Engineering and Technology Journal e-ISSN: 2456-3358
Volume 09 Issue 11 November-2024, Page No.- 5469-5474
DOI: 10.47191/etj/v9i11.02, I.F. – 8.227

Duration Diabetes Survival Times Modelling Using Some Extended Lomax Distribution

Fajar Fadhilah Mondri[1], Rado Yendra[2], Ari Pani Desvina[3], Rahmadeni[4], Muhammad Marizal[5]
[1,2,3,4,5] Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRACT: The precise modeling of survival time in diabetes is essential for accurately estimating the potential lifespan of individuals diagnosed with diabetes. One of the key factors in assessing the potential survival times of diabetes patients in a specific region is the probability distribution of diabetes survival times. Therefore, data on diabetes survival times are necessary to conduct statistical modeling, particularly in determining the most suitable probability distribution. Statistical models are developed to draw conclusions about the probability distribution of diabetic patients at the Mandau Regional General Hospital (RSUD) in Bengkalis Regency, Riau Province. To achieve this, six distributions will be utilized and evaluated to identify the best model for describing diabetes survival times. The primary objective of this research is to identify the most appropriate distribution to represent the survival times of diabetes patients from 50 individuals in the Bengkalis region, using the Lomax (LM) distribution, three parameters modified Lomax distributions (Rayleigh Lomax (RL), Logistics Lomax (LL), New Rayleigh Lomax (NRL)) and four parameters modified Lomax Distribution (Odd Lomax Log Logistics (OL), Novel Extended Power Lomax (PL)). The maximum likelihood method will be employed to estimate the parameter values of the distributions used in this study. Additionally, graphical assessments (density-density plot) and numerical criteria (Akaike's Information Criterion (AIC), -log likelihood (-l)) will be utilized to determine the best-fitting model. In most instances, the results obtained from graphical assessments were consistent but differed from the numerical criteria. The model with the lowest values of AIC and -l was selected as the best fit. Overall, the RL and OL distributions was identified as the most suitable model.

KEYWORDS: Lomax, Rayleigh Lomax, Logistics Lomax, New Rayleigh Lomax, Odd Lomax Log Logistics, Novel Extended Power Lomax, diabetes survival times

INTRODUCTION

The COVID-19 pandemic has indirectly increased death rates due to a rise in diabetes patients. Some reasons for this increase include heightened stress from being confined at home, frequent consumption of fast food, and lack of exercise. Diabetes is a significant public health issue in Indonesia and globally. According to the International Diabetes Federation, 463 million people worldwide have diabetes, with the prevalence expected to reach 9.3% by 2020, representing a 45% increase to 629 million patients annually by 2045. Indonesia has one of the highest rates of diabetes, with 6.2% of the population affected[2]. This equates to over 10.8 million individuals in 2020, projected to increase to 16.7 million by 2045[1]. The Minister of Health of Indonesia reported that diabetes affects the most people in the province of Riau, with a 358.3% increase[3], while in Bengkalis, diabetes affected 10.57% of the population in 2019[4]. Understanding the survival times of diabetic patients is crucial for estimating the risk of death from diabetes. Survival time studies can be conducted using statistical techniques to develop a model that accurately represents survival patterns. Various studies have aimed to determine the optimal probability model for diabetic survival time data. Alka and Gurpit[5] utilized the Weibull distribution to estimate the onset time of nephropathy in type 2 diabetic patients. Gurpit et al. discussed survival function estimates of diabetic nephropathic patients using distributions such as exponential, gamma, Weibull, log-normal, inverse Gaussian, and Rayleigh. Their research concluded that the gamma distribution is the most effective method for predicting the survival function of diabetic nephropathic patients. Previous research has been conducted to determine the best probability model for data on the survival time of diabetic patients. Ummu et al. [6] estimated the duration of diabetes survival time using the Weibull, Gamma, and Log-Normal distributions. The results showed that the Weibull model was the best in approaching the given observational data. This was also supported by numerical models such as AIC and BIC, which provided the smallest values for the two numerical methods compared to other probability models. Furthermore, Manda Lisa Usvita et al. [7] compared three types of distributions, namely Exponential (E), Weibull (W), and Rayleigh-Lomax (RL), applied to survival times of diabetes patients. The Method of Moments was used to obtain the estimated parameters. Based on the smallest Akaike's Information Criterion (AIC) and Bayesian Information

Criterion (BIC) values, and graphical inspection (probability density function (pdf)) of 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. Sutriana et al[8] 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 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. Gurpit et al. [9] discussed the estimation of the 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 log-normal 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 included information on aircraft windshield damage, glass fiber resistance, and carbon fiber tension. Their research indicates that the Rayleigh-Lomax distribution is appropriate for survival data analysis. Selecting the most suitable distribution is a key focus in studying survival times. Therefore, this study aims to identify the best-fitting distribution for the survival times of diabetic patients based on various goodness-of-fit criteria. A preliminary investigation was conducted on the survival times of diabetes in 50 patients from the Bengkalis region. The goal is to propose six distributions: Lomax (LM) distribution[12], three-parameter modified Lomax distributions (Rayleigh Lomax (RL)[11], Logistics Lomax (LL)[13], New Rayleigh Lomax (NRL)[14]), and four-parameter modified Lomax distributions (Odd Lomax Log Logistics (OL)[16], Novel Extended Power Lomax (PL)[17]) to model the duration of diabetes survival time data in Bengkalis. The proposed distributions are compared with existing distribution functions to assess their suitability in describing diabetes characteristics. Unknown parameter estimates were calculated using the Maximum Likelihood Method. Graphical methods such as pdf plots, as well as numerical criteria like AIC and - l, were employed to determine the distribution that best fits the diabetes data. The next section includes the distributions selected for modeling the duration of diabetes survival time data.

MATERIALS

For this study, an independent sample of 50 diabetes patients was observed at Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province. The following table provides initial information on the survival times of diabetes patients. Descriptive statistics for the duration of diabetes, including mean, variance, minimum, and maximum values, are presented in Table 1. The consistency between the variations in data and the means suggests that the duration of diabetes survival times is relatively stable. The data and histogram of diabetes survival times are illustrated in Figure 1.

Table 1: The descriptive statistics for duration of diabetes (years)

Statistics	Mean	Variation	Minimum	Maximum
	3.936	4.828882	0.300	9.300

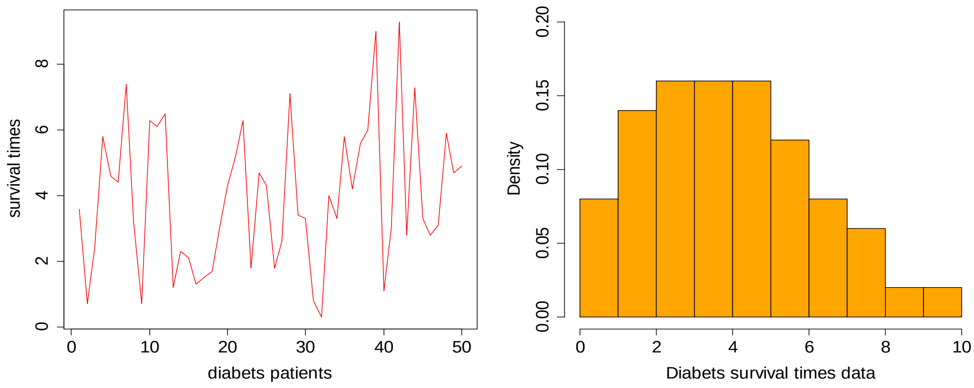

Figure 1. The plot and histogram of the duration of diabetes data, respectively.

METHODS

3.1 Probability Density Function (PDF)

In this study, six probability density functions (PDFs) associated with modeling duration of diabetes, LM, RL, LL, NRL, OL and PL are considered. The equations defining the PDFs for various candidate distributions of interest are provided below. Each distribution considered is listed in Table 2, where x (duration of diabetes) represents the observed values of the random variable for the event of interest. To fit a specific theoretical distribution to the observed distribution of duration of diabetes, parameters are estimated using the maximum likelihood method. The maximum likelihood function for this model is implicit and complex, and detailed discussion is beyond the scope of this paper. The nonlinear equation generated by the maximum log-likelihood function (*l(Ω), where Ω represents parameters*) necessitates a numerical method, specifically the Newton-Raphson method, to obtain the solution. This method is employed iteratively to find the solution. Various initial values have been tested, and if they converge to the same value, it is considered the estimated parameter. The procedure for goodness-of-fit tests for model selection, both numerically and graphically, is also discussed.

Table 2: Presents the Probability Density Function (PDF) of various distributions.

Distribution	Pdf (x)
LM	$f(x;\beta,\lambda)=\lambda\beta(1+\beta x)^{-(\lambda+1)}, x>0,\beta>0,\lambda>0$
RL	$f(x;\ \beta,\lambda,\theta)=\frac{\beta\lambda}{\theta}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda+1} exp\left(-\frac{\beta}{2}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda}\right), x\geq -x,\beta>0,\lambda>0,\theta>0$
LL	$f(x;\ \beta,\lambda,\alpha)=\frac{\lambda\beta\alpha(\beta x+1)^{\lambda-1}\left((\beta x+1)^{\lambda}-1\right)^{\alpha-1}}{(1+((\beta x+1)^{\lambda}-1)^{\alpha})^{2}}, x>0,\beta>0,\lambda>0,\alpha>0$
NRL	$f(x;\ \theta,\lambda,\sigma)=\frac{\theta x}{\lambda\sigma^{2}}\left(1+\frac{x^{2}}{2\lambda\sigma^{2}}\right)^{-(\theta+1)}, x>0,\theta>0,\lambda>0,\sigma>0$
OL	$f(x;\ a,b,c,d)=\frac{\frac{a}{b}\frac{c}{d}\left(\frac{x}{b}\right)^{a-1}}{\left(1+\frac{1}{d}\left(\frac{x}{b}\right)^{a}\right)^{(c+1)}}, x,a,b,c,d>0$
PL	$f(x;\ \alpha,\gamma,\beta,\lambda)=\frac{\log\alpha}{\alpha-1}\gamma\beta\lambda^{\gamma}x^{\beta-1}\left(\lambda+x^{\beta}\right)^{-\gamma-1}\alpha^{1-\lambda^{\gamma}(\lambda+x^{\beta})^{-\gamma}}, x,\alpha,\gamma,\beta,\lambda>0$

3.2 Maximum Likelihood Estimate (MLE) and Goodness of Fit Tests (GOF)

Let $(x_1,x_2,\cdots,x_n)$ be a random sample from LM, RL, LL, NRL, OL and PL distributions. The log-likelihood **(*l(Ω)*)** is presented in Table 3. The maximum likelihood estimate (MLE) is the solution of the equation and thus the solution of the following nonlinear equation. The most appropriate distribution is identified using results based on several goodness-of-fit (GOF) tests. The GOF tests considered are based on graphical inspection (PDF plot), and numerical criteria such as Akaike’s Information Criterion (AIC) and –log likelihood (-*l*) were applied to determine the GOF criteria of the distributions. In most cases, graphical inspection gave the same result, but their AIC results differed. The best fit result was chosen as the distribution with the lowest values of AIC. The formula of numerical methods such as AIC is shown in Table 4.

Table 3: The Log Likelihood of Various Distributions

Distribution	*l(x)* (Log-Likelihood function)
LM	$l(x;\lambda,\beta)=n\log(\lambda)+n\log(\beta)-(\lambda+1)\sum_{i=1}^{n}\log(1+\beta x_i)$
RL	$l(x;\lambda,\beta,\theta)=n\log(\lambda)+n\log(\beta)-n\log(\theta)-(2\lambda-1)\sum_{i=1}^{n}\log\left(\frac{\theta}{\theta+x_i}\right)-\left(\frac{\beta}{2}\left(\frac{\theta}{\theta+x}\right)^{-2\lambda}\right)$
LL	$l(x;\lambda,\beta,\alpha)=n\log(\lambda)+n\log(\beta)+n\log(\alpha)+(\lambda-1)\sum_{i=1}^{n}\log(\beta x_i+1)+(\alpha-1)\sum_{i=1}^{n}\log\left((\beta x_i+1)^{\lambda}-1\right)-2\sum_{i=1}^{n}\log\left(1+\left((\beta x_i+1)^{\lambda}-1\right)^{\alpha}\right)$
NRL	$l(x;\theta,\lambda,\sigma\ )=n\log(\theta)-n\log(\lambda)-2n\log(\sigma)-(\theta+1)\sum_{i=1}^{n}\log\left(1+\frac{{x_i}^{2}}{2\lambda\sigma^{2}}\right)$

OL	$l(x;a,b,c,d) = n\left(\log(a) - \log(b) + \log(c) - \log(d)\right) + (a-1)\sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right) - (c+1)\sum_{i=1}^{n} \log\left(1 + \frac{1}{d}\left(\frac{x_i}{b}\right)^a\right)$
PL	$l(x;\alpha,\gamma,\beta,\lambda) = n\log(\log(\alpha)) - n\log(\alpha-1) + n\log(\gamma+\beta) + n\gamma\log(\lambda) + (\beta-1)\sum_{i=1}^{n} \log(x_i) + (-\gamma-1)\sum_{i=1}^{n} \log\left(\lambda + {x_i}^{\beta}\right) + \log(\alpha)\sum_{i=1}^{n} \log\left(1 - \lambda^{\gamma}\left(\lambda + x^{\beta}\right)^{-\gamma}\right)$

Table 4: presents the formulas of numerical criteria for model evaluation.

Numerical Criteria	Formula
AIC	$-2l + 2p$
$-l$	- log likelihood

l = log likelihood, p = Number of parameters

RESULTS AND DISCUSSION

In this section, we analyze a dataset of diabetes survival times to demonstrate the performance of LM, RL, LL, NRL, OL, and PL distributions in practice. The fitting of these distributions was assessed using the data. The computed parameter values of various probability density functions are presented in Table 3.

Table 3: Computed parameter values of different probability density functions

	LM	RL	LL	NRL	OL	PL
λ	1227.722					
β	0.0002					
θ		0.0503				
λ		0.973				
β		0.0003				
λ			15.353			
β			0.0130			
α			1.904			
θ				89.055		
λ				0.0019		
σ				674.650		
a					1.855	
b					2.334	
c					23309.71	
d					76375.69	
α						1.00011
γ						43.043
β						1.7160
λ						491.022

On the graphical presentation of the modeling of the duration of diabetes survival time data, in other words, on the duration of diabetes survival time histogram, the density function curves for LM, RL, LL, NRL, OL, and PL distribution models are seen in Figures 2, 3, and 4 respectively. When the density functions (pdf) are examined, it was determined that some distributions yield similar results. From these figures, the RL and OL distribution models are able to provide good results for the duration of diabetes survival time data. However, instead of graphical evaluation, Table 4 provides a more meaningful comparison using AIC, and - l values. Table 4 includes AIC, and - l values test statistics for the goodness of fit test for the fitness of the duration of diabetes survival time data based on Maximum Likelihood Estimators for LM, RL, LL, NRL, OL and PL distributions. According to these results, although similar results are obtained for all six distributions, the lowest AIC, and - l values are obtained for the RL and OL distribution. In conclusion, it is seen that the RL and OL distribution provides better modeling in terms of numerical criteria.

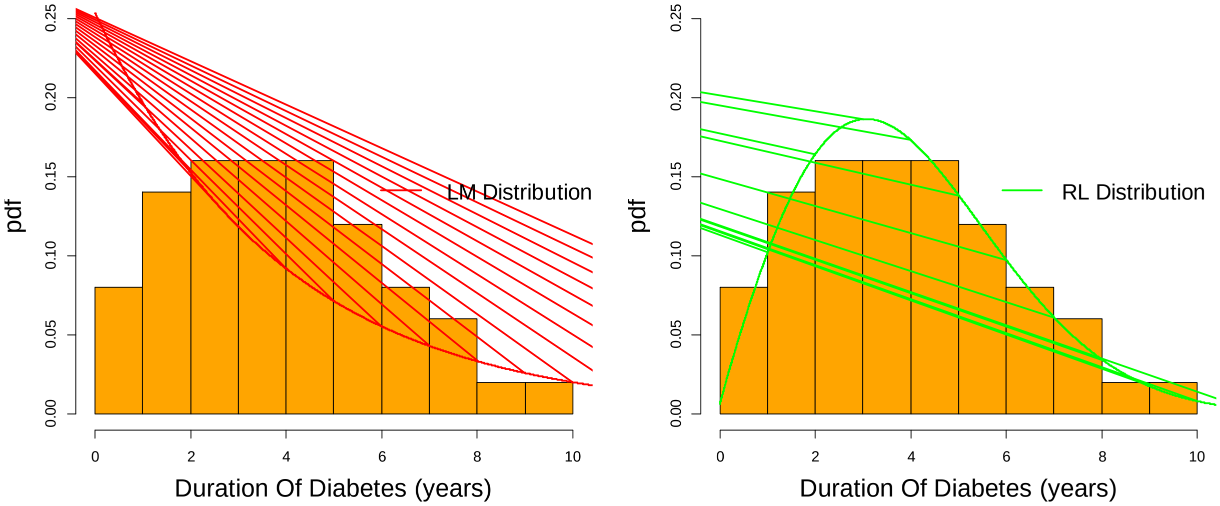

Figure 2. Fitted probability density functions (pdf) for LM and RL distributions, respectively.

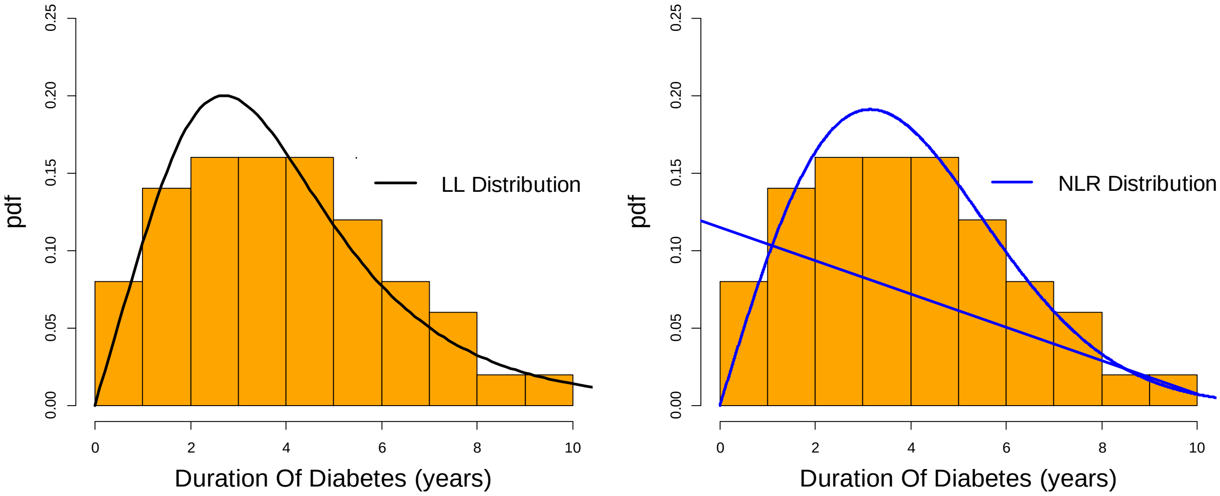

Figure 3. Fitted probability density function (PDF) for LL and NRL distributions, respectively.

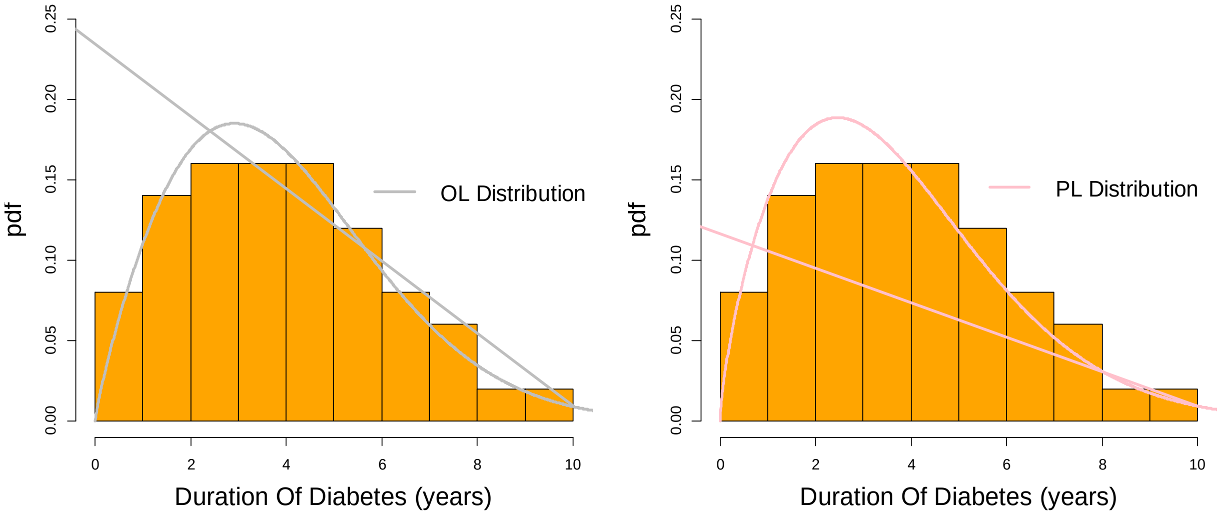

Figure 4. Fitted probability density function (pdf) of OL and PL distributions, respectively.

Table 4: AIC and – Log Likelihood (*l*) function values

	LM	RL	LL	NRL	OL	PL
-log	118.5225	107.3208	109.365	107.522	**107.286**	107.802
AIC	241.045	**220.6416**	224.73	221.044	222.572	223.604

CONCLUSION

In this research, the focus is on determining the best statistical model for the duration of diabetes survival time data. The study considers six distributions: Two-Parameter Lomax Distribution (LM), Three-Parameter Rayleigh Lomax Distribution (RL), Three-Parameter Logistics Lomax Distribution (LL), Three-Parameter New Rayleigh Lomax Distribution (NRL), Four-Parameter Odd Lomax Log Logistics Distribution (OL), and Four-Parameter Novel Extended Power Lomax Distribution (PL). It is shown that conventional probability density functions (PDFs), such as LM, are inadequate; hence, extended functions are used to model the observed duration survival distributions more effectively. Results clearly show that the proposed extended PDFs, RL and OL, provide a viable alternative to other PDFs in describing the duration of diabetes survival time.

REFERENCES

1. E. Pranita, *Naik 6,2 persen selama pandemi pasien di- abetes Indonesia peringkat 7 dunia*, 5 November 2020, https://www.kompas.com/sains/read/2020/11/05/100200923.
2. International Diabetes Federation, International Diabetes Federation 9th ed., Brusseles,(2019).
3. 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.
4. 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.
5. S. Alka dan G. Gurprit, *A parametric approach to estimate survival ti- me of diabetic nephropathy with left truncated and right censored data*, International Journal of Statistics and Probability, 1 (2012), 128-137.
6. 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
7. 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.
8. Sutriana, Rado Yendra, Rahmadeni, Muhammad Marizal (2023). 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. *International Journal of Mathematics And Computer Research,*11(12), 3894-3898
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. Abbas Pak and Mohammad Reza Mahmoudi, *Estimating the parameters of Lomax distribution from imprecise information*, Journal of Statistical Theory and Applications, 2018, 17(1), 122-135
13. Arun Kumar Chaudhary, Vijay Kumar, *The Logistic Lomax Distribution with Properties and Applications,* International Journal of New Technology and Research, 2020, 6(12), 74-80
14. K. Naga Saritha, G. S. Rao And K. Rosaiah, *Survival analysis of cancer patients using a new Lomax Rayleigh distribution,* JAMSI, (2023), 1, 19-45
15. Benson Benedicto Kailembo, Srinivasa Rao Gadde, Peter Josephat Kirigiti, *Application of Odd Lomax log-logistic distribution to cancer data*, Heliyon, 2024, 1-17
16. Maha E. Qura, Mohammed Alqawba, Mashail M. Al Sobhi and Ahmed Z. Afify, *A Novel Extended Power-Lomax Distribution for Modeling Real-Life Data: Properties and Inference*, Journal of Mathematics, 2023, 1-16