UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521

University of Kirkuk Journal For Administrative and Economic Science

Aziz Nakhshin Rafiq & Hussein Mohammad Mahmood Faqe. A COMPARATIVE ANALYSIS OF DE-FUZZIFICATION TECHNIQUES FOR SURVIVAL TIME DATA IN WEIBULL DISTRIBUTION. *University of Kirkuk Journal For Administrative and Economic Science* (2025) 15 (1):201-214.

A COMPARATIVE ANALYSIS OF DE-FUZZIFICATION TECHNIQUES FOR SURVIVAL TIME DATA IN WEIBULL DISTRIBUTION

Nakhshin Rafiq Aziz [1], Mohammad Mahmood Faqe Hussein [2]

[1] *Department of Mathematics, College of Education, University of Sulaimani, Sulaymaniyah, Iraq*
[2] *Department of Informatics and Statistics, College of Administration and Economics, University of Sulaimani, Sulaymaniyah, Iraq*

nakhshin.aziz@univsul.edu.iq
mohammad.faqe@univsul.edu.iq

Abstract: Survival analysis is critical for evaluating the lifetime and performance of systems and goods, especially in sectors where exact failure predictions are required. The goal of this research is to compare the performance and accuracy of various defuzzification techniques in estimating survival times for fuzzy data using the 2 and 3-Parameter Weibull distribution, and to identify the most effective method for improving reliability in survival analysis applications. The fuzzy logic process is divided into three stages: fuzzification, inference language, and de-fuzzification. We applied some basic methods for the De-fuzzification process to convert the fuzzified inputs into crisp outputs, and the Maximum likelihood Method was applied to estimate Weibull distribution's parameters. After that, the real and fuzzy data were applied to calculate the probability density function and survival function, hazard rate of Weibull distribution, and to make comparisons Mean Square Error (MSE) is calculated. This study was focused on survival times of (127) heart disease patients who were admitted to Suleimani Centre for Heart Disease in Sulaymaniyah City during February-October (2024). In conclusion, by using MSE to differentiate between real and fuzzy data, confirming that fuzzified data provided the most accurate results for survival and hazard rate evaluations, and showed an inverse correlation between failure times and survival function.

Keywords: De-fuzzification Process, Fuzzy logics, Maximum Likelihood Method, Survival Analysis, Weibull Distribution.

تحليل مقارن لتقنيات إزالة الضبابية لبيانات وقت البقاء في توزيع ويبل

الباحثة: نخشين رفيق عزيز [١] ، ا.م.د. محمد محمود فقى حسين [٢]

[١] *قسم الرياضيات، كلية التربية-جامعة السليمانية، السليمانية، العراق*
[٢] *قسم الاحصاء والمعلوماتية، كلية الادارة والاقتصاد-جامعة السليمانية، السليمانية، العراق*

المستخلص: يعد تحليل البقاء أمرًا بالغ الأهمية لتقييم عمر وأداء الأنظمة والسلع، وخاصة في القطاعات التي تتطلب تنبؤات دقيقة بالفشل. الهدف من هذه الدراسة هو مقارنة أداء ودقة تقنيات إزالة الضبابية المختلفة في تقدير أوقات البقاء للبيانات الضبابية باستخدام توزيع ويبل ذي ٢ و٣ معلمات، وتحديد الطريقة الأكثر فعالية لتحسين الموثوقية في تطبيقات تحليل البقاء. تنقسم عملية المنطق الضبابي إلى ثلاث مراحل: الضبابية ولغة الاستدلال وإزالة الضبابية. لقد طبقنا بعض الأساليب الأساسية لعملية إزالة الضبابية لتحويل المدخلات الضبابية إلى مخرجات واضحة، واستُخدمت طريقة الاحتمالية القصوى لتقدير معلمات توزيع ويبل. بعد ذلك، تم تطبيق البيانات الفعلية والضبابية لحساب دالة كثافة الاحتمالية ودالة البقاء ومعدل الخطر لتوزيع ويبل، ولإجراء المقارنات تم حساب خطأ المربع المتوسط (MSE). ركزت هذه الدراسة على أوقات البقاء على قيد الحياة لـ (١٢٧) مريضًا بأمراض القلب تم إدخالهم إلى مركز السليمانية لأمراض القلب في مدينة السليمانية خلال الفترة من فبراير إلى أكتوبر (٢٠٢٤). في الاستنتاجات، باستخدام MSE للتمييز بين البيانات الحقيقية والضبابية، تم التأكيد على أن البيانات الضبابية قدمت النتائج الأكثر دقة لتقييمات معدل البقاء على قيد الحياة ومعدل الخطر، وأظهرت ارتباطًا عكسيًا بين أوقات الفشل ووظيفة البقاء على قيد الحياة.

الكلمات المفتاحية: عملية إزالة الضبابية، المنطق الضبابي، طريقة الاحتمال العظمى، تحليل البقاء، توزيع ويبل.

Corresponding Author: E-mail: nakhshin.aziz@univsul.edu.iq

1.Introduction

Survival analysis is a field of statistics that examines the predicted length of time before one event happens. Survival analysis is primarily concerned with the duration of an event in domains such as medicine, engineering, and social sciences, where understanding system failure behavior is critical for risk assessment, maintenance planning, and quality control. Traditional reliability models, particularly those based on Weibull distributions, rely on precise parameter estimations to accurately anticipate a system's odds of failure and survival. Real-world data, on the other hand, is typically ambiguous owing to variables such as measurement constraints, subjective evaluations, or changing environmental circumstances. This is especially important if inaccurate or insufficient data is utilized to predict failure rates.

Conventional models with exact parameters might not work well in certain circumstances, which could result in overly optimistic or incorrect forecasts. Instead of utilizing a single point estimate to describe uncertainty, fuzzy set theory provides a mathematical framework by employing "fuzzy numbers" that span a range of potential values. This method makes it possible to estimate "fuzzy hazard rates," which more accurately account for failure rate uncertainty than deterministic models.

Because it can accurately reflect the time-to-failure of real-world occurrences and is sufficiently easy despite having two parameters, the Weibull distribution is especially favored in survival analysis. This paper focuses on developing a fuzzy survival time based on a two-parameter Weibull distribution. In (2003) Carroll and Kevin [1], This paper explores The Weibull model can be advantageous when analyzing survival data in clinical trials, highlighting its practical benefits. In (2011) Zhu, et al. [2], used Weibull distribution to analyse the prognostic factors in patients with gastric cancer. In (2022),_Roohanizadeh,et al,[3], explored various estimation methods for two-parameter Weibull distribution by using intuitive fuzzy lifetime data with maximum likelihood and Bayesian estimation methods. In (2022), Jain, et al., [4] examined various methods for de-fuzzifying fuzzy output, which involves converting a value quantity into a concrete value. In (2010). Hadi-Vencheh and Allame,[5] studied the method of determining the centroid and left/right spread of an unknown fuzzy number given its information.

This article's goal is to model fuzzily hazardous rates in survival analysis. Applying Weibull distribution: The survival and hazard functions are estimated after the parameter of the Weibull distribution is estimated using the maximum likelihood method. The actual data has been fuzzified using trapezoidal membership based on the process of fuzzification and linguistic inference then be converted to a crisp data using some basic techniques of defuzzification process. Lastly, the survival and hazard functions are determined for each of the real and fuzzy data.

2. Methodology

A. Weibull Distribution

(1) 2-Parameter Weibull Distribution

Survival and Reliability functions are one of the most important probability distributions, this distribution was first found by the Swedish physicist Walddi Weibull in 1939 [6].
In this section, we will provide the probability density function(pdf) of Weibull distribution, which is as follows:

$$f(t,\sigma,\kappa) = \begin{cases} \frac{\sigma}{\kappa}(\frac{t}{\kappa})^{\sigma-1} e^{-\left(\frac{t}{\kappa}\right)^{\sigma}}, & t > 0 \\ 0, & \text{o.w} \end{cases} \quad (1)$$

σ(shape), κ(scale) are parameters > 0
The Cumulative Distribution Function (CDF) is:

$$F_T(t,\sigma,\kappa) = P(T \le t) = \int_0^t f(u)du = 1 - e^{-\left(\frac{t}{\kappa}\right)^{\sigma}} \quad (2)$$

Where S(t) is the survival function, which defined by:

$$S(t) = 1 - F(t) = e^{-\left(\frac{t}{\kappa}\right)^{\sigma}} \quad (3)$$

The hazard function of Weibull distribution is:

$$h(t) = \lambda = \frac{f(t)}{S(t)} = \frac{\sigma}{\kappa}(\frac{t}{\kappa})^{\sigma-1} \quad (4)$$

Mean and Variance of the distribution are given respectively:

$$E(t) = \mu = \kappa\Gamma\left(1 + \frac{1}{\sigma}\right)$$

$$var(t) = \sigma^2 = \kappa^2\left[\Gamma\left(1 + \frac{2}{\sigma}\right) - \left(\Gamma(1 + \frac{1}{\sigma})\right)^2\right]$$

(2) 3-Parameter Weibull Distribution

Let T be a random variable represents time to failure with probability density function (t), where (t) is the pdf of the 3-parameter Weibull distribution [7]:

$$f(t,\sigma,\kappa,\zeta) = \begin{cases} \frac{\sigma}{\kappa}(\frac{t-\zeta}{\kappa})^{\sigma-1} e^{-\left(\frac{t-\zeta}{\kappa}\right)^{\sigma}}, & t > 0 \\ 0, & t \le 0 \end{cases} \quad (5)$$

where σ(shape), κ(scale) are parameters > 0 and ζ(location) $\le t$.
The (CDF) is:

$$F_T(t,\sigma,\kappa,\zeta) = P(T \le t) = \int_0^t f(u)du = 1 - e^{-\left(\frac{t-\zeta}{\kappa}\right)^{\sigma}} \quad (6)$$

Where S(t) is given by:

$$S(t) = 1 - F(t) = e^{-\left(\frac{t-\zeta}{\kappa}\right)^{\sigma}} \quad (7)$$

The h(t) of the distribution is:

$$h(t) = \lambda = \frac{f(t)}{S(t)} = \frac{\sigma}{\kappa}(\frac{t-\zeta}{\kappa})^{\sigma-1} \quad (8)$$

The mean (μ) and variance (σ^2) of the distribution are given respectively:

$$E(t) = \mu = \zeta + \kappa\Gamma\left(1 + \frac{1}{\sigma}\right)$$

$$var(t) = \sigma^2 = \kappa^2\left[\Gamma\left(1 + \frac{2}{\sigma}\right) - \left(\Gamma(1 + \frac{1}{\sigma})\right)^2\right]$$

B. Estimation Method

Maximum Likelihood Estimation (MLE)

The most crucial way for estimating the parameters in any probability function is the maximum likelihood approach. This method's concept is aimed at discovering the parameter value that maximizes the likelihood function for any distribution. In this section we have separately estimated (2 and 3) parameters of Weibull Distribution.

(1) MLE for 2-Parameter Weibull distribution

The 2-weibull distribution's parameters likelihood function is:

$$f(t;\sigma,\kappa)=\frac{\sigma}{\kappa}\left(\frac{t}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t}{\kappa}\right)^{\sigma}} \quad ,t\geq 0$$

$$\prod_{i=1}^{n} f(t;\sigma,\kappa)=\frac{\sigma}{\kappa}\left(\frac{t_1}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t_1}{\kappa}\right)^{\sigma}} * \frac{\sigma}{\kappa}\left(\frac{t_2}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t_2}{\kappa}\right)^{\sigma}} * \dots * \frac{\sigma}{\kappa}\left(\frac{t_n}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t_n}{\kappa}\right)^{\sigma}}$$

$$Lf(\sigma,\kappa;t_1,t_2,\dots,t_n)=\prod_{i=1}^{n}\frac{\sigma}{\kappa}\left(\frac{t_i}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t_i}{\kappa}\right)^{\sigma}}$$

$$\ln Lf(\sigma,\kappa;t_1,t_2,\dots,t_n)=n\ln\left(\frac{\sigma}{\kappa}\right)+(\sigma-1)\prod\ln\left(\frac{t_i}{\kappa}\right)-\sum\left(\frac{t_i}{\kappa}\right)^{\sigma}$$

$$=n\ln(\sigma)-n\ln(\kappa)+\sigma\sum\ln\left(\frac{t_i}{\kappa}\right)-\sum\ln\left(\frac{t_i}{\kappa}\right)-\sum\left(\frac{t_i}{\kappa}\right)^{\sigma}$$

$$=n\ln(\sigma)-n\ln(\kappa)+\sigma\sum(\ln(t_i)-\ln(\kappa)-\sum\left(\frac{t_i}{\kappa}\right)^{\sigma} \qquad (9)$$

differentiating eq(9) w.r.to σ:

$$\frac{\partial \ln L}{\partial\sigma}=\frac{n}{\sigma}+\sum\ln\left(\frac{t_i}{\kappa}\right)-\sum\left(\frac{t_i}{\kappa}\right)^{\sigma}\ln\left(\frac{t_i}{\kappa}\right) \quad \Rightarrow \frac{\partial \ln L}{\partial\sigma}=0$$

$$\frac{n}{\sigma}+\sum\ln\left(\frac{t_i}{\kappa}\right)-\sum\left(\frac{t_i}{\kappa}\right)^{\sigma}\ln\left(\frac{t_i}{\kappa}\right)=0 \qquad (10)$$

differentiating eq(6) w.r.to κ:

$$\frac{\partial \ln L}{\partial\kappa}=-\frac{n}{\kappa}+\sigma\sum\frac{1}{\kappa}+\sum\frac{1}{\kappa}-\sum t_i^{\sigma}\left(-\sigma\kappa^{-(\sigma+1)}\right)=0$$

$$-\frac{n}{\kappa}-\frac{n\sigma}{\kappa}+\frac{n}{\kappa}+\sigma\sum\frac{t_i^{\sigma}}{\kappa^{(\sigma+1)}}=0$$

$$-\frac{n\sigma}{\kappa}+\frac{\sigma}{\kappa}\sum\left(\frac{t_i}{\kappa}\right)^{\sigma}=0$$

$$n=\sum\left(\frac{t_i}{\hat\kappa}\right)^{\sigma} \quad \Rightarrow \hat\kappa^{\sigma}=\sum\frac{t_i^{\sigma}}{n}$$

$$\hat\kappa=\left(\sum\frac{t_i^{\sigma}}{n}\right)^{\frac{1}{\sigma}} \qquad (11)$$

Since it is difficult to separate σ and κ from equations (10)&(11) so, we will use iterative methods like Newton-Raphson-Method, to obtain the parameters of (σ,κ):

$$t_{i+1}=t_i-\frac{f(t_i)}{f'(t_i)}$$

consider $f(t_i)$ is $f(\hat\sigma)$

$$f(\hat\sigma)=\frac{n}{\hat\sigma}+\sum\ln\left(\frac{t_i}{\kappa}\right)-\sum\left(\frac{t_i}{\kappa}\right)^{\hat\sigma}\ln\left(\frac{t_i}{\kappa}\right)$$

And, $f'(t_i)$ is $f'(\hat\sigma)$

$$f(\hat\sigma)=\frac{\partial f(\hat\sigma)}{\partial\hat\sigma}=-\frac{n}{\hat\sigma^2}-\sum\left(\frac{t_i}{\kappa}\right)^{\hat\sigma}\ln\left(\frac{t_i}{\kappa}\right)\ln\left(\frac{t_i}{\kappa}\right)$$

Through the numerical solution we can obtain the parameters.

(2) MLE for 3-parameter Weibull distribution

The 3-weibull distribution's parameters likelihood function is:

$$Lf(\sigma,\kappa,\zeta;t_1,t_2,\dots,t_n)=\prod_{i=1}^{n}\frac{\sigma}{\kappa}\left(\frac{t-\zeta}{\kappa}\right)^{\sigma-1}e^{-\left(\frac{t-\zeta}{\kappa}\right)^{\sigma}} \qquad (12)$$

Taking logarithms of both sides eq(12):

$$\ln L=n\ln\left(\frac{\sigma}{\kappa}\right)+(\sigma-1)\prod\ln\left(\frac{t_i-\zeta}{\kappa}\right)-\sum\left(\frac{t_i-\zeta}{\kappa}\right)^{\sigma}$$

$$=n\ln(\sigma)-n\ln(\kappa)+\sigma\sum\ln\left(\frac{t_i-\zeta}{\kappa}\right)-\sum\ln\left(\frac{t_i-\zeta}{\kappa}\right)-\sum\left(\frac{t_i-\zeta}{\kappa}\right)^{\sigma} \qquad (13)$$

taking the partial derivatives of equation (13) for σ,κ and ζ , we obtain the following estimated equations:

$$\frac{\partial \ln L}{\partial\sigma}=\frac{n}{\sigma}+\sum\ln\left(\frac{t_i-\zeta}{\kappa}\right)-\sum\left(\frac{t_i-\zeta}{\kappa}\right)^{\sigma}\ln\left(\frac{t_i-\zeta}{\kappa}\right)=0 \qquad (14)$$

$$\frac{\partial \ln L}{\partial\kappa}=-\frac{n\sigma}{\kappa}+\sigma\frac{\sum_{i=1}^{n}(t_i-\zeta)^{\sigma}}{\kappa^{\sigma+1}}=0 \qquad (15)$$

$$\frac{\partial \ln L}{\partial\zeta}=-(\sigma-1)\sum_{i=1}^{n}\frac{1}{t_i-\zeta}+\frac{\sigma}{\kappa}\sum_{i=1}^{n}(t_i-\zeta)^{\kappa-1}=0 \qquad (16)$$

It can be noted that the equations (13),(14),(15) are nonlinear, so Numerical solutions can be used to get the values of the parameters.

3. Evaluation metrics

(A) Mean Squared Error (MSE)

$$MSE(\widehat{R(t_i)})=\sum_{i=1}^{n}\frac{\left[\widehat{R(t_i)}-R(t_i)\right]^2}{n} \qquad (17)$$

Where $\widehat{R(t_i)}$ is estimated survival function, $R(t_i)$ is empirical survival s.t. $R(t_i)=\frac{i-0.5}{n}$

(1) Akaike's Information Criterion (AIC)

The most practical methods widely used in statistical modeling is the Akaike information criterion (AIC). In 1973, Hirotugu Akaike developed the technique as an extension of MLE principle. AIC was the first criteria for selecting models to become widely accepted.

$$AIC=-2\ln f(y|\hat{\theta})+2k \qquad (18)$$

Where $f(y|\hat{\theta})$ is goodness of fit (ML), k number of parameters.

(2) Bayesian information criterion (BIC)

The BIC was found by Gideon E. Schwarz and published in 1978 a paper.

$$BIC=-2\ln f(y|\hat{\theta})+2k\ \ln n \qquad (19)$$

4. Fuzzy set theory

Realistic and exciting life requires the creation of fuzzy sets, which are an extension of crisp sets in traditional set theory. A fuzzy set is a mathematical construct that generalizes the classical notion of a set. Unlike classical sets, which assign a binary membership (either 0 or 1) to an element with respect to the set, fuzzy sets allow for partial membership. This indicates that an element can belong to a fuzzy set to a certain degree between 0 and 1. [8]

This section includes some basic mathematical operations described in fuzzy sets, as well as some basic definitions and basic concepts connected to fuzzy set theory. To provide a comparison with fuzzy sets and to explain why fuzzy sets were introduced, we begin this section with defining ordinary or non-fuzzy sets. [9]

Definition (1)[12]

Let X be a universal set of classical objects, whose components are represented by x. A characteristic function χ from X into {0, 1} is often used to describe membership in a classical subset S of X, such that:

$$\chi(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases} \qquad (20)$$

The set {0, 1} is called a valuation set.
The symbol "~" is used to distinguish fuzzy sets from non-fuzzy sets, i.e. fuzzy sets will be abbreviated as S and non-fuzzy sets are denoted by $\tilde{S}$.

Definition (2)[12]

Let X be the universal set and $\tilde{S}$ be any subset of X, then $\tilde{S}$ is called fuzzy subset of X, which is characterized by a membership function $\mu_{\tilde{S}}: X \longrightarrow [a, b]$, where $a, b \in \mathbb{R}$ and in special case $\mu_{\tilde{S}}: X \longrightarrow [0,1]$, i.e.,

$$\tilde{S} = \{(x, \mu_{\tilde{S}}(x)) | x \in X, 0 \leq \mu_{\tilde{S}}(x) \leq 1\} \qquad (21)$$

Definition (3)[12]

α-Cuts or α-Sets

α-Cut sets can be considered as a transitional set that bridges the gap between normal and fuzzy sets. S_α represents the ordinary set of these elements, which are known as the α-cut sets of $\tilde{S}$.

$$S_\alpha = \{x \in X: \mu_{\tilde{S}}(x) \geq \alpha, \alpha \in (0,1]\}$$

And the strong α-cut set:

$$S_{\alpha+} = \{x \in X: \mu_{\tilde{S}}(x) > \alpha, \alpha \in (0,1]\}$$

Definition (4)[10]

The trapezoidal membership function is specified by four parameters $\tilde{S} = (a, b, c, d)$ which is given by :

$$\mu_{\tilde{S}}(x) = \begin{cases} \dfrac{x-a}{b-a} & a \leq x < b \\ 1 & b \leq x < c \\ \dfrac{x-d}{c-d} & c \leq x \leq d \\ 0 & \text{o.w.} \end{cases}$$

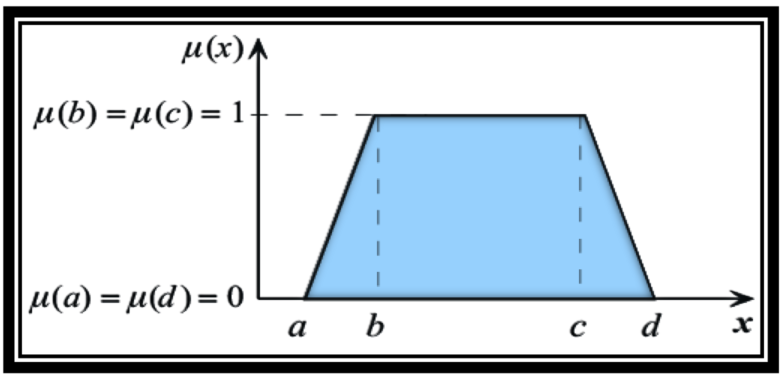

Figure (1): Graph of Trapezoidal membership function.

A. The procedure of fuzzy logic

The procedure of fuzzy logics can be divided into four main parts:

(1) Fuzzification: It is used to transform inputs i.e. crisp numbers into fuzzy sets. In this step we need to define linguistic variables (e.g., "survival time" as low, medium, high) and then choosing an appropriate membership functions (e.g., triangular, trapezoidal, Gaussian) to show the degree of membership for each fuzzy set. There are some basic methods used in this stage:

(a) Inference
(b) Order Ranking
(c) Intuition
(d) Neural Networks

(2) Rule base: Creation set of IF-THEN rules to define the relationship between input and output.

(3) Inference Engine: Since the Inference Engine processes all of the data, it is a crucial part of any fuzzy logic system (FLS). It enables users to determine the degree of correspondence between the rules and the current fuzzy input.

(4) A module or component called defuzzification converts the fuzzy set inputs produced by the inference engine into a distinct value. It is the final stage of a fuzzy logic system's operation.

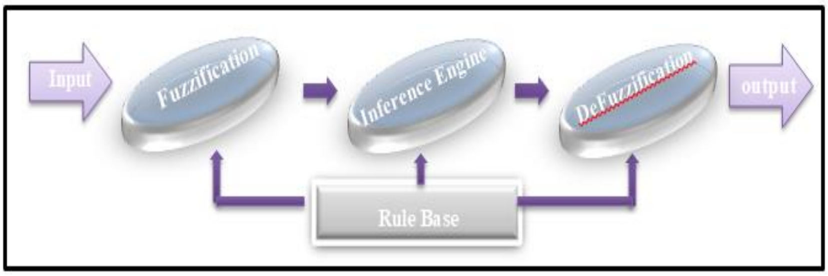

Figure (2): The process of fuzzy logics.

B. Techniques used in the defuzzification process [4]

(1) Centre of Area Method:

Center of Area (COA) method which is also known as Center of gravity. This method was created by Sugeno in 1985. The most popular approach is this one. This method's computational difficulty for complex membership functions is its sole disadvantage Of all the known defuzzification techniques, the centroid method's algebraic formulation is the most aesthetically beautiful. [11,12]

$$x_{COA} = \frac{\int \mu_S(x) x dx}{\int \mu_S(x) dx} \qquad (22)$$

Where x_{COA} is the crisp output, $\mu_S(x)$ is the aggregated membership function and the output variable is x.

For more clarification of the method, we solved an example:

Let t=5 and fuzzy set= [3, 5, 7, 20] where the trapezoidal membership function $\mu_S(x)$ is defined by :

$$\mu_S(x) = \begin{cases} \frac{x-3}{5-3} & \text{for } 3 \le x \le 5 \\ 1 & \text{for } 5 \le x \le 7 \\ \frac{20-x}{20-7} & \text{for } 7 \le x \le 20 \\ 0 & \text{for } x < 3 \text{ or } x > 20 \end{cases}$$

$$x_{COA} = \frac{\int_3^{20} \mu_S(x) x dx}{\int_3^{20} \mu_S(x) dx} = \frac{\int_3^5 \left(\frac{x-3}{2}\right) x dx + \int_5^7 x dx + \int_7^{20} \left(\frac{20-x}{13}\right) x dx}{\int_3^5 \left(\frac{x-3}{2}\right) dx + \int_5^7 dx + \int_7^{20} \left(\frac{20-x}{13}\right) dx} \cong 9.47$$

According to centroid method the crisp value for the fuzzifed set [3, 5, 7, 20] is 9.49.

(2)Bisector Method:

The area under the curve where area on both sides are equal, is calculated using this approach. The action that divides the area into two areas of the same size is generated by the BOA.

$$\int_a^{x^*} \mu_S(x)dx = \int_{x^*}^{d} \mu_S(x)dx \qquad (23)$$

(3)Mean of Maximum Method:

The greatest membership values of an element in this procedure is the defuzzified value. The mean value of the maxima is determined when multiple elements have maximum membership values.

$$x^* = \frac{\sum x_i \in \mu_S(x_i)}{|M|} \qquad (24)$$

(4)Largest of Maximum method:

gives the maximum value of the domain with largest membership value.

$$\mu_{max} = \max_x \mu_S\,(x) \qquad (25)$$

Where $\mu_S(x)$ is the membership function value at any point x, and μ_{max} is the maximum membership value in the fuzzy set.

(5)Smallest of Maximum method:

This technique determines the smallest value of the domain with maximum membership value.

$$x_{SoM} = \min\,(X_{\mu_{max}}) \qquad (26)$$

This is the defuzzied value using the SoM method.

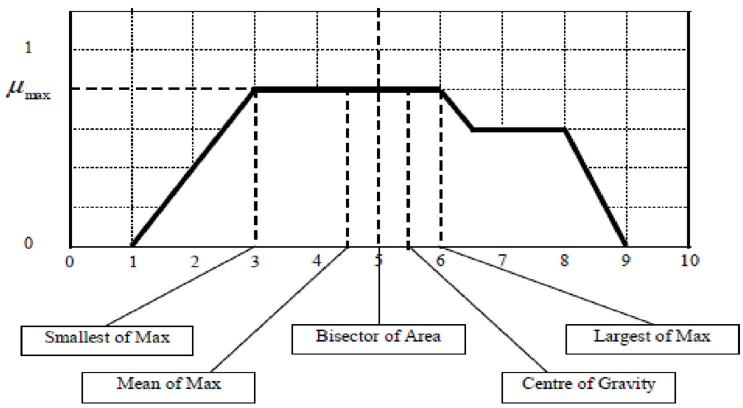

Figure (3): Defuzzification methods

5. Evaluation of Data and Interpretation of Outcomes

The analysis in this study was conducted using Python, Easy Fit, and Stata software. The data was collected from Suleimani Centre for Heart Disease in Sulaymaniyah City. The study period spanned from February 1, 2024, to October 15, 2024, during which a total of 127 patients were included. The survival times of these patients were recorded in **hours**.

T=[5,10,11,11,12,13,…,346,350,359,367,432,480]

A. Results

(1) Fit Assessment of Raw Data to Various Distributions

To know which distribution is best fits the actual data, we evaluate the measures of Log-likelihood, (AIC) and (BIC):

Table 1: Evaluating Model Performance Using AIC and BIC

Distribution	Loglikelihood	AIC	BIC
Exponential 1-Parameter	-740.7307	1483.4614	1486.305587
Exponential 2-Parameters	-14709.50313	29423.006	29428.695
Gamma 2-Parameters	-736.5284	1477.0568	1482.745174
Gamma 3-Parameters	-734.657	1475.314	1483.847
Lognormal 2-Parameters	-741.7201	1487.4402	1493.129
Lognormal 3-parameters	-740.4305	1486.861	1495.394
Weibull 2-Parameters	-736.3229	1476.6458	1482.3342
Weibull 3-Parameters	-734.1967	1474.3934	1482.925961

The table above shows that the 3-parameter Weibull distribution has the lowest loglikelihood and AIC values, making it the best-fitting model for the data. Therefore, the optimal model based on these criteria is the 3-parameter Weibull distribution.

(2) Methods of Defuzzification process

From the diagram of fuzzy logic process, first the survival times classified into three trapezoidal membership functions: Low, Medium, and High. From The definition of trapezoidal function for each survival time [a, b, c, d] assigns trapezoidal parameters based on its value, ensuring proper ordering. For survival times up to 20, the fuzzy set is classified as Low; between 20 and 100, it is Medium; and beyond 100, it is High. Then, actual data was transformed into fuzzy data using the five defuzzification techniques as shown in table (3):

Table 2: Computational defuzzification methods

Raw data	Centroid	Bisector	Mean of Maximum	Smallest of Maximum	Largest of Maximum
5	10.02369	9.418812	5.475475	5.475475	5.950951
10	12.79619	12.379537	9.992492	9.992492	10.70571
11	13.44309	13.071121	10.94344	10.94344	11.65666
11	13.44309	13.071121	10.94344	10.94344	11.65666
12	14.0922	13.766896	11.89439	11.89439	12.60761
			. .		
350	268.4966	278.7166	349.9575	349.9575	352.5726
359	274.5203	285.1057	358.9915	358.9915	361.6066
367	279.9091	290.8214	367.0746	367.0746	369.6897
432	323.3423	336.8896	431.977	431.977	434.8298
480	353.3254	368.6921	478.5736	478.5736	480

The table provides a comparative view of different defuzzification techniques, showing how each method produces different crisp values from the fuzzy set based on the given actual data. By comparing these values, one can assess the sensitivity of each defuzzification method and decide which one best represents the underlying fuzzy system or the specific requirements of the application.

(3) Comparing Defuzzification Methods Using Mean Squared Error (MSE)

Since for the data of each method has been changed, so their distributions found by using (easy fit) program which are shown below:

Table 3: Best-Fitting Distribution Models for Defuzzification Methods

Methods	Centroid	Bisector	Mean of Maximum	Smallest of Maximum	Largest of Maximum
Best fit distribution	3-Parameter Weibull	3-Parameter Weibull	2-Parameter Weibull	2-Parameter Weibull	2-Parameter Weibull

The table is showing the relationship between different defuzzification methods and their respective best-fit distributions. It suggests that the **Centroid** and **Bisector** methods work better with the **3-parameter Weibull distribution**, while the **Mean of Maximum, Smallest of Maximum**, and **Largest of Maximum** methods align more with the **2-parameter Weibull distribution.**

(4) Evaluating Defuzzification Techniques: A Comparison Based on MSE

Determine which defuzzification technique produces the least error (lowest MSE) and is therefore the most reliable or effective method for a given application, dataset, or fuzzy system. This comparison helps in choosing the best-suited defuzzification method for real-world scenarios, based on objective performance metrics.

Table 4: Evaluating Defuzzification Methods Through MSE

Methods	Centroid	Bisector	Mean of Maximum	Smallest of Maximum	Largest of Maximum
MSE	0.33410645	0.335393791	0.343493885	0.341725019	0.344167389

Since from table (4), we see that the smallest mean square error of the methods is centroid method, therefore we later use the fuzzy numbers of the method to analyse the probability density function and survival function, hazard function.

(5) Assessing the Differences Between Raw and Fuzzy Data

Since the fuzzy data fitted with 3- Parameter Weibull distribution and the classical data also fitted with 3-Parameters Weibull distribution. In this section we showed comparison based on f(t), s(t) and h(t), for this reason the parameters of both distributions was estimated.

Table 5: Maximum Likelihood Estimation for Parameter Estimation

Parameters of Weibull distribution	Raw data	Fuzzy data
σ	1.1396	1.2935
κ	126.33	108.24
ζ	4.7852	9.1309

<u>**Interpretation of the table**</u>:

- **Raw Data**: The 3-Parameter Weibull distribution describes data where failures start at time zero, with a shape parameter of 1.1396 and a characteristic life (κ) of 126.33 and location parameter (ζ) of 4.7852.
- **Fuzzy Data**: The 3-Parameter Weibull distribution introduces the additional factor of a location parameter ($\zeta = 9.1309$), which suggests that failures only occur after a certain point in time. The scale parameter ($\kappa = 108.24$) is lower, meaning the characteristic life is shorter than in the raw data case, and the shape parameter ($\sigma = 1.2935$) suggests a slightly steeper increase in the failure rate.

(a) Calculating Reliability Metrics: Probability Density $f(t)$, Survival $s(t)$, and Hazard $h(t)$ Functions in Weibull Distribution for fuzzy data and raw data

Calculating Reliability Metrics: Probability Density f(t)**, Survival** s(t)**, and Hazard** h(t)**Functions in Weibull Distribution** refers to the process of computing three key reliability functions used in survival analysis and reliability engineering. These functions—**Probability Density Function (PDF)**, **Survival Function**, and **Hazard Function**—are essential in describing the behavior of systems or components over time in terms of their failure characteristics.

1) Raw data

Table 6: Calculation of $f(t)$, $s(t)$, and $h(t)$ for Raw Data Analysis

T	f(t)	s(t)	h(t)
5	0.003701046	0.999302157	0.00370363
10	0.006165764	0.973893086	0.006331048
11	0.006211876	0.968208133	0.006415848
11	0.006211876	0.968208133	0.006415848
12	0.006250252	0.962427834	0.006494255
13	0.006282012	0.956569789	0.006567229
13	0.006282012	0.956569789	0.006567229
15	0.006329061	0.944672529	0.00669974
..			
.			
337	0.000510046	0.049303375	0.010345054
346	0.00046642	0.044920685	0.010383187
350	0.000448184	0.043095238	0.010399861
359	0.000409592	0.039245026	0.010436787
367	0.000377953	0.036102284	0.010468947
432	0.000194454	0.018156288	0.010709993
480	0.000117651	0.010824738	0.010868683

2) Fuzzy data

Table 7: Calculation of $f(t)$, $s(t)$, and $h(t)$ for Fuzzy Data Analysis

T	f(t)	s(t)	h(t)
10.02369	0.003610138	0.975144072	0.003702158
12.79619	0.004010127	0.964535484	0.004157573
13.44309	0.004092875	0.961942357	0.004254802
13.44309	0.004092875	0.961942357	0.004254802
14.0922	0.004173805	0.959255521	0.004351088
14.74415	0.004251779	0.956517051	0.004445063
14.74415	0.004251779	0.956517051	0.004445063
16.06099	0.004401668	0.950804573	0.004629415
.			
259.9333	0.000816987	0.047185917	0.017314215
265.9571	0.000743607	0.04248365	0.017503375
268.4966	0.00071441	0.040632178	0.017582379
274.5203	0.000649081	0.03653075	0.017768067
279.9091	0.000594997	0.033179777	0.017932513
323.3423	0.000284347	0.014809092	0.019200851
353.3254	0.000164668	0.008223156	0.020024965

<u>Interpret the result:</u>

Both tables (6) and (7) provide an in-depth look at how the system's reliability changes over time. The PDF f(t), Survival Function s(t), and Hazard Function h(t) together offer valuable insights into the failure dynamics of the system for both fuzzy and raw data:

- Early Times: High survival probability, low failure rate, and low hazard rate.
- Later Times: Decreasing survival probability, increasing failure rate, and increasing hazard rate as the system ages and becomes more likely to fail.

<u>At T=480 for (raw data) and T=353.3254 for (fuzzy data):</u>

- For raw data $f(t) = 0.000117651$ while for fuzzy data $f(t) = 0.000164668$: In both the probability density is very low, meaning failure is unlikely at this specific time.
- For raw data $s(t) = 0.010824738$while for fuzzy data $s(t) = 0.008223156$: The survival probability is extremely low (1.08% and 0.8%, respectively), indicating that the system will almost certainly fail before reaching this point in both.
- $h(t) = 0.010868683$ (real data), $h(t) = 0.020024965$(fuzzy data): The hazard rate for both data sets is the highest in the tables (6,7), indicating the system's failure rate is at its peak at this time.

(b) Comparative Analysis of Crisp and Fuzzy Data in Statistical Modeling

The fuzzy data's distribution was determined to be three parameters of the Weibull distribution after the fuzzy logic method was applied to the classical data. Based on the data's computed mean squares error, the fuzzy data's MSE is 0.332007034, which can be used to determine which data is accurate and appropriate to study.

Table 8: Assessing the Accuracy of Crisp and Fuzzy Data with MSE

Algorithm	Parameters			MSE
	σ	κ	ζ	
Crisp data (3- Parameters Weibull distribution)	1.1396	126.33	4.7852	0.340958536
Fuzzy data (3-Parameters Weibull distribution)	1.2935	108.24	9.1309	0.33410645

The MSE for fuzzy data **(**0.33410645**)** is slightly smaller than the MSE for crisp data **(**0.340958536**),** this implies that a slightly better fit to the observed data is offered by the fuzzy data technique. The fuzzy model is more appropriate for situations when imprecision is present in the data since it can manage uncertainty and variability better than the crisp approach, which may account for its increased accuracy. The figure below shows the difference between the crisp data and the fuzzy data.

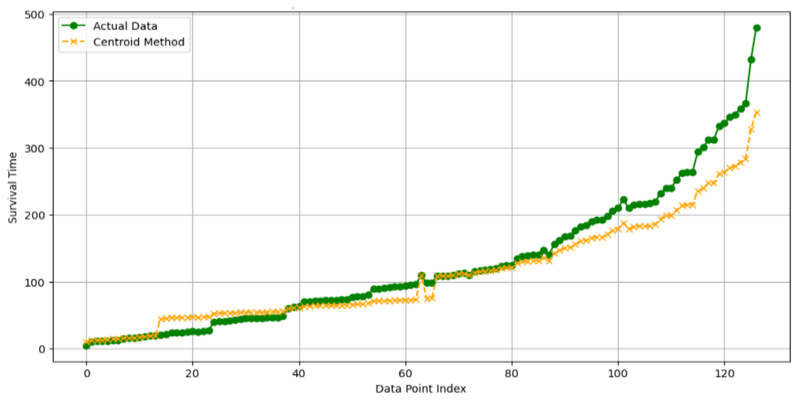

Figure 4: Comparison between the crisp and fuzzy data

6. Conclusions and Recommendations

A. Conclusions

Based on the analytical part of the study, we got a number of conclusions, as follows:
1. This research presents a fuzzy logic method for changing survival time data for patients with heart illnesses. By using defuzzification techniques and MSE to evaluate survival and hazard rates with Weibull distribution parameters.
2. MSE was used to determine the differences between real and fuzzy data, revealing that fuzzy data provided the best accurate findings for survival and hazard rate assessments.
3. The study found an inverse association between failure times and survival function for heart disease patients, with predicted survival function values decreasing as failure times increased, this is existed to both actual and fuzzy data.

B. Recommendations

Based on the conclusions, we reached a set of recommendations, as follows:

1. Healthcare workers can use fuzzy logic methodologies to investigate survival times and hazard rates since they are both interpretable and accurate.
2. Patients with heart disease can benefit from individualized risk assessment and treatment planning when fuzzy survival models are included in healthcare systems.
3. Future studies should examine the robustness of this method in other chronic conditions, such as diabetes or cancer, to demonstrate its effectiveness across multiple domains.
4. Researchers might explore additional parametric models in different disciplines for further investigation.

References

1- Algamal, Z. Y., & Basheer, G. (2021). Reliability estimation of three parameters Weibull distribution based on particle swarm optimization. *Pakistan Journal of Statistics and Operation Research*, 35-42.
2- Carroll, K. J. (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled clinical trials*, *24*(6), 682-701.
3- Hadi-Vencheh, A., & Allame, M. (2010). On the relation between a fuzzy number and its centroid. *Computers & Mathematics with Applications*, *59*(11), 3578-3582.
4- Jain, D., Sharma, S. K., & Dhiman, P. (2022). Comparative Analysis of Defuzzification Techniques for Fuzzy Output. *Journal of algebraic statistics*, *13*(2), 874-882.
5- P. Dhiman and A. Kumar. (2018). “Defuzzification technique for fuzzy output,” *International Journal of Research and Analytical Reviews*, [Online]. Available: www.ijrar.org]
6- Pujaru, K., Adak, S., Kar, T. K., Patra, S., & Jana, S. (2024). A Mamdani fuzzy inference system with trapezoidal membership functions for investigating fishery production. *Decision Analytics Journal*, *11*, 100481.
7- Roohanizadeh, Z., Baloui Jamkhaneh, E., & Deiri, E. (2022). Parameters and reliability estimation for the Weibull distribution based on intuitionistic fuzzy lifetime data. *Complex & Intelligent Systems*, *8*(6), 4881-4896.
8- Sugeno, M. (1985). An introductory survey of fuzzy control. *Information sciences*, *36*(1-2), 59-83.
9- Weibull, W. (1939). A statistical theory of strength of materials. *IVB-Handl.*
10- Yager, R. R. (1981). A procedure for ordering fuzzy subsets of the unit interval. *Information sciences*, *24*(2), 143-161.
11- Zhu, H. P., Xia, X., Yu, C. H., Adnan, A., Liu, S. F., & Du, Y. K. (2011). Application of Weibull model for survival of patients with gastric cancer. *BMC gastroenterology*, *11*, 1-6.
12- Zimmermann, H. J. (2010). Fuzzy set theory. *Wiley interdisciplinary reviews: computational statistics*, *2*(3), 317-332.

Appendix

The following code shows the process of Defuzzification methods in Python.

```
for point in data:
    # Fuzzify the data point
    a, b, c, d = fuzzify_trapezoidal(point)

    # Generate membership function
    mfx = fuzz.trapmf(x, [a, b, c, d])

    # Apply defuzzification methods
    defuzz_centroid = fuzz.defuzz(x, mfx, 'centroid')
    defuzz_bisector = fuzz.defuzz(x, mfx, 'bisector')
    defuzz_mom = fuzz.defuzz(x, mfx, 'mom')
    defuzz_som = fuzz.defuzz(x, mfx, 'som')
    defuzz_lom = fuzz.defuzz(x, mfx, 'lom')

    # Store results
    results.append({
        "data_point": point,
        "centroid": defuzz_centroid,
        "bisector": defuzz_bisector,
        "mom": defuzz_mom,
        "som": defuzz_som,
        "lom": defuzz_lom,
    })
```