STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, Month 2025, pp 0–22.
Published online in International Academic Press (www.IAPress.org)

Choosing Best Robust Estimation Parameters among Some Exponential Family Distributions

Hozan Taha Abdalla*, Samira Muhamad Salh

Department of Statistics and Informatics, College of Administration and Economics, Sulaimani University, Sulaimani, Kurdistan Region – Iraq

Abstract Survival analysis is statistical route deliberated to explore the time until an event appears. Most research on the reliability of the survival function has some shortcomings in the process of accurate statistical analysis that aims to obtain highly efficient estimates. The necessity for these efficient estimation methods is called robust methods which becomes important when the data for the phenomenon being studied is contaminated, i.e., there are outliers in the observations, which results in estimates that lead to an increase or decrease in the mean square error (MSE), which leads to inaccurate statistical inference. A breast cancer is commences from the abnormal growth of the cells in the mammary gland, ductal carcinoma or lobular carcinoma of the breast the data that has been used has 4024 observations and the independent variables are (Age, Tumor Size, Tumor Stage, and Node Stage). The study focusses on comparing different methods such as (AFT), Regression model, Robust AFT with Tukey's Biweight function, Robust AFT with Median function and Robust Regression model with Tukey's Biweight function for each of (Exponential, Weibull and Lognormal) distributions and non-parametric bootstrap is used to derive the standard error, z-values and p-values for each of the classical and robust methods, ensuring robust inference free from asymptotic assumptions. The Aim of this study is to choose the best method from each of (AFT), Regression model, Robust AFT with Tukey's Biweight function, Robust AFT with Median function and Robust Regression model with Tukey's Biweight function for each of (Exponential, Weibull and Lognormal) distributions and to choose those parameters that affect the survival time. A comparison was conducted among AFT, Regression model, Robust AFT with Tukey's Biweight function, Robust AFT with Median function and Robust Regression model with Tukey's Biweight function. We conclude that that the Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function is the best method, the Robust AFT Weibull Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function is also the best model. As well as the Robust AFT for (Exponential, Weibull and Lognormal) by using Maximum likelihood-Type Estimator with Median weight function shows the best result.

Keywords Index Terms-Survival, Censoring, AFT, Estimation, Outliers, Robust

AMS 2010 subject classifications 62M10, 93A30

DOI: 10.19139/soic-2310-5070-3169

1. Introduction

The increasing interest of many studies in reliability is the result of the important role that this science plays in dealing with time, or more precisely with the ages of both equipment and living organisms, such as the probability of survival and average life. Both reliability theory and survival theory share a measure of life span, the first for equipment and machines and the second for living organisms. [2, 27, 31]
Because the theory of reliability or survival is considered one of the important branches of statistics, which has a fundamental role in studying and analyzing any phenomenon based on the statistical data available about this

*Correspondence to: Hozan Taha Abdalla (Email: hozan.abdulla@univsul.edu.iq). Department of Statistics and Informatics, College of Administration and Economics, Sulaimani University, Sulaimani, Kurdistan Region – Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

phenomenon on one hand, and on the other hand the purity of the data or its lack of anomalies on the other hand (Outliers). [7, 31] Outliers are defined as random points in nature, i.e., a group of observations that appear to deviate significantly from the rest of the observations in the sample. In statistical terms, they are observations that come from a population that differs from the population under study. This difference may arise as a result of several factors, including inherited variability, measurement error, or implementation error.

The first factor, in the case of a population, for example, refers to uncontrolled variations in numbers, which are considered a natural phenomenon. The second factor refers to the incorrect use of natural measurements, in addition to errors in measuring methods, reading, or recording, etc. The third factor refers to a deficiency in data collection, i.e., the failure to obtain a representative sample.[31]

Therefore, when performing the estimation process to obtain good estimators that have the required characteristics in an estimator the appropriate estimation method must be chosen so that can be relied upon to achieve more accurate results. The most common methods used to estimate the parameters of statistical models are: Maximum Likelihood (ml), Ordinary Least Squared (ols) and Method of Moments (MOM). And etc. However, the capabilities of these methods may be inefficient when the data contains outliers. Therefore, in recent decades, statisticians have focused on dealing with outliers in data, or more precisely, on how to handle data that contains outliers. [25] This is done through what are called robust estimation methods. [24]

Where robustness is obtained more efficiently than conventional methods in the presence of anomalies, as it is assumed that it will be very close to the conventional methods in the absence of anomalies. [31]

Over recent years, the evolution of survival analysis has seen increased interest. The effectiveness of survival analysis in real dataset gained significant attention. Abdullahi et al. (2025) utilized parametric Accelerated Failure Time (AFT) models, specifically the Exponential and Weibull distributions, to examine the survival time of tuberculosis patients in Nigeria. Using data from hospitals, they looked at demographic and clinical factors that could predict things like age, HIV status and the type of treatments. AIC and log-likelihood values were used to compare models, and the Weibull AFT model was found to be the best fit because it captured the non-constant hazard that is common in TB progression. The research underscores the efficacy of parametric AFT methodologies in estimating survival time ratios and comprehending disease dynamics beyond proportional hazard assumptions. However, it is still limited by the fact that it only looks at one center and doesn't have any outside validations. This means that future research needs to use a bigger dataset and more flexible parametric modeling.[3] Sabila et al. (2025) examine a Kaggle dataset on breast cancer (n = 317), and they provide descriptive statistics, a brief discussion of risk factors (age, stage, tumor size, cancer type and operation type), and Kaplan-Meier estimates of survival over time. The authors indicate that they intend to use Cox Regression, but they do not provide any comprehensive multivariable results. They also used Exponential distribution but the assumptions did not meet with the dataset. The paper reports a survival table and curve as well as frames the Kaplan-Meier method as the main analytical tool.[28] In order to enhance long-term extrapolation in situations where follow-up data is scarce, Cooney and White (2023) presented a technique for directly integrating expert opinion into parametric survival models. Using a penalized-likelihood or Bayesian framework, their method combines trial data with expert opinions regarding survival probabilities at particular time points. The approach works with flexible Royston-Parmar spline models and standard distributions like Weibull, exponential, Gompertz, log-normal, log-logistic and generalized gamma. While maintaining a good model fit, incorporating expert opinion decreased the uncertainty of survival projections in an applied leukemia example. The study emphasizes how expert-informed priors can stabilize extrapolation in evaluations of health technology.[9] Tasfa Marine and Mengistie (2023) utilized parametric Accelerated Failure Time (AFT) models to examine survival rates among 552 Ethiopian women with breast cancer receiving treatment at Jimma University Medical Center. They used AIC and BIC to compare the exponential, weibull, log-normal and log-logistic distributions to find the best fitting model. They chose the Weibull AFT model because it fit the data better. The analysis indicated that age, residence, comorbidity, tumor size, nodal status, stage and treatment significantly impacted the survival time. The study illustrated the utility of parametric AFT methodologies for estimating survival and pinpointing prognostic factors in resource-constrained environments, highlighting that the Weibull model offers a versatile yet comprehensible framework for breast cancer outcomes.[33]

Parametric Accelerated Failure Time (AFT) models were used by Birhanu et al. (2022) to investigate breast cancer mortality predictors among patients receiving treatment at a tertiary hospital in Ethiopia. In order to find the best fit, they used log-likelihood, AIC and BIC to compare a number of parametric distributions, including exponential, Weibull, log-normal and log-logistic. In the end, they chose the Weibull AFT model. Age, tumor stage, lymph node status, metastasis and treatment type were all significant predictors and were all highly correlated with a shorter survival period. The study shows that the Weibull distribution offers a flexible yet interpretable framework for mortality prediction in breast cancer patients in resource-constrained settings, and it emphasizes the usefulness of parametric AFT modeling in quantifying the direct effects of clinical covariates on survival time.[32]
In this paper parametric survival models have been used such as (Exponential, Weibull and Lognormal) distributions and for each distribution Accelerated Failure Time (AFT), Regression model, Robust Accelerated Failure Time with Tukey's Biweight function, Robust Accelerated Failure Time with Median function and Robust Regression model with Tukey's Biweight function have been used. The information criteria's such as AIC and BIC is used to compare between the classical with robust methods, as well as the goodness of fit test applied to see which distribution fits the data best.

2. The Theoretical Part

2.1. *Survival Function, S(t)*

The survival function is well defined in medicine and biology, but it is different from the reliability function. It is known as the survival of a certain quantity at that moment t at least ($t > 0$). [23, 32]
Its expression is as follows:

$$S(t) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - F(t) \tag{1}$$

2.2. *Failure Rate $\lambda(t)$:*

It is also known as Hazard Rate $\lambda(t)$. It is the probability of failing an individual under study during the period ($t < T < t + \Delta T$), given that it was working until time t. [13, 32] The mathematical expression is as follow:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr[t < T < t + \Delta T \mid T > t]}{\Delta t} \tag{2}$$

2.3. *Some properties of the survival function:*

One of the most important properties of the above concepts is that there is a relationship between the death function $f(t)$ and the survival function $S(t)$ with the risk or failure function, and this is expressed mathematically as follows:

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{3}$$

From this important relationship, we can say that obtaining any of these three functions leads to obtaining the other two functions.[31]

2.4. *Censoring*

We won't be able to explain an observation that is censored until we carefully identify an uncensored observation. [18] This case may seem to be obvious. But in applied it will confuse us about a censoring, censoring might not be because of the fact of incomplete observations, but maybe due to the result of an unclear survival time definition. [28]
The types of censoring are

- Type I Censoring

At a precise time point or during research if the patients are tested multiple times after a specific amount of time has passed from the starting point of the experiments, the study comes to an end. [12]

- Type II Censoring (Left Censoring, Interval Censoring and Right Censoring). [14, 15, 16]

Failure time data differs from other data types in that it is censored. This is predicated on the idea that we have researched the death rates of people who suffer from a particular illness. Following the trial, some patients typically remain alive. Therefore, it is known that their failure times are longer than the time it takes to enroll patients and finish a study. Because of this censoring, statistical methods other than basic linear regression are needed for survival analysis. [11]

2.5. *Accelerated Failure Time (AFT)*

To analyze the survival data another popular regression model is used which is AFT, the AFT model connects the lifetime distribution to the regressors variable (stress, covariate).
With reference to T_i as a random variable representing the survival time of the i^{th} unit, since T_i must be a non-negative value, also it should be considered modeling its logarithm using a customary linear model: [30]

$$\log T_i = x_i'\alpha + \epsilon_i \tag{4}$$

Where:
ϵ_i is the error term and x_i is the covariate factor, T_i is survival time.
The survival time by an intercept factor accelerated or decelerated if the covariates are compared in different levels. [6, 14, 22]
The AFT model can be enumerated as follow:

$$\delta(t|\varphi) = \varphi\delta_0(\varphi t) \tag{5}$$

Where φ is the covariates.

$$\varphi = \exp(-[\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_p X_p]) \tag{6}$$

2.6. *Parametric Distribution*

2.6.1. Exponential Distribution The simplest survival distribution is acquired by presuming a constant risk over time, so the hazard is constant.[10, 32]

$$\lambda(t) = \gamma \tag{7}$$

For all t, the counterpart survival function is [26]

$$S(t) = \exp^{-\gamma t} \tag{8}$$

The density function can be obtained from multiplying both of survival function and the hazard rate as follows:[2]

$$f(t) = \gamma \exp^{-\gamma t} \tag{9}$$

The Maximum Likelihood function of Exponential Distribution is as follows:[17]

$$L(\gamma, t_1, t_2, t_3, ..., t_n) = \prod_{i=1}^{n} \gamma \exp^{-\gamma t_i} = \gamma^n \mathrm{e}^{-\gamma \sum_{i=0}^{n} t_i} \tag{10}$$

2.6.2. Weibull Distribution The Weibull distribution is one of the most crucial distributions that is widely used in reliability applications and life tests. The following equation is the two-parameter Weibull distribution model:

$$f(t, \alpha, \delta) = \frac{\alpha}{\delta} t^{\alpha-1} e^{-\frac{t^\alpha}{\delta}}, \quad t > 0 \tag{11}$$

So:
The shape parameter $\alpha > 0$
The scale parameter $\delta > 0$
So the cumulative (non-reliable) distribution function is as follows:

$$F(t) = \Pr(T \leq t) = \int_0^t f(u)du = 1 - \exp\left[-\frac{t^\alpha}{\delta}\right] \tag{12}$$

The survival function is as follows:

$$S(t) = 1 - F(t) = \exp\left[-\frac{t^\alpha}{\delta}\right] \tag{13}$$

But the hazard function is as follows:

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\alpha}{\delta}t^{\alpha-1} \tag{14}$$

Note that:

1. When $\alpha > 1$, the risk function is increasing with time t.
2. When $\alpha < 1$, the risk function is decreasing with time t.
3. When $\alpha = 1$, the risk function is constant. [2, 4, 16, 32]

The Likelihood function for Weibull Distribution is as follows:[30]

$$L(\alpha, \delta, t_1, t_2, t_3, ..., t_n) = \left(\frac{\alpha}{\delta}\right)^n \prod_{i=1}^{n} (t_i)^{\alpha-1} \mathrm{e}^{-\sum_{i=1}^{n} \frac{t_i^\alpha}{\delta}} \tag{15}$$

2.6.3. Log-Normal Distribution If X is a random attribute which has lognormal distribution, the probability density function and the survival function is given as follows:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{-1}{2\sigma^2}(\ln t - \omega)^2} \quad \text{for } t > 0, \sigma > 0 \tag{16}$$

$$F(t) = 1 - \vartheta\left[\frac{\ln(at)}{\sigma}\right] \tag{17}$$

Where $a = e^{-\omega}$ and the cumulative distribution function for the standard normal is $\vartheta(.)$. [4, 29, 32]
The failure rate is given as follows:

$$\lambda(t) = \frac{\frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(\ln(\alpha t))^2}{2\sigma^2}}}{1 - \vartheta\left[\frac{\ln(\alpha t)}{\sigma}\right]} \tag{18}$$

The Maximum Likelihood function for Lognormal Distribution is as follows:[20]

$$L(\omega, \sigma, t_1, t_2, t_3, ..., t_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \frac{1}{\sum_{i=0}^{n} t_i} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln t_i - \omega)^2} \tag{19}$$

3. Estimation Methods

There are many methods for estimating the parameters of the Exponential, Weibull and Lognormal models and the survival functions. We will discuss some of them as follows:

3.1. Maximum Likelihood Method (MLE)

Maximum Likelihood is a popular estimation technique with application in time series modeling, panel data, discrete data, and even machine learning. Maximum likelihood estimation is a statistical technique for estimating a model's parameters. The parameters in maximum likelihood estimation are selected to increase the probability that the observed data will be produced by the assumed model. [17] It is one of the most important estimation methods that aims to bring the likelihood function of random variables to its maximum.[21, 24] The likelihood function is as follows:

$$L(\varphi) = P(T_1 = t_1, T_2 = t_2, ..., T_n = t_n) = f(t_1, \varphi) \cdot f(t_2, \varphi) \cdots f(t_n, \varphi) = \prod_{i=1}^{n} f(t_i, \varphi) \tag{20}$$

The maximum likelihood estimator:

$$\hat{\varphi}(t) = \arg\max_{\varphi} L(\varphi|t) \tag{21}$$

3.2. Robust Estimation Methods

3.2.1. M - Estimate It is known that the idea of the least squares method depends on minimizing the sum of squared errors as small as possible, i.e.:

$$\text{Min}_{a,b} \sum_{i=1}^{n} [Y_i - a - bX_i]^2 \tag{22}$$

The method adopted by the M-estimate method is to minimize the following amount:

$$\text{Min}_{a,b} \sum_{i=1}^{n} \rho[Y_i - a - bX_i] \tag{23}$$

Since the ρ function from which we obtain robust estimators, to reduce the value in the above formula, we take the partial derivative of the parameters and set it equal to zero. [5, 19]

There are many types of Weight Function:

1. Huber	2. Cauchy	3. Fair
4. Hampel	5. Andrews	6. Welsch
7. Median	8. Tukey's Biweight	

The weight function of Tukey's Biweight has the following form: [24]

$$Z(x, a) = \begin{cases} (1 - \left(\frac{x}{a}\right)^2)^2 & \text{if } |x| < a \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Where

a : is the value of tuning constant (typically $a = 4.685$)

The weight function of Median has the following form: [1]

$$Z(x, a) = \begin{cases} \frac{1}{a} & \text{if } x = 0 \\ \frac{1}{|x|} & \text{otherwise} \end{cases} \tag{25}$$

4. Bootstrap Method

One way to do re-sampling is with bootstrap methods. By resampling the initial data sets, bootstrap allows us to approximate the full sampling distributions. When the initial sample size is small and the normalcy assumption is not met. The bootstrap method is utilized to estimate the sampling distribution, calculate the measures of accuracy to sample estimates, and estimate the regression parameters and their standard errors as well as the confidence interval,; It can also be applied to the creation of hypothesis tests.[8]

5. Study on a Real Dataset

In this study, we used medical data from Kaggle platform. It is a breast cancer dataset, and the dataset includes 4 attributes and 4024 instances. The number of deaths are (616) patients with a censoring rate (0.153), and the R Language Program is used in the application where the dependent variable is the survival months of each patient and the explanatory variables are the (Age, Tumor Size, Tumor Stage and Node Stage). The link of the breast cancer dataset in Kaggle platform is (https://www.kaggle.com/code/jingrini/breast-cancer-dataset).

Let's denote each variable as:

$x_1 =$ Age $x_2 =$ Tumor Size

$x_3 =$ Tumor Stage $x_4 =$ Node Stage

5.1. Discussion

1. In this dataset for breast cancer with 4024 observations four different methods were applied to see the differences between each method and to choose the best parameters that affect the survival time for each patient.
2. We fortified the parametric analysis by employing the non-parametric bootstrap. This approach provides the standard error, z-values and p-values that are not contingent on the assumed models asymptotic properties, but are instead empirically derived from the data's own distributions.
3. The first method is parametric Accelerated Failure Time AFT for three different distributions such as (Exponential, Weibull and Lognormal) in all the three distributions the attribute Tumor Size was not significant since the value in three distributions respectively were (0.9429, 0.8878 and 0.7385) greater than the $\alpha = 0.05$.
4. The Second method is Parametric Survival Regression for (Exponential, Weibull and Lognormal) Distributions in Survival Analysis with base of AFT. For each of the three distributions the variable Tumor Size respectively also is not significant with $\alpha = 0.05$, the values are (0.9429, 0.8878 and 0.7385) as it is obvious the values are greater than the α.
5. The Third method is Robust AFT for (Exponential, Weibull and Lognormal) distributions in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function, all the variables (Age, Tumor Size, Tumor Stage and Node Stage) were significant to compare them with $\alpha = 0.001$
6. The last method is Robust Parametric Survival Regression for (Exponential, Weibull and Lognormal) Distributions in Survival Analysis with base of AFT by using Tukey's Biweight Function. Both of the Exponential and Weibull distribution's variables were significant to compare it with $\alpha = 0.001$ but in the Lognormal Distribution the Tumor Size with 0.92 and the T Stage with 0.111 were not significant with the $\alpha = 0.05$.
7. From these comparisons we can say that the best method is Robust AFT for (Exponential, Weibull and Lognormal) Distributions in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function.

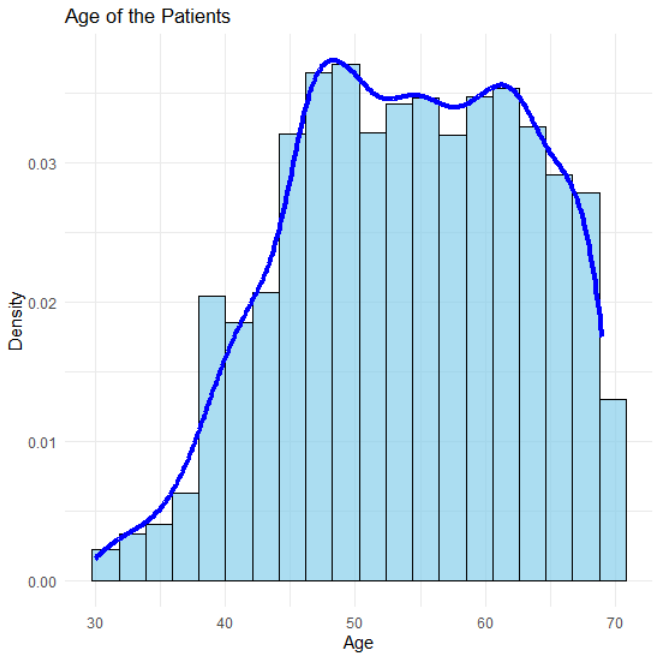
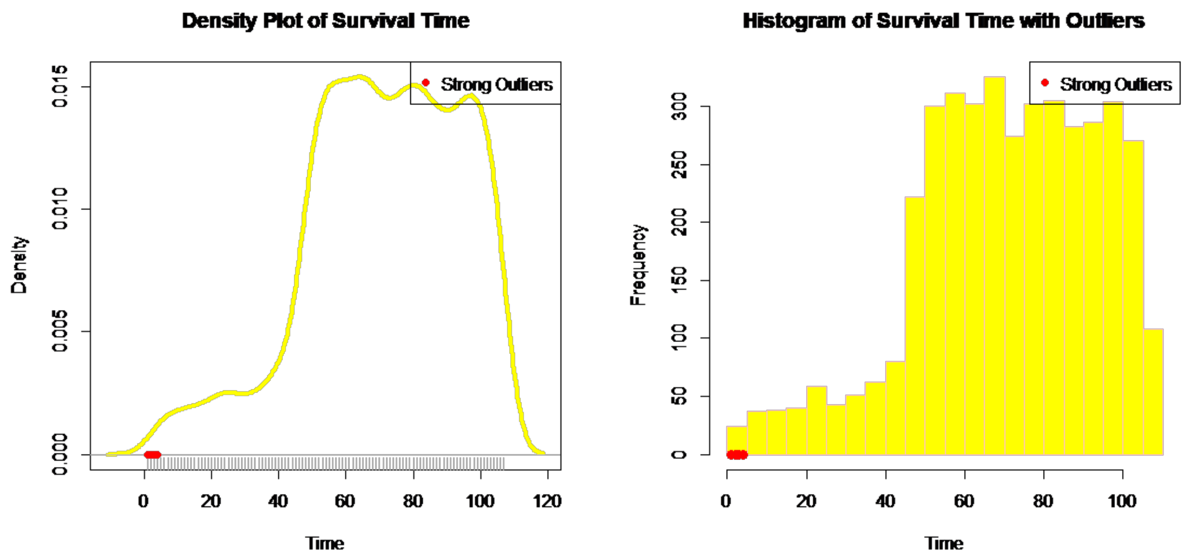


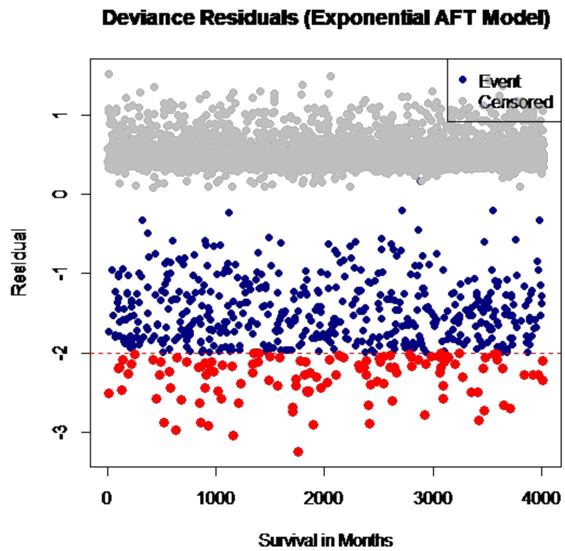
Figure 1. Represents the Age of the Breast Cancer Patients



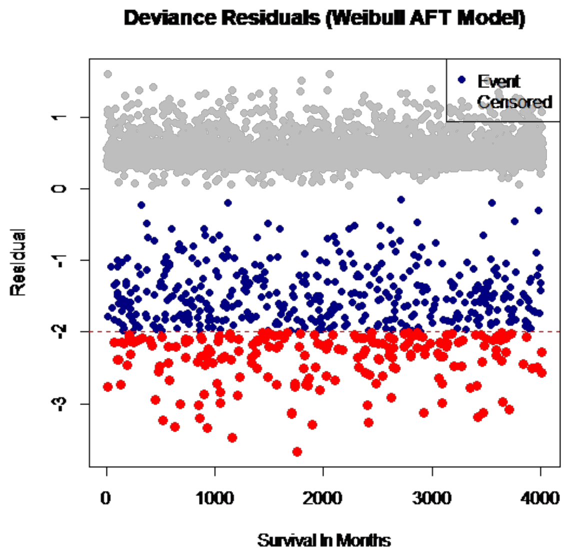
(a) (b)

Figure 2. Represents the survival months of each breast cancer patients

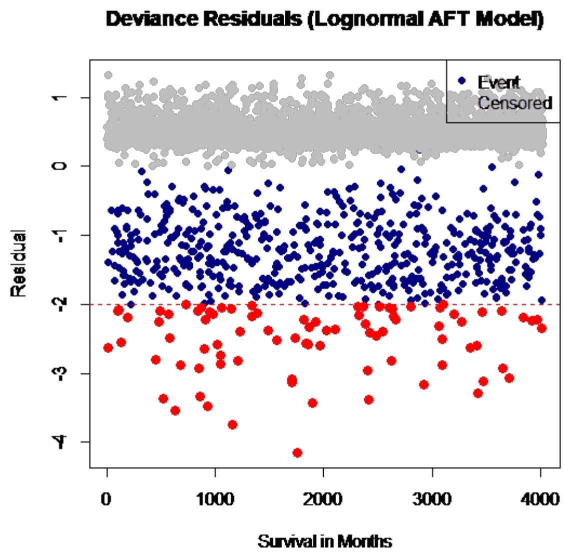
The survival months of each patients in histogram and density plot with outliers.



(a)



(b)



(c)

Figure 3. Represents the appearance of outliers in (Exponential, Weibull and Lognormal) distributions

In the figure above, outliers are shown in each survival model by using deviance residuals plot.
The Density Plot in Figure 2 and the Deviance Residuals Plot in Figure 3 show the presence of the outliers in the dataset which is the reason that the robust parametric survival models are showing more accurate results than the classical parametric survival models.

Table 1. Represents the Goodness of fit Test for (Exponential, Weibull and Lognormal) distributions.

Distribution	Kolmogorov-Smirnov	Log Likelihood
Weibull	0.054059	-18394.1
Lognormal	0.15728	-19643.0
Exponential	0.372397	-21193.9

Table 1 presents the Kolmogorov-Smirnov and Log likelihood test for testing the goodness of fit. The smaller value of Kolmogorov-Smirnov indicate the better fitting of the model. According to the Table 1, based on the Kolmogorov-Smirnov statistic the Weibull distribution has the best performance, but the exponential distribution is not a good fit over the data.

Table 2. Represents the Parametric Survival Regression for Exponential Distribution in Survival Analysis with base of AFT.

Var	Estimate	S. Error	Z	P
x_1	0.017737	0.004663	3.7523	0.00017
x_2	0.00018	0.002662	0.0715	0.94295
x_3	0.283223	0.07592	3.7304	0.00019
x_4	0.66401	0.05048	13.151	0.00000

Table 3. Represents the Parametric AFT model for Exponential Distribution in Survival Analysis.

Var	Estimate	S. Error	Z	p
x_1	-0.017374	0.004630	-3.75	0.00018
x_2	-0.000188	0.002625	-0.07	0.94296
x_3	-0.283237	0.075927	-3.73	0.00019
x_4	-0.664014	0.050488	-13.15	$<$ 2e-16

Table 4. Represents the Robust Parametric Survival Regression for Exponential Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function.

Var	Estimate	S. Error	Z	P
x_1	-2.93983	0.02226	-132.036	$<$ 1e-16
x_2	-1.64045	0.02267	-72.3416	$<$ 1e-16
x_3	-0.09682	0.00096	-100.389	$<$ 1e-16
x_4	-0.07711	0.00082	-93.0857	$<$ 1e-16

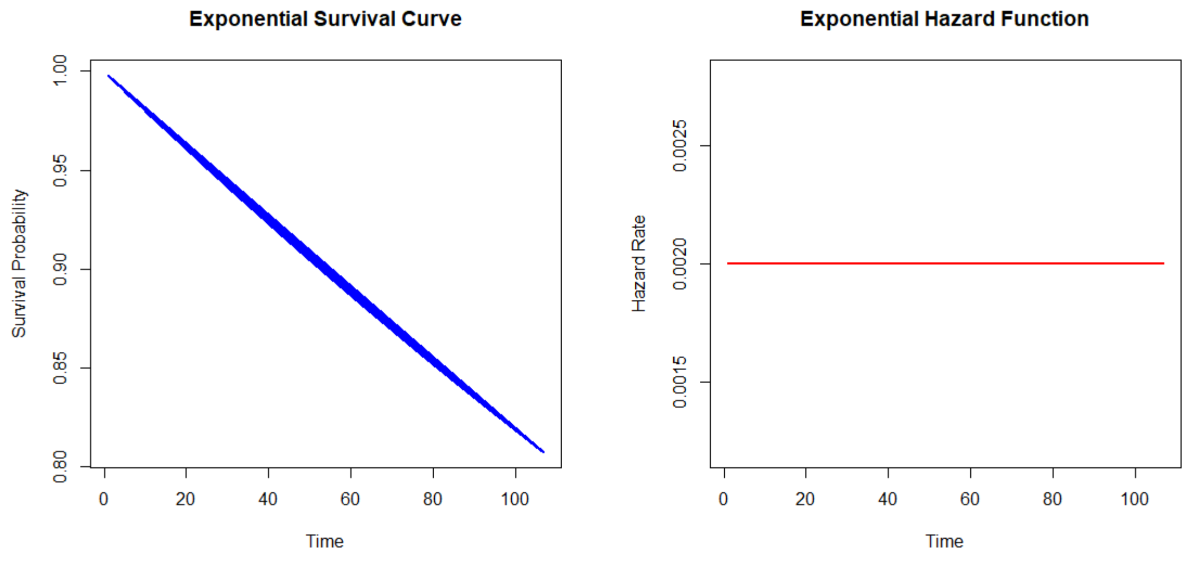
Table 5. Represents the Robust AFT Exponential Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight Function.

Var	Estimate	S. Error	Z	P
x_1	-2.93983	0.02226	-132.0369	$<$ 1e-16
x_2	-1.64045	0.022676	-72.34163	$<$ 1e-16
x_3	-0.09682	0.000964	-100.3897	$<$ 1e-16
x_4	-0.07711	0.000828	-93.08570	$<$ 1e-16

Table 6. Represents the Robust AFT Exponential Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Median Weight Function.

Var	Estimate	S. Error	P	Sig
x_1	23.1540	(0.0000)	< 0.001	***
x_2	-37.3203	(0.0000)	< 0.001	***
x_3	109.4631	(0.0000)	< 0.001	***
x_4	-12.8052	(0.0000)	< 0.001	***

From tables (2 - 6) for exponential parametric model four different methods have been used. In the first method which is Parametric Survival Regression for Exponential Distribution in Survival Analysis with base of AFT and the second method Parametric AFT model for Exponential Distribution in Survival Analysis the Tumor Size was not significant. The last two methods are the Robust Parametric Survival Regression for Exponential Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function and the Robust AFT Exponential Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function where all the variables are significant to compare them with α=0.001. In table (2 and 5) the x_1 = Age variable coefficient is 0.017 and -2.93 respectively, the Age has a long survival time in table 2 but in table 5 it has a short survival time. Also the results are the same in table (4 and 5) since the hazard rate is constant. But in table 6 Median weight function is used so the results are all significant.



(a) (b)

Figure 4. Represents the Survival Function of Exponential Parametric Model and the Hazard Function of Exponential Parametric Model

From Figure 4 it is obvious that the survival rate for the patients are decreasing by increasing time. But the hazard function is stable since the hazard rate is constant.

Table 7. Represents the Parametric Survival Regression for Weibull Distribution in Survival Analysis with base of AFT.

Var	Estimate	S. Error	Z	P
x_1	-0.01268	0.003359	-3.7771	0.000158
x_2	-0.00026	0.001892	-0.1410	0.88778
x_3	-0.2033	0.05522	-3.6824	0.000231
x_4	-0.4858	0.04004	-12.1324	0.000000

Table 8. Represents the Parametric AFT model for Weibull Distribution in Survival Analysis.

Var	Estimate	Std. Error	Z	p
x_1	-0.012688	0.003359	-3.78	0.00016
x_2	-0.000267	0.001893	-0.14	0.88785
x_3	-0.203352	0.055222	-3.68	0.00023
x_4	-0.485883	0.040048	-12.13	$<$ 2e-16

Table 9. Represents the Robust Parametric Survival Regression for Weibull Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function.

Var	Estimate	S. Error	Z	P
x_1	-2.87340	0.02302	-124.799	$<$ 1e-16
x_2	-1.59487	0.02263	-70.4573	$<$ 1e-16
x_3	-0.09432	0.00097	-96.4714	$<$ 1e-16
x_4	-0.07484	0.00082	-90.2556	$<$ 1e-16

Table 10. Represents the Robust AFT Weibull Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function.

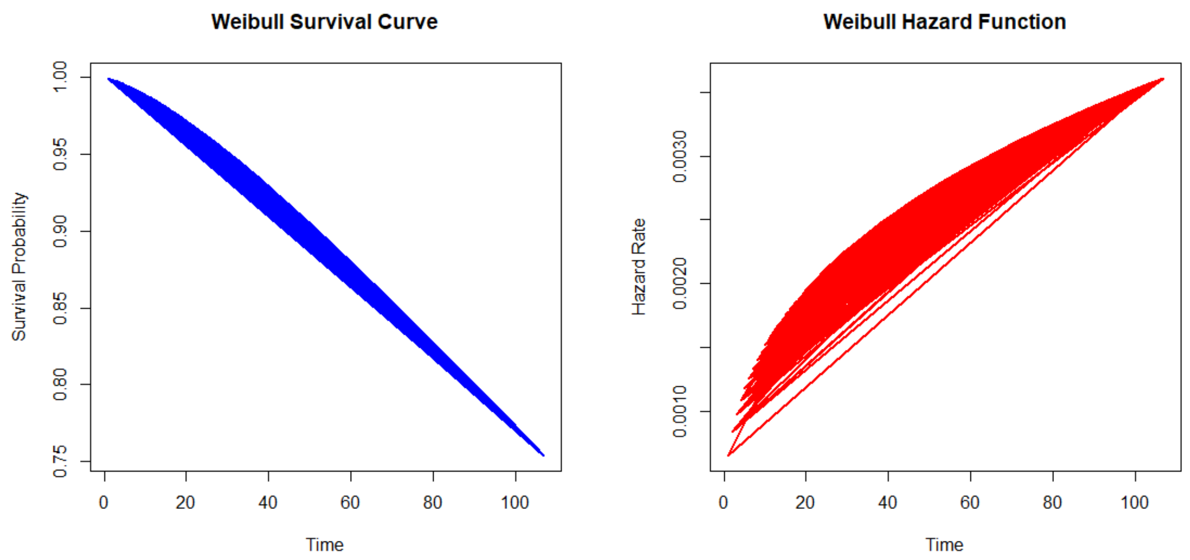
Var	Estimate	S. Error	Z	P
x_1	-33.78109	1.32376	-25.5189	$<$ 1e-16
x_2	-23.03537	0.89615	-25.7045	$<$ 1e-16
x_3	-1.271103	0.04788	-26.5474	$<$ 1e-16
x_4	-1.147320	0.03790	-30.2722	$<$ 1e-16

Table 11. Represents the Robust AFT Weibull Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Median Weight Function.

Var	Estimate	S. Error	P	Sig
x_1	550.5424	(0.0000)	$<$ 0.001	***
x_2	347.0798	(0.0000)	$<$ 0.001	***
x_3	-4.8246	(0.0000)	$<$ 0.001	***
x_4	18.1461	(0.0000)	$<$ 0.001	***

In tables (7- 11) Weibull parametric model have been used with four different methods both of the two methods Parametric Survival Regression for Weibull Distribution in Survival Analysis with base of AFT and the Parametric AFT model for Weibull Distribution in Survival Analysis, the tumor size was not significant. The last two methods are the Robust Parametric Survival Regression for Weibull Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function and the Robust AFT Weibull Distribution in Survival Analysis by using

Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function where all the variables are significant to compare them with α=0.001 but the best method is the Robust AFT Weibull Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Median weight Function. For the Five tables above we used Weibull parametric model in survival analysis. It is clear that the best method is the Robust Parametric Survival Regression with base of AFT by using Tukey's Biweight Function and Median weight function.



(a) (b)

Figure 5. Represents the Survival Function of Weibull Parametric Model and the Hazard Function of Weibull Parametric Model

From Figure 5 it is obvious the survival rate for the patients are decreasing by increasing time. But the hazard rate is increasing with increasing of time.

Table 12. Represents the Parametric Survival Regression for Lognormal Distribution in Survival Analysis with base of AFT.

Var.	Estimate	S. Error	Z	P
x_1	-0.01563	0.00371	-4.20398	0.000026
x_2	-0.00078	0.00236	-0.33381	0.73852
x_3	-0.22642	0.06748	-3.35542	0.000792
x_4	-0.55886	0.04665	0.00000	0.000000

Table 13. Represents the Parametric AFT model for Lognormal Distribution in Survival Analysis.

Var.	Estimate	S. Error	Z	P
x_1	-0.015631	0.003717	-4.21	0.000026
x_2	-0.000789	0.002364	-0.33	0.73851
x_3	-0.226426	0.067484	-3.36	0.00079
x_4	-0.558866	0.046657	-11.98	$<$ 2e-16

Table 14. Represents the Robust Parametric Survival Regression for Lognormal Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function.

Var.	Estimate	S. Error	Z	P
x_1	0.003459	0.00122	2.8238	0.00475
x_2	0.000091	0.00090	0.1010	0.92
x_3	0.038231	0.02397	1.5944	0.111
x_4	0.15218	0.02073	7.3404	2.13e-13

Table 15. Represents the Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function.

Var.	Estimate	S. Error	Z	P
x_1	-15.26562	0.12174	-125.3938	$<$ 1e-16
x_2	-8.632581	0.12256	-70.43745	$<$ 1e-16
x_3	-0.505403	0.005290	-95.52344	$<$ 1e-16
x_4	-0.404665	0.004613	-87.71001	$<$ 1e-16

Table 16. the Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Median Weight Function.

Var.	Estimate	S. Error	P	Sig
x_1	56.7226	(0.0000)	$<$ 0.001	***
x_2	86.2522	(0.0000)	$<$ 0.001	***
x_3	7.5748	(0.0000)	$<$ 0.001	***
x_4	2.2308	(0.0000)	$<$ 0.001	***

For tables (12 - 16) Lognormal parametric model is used with four different methods the Parametric Survival Regression for Lognormal Distribution in Survival Analysis with base of AFT and the Parametric AFT model for Lognormal Distribution in Survival Analysis, the Tumor Size was not significant, in the Robust Parametric Survival Regression for Lognormal Distribution in Survival Analysis with base of AFT by using Tukey's Biweight Function the variables (Tumor Size and Tumor Stage) were not significant. But in the Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function all the variables were significant to compare it with α=0.001. In the above - five tables above we used Lognormal parametric model in survival analysis. It is clear that the best method is the Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function and Median weight function.

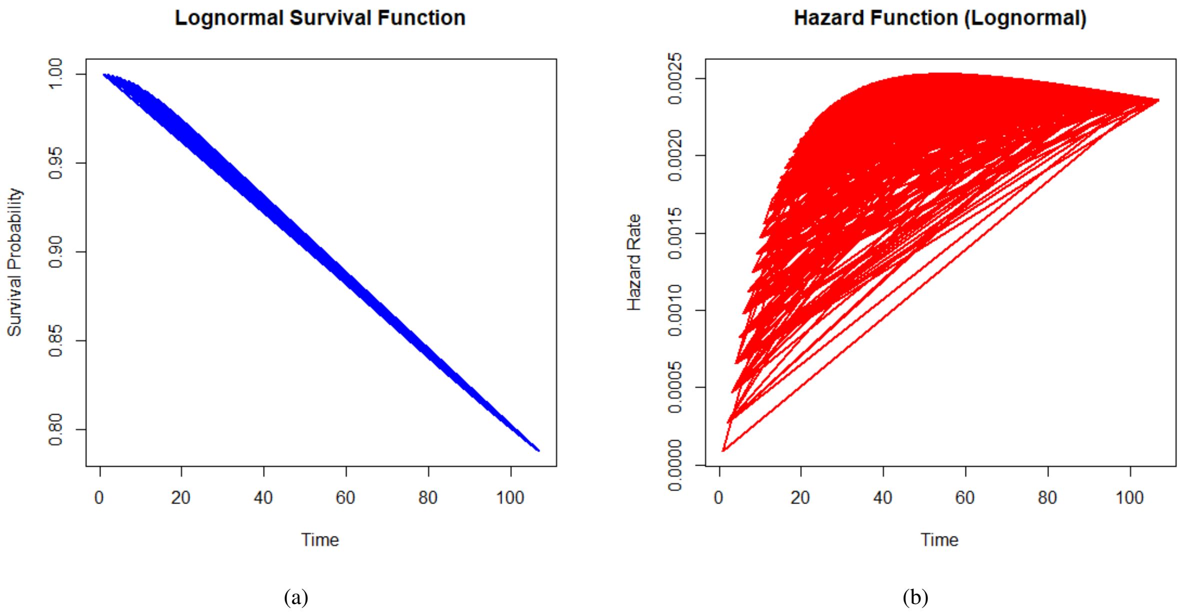


Figure 6. Represents the Survival Function of Lognormal Parametric Model and the Hazard Function of Lognormal Parametric Model

From Figure 6 it is obvious the survival rate for the patients are decreasing by increasing time. But the hazard rate is increasing with increasing of time.

Table 17. Represents the AIC and BIC for the Classical and Robust Survival Models.

	Classical Models			**Robust Models**		
Information Criteria	**Exponential**	**Weibull**	**Lognormal**	**Exponential**	**Weibull**	**Lognormal**
AIC	8544.899	8477.017	8488.623	313.6011	46.5376	45.9373
BIC	8576.399	8514.817	8526.423	338.8012	78.0378	77.4375

Table 17 presents the AIC and BIC for the classical and robust parametric models in the classical parametric models Weibull distribution is the best model since it has the minimum AIC and BIC as follows (8477.017 and 8514.817) but in the robust parametric models the Lognormal distribution is the best model with the smallest AIC and BIC (45.9373 and 77.4375).

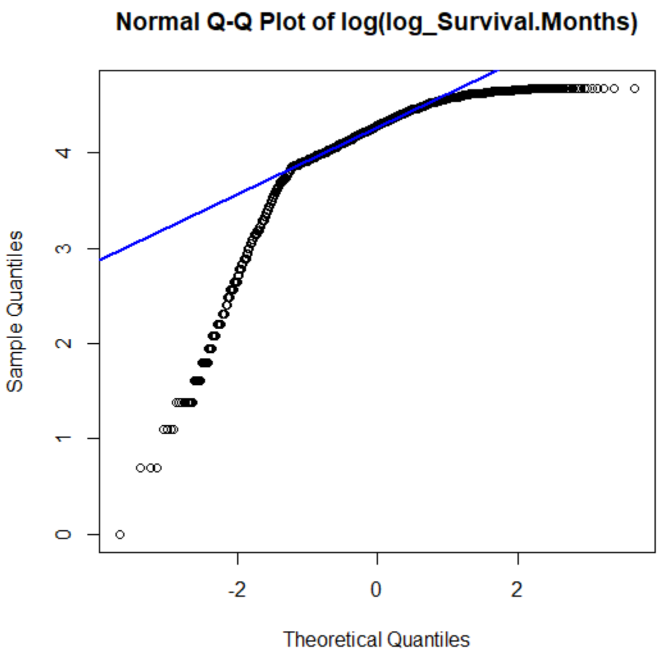


Figure 7. Represents the Survival Months of each patients in breast cancer

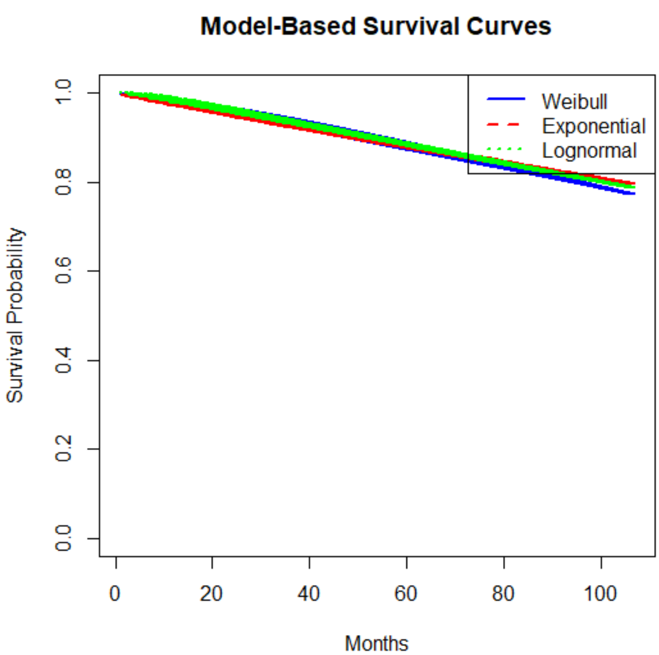


Figure 8. Represents the Survival curve for the (Exponential, Weibull and Lognormal) parametric Models in breast cancer Patients

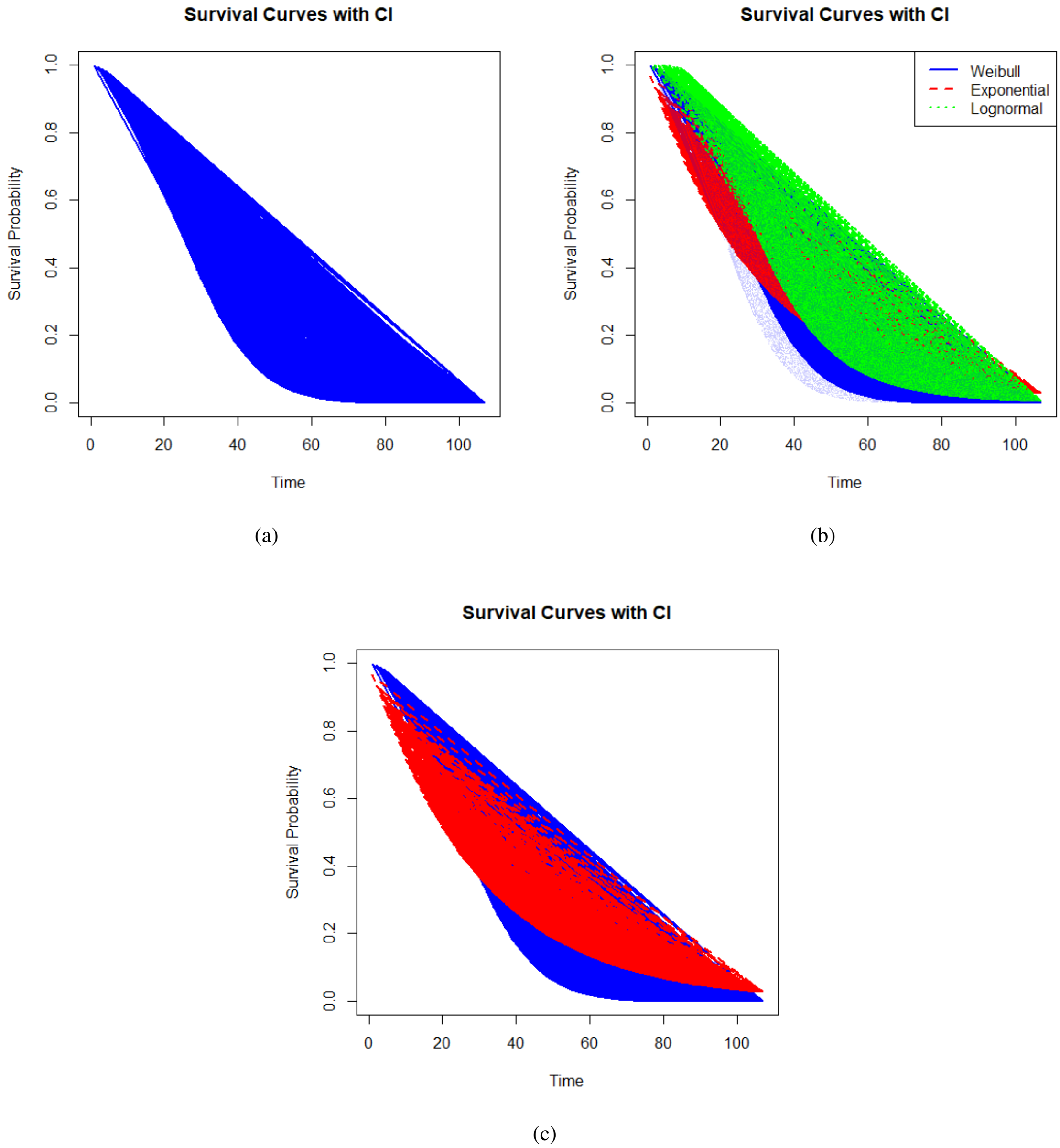


Figure 9. Represents the Survival curve with Confidence Interval for the (Exponential, Weibull and Lognormal) parametric Models in breast cancer Patients

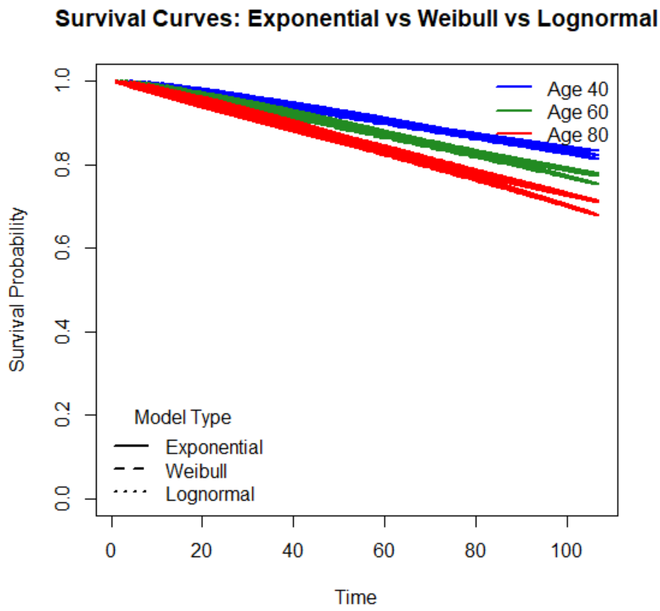
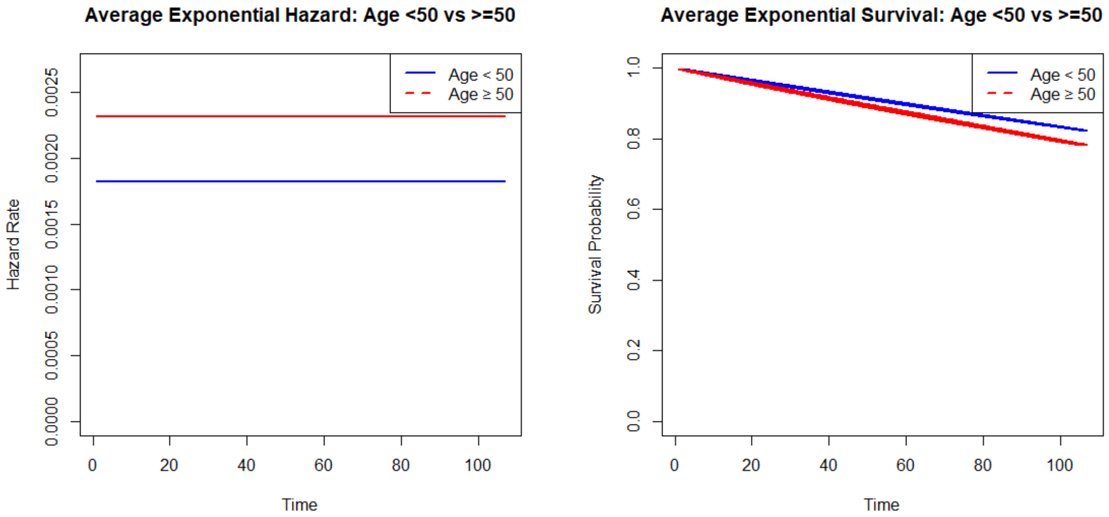


Figure 10. Represents the Survival curve for the (Exponential, Weibull and Lognormal) parametric Models in breast cancer Patients in Ages (40, 60 and 80)

Table 18. Represents the Exponential distribution for grouped age variable with (Age $<$ 50 and Age $>=$ 50)

	Age: Less than 50		Age: Greater than or Equal 50	
	$S(t)$	$\lambda(t)$	$S(t)$	$\lambda(t)$
1	0.896	0.00182	0.870	0.00231
2	0.893	0.00182	0.866	0.00231
3	0.872	0.00182	0.840	0.00231
4	0.858	0.00182	0.823	0.00231
5	0.912	0.00182	0.8907	0.00231
6	0.850	0.00182	0.813	0.00231
⋮				
4022	0.855	0.00182	0.852	0.00231
4023	0.965	0.00182	0.852	0.00231
4024	0.833	0.00182	0.846	0.00231

The table above is the grouped ages for exponential distribution and $h(t) = 0.00231$ which means that the hazard rate for all the patients are the same since the exponential distribution's hazard rate is constant. [10, 32].



(a) (b)

Figure 11. Represents the Survival and Hazard curve for the Exponential parametric Model in breast cancer Patients in Ages (Less than 50 and Greater than or Equal to 50)

Table 19. Represents the Weibull distribution for grouped age variable with (Age $<$ 50 and Age $>=$ 50)

	Age: Less than 50		**Age: Greater than or Equal 50**	
	$S(t)$	$\lambda(t)$	$S(t)$	$\lambda(t)$
1	0.905	0.00226	0.880	0.00289
2	0.901	0.00229	0.875	0.00292
3	0.873	0.00246	0.841	0.00313
4	0.854	0.00256	0.817	0.00327
5	0.925	0.00212	0.905	0.00270
6	0.843	0.00261	0.804	0.00334
⋮				
4022	0.849	0.00258	0.857	0.00304
4023	0.979	0.00148	0.857	0.00304
4024	0.818	0.00273	0.849	0.00309

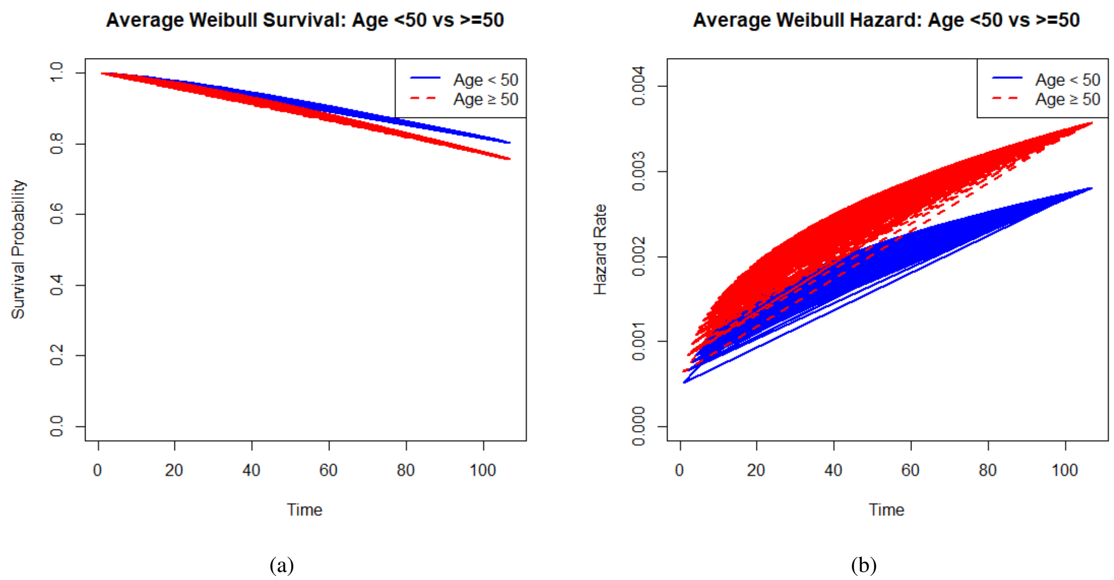
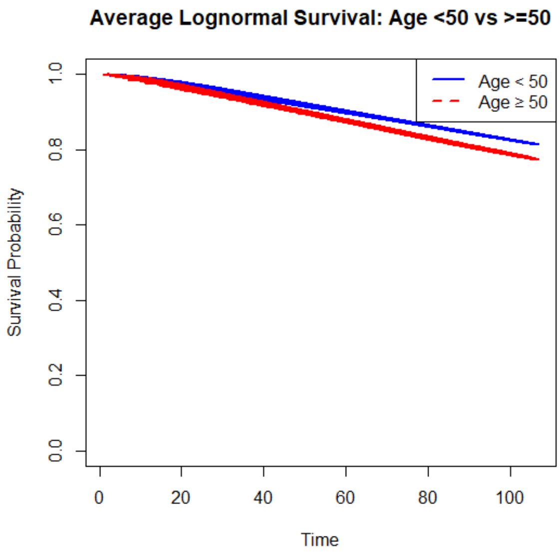


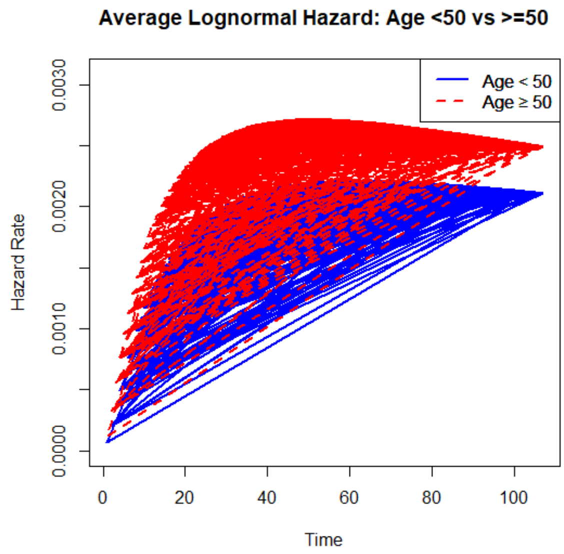
Figure 12. Represents the Survival and Hazard curve for the Weibull parametric Model in breast cancer Patients in Ages (Less than 50 and Greater than or Equal to 50)

Table 20. Represents the Lognormal distribution for grouped age variable with (Age $<$ 50 and Age $>=$ 50)

	Age: Less than 50		**Age: Greater than or Equal 50**	
	$S(t)$	$\lambda(t)$	$S(t)$	$\lambda(t)$
1	0.900	0.00220	0.874	0.00269
2	0.896	0.00220	0.869	0.00269
3	0.871	0.00219	0.839	0.00264
4	0.854	0.00217	0.820	0.00260
5	0.920	0.00218	0.898	0.00270
6	0.845	0.00216	0.809	0.00257
⋮				
4022	0.850	0.00217	0.853	0.00266
4023	0.980	0.00170	0.853	0.00266
4024	0.825	0.00213	0.846	0.00265



(a)



(b)

Figure 13. Represents the Survival and Hazard curve for the Lognormal parametric Model in breast cancer Patients in Ages (Less than 50 and Greater than or Equal to 50)

Conclusion

In this study applied the Accelerated Failure Time, Regression model, Robust Accelerated Failure Time with (Tukey's Biweight function and Median Weight function) and Robust Regression model with Tukey's Biweight function through exponential, weibull and lognormal parametric survival models for their comparative analysis, to show the best result from these four different methods for each of (Exponential, Weibull and Lognormal) distributions.
We conclude that:

1. The Robust AFT Lognormal Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function is the best method since the p value is less than the $\alpha = 0.001$.
2. The Robust AFT Weibull Distribution in Survival Analysis by using Maximum likelihood-Type Estimator with Tukey's Biweight function is also the best model.
3. The Robust AFT for (Exponential, Weibull and Lognormal) by using Maximum likelihood-Type Estimator with Median weight function shows the best result since the P value is less than the $\alpha = 0.001$.

In the goodness of fit test by depending on the Kolmogorov-Smirnov statistic the best distribution that fits the data is the weibull distribution and the distribution that is not a good fit for the data is the exponential distribution.

REFERENCES

1. Abdalla, H. T., *Comparison between biexponential and robust biexponential nonlinear models, using simulation [Master's thesis, University of Sulaimani]*, 2016.
2. Abdullah, M. A. A., Long, K. S., & Jamil, S. A. M., *Analysis of survival in breast cancer patients by using different parametric models*, IOP Conference Series: Journal of Physics: Conference Series, 890, 012169, 2017.
3. Akor, A., Sadiq, T. A., Usman, A., Doguwa, S. I., & Akor, L. O., *Advanced survival modeling of tuberculosis patients: Insights from exponential and Weibull AFT models*, FUDMA Journal of Sciences, 9(3), 169–182, 2025. https://doi.org/10.33003/fjs-2025-0903-3350
4. Aktürk Hayat, E., Suner, A., Uyar, B., Dursun, Ö., Orman, M. N., & Kitapçıoğlu, G., *Comparison of five survival models: Breast cancer registry data from Ege University Cancer Research Center*, Turkiye Klinikleri Journal of Medical Sciences, 30(5), 1665–1674, 2010. https://doi.org/10.5336/medsci.2009-16200
5. Almongy, H. M., & Almetwally, E. M., *Robust estimation methods of generalized exponential distribution with outliers*, Pakistan Journal of Statistics and Operation Research, 16(3), 545–559, 2020. https://doi.org/10.18187/pjsor.v16i3.3016
6. Awodutire, P. O., Kolawole, O. A., & Ilori, O. R., *Paramometric modeling of survival times among breast cancer patients in a teaching hospital, Osogbo*, Journal of Cancer Treatment and Research, 5(4), 81–85, 2017.
7. Ayitey, E., *Survival analysis of breast cancer patients in Ghana: Using the non-parametrics and parametrics approaches [Master's thesis, University of Ghana]*, University of Ghana Digital Collections, 2019. http://ugspace.ug.edu.gh
8. Barznji, N. S., Sedeeq, B. S., Taha, K. S., & Abullah, D. F., *Utilizing Cox regression analysis and bootstrapping test for indicating the risk of smoking on health in Kurdistan*, Tikrit Journal of Administrative and Economic Sciences, 20(65), 331–347, 2024. https://doi.org/10.25130/tjaes.20.65.1.19
9. Cooney, P., & White, A., *Direct incorporation of expert opinion into parametric survival models to inform survival extrapolation*, Medical Decision Making, 43(3), 325–336, 2023.
10. Cooney, P., & White, A., *Extending beyond Bagust and Beale: Fully parametric piecewise exponential models for extrapolation of survival outcomes in health technology assessment*, Value in Health, 26(10), 1510–1517, 2023. https://doi.org/10.1016/j.jval.2023.06.007
11. Dey, T., Mukherjee, A., & Chakraborty, S., *A practical overview and reporting strategies for statistical analysis of survival studies*, Chest, 158(1), S39–S48, 2020. https://doi.org/10.1016/j.chest.2020.04.011
12. Ekman, A., *Variable selection for the Cox proportional hazards model: A simulation study comparing the stepwise, lasso and bootstrap approach [Master's thesis, Uppsala University]*, DiVA Portal, 2017. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-325363
13. Farida, Y., Maulida, E. A., Desinaini, L. N., Utami, W. D., & Yuliati, D., *Breast cancer survival analysis using Cox proportional hazard regression and Kaplan-Meier method*, Jurnal Teori dan Aplikasi Matematika (JTAM), 5(2), 340–358, 2021.
14. Gail, M., Krickeberg, K., Samet, J., Tsiatis, A., & Wong, W., *Survival analysis: A self-learning text (2nd ed.)*, Springer, 2005.
15. Ghosh, A., & Basu, A., *Robust and efficient estimation in the parametric Cox regression model under random censoring*, arXiv, 2018. https://doi.org/10.48550/arXiv.1810.02991
16. Güvercin, A. C. Y., Tekindal, M. A., Kaymaz, Ö., & Güvercin, C. H., *Comparison of nonparametric, semiparametric and parametric survival analysis methods in right-censored medical data*, Biomedical Research, 28(10), 4360–4366, 2017.
17. Hamad, A. M., *Estimation of the parameter of an exponential distribution when applying maximum likelihood and probability plot methods using simulation*, Ibn Al-Haitham Journal for Pure and Applied Science, 25(1), 1–10, 2012.
18. Hosmer, D. W., Lemeshow, S., & May, S., *Applied survival analysis: Regression modeling of time-to-event data (2nd ed.)*, Wiley, 2008.

19. Huber, P. J., *Robust statistical procedures (2nd ed.)*, SIAM, 1997.
20. Kareem, R. S., *Comparing three different estimators of reliability function of lognormal distribution*, Al-Mansour Journal, (30), 101–114, 2018.
21. Karagöz, D., & Ata Tutkun, N., *Robust estimations of survival function for Weibull distribution*, Statistica, 81, 3–20, 2021.
22. Kleinbaum, D. G., & Klein, M., *Survival analysis: A self-learning text (3rd ed.)*, Springer, 2012. https://doi.org/10.1007/978-1-4419-6646-9_1
23. Manzo, G., D'Amato, M., Mazzarella, G., Montella, M., & Arienzo, C., *Breast cancer survival analysis agents for clinical decision support*, Computer Methods and Programs in Biomedicine, 231, 107373, 2023. https://doi.org/10.1016/j.cmpb.2023.107373
24. Maronna, R. A., Martin, R. D., & Yohai, V. J., *Robust statistics: Theory and methods*, John Wiley & Sons, 2006.
25. Nabeel, M., Tariq, M., Azam, M., Ayub, S., & Imran, M., *Robust surveillance schemes based on proportional hazard model for monitoring reliability data*, Mathematics, 11(11), 2480, 2023. https://doi.org/10.3390/math11112480
26. Potter, M., Maxenti, S., & Everett, M., *Survival analysis with adversarial regularization*, arXiv, 2024. https://doi.org/10.48550/arXiv.2312.16019
27. Rasheed, I., Cannings, T. I., Sell, T., & dos Reis, G., *Robust survival model for the prediction of Li-ion battery lifetime reliability and risk functions*, Energy and AI, 100465, 2024. https://doi.org/10.1016/j.egyai.2024.100465
28. Sabila, B. A., Julianto, A. H., Kurniawan, A., & Rifada, M., *Modeling survival analysis and risk factors in breast cancer patients*, International Journal of Multidisciplinary Research and Publications (IJMRAP), 7(1), 1–6, 2025.
29. Sadiq, M., & Mehmood, T., *A flexible and robust approach to analyze survival systems in the presence of outliers*, Mathematical Problems in Engineering, 2021, 9927377, 2021. https://doi.org/10.1155/2021/9927377
30. Salih, D. T. M., Salih, S. M., & Fage, M. M., *Estimate parametric regression models depend on (time-to-event) survival time distributions with application*, Al-Mustansiriyah Journal of Science, 33(4), 270–290, 2022.
31. Shoraim, M. H. A., *A comparative study of the robust estimation methods of survival function with practical application on blood cancer patients in Yemen [Doctoral dissertation, University of Baghdad]*, 2005.
32. Taketomi, N., Yamamoto, K., Chesneau, C., & Emura, T., *Parametric distributions for survival and reliability analyses: A review and historical sketch*, Mathematics, 10(22), 3907, 2022.
33. Tasfa Marine & Mengistie, *Application of parametric survival analysis to women patients with breast cancer at Jimma University Medical Center*, 2023.