International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 11 Issue 12 December 2023, Page no. – 3894-3898

Index Copernicus ICV: 57.55, Impact Factor: 7.362

DOI: 10.47191/ijmcr/v11i12.02

The Comparison Duration Diabetes Survival Times Modelling Using Lindley (LIN), Weighted Lindley Exponential (WLE), Power Modified Lindley (PML), Lindley Half-Cauchy (LHC) and Rayleight Lomax (RL) Distributions

Sutriana[1], Rado Yendra[2], Rahmadeni[3], Muhammad Marizal[4]

[1,2,3,4]Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau

ARTICLE INFO

Published Online: 06 December 2023

Corresponding Author: **Sutriana**

ABSTRACT

Accurate diabetes survival time modeling is critical in estimating diabetes survival time potential for effectively. One of the bases for assessment of diabetes survival times potential for a specified region is the probability distribution of diabetes survival times, therefore diabetes survival times data is needed to produce statistical modeling, especially in determining the best probability distribution. Statistical models are designed to facilitate conclusions about the occurrence probability distribution of diabetic patient in the Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province. For this purpose, five distributions will be used and tested to determine the best model to describe diabetes survival times. The main goal of this study is to find the best fitting distribution to the duration survival times diabetes of 50 patients measured over Bengkalis region by using lindley (LIN) distribution, three modified lindley distributions such as weighted Lindley exponential (WLE), Power Modified Lindley (PML), Lindley half-Cauchy (LHC) and Rayleight Lomax distribution (RL). The maximum likelihood method will be used to get the estimated parameter value from the distribution used in this study. Furthermore, the graphical inspection (density-density plot and cumulative plot) and numerical criteria (Akaike's information criterion (AIC), Bayesian Information Criteria (BIC), - log likelihood (- l) were used to determine the best fit model. In most cases, the results produced by the graphical inspection were similar, and differed from the numerical criteria . The best fit result was chosen as the distribution with the lowest values of AIC, BIC and - l. In general, the Rayleight Lomax (RL) distribution has been selected as the best model.

KEYWORDS: lindley, weighted Lindley exponential, Power Modified Lindley, Lindley half-Cauchy, Rayleight Lomax, diabetes survival times

I. INTRODUCTION

Diabetes is a very serious disease for humans even though it is not contagious. The number of patients suffering from diabetes continues to increase indirectly due to changes in passive lifestyles, unhealthy eating patterns, smoking habits and high levels of stress. Globally, the number of diabetes sufferers increases every year. Based on 2019 data from the International Diabetes Federation, diabetes sufferers worldwide reached 9.3% in 2019 and are expected to increase to 10.2% in 2030 and 10.9% in 2045. Indonesia ranks 7th in the world in the number of diabetes patients in 2019 [4]. Based on data released by the Minister of Health of the Republic of Indonesia, Riau Province was in the top ranking with a very significant increase of 358.3% [5], while according to data from the Bengkalis District Health Service (Diskes), diabetes sufferers in 2019 were 10 .57% [6]. The thing that diabetes sufferers most want to know is how big their chances of survival are or how dangerous diabetes is in terms of causing the risk of death, so data analysis is needed to find out the problem. The analysis used is called survival analysis. Survival analysis requires data in the form of an individual's survival time which is usually measured in days, weeks, months and years [1, 2, 3]. In survival analysis, there is a model that can be used to analyze survival data, namely the parametric model. Parametric models are analyzes based on data distribution. Parametric models assume that the underlying distribution of survival times follows a certain distribution, for example Weibull, exponential, log-normal, Rayleigh distribution. Moreover, data on the survival time of diabetics is very important to be investigated, especially concerning the determination of the probability of death. This survival time analysis is carried out using statistical techniques, especially to determine the best probability model that can describe the pattern of survival time. Previous research has been conducted to determine the best probability model for data on the survival time of diabetic patients. Ummu et al [7] estimated the duration diabetes survival time using the Weibull, Gamma and Log Normal distribution and the results showed that the Weibull model was the best in approaching the given observation data. This was also supported by numerical models such as AIC and BIC, by giving the smallest values for the two numerical methods compared to other probability models. Furthermore, Manda Lisa Usvita et al. [8] compare the three kinds of distribution, namely Exponential (E), Weibull (W), and Rayleigh-Lomax (RL) were applied to survival times of diabetes patients. Method of Moments was used to obtain the estimated parameter. Based on the smallest Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, and graphical inspection (probability density function (pdf)) to survival times of diabetes patients, the study has shown that RL is the best fit distribution in modeling survival times for diabetes patients in the Mandau RSUD, Bengkalis Regency, Riau Province. Gurpit et al. [9] discussed the estimation of survival function in diabetic nephropathy patients with exponential, gamma, Weibull, log-normal, inverse Gaussian, and Rayleigh distributions, where the gamma distribution was discovered as the best. Marvasti et al. [10] compared the Cox and the parametric models to analyze the effective time factor of occurrence in patients with type 2 nephropathy using the lognormal distribution. The results showed that the log-normal distribution was suitable for this case. Based on this description, the distributions that are commonly used in survival analysis for diabetic patient data with parametric models are Weibull, exponential, gamma, Rayleigh, and log-normal distributions. Fatima et al. [11] conducted research related to the introduction of developing a new distribution, the Rayleigh-Lomax distribution, and applied this distribution to the survival data. The data used is data on aircraft windshield damage, glass fiber resistance data, and carbon fiber tension data. Their research shows that the Rayleigh-Lomax distribution is suitable for survival data. The selection of the best fitting distribution is always a main interest in study of survival times analysis. Therefore, in this study, we would like to find the best fiting distribution for duration (survival times) diabetes patient based on several goodness of fit criteria. In this study, preliminary study on duration survival times diabetes of 50 patients measured over Bengkalis region was evaluated. The objective of this study is to propose five distributions such as lindley (LIN) distribution, weighted Lindley exponential (WLE), Power Modified Lindley (PML), Lindley half-Cauchy (LHC) and Rayleight Lomax distribution (RL) were used for modelling the duration diabetes survival time data of Bengkalis. Comparison of the proposed distributions with existing distribution functions is done to demonstrate their suitability in describing duration diabetes characteristics. Unknown parameter estimations were obtained with the Maximum Likelihood Method. Graphical methods such as pdf and cdf plot and also Numerical criteria such as AIC, BIC, and - *l* were used to obtain the distribution which provides the best fit the wind speed data. In the following section, the distributions to be used in modelling the duration diabetes survival time data are included.

II. MATERIALS AND DATA

For this study an independent samples from 50 diabetes patients were observed from Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province. The following table describes the initial information on duration survival times of patients diabetes. The descriptive statistics for duration of diabetes such as mean, varians, minimum and maximum can be seen in the descriptive statistics are presented in Table 1. The variations of data is not significantly different with the means indicate that this the duration diabetes survival times is quite stable. The data and the histogram or characteristic diabetes survival times are presented in Figure 1.

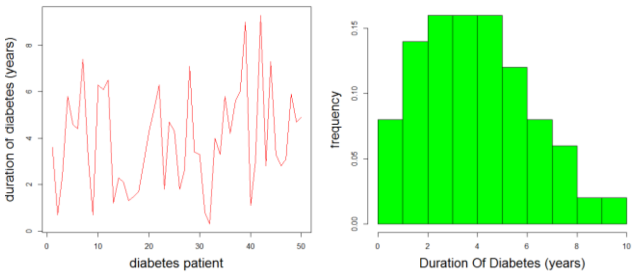

Fig. 1. Plot and histogram daily wind speed data on Pekanbaru respectively

Table 1: The descriptive statistics for daily wind speed.

Statistics	***Velue***
Mean	3.936
Varians	4.828
Minimum	0.300
Maximum	9.300

III. METHODOLOGY

In this study five models distribution were tested. When fitting a fully distributions model, the survival times are assumed to follow a statistical distribution. Several different distributions have been proposed, and the identification of a suitable one is a crucial step. There are lindley (LIN) distribution, weighted Lindley exponential (WLE), Power Modified Lindley (PML), Lindley half-Cauchy (LHC) and Rayleight Lomax distribution (RL). The probability density function f(x) and distribution function F(x) are respectively defined as follows:

A. Lindley Distribution (Lin).

We define the one parameters Lindley distribution (θ) denoted by LIN (θ) with density and cumulative distribution [12] are respectively given by

$$f(x) = \frac{\theta^2}{\theta+1}(1+x)exp(-\theta x),\ \ \theta > 0,\ \ x > 0$$

$$F(x) = 1 - \frac{\theta+1+\theta x}{\theta+1}exp(-\theta x)$$

B. Weighted Lindley Exponential Distribution (WLE)

We define the three parameters weighted Lindley exponential distribution $(\theta, \alpha, \lambda)$ denoted by WLE $(\theta, \alpha, \lambda)$ with density and cumulative distribution [13] are respectively given by

$$f(x) = \frac{\theta^2(\theta+\alpha\lambda)^2}{\alpha\lambda\big(\theta(\theta+2)+\alpha\lambda(1+\theta)\big)}(1+x)exp(-x\theta)\big(1 - exp(-x\alpha\lambda)\big),\qquad \theta,\alpha,\lambda > 0, x > 0$$

$$F(x) = \frac{\theta^2(\theta+\alpha\lambda)^2}{\alpha\lambda\big(\theta(\theta+2)+\alpha\lambda(1+\theta)\big)}\left(\frac{1-exp(-x\theta)}{\theta} + \frac{exp(-x\theta)(-1+exp(x\theta)-x\theta)}{\theta^2} + \frac{exp\big(-x(\theta+\alpha\lambda)\big)-1}{\theta+\alpha\lambda} + \frac{exp\big(-x(\theta+\alpha\lambda)\big)\Big(1-exp\big(x(\theta+\alpha\lambda)\big)+x(\theta+\alpha\lambda)\Big)}{(\theta+\alpha\lambda)^2}\right)$$

C. Reyleight Lomac Distribution (RL)

The density fuction and cumulative distribution function of the Reyleight Lomax distribution, denoted by RL [14] are respectively given by:

$$f(x) = \frac{\beta\lambda}{\theta}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda+1} exp\left(-\frac{\beta}{2}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda}\right),\qquad \lambda,\beta,\theta > 0, x > -\theta$$

$$F(x) = 1 - exp\left(-\frac{\beta}{2}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda}\right)$$

D. The Power Modified Lindley Distribution (PML)

The Power Modified Lindley Distribution (PML) distribution has two parameters θ and α . The density and cumulative distribution function of the PML (θ,α) [15] are respectively defined as follows:

$$f(x) = \frac{\alpha\theta}{1+\theta}\big((1+\theta)x^{\alpha-1}exp(\theta x^\alpha) + 2\theta x^{2\alpha-1} - x^{\alpha-1}\big)exp(-2\theta x^\alpha), x > 0, \theta, \alpha > 0$$

$$F(x) = 1 - \left(1 + \frac{\theta x^\alpha}{1+\theta}exp(-\theta x^\alpha)\right)exp(-\theta x^\alpha)$$

E. The Lindley half-Cauchy Distribution (LHC)

The Lindley half-Cauchy Distribution (LHC) distribution has two parameters θ and . The density function and cumulative distribution function of the LHC (θ,λ) [16] are respectively defined as follows:

$$f(x) = \frac{2}{\pi}\left(\frac{\theta^2}{1+\theta}\right)\left(\frac{\lambda}{\lambda^2+x^2}\right)\left(1-\frac{2}{\pi}tan^{-1}\left(\frac{x}{\lambda}\right)\right)^{\theta-1}\left(1 - log\left(1-\frac{2}{\pi}tan^{-1}\left(\frac{x}{\lambda}\right)\right)\right), x > 0, \theta, \alpha > 0$$

$$F(x) = 1 - \left(1-\frac{2}{\pi}tan^{-1}\left(\frac{x}{\lambda}\right)\right)^{\theta}\left(1 - \left(\frac{\theta}{1+\theta}\right)log\left(1-\frac{2}{\pi}tan^{-1}\left(\frac{x}{\lambda}\right)\right)\right)$$

The procedure of goodness of fit tests for model selection, both graphically and numerically. The formulations of Numerical criteria such as AIC and BIC for model evaluation are given in Table 2.

Table 2. The formulas of numerical criteria for model evaluation

Numerical Criteria	Formula
AIC	-2*l* + 2*q*
BIC	-2*l* + *q* log (*n*)

l = log likelihood, *q* = Number of parameters, *n* = number of data

IV. RESULT AND DISCUSSION

In this section, we analyze a duration diabetes survival times data set to demonstrate the performance of the LIN, WLE, RL, PML and LHC distributions in practice. The fitting of that distributions was considered using the data. Computed parameter values of different probability density function are presented in Table 3

Table 3. Computed parameter values of different probability density functions

	LIN	WLE	RL	PML	LHC
θ	0.4315	-	-	-	-
θ	-	0.6964	-	-	-
α	-	0.2177	-	-	-
λ	-	0.0006	-	-	-
θ	-	-	0.0503	-	-
λ	-	-	0.9735	-	-
β	-	-	0.0003	-	-
θ	-	-	-	0.1735	-
α	-	-	-	1.3571	-
θ	-	-	-	-	5.8607
λ	-	-	-	-	13.841

On the graphical presentation of modelling of the duration diabetes survival time data, in other words on duration diabetes survival time histogram, the density function curve for LIN, WLE, RL, PML and LHC distributions models are seen in Figure 2a, 3a, 4a, 5a, and 6a are respectively. When the density functions (pdf) are examined, it was determined that some distributions have yield similar results. Figure 2b, 3b, 4b, 5b, and 6b shows the fitted five distributions, based on cumulative distribution function (cdf) . From this figure RL and PML distributions model is able to provide a good result for duration diabetes survival time data. However, instead of graphical evaluation, Table 4 provides a more meaningful comparison using AIC, BIC, and – l values. Table 4 includes AIC, BIC and - l values test statistics values of the goodness of fit test for the fitness of duration diabetes survival time data based on Maximum Likelihood Estimators for LIN, WLE, RL, PML and LHC distributions. According to these results, although similar results are obtained for all four distributions, lowest AIC, BIC and - l value are obtained for RL distribution. As a conclusion, it is seen that RL distribution provides better modelling in terms of Numerical criteria.

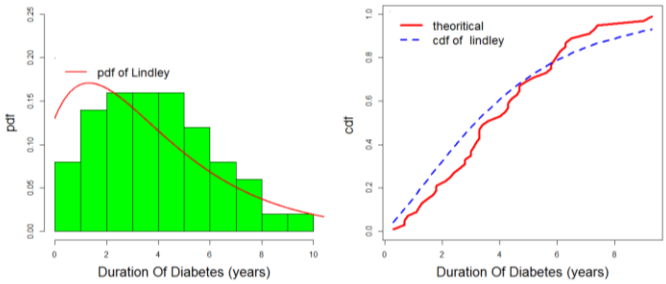

Fig. 2 Fitted pdf and cdf plots of LIN distribution are repectively.

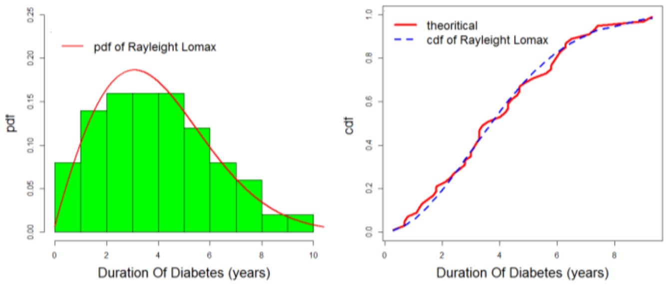

Fig. 3 Fitted pdf and cdf plots of RL distribution are repectively

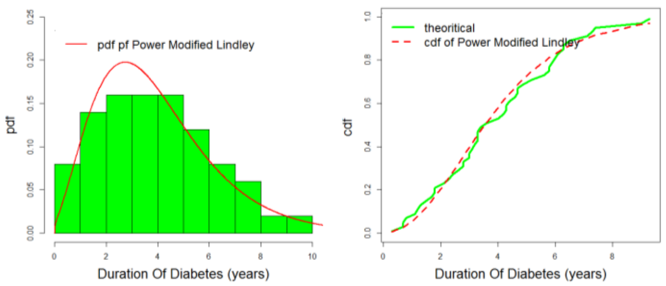

Fig. 5 Fitted pdf and cdf plots of PML distribution are repectively

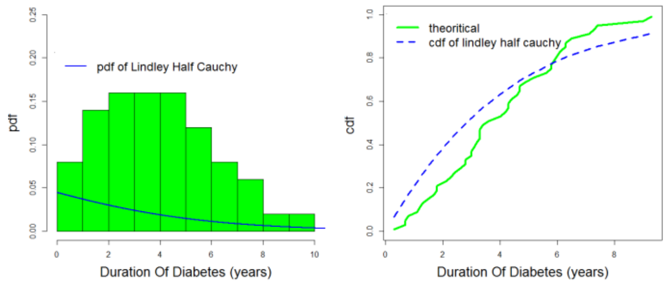

Fig. 6 Fitted pdf and cdf plots of LHC distribution are repectively

Table 4. AIC, BIC, and – Log Likelihood (*l*) function values

	LIN	WLE	RL	PML	LHC
AIC	287.8799	222.7924	**220.6416**	285.6562	250.0941
BIC	289.7919	228.5285	**226.3776**	289.4803	253.9181
- *l*	142.9399	108.3962	**107.3208**	140.8281	123.047

V. CONCLUSION

In the present article, a comparison of distribution models has been undertaken for describing duration diabetes survival times (years). Common conventional PDFs LIN and the modified PDFs along with three proposed PDFs, viz., WLE, PML, and LHC also three parameters RL distribution are used for duration diabetes survival time modeling. It is shown that conventional PDFs, such as LIN, is inadequate; hence, extended functions are used to model the observed duration survival distributions better. Results show clearly that proposed extended PDFs, RL provide viable alternative to other PDFs in describing duration diabetes survival time

REFERENCES

1. D. Collet, Modeling survival data in medical research, 2nd ed., Chapman and Hall/CRC, London, (2003).
2. D. G. Kleinbaum and M. Klein, Survival Analysis, A Self-Learning Text 3rd Ed, Springer, New York, (2012).
3. E. T. Lee dan J.W. Wang, Statistical methods for survival data analysis, Third Ed., Johns Wiley and Sons, New Jersey, 2003.
4. International Diabetes Federation, International Diabetes Federation 9th ed., Brusseles,(2019).
5. V. Lusiana, Menkes sebut diabetes paling banyak "serang" warga Riau ini sebabnya, 25 Maret 2019, https://riau.antaranews.com/berita/112025, accessed 8 juni 2019.
6. Juli, Kencing Manis dan Hipertensi Masuk 10 Pola Penyakit Utama di Bengkalis, 3 November 2019, http ://infopublik.id/ kategori/ nusantara/ 384135, accessed 11 januari 2020.

7. Ummu Athifah, Rado Yendra, Muhammad Marizal, and Rahmadeni, Survival Time Data Modeling for Diabetics During the Covid 19 Pandemic Using Several Sensitive Distributions (Weibull, Gamma, and Normal Logs), Applied Mathematical Sciences, 15(2021), 717-724
8. Manda Lisa Usvita, Arisman Adnan, Rado Yendra (2021). The Modelling Survival Times for Diabetes Patient Using Exponential, Weibull and Rayleigh-Lomax Distribution International Journal of Mathematics Trends and Technology, 109-112.
9. G. Gurpit, S. Alka, dan M. Juhi, An application of gamma generalized linear model for estimation of survival function of diabetic nephropathy patient, International Journal of Statistics in Medical Research, 2 (2013), 209-219.
10. S. K. Marvasti, S. Rimaz, J. Abolghasemi, dan I. Heydari, Comparing of Cox model and parametric models in analysis of effective factors on eventtime of neuropathy in patients with type 2 diabetes, Journal of Research in Medical Science, (2017).
11. K. Fatima, U. Jan, dan S. P. Ahmad, Statistical Properties of Rayleigh Lomax Distribution with Applications in Survival Analysis, Journal of Data Science, 3(2018), 531-548.
12. Shanker R, Fesshaye H (2016) On Modeling of Lifetime Data Using Akash, Shanker, Lindley and Exponential Distributions. Biom Biostat Int J 3(6): 00084. DOI: 10.15406/bbij.2016.03.00084
13. Doaa Basalamah and Bader Alruwaili, The weighted Lindley exponential distribution and its related properties, AIMS Mathematics, 8(2023), 24984–24998
14. Kawsar Fatima, Uzma Jan and S.P Ahmad, Statistical Properties of Rayleigh Lomax distribution with applications in Survival Analysis, Journal of Data Science, 16(2018), 531-548
15. Omid Kharazmi, Devendra Kumar, and Sanku Dey, Power Modi_ed Lindley Distribution: Properties,Classical and Bayesian Estimation and Regression Model with Applications, Austrian Journal of Statistics, 52(2023), 71-95
16. Arun Kumar Chaudhary and Vijay Kumar, Lindley Half Cauchy Distribution: Properties and Applications, International Journal For Research in Applied Science & Engineering Technology, 8(2020),1233-1242