ISSN 2664-4150 (Print) & ISSN 2664-794X (Online)
South Asian Research Journal of Engineering and Technology
Abbreviated Key Title: South Asian Res J Eng Tech

| Volume-7 | Issue-1 | Jan-Feb- 2025 | **DOI: https://doi.org/10.36346/sarjet.2025.v07i01.002**

Original Research Article

The Impact of Random Variable Transformation on the Lindley and Sujatha Distribution Probability Models in Modeling Diabetes Survival Data

Nanda Saputra Siregar[1*], Rado Yendra[1], Muhammad Marizal[1], Ari Pani Desvina[1]
[1]Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

***Corresponding Author:** Nanda Saputra Siregar
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Article History
Received: 15.11.2024
Accepted: 21.12.2024
Published: 03.01.2025

Abstract: The probability models of two and three mixed gamma distributions, specifically the Lindley and Sujatha distributions, will be enhanced through the application of random variable transformation techniques, resulting in the Power Lindley and Power Sujatha probability models. This study employs four probability models: Lindley, Sujatha, Power Lindley, and Power Sujatha, to analyze the survival time of diabetic patients. All probability models in this study will utilize the maximum likelihood method for parameter estimation. The optimal model will be determined based on a goodness-of-fit test, which will incorporate both graphical methods (density and cumulative distribution graphs) and numerical methods (Akaike's Information Criterion (AIC) and negative log-likelihood). The results of the goodness-of-fit test indicate that the model derived from the random variable transformation yields a superior probability model compared to its original form.

Keywords: Lindley Distribution, Sujatha Distribution, Random Variable Transformation Techniques, Power Lindley Distribution, Power Sujatha Distribution.

INTRODUCTION

Probability modeling of survival data can be conducted using probability models that vary in the number of parameters, ranging from one to four. Models with a single parameter, such as the Akash [1], Shanker [2], Aradhana [3], Devya [4], Shambhu [5], and Rama [6], and Amarendra [7], distributions, have been utilized in this context. Several studies have sought to identify the optimal probability model for analyzing diabetes survival time data. Most of the probability models employed contain two or more parameters. For instance, Alka and Gurpit [8], utilized the Weibull distribution to estimate the time to onset of nephropathy in patients with type 2 diabetes. Gurpit *et al*., examined the estimation of survival functions in patients with diabetic nephropathy using various distributions, including exponential, gamma, Weibull, log-normal, inverse Gaussian, and Rayleigh. The authors concluded that the gamma distribution is the most effective predictor of survival in patients with diabetic nephropathy. Ummu *et al*., [9], estimated the duration of diabetes survival using the Weibull, Gamma, and Log-Normal distributions. The results indicated that the Weibull model provided the best approximation of the observational data. This finding was further supported by numerical models, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which yielded the smallest values for both criteria compared to other probability models. Additionally, Manda Lisa Usvita *et al*., [10], compared three distributions Exponential (E), Weibull (W), and Rayleigh-Lomax (RL) applied to the survival time of patients with diabetes. The Method of Moments was employed to estimate the parameters. Based on the smallest AIC and BIC values, along with graphical examinations of the probability density function (pdf) of diabetes patient survival time, this study demonstrated that the Rayleigh-Lomax distribution is the most suitable model for diabetes patient survival time at Mandau Regional Hospital in Bengkalis Regency, Riau Province. Sutriana *et al*., [11], utilized the Lindley distribution (LIN) along with three modified versions: the Weighted Lindley Exponential (WLE), the Power Modified Lindley (PML), and the Lindley Half-Cauchy

CITATION: Nanda Saputra Siregar, Rado Yendra, Muhammad Marizal, Ari Pani Desvina (2025). The Impact of Random Variable Transformation on the Lindley and Sujatha Distribution Probability Models in Modeling Diabetes Survival Data. *South Asian Res J Eng Tech*, *7*(1): 11-16.

(LHC), as well as the Rayleigh–Lomax (RL) distribution. The most suitable results were achieved with the distributions that exhibited the lowest values of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and - log-likelihood (- l). Overall, the Rayleigh–Lomax (RL) distribution was identified as the optimal model. The latest research conducted in 2024 utilized survival data from individuals with diabetes to perform probability modeling. This study employed several probability models, including the Lomax (LM) distribution, three modified Lomax distributions (Rayleigh Lomax (RL), Logistics Lomax (LL), and New Rayleigh Lomax (NRL)), as well as a four-parameter modified Lomax distribution (Odd Lomax Log Logistics (OL) and Novel Extended Power Lomax (PL)). The model with the lowest values of Akaike Information Criterion (AIC) and - log-likelihood (-l) was selected as the best fit. Overall, the Rayleigh Lomax (RL) and Odd Lomax Log Logistics (OL) distributions were identified as the most suitable models [12]. This probability models generally consist of two or more variables. This study examines the application of a one-parameter probability model to the survival data of patients with diabetes. This model consists of a mixture of two gamma density functions, referred to as the Lindley [13], probability model, and a mixture of three gamma density functions, known as the Sujatha [14], probability model. To improve the model's goodness of fit, we will employ random variable transformation techniques to extend the parameters of both the Lindley and Sujatha probability models to two parameters. The resulting two-parameter probability models are termed the Power Lindley distribution and the Power Sujatha distribution. The objective of this study is to propose four distributions: the Lindley (LD) distribution, the Sujatha distribution (SD), the Power Lindley distribution (PL) [15], and the Power Sujatha distribution (PS) [16], to model the duration of diabetes survival time data in Bengkalis. The proposed distributions are compared with existing distribution functions to evaluate their effectiveness in characterizing diabetes-related data. Unknown parameter estimates were calculated using the Maximum Likelihood Method. Graphical methods, such as probability density function (PDF) plots, along with numerical criteria like the Akaike Information Criterion (AIC) and log-likelihood (−l), were employed to identify the distribution that best fits the diabetes data. The following section presents the distributions selected for modeling the duration of diabetes survival time data.

Materials

For this study, Duration 50 patients with diabetes (years) was collected from Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province, are presented in Table 1.

Table 1: Duration 50 patients with diabetes (years)

3.6	0.7	2.4	5.8	4.6	4.4	7.4	3.2	0.7	6.3
6.1	6.5	1.2	2.3	2.1	1.3	1.5	1.7	3.0	4.3
5.2	6.3	1.8	4.7	4.3	1.8	2.6	7.1	3.4	3.3
0.8	0.3	4.0	3.3	5.8	4.2	5.6	6.0	9.0	1.1
3.0	9.3	2.8	7.3	3.3	2.8	3.1	5.9	4.7	4.9

METHODS

Probability Density Function (pdf) and Cumulative Distribution Function (CDF)

In this study, four probability density functions (PDFs) associated with modeling the duration (in years) of diabetes in 50 patients such as LD, SD, PL, and PS are considered. The equations defining the probability density functions (PDFs) and cumulative distribution functions (CDFs) for the various candidate distributions of interest are provided below.

Lindey Distribution (LD)

One-Parameter distribution known as Lindley distribution is defined by Probability Density Function (PDF) and Cumulative Distribution Function (CDF) as

$$f(y) = \frac{\theta^2}{\theta+1}(1+y)e^{-y\theta}, y, \theta \geq 0$$
$$F(y) = 1 - \left(1 + \frac{\theta y}{\theta+1}\right)e^{-y\theta}$$

The Lindley distribution is a two-component mixture of distributions such as a gamma (1, θ) and a gamma (2, θ) with their mixing proportions of $\frac{\theta}{\theta+1}$ and $\frac{1}{\theta+1}$, respectively. Where the density function of gamma (1, θ) and gamma (2, θ) are $f(y) = \theta e^{-\theta y}$ and $f(y) = \theta^2 y e^{-\theta y}$, respectively.

Sujatha Distribution (SD)

This distribution is Additionally, a three-component mixture of an gamma (1, θ), a gamma (2, θ) and a gamma (3, θ) with Mixing Proportions $\frac{\theta^2}{\theta^2+\theta+2}$, $\frac{\theta}{\theta^2+\theta+2}$ and $\frac{2}{\theta^2+\theta+2}$, respectively. Where the density function of gamma (3, θ) is

$f(y) = \frac{\theta^3}{2} y^2 e^{-\theta y}$. Sujatha distribution defined by probability density function (pdf) and cumulative distribution function (cdf) as

$$f(y) = \frac{\theta^3}{\theta^2+\theta+2}(1 + y + y^2)e^{-y\theta}, y, \theta \geq 0$$
$$F(y) = 1 - \left(1 + \frac{\theta y(\theta y + \theta + 2)}{\theta^2+\theta+2}\right) e^{-y\theta}$$

Power Lindley Distribution (PL)

Assuming the power transformation $= X^{1/\alpha}$, where X is random variable of Lindley Distribution. Using the random variable transformation technique, the pdf and cdf of the random variable Y of Power Lindley Distribution can be obtained as

$$f(y) = \frac{\alpha\theta^2}{\theta+1}(y^{\alpha-1} + y^{2\alpha-1})e^{-\theta y^\alpha}, y > 0, \alpha, \theta > 0$$

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

Power Sujatha Distribution (PS)

The transformation $= X^{1/\alpha}$, where X is random variable of Sujatha Distribution. Using the random variable transformation technique, the pdf and cdf of the random variable Y of Power Sujatha Distribution can be obtained as

$$f(y) = \frac{\alpha\theta^3}{\theta^2+\theta+2}(y^{\alpha-1} + y^{2\alpha-1} + y^{3\alpha-1})e^{-\theta y^\alpha}, y > 0, \alpha, \theta > 0$$

$$F(y) = 1 - \left(1 + \frac{\theta^2(y^\alpha + y^{2\alpha}) + \theta(2y^\alpha)}{\theta^2+\theta+2}\right) e^{-\theta y^\alpha}$$

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

Let $(y_1, y_2, \cdots, y_n)$ be random samples from the LD, SD, PL and PS distributions. The log likelihood function of LD, SD, PL and PS are presented in Table 2

Table 2: The Log Likelihood of Various Distributions

	l(y) (Log-Likelihood function)
LD	$l(\theta, y) = 2n\log(\theta) - n\log(\theta+1) + \sum_{i=1}^{n}\log(1+y_i) - \theta\sum_{i=1}^{n} y_i$
SD	$l(\theta, y) = 3n\log(\theta) - n\log(\theta^2+\theta+2) + \sum_{i=1}^{n}\log(1+y_i+{y_i}^2) - \theta\sum_{i=1}^{n} y_i$
PL	$l(\theta, \alpha, y) = n\log(\alpha) + 2n\log(\theta) - n\log(\theta+1) + \sum_{j=1}^{n}\log\left({x_j}^{\alpha-1} + {x_j}^{2\alpha-1}\right) - \theta\sum_{j=1}^{n}{x_j}^{\alpha}$
PS	$l(\theta, \alpha, y) = n\log(\alpha) + 3n\log(\theta) - n\log(\theta^2+\theta+2) + \sum_{j=1}^{n}\log\left({x_j}^{\alpha-1} + {x_j}^{2\alpha-1} + {x_j}^{3\alpha-1}\right) - \theta\sum_{j=1}^{n}{x_j}^{\alpha}$

The MLE $\widehat{\Omega}\ of\ \Omega$ is the solution of the equation $\frac{\partial l(\Omega)}{\partial \Omega} = 0$ and thus it is the solution of the following nonlinear equation. The most appropriate distribution was identified using results from several GOFs. The GOF tests considered were based on graphical inspection (pdf plot) and (cdf plot). Numerical criteria like Akaike's information criterion (AIC) and -2*Log like Lihood (-2*l) was 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 fitting result was selected as the distribution with the lowest AIC value. The formula of numerical methods such as AIC is exhibited in the following Table 3

Table 3: The formulas of numerical criteria for model evaluation

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

l = log likelihood, *p* = Number of parameters

RESULT

In this section, the parameter values of the LD, SD, PL, and PS probability models, obtained through the maximization of the log-likelihood function, will be presented as shown in Table 2. These parameters are detailed in Table 4, which also includes the goodness-of-fit test results, such as the AIC value and the -2*log-likelihood value. The formula for this goodness-of-fit test is provided in Table 3.

Table 4: Computed parameter, AIC and -2* Log Likelihood (-2*l)

	θ	α	AIC	-2*l
Lindley	0.431542		227.1103	225.1103
Sujatha	0.640519		220.7192	218.7192
Power Lindley	0.242325	1.403874	219.1187	215.1187
Power Sujatha	0.504769	1.181096	220.0599	216.0599

In Table 6, it is evident that the probability model derived from the random variable transformations, specifically PL and PD, outperforms the two models that were analyzed prior to these transformations, namely LD and SD. The comparison of the AIC values indicates that the PL probability model has the lowest AIC value, which indirectly suggests that the survival data for individuals with diabetes is effectively modeled by the PL approach. The goodness-of-fit test for the graphical model was conducted using both density plots (PDF plots) and cumulative distribution function plots (CDF plots) in this study. Figure 1 illustrates that the PDF plot for the LD distribution does not adequately capture the histogram of the data, while the CDF plot indicates that the CDF model is not perfectly aligned with the thoritical CDF$\left(F = \frac{i-0.5}{n}, i = 1,2,\cdots,n\right)$, with n is number of data. Almost identical results were also demonstrated by the graphical goodness-of-fit test for the SD probability model, as illustrated in Figure 2.

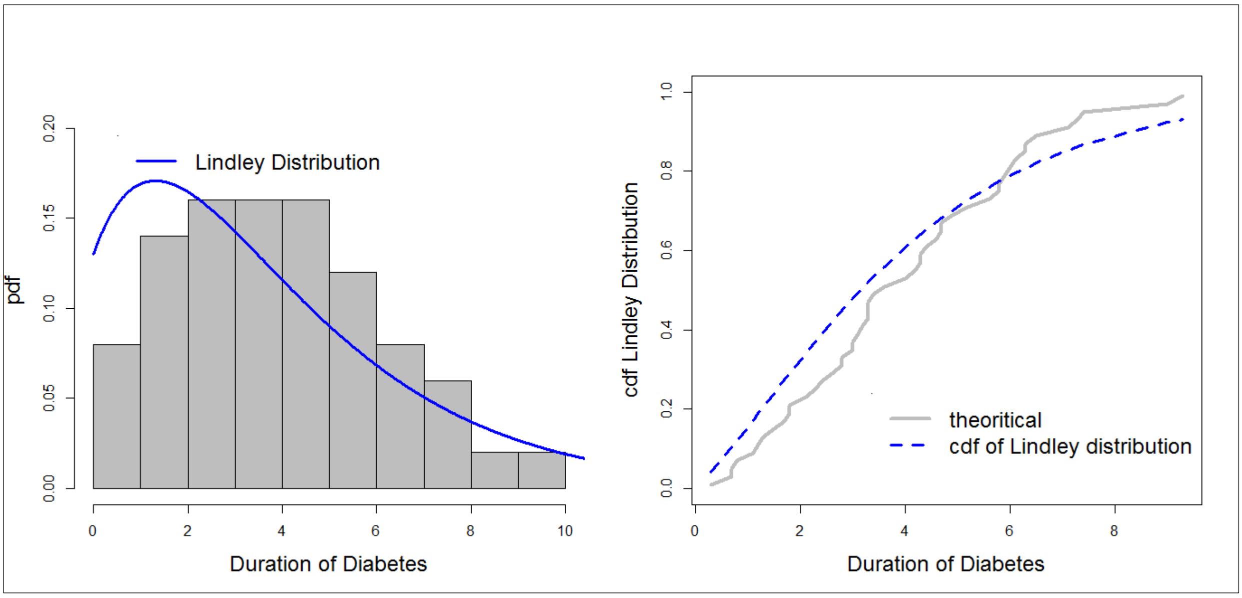

Figure 1: fitted pdf and cdf of LD distributions, respectively

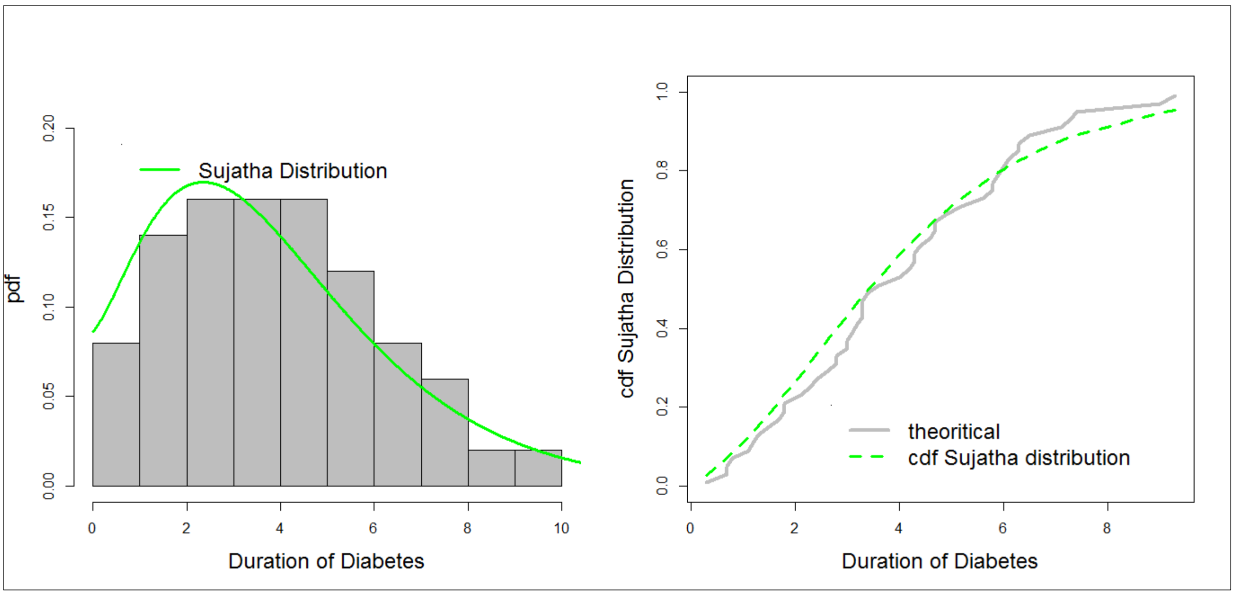

Figure 2: fitted pdf and cdf of SD distributions, respectively

The substantial difference in the goodness-of-fit test for this graphical model is illustrated in Figures 3 and 4, which present the probability density function (PDF) plot and cumulative distribution function (CDF) plot for the PL and PS probability models, respectively. The PDF plot of both models effectively captures the histogram of the data, while the CDF plot demonstrates that the model's CDF closely approximates the theoretical CDF.

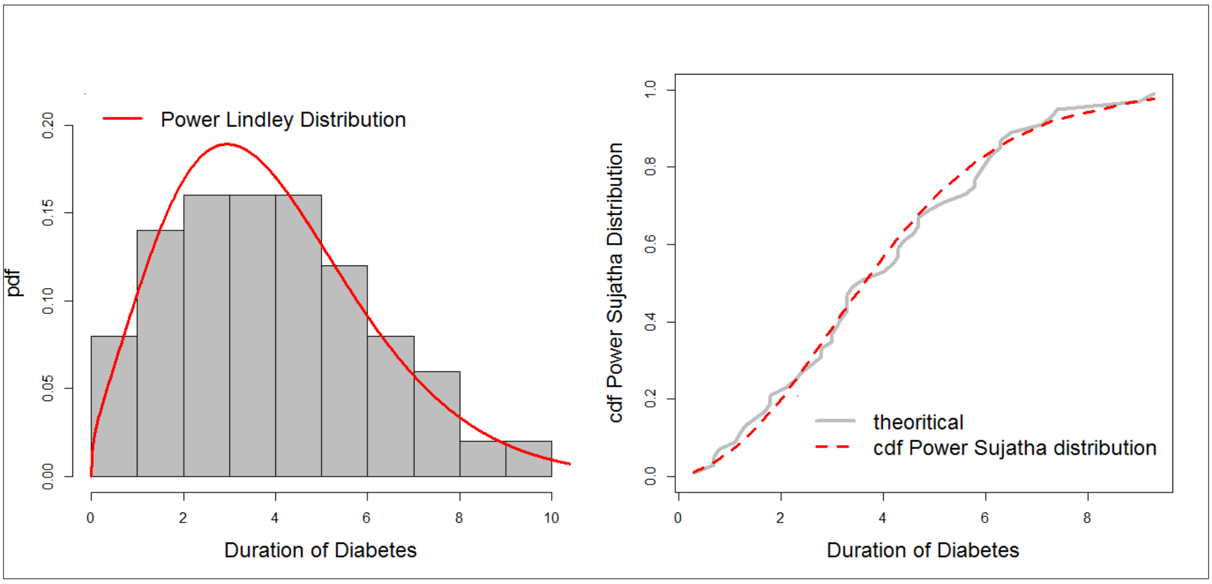

Figure 3: fitted pdf and cdf of PL distributions, respectively

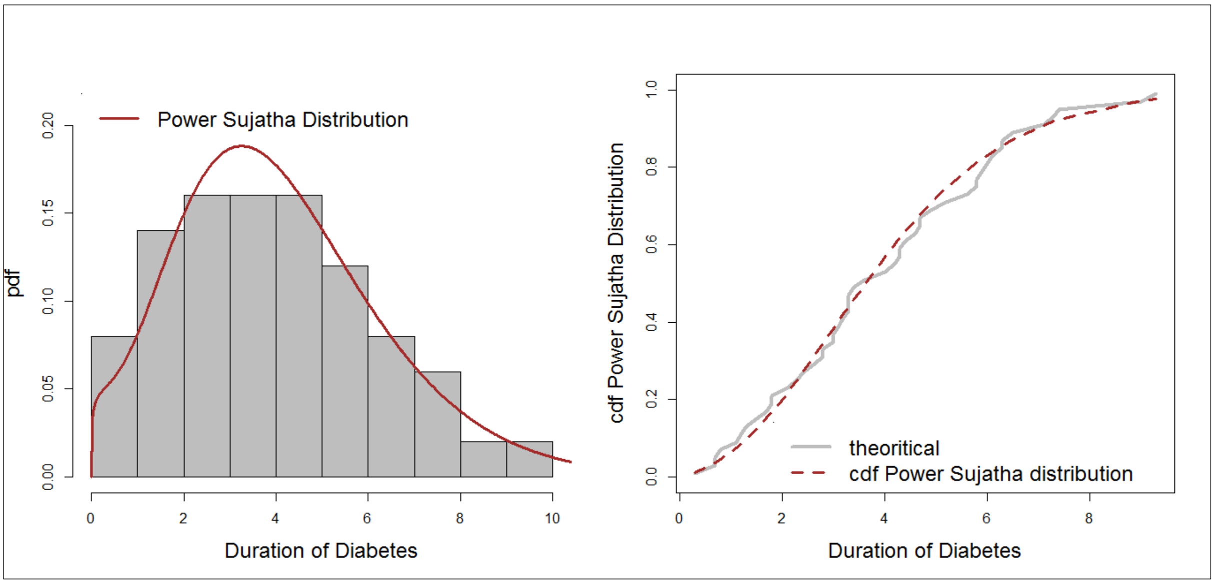

Figure 4: fitted pdf and cdf of PS distributions, respectively

CONCLUSION

The inclusion of additional parameters in a probability model, resulting from the transformation of random variables, can enhance the model's goodness of fit. This improvement is evidenced by a decreasing Akaike Information Criterion (AIC) value, as well as by the enhanced probability density function (pdf) and cumulative distribution function (cdf) plots, which better capture the histogram and approximate the theoretical cdf, respectively.

REFERENCES

1. Shanker, R. (2016). Amarendra Distribution and Its Applications. *American Journal of Mathematics and Statistics*, *6*(1), 44-56.
2. Shanker, R. (2016). Devya Distribution and Its Applications. *International Journal of Statistics and Applications*, *6*(4), 189-202 DOI: 10.5923/j.statistics.20160604.01

3. Shanker, R. (2016). Shambhu distribution and its Applications, *International Journal of Probability and Statistics*, *5*(2), 48 -63.
4. Shanker, R. (2015). Akash distribution and its Applications, *International Journal of Probability and Statistics*, *4*(3), 65-75.
5. Shanker, R. (2015). Shanker distribution and its Applications, *International Journal of Statistics and Applications*, *5*(6), 338-348.
6. Shanker, R. (2016). Aradhana distribution and its Applications, *International Journal of Statistics and Applications*, *6*(1), 23-34.
7. Shanker, R. (2017). Rama distribution and its Application, *International Journal of Statistics and Applications*, *7*(1) 26-35.
8. Grover, G., & Sabharwal, A. (2012). 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), 128.
9. Ummu, A., Rado, Y., Muhammad, M., & Rahmadeni, S. (2021). Time Data Modeling for Diabetics During the Covid 19 Pandemic Using Several Sensitive Distributions (Weibull, Gamma, and Normal Logs), *Applied Mathematical Sciences*, *15*, 717-724.
10. Usvita, M. L., Adnan, A., & Yendra, R. (2021). The Modelling Survival Times for Diabetes Patient Using Exponential, Weibull and Rayleigh-Lomax Distribution. *International Journal of Mathematics Trends and Technology-IJMTT*, *67*.
11. Sutriana, R. Y., & Rahmadeni, M. M. (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
12. Fajar, F. M., Rado, Y., Ari Pani, D., & Rahmadeni, M. M. (2024). Duration Diabetes Survival Times Modelling Using Some Extended Lomax Distribution. *Engineering and Technology Journal*, *09*(11), 5469-5474
13. Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102-107.
14. Shanker, R. (2016). Sujatha distribution and Its Applications. *To appear in "Statistics in Transition new Series" 17*(3).
15. Ghitany, M. E., Al- Mutairi, D. K., Balakrishnan, N., & Al-Enezi, L.J. (2013). Power Lindley distribution and Associated Inference. *Computational Statistics and Data Analysis*, *64*, 20-33.
16. Shanker, R., & Shukla, K. K. (2019). A Two- Parameter Power Sujatha distribution with Properties and Application. *International Journal of Mathematics and Statistics*, *20*(3), 11–22.