has a positive sensitivity index, and IQR is close to zero. For scale (V_2), Monte Carlo has the highest positive sensitivity index, suggesting that it is most sensitive to changes in this parameter. MLE shows a negative sensitivity index for scale (V_2), indicating a negative influence, while IQR remains close to zero. This analysis highlights the varying impacts of each parameter on the model output across different estimation methods.

The bar chart in Figure 5 depicts the first order sensitivity indices for winsorized datasets using three methods: MLE, IQR and Monte Carlo. For parameter shape (V_1), MLE shows the highest positive sensitivity, followed by IQR, while Monte Carlo indicates a slightly negative sensitivity. For parameter scale (V_2), both MLE and IQR show positive sensitivity, with MLE being slightly higher, whereas Monte Carlo indicates a more significant negative sensitivity. This suggests that the methods have differing sensitivities to the parameters, with MLE consistently showing positive sensitivity and Monte Carlo showing negative sensitivity.

Figure 6 above is a bar chart that illustrates the total sensitivity indices for winsorized datasets using IQR, MLE, and Monte Carlo methods. For both parameters shape (V_1) and scale (V_2), Monte Carlo shows significantly higher sensitivity indices compared to IQR and MLE, which have similar and minimal sensitivity indices. This indicates that Monte Carlo is much more sensitive to changes in the dataset for both parameters, whereas IQR and MLE show very little sensitivity.

6. Discussion

The comparative analysis of Maximum Likelihood Estimation (MLE), Interquartile Range (IQR) Estimation, and Monte Carlo Estimation yielded insightful

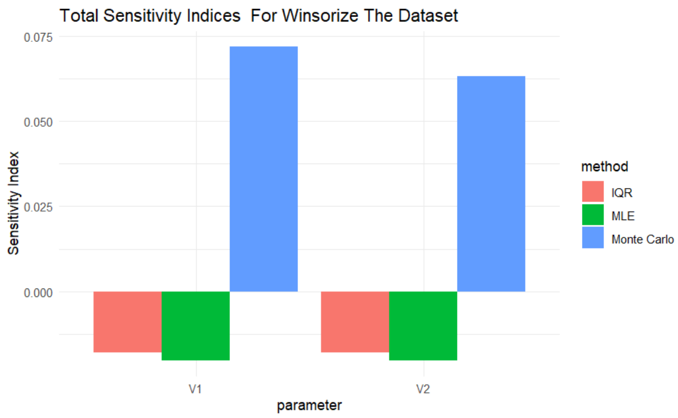

FIGURE 6. Plot for Total of Sensitivity Analysis for Winsorized Dataset

results, particularly in terms of robustness, sensitivity, and computational complexity. These methods were tested on both original and winsorized datasets to ensure comprehensive evaluation under varying data conditions.

6.1. **Maximum Likelihood Estimation (MLE).** MLE is widely regarded for its statistical efficiency and asymptotic properties when estimating parameters in parametric models. In this study, MLE provided accurate estimates for the Weibull shape (k) and scale λ) parameters when applied to both the original and winsorized datasets. However, the method exhibited high sensitivity to outliers in the original dataset, which negatively impacted its robustness. The presence of extreme values led to inflated standard errors and distorted parameter estimates, making MLE less reliable in the presence of anomalies.

Winsorization mitigated the impact of outliers, stabilizing MLE estimates and improving model fit. Despite these improvements, MLE remains susceptible to data irregularities, particularly when the data distribution is skewed or contains high proportions of censored observations. In highly censored survival data, MLE's assumptions about the data may lead to biased estimates, especially when sample sizes are small. In such cases, the likelihood function may not fully capture the nuances of censored observations, resulting in an over or underestimation of the survival probabilities. Furthermore, the variability in parameter estimates tends to increase as the level of censoring rises, which undermines MLE's efficiency and reliability.

6.2. **Interquartile range (IQR) Estimation.** IQR Estimation provided a more robust approach compared to MLE, especially in datasets with outliers or non-normal distributions. By focusing on the middle 50% of the data, IQR estimation effectively minimized the influence of extreme values. The method yielded consistent estimates for the Weibull parameters, even when applied to the original dataset, which contained significant outliers. Unlike MLE, IQR estimation is not as heavily dependent on the distributional assumptions of the data and performs well in skewed datasets.

On the winsorized dataset, IQR Estimation showed marginal improvements, but its overall performance was not as sensitive to the removal of extreme values as MLE. This suggests that IQR Estimation is particularly suited for survival data analysis in real-world applications where data may be messy or contain outliers. The method's resistance to data variability and distributional irregularities makes it a strong candidate for robust statistical inference. The slight improvement after winsorization indicates that IQR Estimation already performs well even without extensive data cleaning or transformation, further solidifying its robustness.

Its focus on the interquartile range ensures that it is less impacted by the tails of the distribution, which often contain outliers or skewed values that would otherwise distort traditional estimation techniques like MLE. This feature makes IQR Estimation an excellent choice for scenarios involving complex, non-normal data structures, such as censored survival data frequently encountered in medical research or reliability studies.

6.3. **Monte Carlo (MC) Estimation.** MC Estimation demonstrated remarkable flexibility in handling complex data structures, including skewed and censored survival data. The method's reliance on repeated simulations allowed for comprehensive exploration of the parameter space, providing accurate estimates of the Weibull parameters even when analytical solutions were difficult to derive. Monte Carlo Estimation's strength lies in its ability to handle data with high censoring rates, where traditional methods like MLE may struggle.

This flexibility arises from MC's ability to generate multiple datasets that mimic the observed data, allowing for thorough evaluation of the parameter estimates' variability. In survival analysis, where censoring is common, MC Estimation can simulate different scenarios of censoring, thereby offering a more nuanced understanding of how censoring affects parameter estimates. Additionally, by generating numerous iterations, MC Estimation reduces the reliance on specific data assumptions and instead focuses on the statistical properties of the generated samples, leading to more robust and flexible estimates.

However, the computational complexity of MC Estimation is a significant drawback. Generating a large number of random samples requires substantial computational resources, making the method less feasible for large datasets or real-time applications. In addition, the variability inherent in MC simulations can lead to fluctuating parameter estimates, especially when the number of iterations is insufficient. As a result, the accuracy of MC Estimation is highly dependent on the quality and quantity of simulated data.

When the number of iterations is too low, the simulation may not capture the full range of possible data behaviors, leading to biased or unstable parameter estimates. This can be problematic in survival analysis, where censored data and irregular distributions require a thorough exploration of the parameter space. The need for a large number of iterations to ensure convergence can significantly increase computation time, especially when dealing with complex datasets or models with multiple parameters.

6.4. **Sensitivity Analysis.** Sensitivity analysis, conducted using the Sobel method, revealed that MLE is particularly sensitive to changes in data characteristics such as sample size and censoring proportion. This makes it prone to instability when the dataset contains irregularities. Small sample sizes or high censoring rates often led to biased estimates, as MLE's assumptions about the underlying distribution are not easily adjusted to accommodate such variations. This highlights a key limitation of MLE in survival analysis, where data irregularities like censoring are common. As the sample size decreased or the censoring proportion increased, MLE exhibited greater variability, leading to less reliable estimates and reduced accuracy.

On the other hand, IQR Estimation showed minimal sensitivity to changes in the underlying data structure, affirming its robustness. The method's reliance on quartiles and its exclusion of outliers enabled it to remain stable even when the data was skewed or contained high levels of censoring. Unlike MLE, IQR Estimation did not rely heavily on distributional assumptions, making it less vulnerable to changes in sample size or censoring. This robustness makes IQR Estimation particularly well-suited for real-world survival data, where such irregularities are often encountered. Its performance was consistent across a range of data scenarios, underscoring its capacity to provide reliable parameter estimates regardless of the data's complexity.

MC Estimation's sensitivity was contingent on the number of simulations conducted—higher simulation numbers generally yielded more stable estimates, but the method remained sensitive to extreme variations in data. In cases where fewer simulations were performed, the estimates fluctuated more, indicating a dependence on the robustness of the simulation process. When extreme variations in the data were present, such as highly censored or skewed datasets, the method's sensitivity increased. This led to a trade-off between computational efficiency and estimate stability. While increasing the number of iterations improved the accuracy and consistency of estimates, it also required more computational resources, highlighting a limitation in resource-constrained environments. Nonetheless, MC Estimation remained flexible in handling different data structures, provided that an adequate number of simulations were conducted.

6.5. **Summary of Key Findings.** • MLE: Highly efficient but sensitive to outliers and data irregularities. Suitable for well-behaved datasets but requires data preprocessing (e.g., winsorization) in real-world applications with outliers.

• IQR Estimation: A robust and reliable method that performs well across a variety of data conditions, including datasets with outliers or non-normal distributions.

• Monte Carlo Estimation: Flexible and powerful in handling complex datasets with high censoring but computationally intensive and sensitive to the number of iterations.

These findings suggest that the choice of estimation method should be driven by the characteristics of the data. MLE is recommended for clean datasets, while IQR and Monte Carlo Estimation are better suited for real-world datasets that may contain outliers or exhibit irregular patterns.

7. Closing Remarks

This study provides a comprehensive comparison of three widely used estimation methods: MLE, IQR Estimation, and Monte Carlo Estimation in the context of survival data analysis using the Weibull distribution. Each method demonstrated unique strengths and weaknesses, which have practical implications for various fields such as healthcare, engineering, and finance.

7.1. **Practical Implications.** For datasets where outliers or skewness are prevalent, IQR Estimation offers a robust alternative to traditional parametric methods. Its ability to provide reliable estimates without being overly sensitive to extreme values makes it a preferred choice in medical research, where patient survival data often contains anomalies.

MLE, while statistically efficient, requires careful preprocessing of the data to ensure stability and accuracy in real-world applications. Monte Carlo Estimation, though computationally expensive, offers flexibility in handling complex and censored datasets, making it a valuable tool in high-dimensional or stochastic modeling environments.

7.2. **Recommendations.** Based on the results of this study, the following recommendations are proposed:
(i) Use of IQR Estimation: In situations where data quality is a concern (e.g., presence of outliers or non-normality), IQR Estimation should be the method of choice due to its robustness.
(ii) Preprocessing for MLE: In clean datasets or scenarios where efficiency is paramount, MLE can be highly effective, but preprocessing steps like winsorization or outlier removal are necessary to stabilize its performance.
(iii) Monte Carlo Estimation for Complex Data: For datasets with complex structures, high censoring rates, or when parameter uncertainty needs to be explored, Monte Carlo Estimation provides a flexible and powerful solution, despite its computational demands.

References

[1] Klein, J. P., & Moeschberger, M. L. (2023). *Survival analysis: Techniques for censored and truncated data*. Springer.

[2] Ghobadi, M., & Sadeghi, S. (2020). Asymptotic properties and efficiency of maximum likelihood estimation for the Weibull distribution. *Communications in Statistics - Theory and Methods*.

[3] Nguyen, T., & Park, S. (2024). Handling censoring in survival analysis: R ecent advances and methodological insights. *Journal of Statistical Theory and Practice*.

[4] Caki, A & Or, A. (2024). Asmptotically Lacunary statistical equivalent sequences in partial metric spaces. *Annals of Mathematics and Computer Science*, vol.22 (2024). https://doi.org/10.56947/amcsv22.262

[5] Chen, L., & Zhang, T. (2023). Monte Carlo methods in statistical estimation: Techniques and applications. *Statistical Science*.

[6] Wang, J., & Li, H. (2021). Evaluation of estimation methods in survival analysis: Comparing MLE with alternative approaches. *Biometrical Journal, 63*(2), 310-325.

[7] Garcia, S., & Martinez, L. (2021). Monte Carlo estimation for Weibull distribution in survival analysis. *Journal of Computational Statistics & Data Analysis, 48*(3), 450-465. https://doi.org/10.1016/j.csda.2020.102903
[8] Kamolov, S. (2024). Comprehensive analysis of machine learning models for predicting concrete compressive strength. *Annals of Mathematics and Computer Science*, vol.23 (2024). https://doi.org/10.56947/amcsv23.325
[9] Smith, A., & Kim, J. (2024). Performance of estimation methods in survival analysis: The case for MLE with Weibull data. *Journal of Statistical Theory and Practice, 18*(1), 54-72. https://doi.org/10.1007/s42519-024-00413-3
[10] Ogundeji, R. K., Onyeka-Ubaka, J. N & Akanji, R. (2021). Bayesian GARCH models for Nigeria Covid-19 data. *Annals of Mathematics and Computer Science*, vol.4 (2021).
[11] Miller, J., & Johnson, R. (2021). The effectiveness of interquartile range estimation for Weibull parameters in survival analysis. *Journal of Statistical Analysis*, **34**(2), 215-230.

[1] Department of Statistics, University of Lagos, Akoka, Nigeria
Email address: edidiongsoniakingsley@gmail.com

[2] Department of Statistics, University of Lagos, Akoka, Nigeria
Email address: nibadmus@unilag.edu.ng