International Journal of Mathematics Trends and Technology
ISSN: 2231 – 5373 /doi:10.14445/22315373/IJMTT-V67I5P513

Volume 67 Issue 5, 109-112, May, 2021

The Modelling Survival Times for Diabetes Patient Using Exponential, Weibull and Rayleigh-Lomax Distribution

Manda Lisa Usvita[1], Arisman Adnan[1], Rado Yendra[2]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Riau, Pekanbaru, 28293, Indonesia*
[2]*Department of Mathematics, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia*

Abstract — *Diabetes is a complex, slowly progressive silent killer disease. This destroys multiple organs by damaging and clogging the small capillaries – micro vascular system. Survival functions (probability density function) of diabetes patients are estimated in this paper. Conclusions about the occurrence probability distribution of diabetes patient in the Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province will be easier to use statistical models. For this purpose, 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.*

Keywords — *Diabetes, exponential distribution, Rayleigh-Lomax distribution, survival times, Weibull distribution.*

I. INTRODUCTION

Diabetes is a non-contagious degenerative disease which is a serious problem for human being. The increase in diabetes cases is indirectly due to passive lifestyle changes, unhealthy eating patterns, smoking habits and high stress levels. Globally, the number of diabetics has increased every year. Based on the data from the International Diabetes Federation in 2019, it was stated that the prevalence of diabetes worldwide reached 9.3% in 2019 and it is estimated to increase to 10.2% in 2030 and 10.9% in 2045. Diabetes in Indonesia ranks 7th largest in the world in 2019 [7]. The Minister of Health of the Republic of Indonesia stated that diabetes is in the top rank among residents in Riau Province with a very significant increase, namely 358.3% [13], while according to data from the Bengkalis District Health Service (Diskes), diabetes sufferers in 2019 were 10, 57% [8]. Most people with diabetes want to know what their chances of survival or how dangerous the diabetes is in posing a risk of death, so the data analysis is needed to find out the problem. The analysis used is called a survival analysis. Method used to analyzed data relating to the time or length of time until certain events occur is called a survival analysis. Survival analysis requires the 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 a parametric model. Parametric model is an analysis based on data distribution. The parametric model assumes that the distribution underlying the survival time follows a certain distribution, for example Weibull, exponential, log-normal, Rayleigh distribution.

Several researchers have discussed survival analysis in diabetes. Alka and Gurpit [12] have conducted a study to estimate the time of onset (time between drug administration and effect) of nephropathy arising from type 2 diabetic nephropathy patients using the Weibull distribution. Their research shows that the Weibull distribution is suitable for this case. Gurpit et al. [5], who discussed the estimated survival function of diabetic nephropathy patients with gamma, Weibull, exponential, log-normal, Rayleigh inverse, and Gaussian distributions. They found that the Gamma distribution is the best model for estimating the survival function of diabetic nephropathy patients. Marvasti et al. [11] conducted a study comparing the Cox model and the parametric model to analyze the time-effective factor of occurrence in type 2 nephropathy patients. The parametric model used was the Lognormal distribution. Their research showed that the Lognormal distribution is suitable for this case. Based on this description, the distribution commonly used in survival analysis for diabetic patient data with a parametric model is the Weibull, exponential, gamma, Rayleigh and log-normal distribution. Furthermore, Fatima et al. [10] conducted research related to introduction of developing a new distribution, the Rayleigh-Lomax distribution, and then applied this distribution to the survival data. Theu used aircraft windshield damage, glass fiber resistance data, and carbon fiber tension data. Their research shown 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 fitiing distribution for duration (survival times) diabetes patient based on several goodness of fit criteria. The scope of this study is to focussed on the exponential, Weibull and Rayleigh-Lomax distribution.

II. MATERIALS AND METHODS

For this study an independent samples from 50 diabetes patients were observed from Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province. The mean, standard deviation, kurtosis and skewness survival times of diabetes patients in years the following table describes

TABLE 1. SUMMARY STATISTICS SURVIVAL TIMES OF DIABETES PATIENTS

n	Mean	Standard Deviation	Skewness	Excess Kurtosis
50	3.936	2.197	0.569	1.13

A. Model Selection for Survival Times of Diabetes

In this study three models distribution were tested. The survival times are assumed to follow a statistical distribution. There are exponential, Weibull and Rayleigh-Lomax distributions. The distribution function, probability density function, survival function, and the likelihood function of respective distributions are defined as follows:

1. ***Exponential Distribution:***

$F(t;\theta) = 1 - \exp(-\theta t).$

$f(t;\theta) = \theta \exp(-\theta x)\,; t \geq 0 \ \& \ \theta > 0.$

$s(t;\theta) \ = \exp(-\theta t).$

$$L(t|\theta) = \theta^n exp\left(\sum_{i=1}^{n} -\theta t_i\right).$$

2. ***Two Parameters Weibull Distribution:***

$F(t;\alpha,\beta) = 1 - \exp\left((-\alpha t)^{\beta}\right).$

$f(t;\alpha,\beta) = \alpha\beta(\alpha t)^{\beta-1} \exp\left((-\alpha t)^{\beta}\right); \quad t \geq 0 \ \& \ \alpha,\beta > 0.$

$s(t;\alpha,\beta) = \exp\left((-\alpha t)^{\beta}\right).$

$$L(t|\alpha,\beta) = (\alpha\beta)^n \prod_{i=1}^{n} (\alpha t_i)^{\beta-1} \, exp\left(-\sum_{i=1}^{n} (\alpha t_i)^{\beta}\right).$$

3. ***Three Parameters Rayleigh-Lomax Distribution:***

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

$$f(t;\lambda,\theta,\beta) = \frac{\beta\lambda}{\theta}\left(\frac{\theta}{\theta+t}\right)^{-2\lambda+1} exp\left(-\frac{\beta}{2}\left(\frac{\theta}{\theta+t}\right)^{-2\lambda}\right); t \geq -\theta.$$

$$s(t;\lambda,\theta,\beta) = exp\left(-\frac{\beta}{2}\left(\frac{\theta}{\theta+t}\right)^{-2\lambda}\right).$$

$$L(t|\lambda,\theta,\beta) = \left(\frac{\beta\lambda}{\theta}\right)^n \prod_{i=1}^{n}\left(\frac{\theta}{\theta+t_i}\right)^{-2\lambda+1} exp\left(-\frac{\beta}{2}\sum_{i=1}^{n}\left(\frac{\theta}{\theta+t_i}\right)^{-2\lambda}\right).$$

B. Method of Moments

In this section we describe method of moments (MME) for estimating the parameters from exponential, Weibull, and Rayleigh-Lomax distributions. Based on Casela and Berger [4], Let $X_1, \cdots, X_n$ be a sample from a population with probability density function $(x|\theta_1, \cdots, \theta_k)$. Method of moments estimator are found by equating the first k sample moments (m_k) to the corresponding k population moments (μ'_k) and solving the resulting system of simultaneous equations.

$$m_k = \mu'_k(\theta_1, \cdots, \theta_k)$$

where

$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,k}$, and $\mu'_k = EX^k$.

Population moments of respective distributions are defined as follows:

1. ***Exponential Distribution:***

$$EX^k = \left(\frac{1}{\theta}\right)^k \Gamma(k+1).$$

2. ***Two Parameters Weibull Distribution:***

$$EX^k = \left(\frac{1}{\alpha}\right)^k \Gamma\left(\frac{k}{\beta}+1\right).$$

3. ***Three Parameters Rayleigh-Lomax Distribution:***

$$EX^k = \sum_{j=0}^{k} \binom{k}{j} \theta^j \left(\frac{2}{\beta}\right)^{\frac{j}{2\lambda}} (-\theta)^{k-j}\, \Gamma\left(\frac{j}{2\lambda} + 1\right).$$

By equating the first k sample moments to the corresponding k population moments and solving the resulting system of simultaneous equations, the moments estimation (MME) of the parameters are obtained. The moments equation for this distribution is in implicit form and complicated and we will not discuss in details. However, Newton-Rhapson method [9] has been employed in the iteration procedure to solve equation. Different initial values have been tested for this procedure. If the initial values converge to same values, it is considered to be the chosen estimated parameter.

C. Goodness of Fit Test

Statistical modeling that performs comparisons of several models is usually followed by a model fit test, either numerically or graphically. This is done to get the best model in modeling the data used.The smallest AIC and BIC values were selected as the best model for the data. The AIC and BIC are calculated using the following expression:

$$AIC = 2k - 2log\big(L(\theta)\big).$$

$$BIC = k\log(n) - 2log\big(L(\theta)\big).$$

where, n is the number of data , k is the number of parameters from each distribution and $L(\theta)$ is the likelihood function [6].

III. ANALYSIS AND RESULTS

Figure 1 has shown that the distributional nature of the duration of diabetes patient dataset is first examined through a histogram, and it clearly depicts that dataset is approximately normal to positively skewed. The visual technique of plotting data conducted for selecting a probability density function, this includes examining a histogram with the distribution overlaid and comparing the empirical model to the theoretical model. Hence, the study considered distributions such as exponential, Weibull, and Rayleigh-Lomax to identify the best fit.

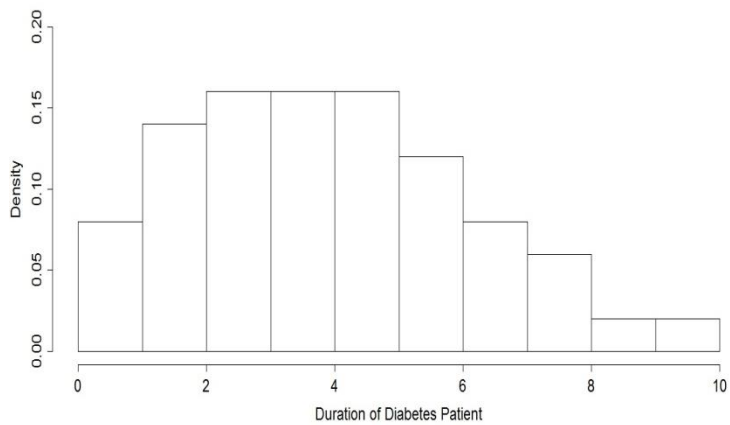

Figure 1. Histogram of survival times diabetes patient.

Table 2 depicts parameter estimates for the E, W and RL distributions, using the MME approach and the goodness-of-fit tests based on AIC and BIC. Based on the smallest AIC and BIC values, and graphical inspection (probability density function (PDF)) to survival times of diabetes patients, the study has shown that RL is best fit distribution in modeling survival times for diabetes patients.

TABLE 2. PARAMETERS, AIC AND BIC FOR SURVIVAL TIMES OF DIABETES PATIENT

Distributions	Parameters	AIC	BIC
Exponential	θ = 0.254	239.016	240.928
Weibull	α = 0.225 β = 1.880	218.585	222.409
Rayleigh-Lomax	λ = 1.195 β = 0.029 θ = 0.949	121.908	127.644

Figure 2 presents the histogram for the data with the fitted densities and we can observe that there are atypical observations and the right tail of the RL model is slightly heavier. Figure 3 presented the empirical and fitted cumulative distribution functions and we can see that the fitted cdf of the RL model is closer to the empirical distribution.

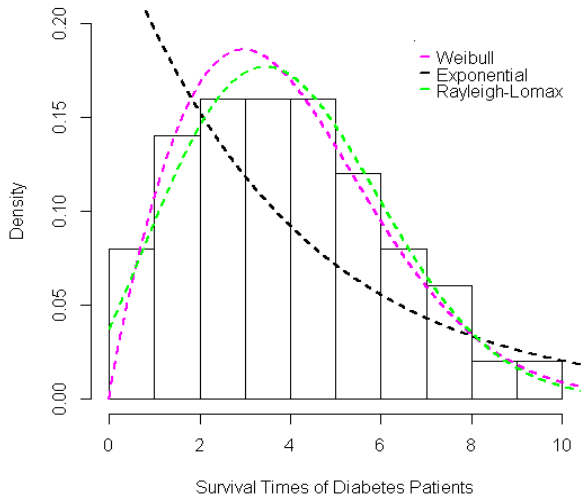

Figure 2. Estimated densities of E, W, and RL Distribution for the duration diabetes patient.

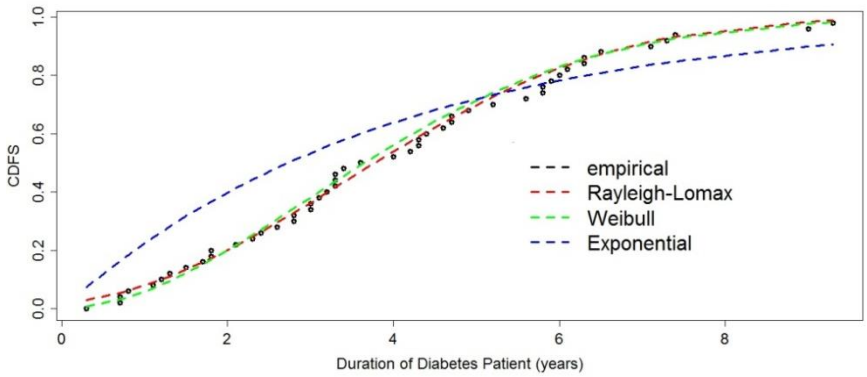

Figure 3. Empirical and fitted CDFs of E, W, and RL Distribution for the duration diabetes patient.

VI. CONCLUSIONS

With regard to the implementation of the three models distribution namely exponential, Weibull and Rayleigh-Lomax to survival times of diabetes patient, we can conclude based on the graphical methods (PDF and CDF) and numerical methods (AIC, BIC), that the Rayleigh-Lomax distribution is better than the exponential and Weibull distribution for the diabetes patient data form Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province. Therefore we recommend that to implementation of the Rayleigh-Lomax distribution to model survival times diabetes patient.

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 and J.W. Wang, Statistical methods for survival data analysis, Third Ed., Johns Wiley and Sons, New Jersey, (2003).
[4] G. Casela and R. L. Berger, Statistical Inference, 2nd ed., Duxbury, New York, (2002).
[5] G. Gurpit, S. Alka, and M. Juhi, An Application of Gamma Generalized Linear Model for Estimation of Survival Function of Diabetic Nephropathy Patiens, International Journal of Statistics in Medical Research, 2(2013), 209-219.
[6] H. Bozdogan, Akaike's information criterion and recent developments in information complexity, Journal of Mathematical Psychology, 44 (2000), 62-91.
[7] International Diabetes Federation, International Diabetes Federation 9th ed., Brusseles,(2019).
[8] 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.
[9] K. Atkinson and W. Han, Elementary Numerical Analysis, 3rd ed., John Wiley and Sons, New Jersey, (2004).
[10] K. Fatima, U. Jan, and S. P. Ahmad, Statistical Properties of Rayleigh Lomax Distribution with Applications in Survival Analysis, Journal of Data Science, 3(2018), 531-548.
[11] S. K. Marvasti, S.Rimaz, J. Abolghasemi, and I. Heydari, Comparing of Cox model and parametric models in analysis of effective factors on event time of neuropathy in patients with type 2 diabetes, Journal of Research in Medical Science, (2017).
[12] S. Alka dan G. Gurprit, A Parametric Approach to Estimate Survival Time of Diabetic Nephropathy with Left Truncated and Right Censored Data, International Journal of Statistics and Probability, 1(1) (2012) 128-137.
[13] 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.