

Article

# Log-generalized exponential-Poisson regression model and its application to real data

Kaihong Dong<sup>1\*</sup>, Renwang Liao<sup>2</sup>, Mustapha Muhammad<sup>3\*</sup>, and Jinsen Xiao<sup>4</sup>

<sup>1</sup> Department of Mathematics, Guangdong University of Petrochemical Technology (GDUPT), Maoming, China; 1198575671@qq.com

<sup>2</sup> Department of Mathematics, Guangdong University of Petrochemical Technology, Maoming, China; liaorenwang0624@gmail.com

<sup>3</sup> Department of Mathematics, Guangdong University of Petrochemical Technology, Maoming, China; mmuhammad.mth@buk.edu.ng

<sup>4</sup> Department of Mathematics, Guangdong University of Petrochemical Technology, Maoming, China; jinsenzhao@gdupt.edu.cn

+ This work is a project submitted to the Department of Mathematics, GDUPT, by the first author.

\* Correspondence: [mmmahoud12@sci.just.edu.jo](mailto:mmmahoud12@sci.just.edu.jo), <https://orcid.org/0000-0003-0329-5898>

**Abstract:** In this paper, we propose a log-generalized exponential Poisson (LGEP) regression model based on generalized exponential Poisson distribution, and some of its structural properties are derived, including distribution function, density function, quantile function, and survival function. Secondly, we developed maximum likelihood estimation (MLE) for the model parameters estimation, also, performed some simulation analysis for different parameter settings and sample sizes to evaluate the performance of the MLEs by analyzing the mean squared error of the MLEs. Finally, we use two set of experimental data, one on the anesthesia response of guinea pigs, and the second on the percentage body fat in human to examine the performance of the LGEP regression model. The results show that the LGEP regression model provides more reliable results and has a better fit than the log-exponential Poisson, log-generalize exponential, and log-generalize extended exponential regression models. In the first data the proposed model reveals that the dosage of the anesthetic drug was shown to exert a positive effect on guinea pigs sleep duration once it reached an adequate level; also, for the second data, the proposed model suggests that both gender and BMI positively influence the percentage body fat across the subjects.

**Keywords:** generalized exponential Poisson distribution; log-linear regression models; maximum likelihood estimation; simulation; residual analysis.

Received: 27 October 2025; Revised: 22 November 2025; Accepted: 19 January 2026; Published: 2 February 2026



Copyright: ©2026 the Author(s). Published by JSSCI. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Journal Abbreviation: J. Stat. Sci. Comput. Intell.

## 1. Introduction

Regression analysis involves the use of regression models and helps us make predictions by establishing the relationship between dependent (response) variables and independent variables, enabling us to estimate future outcomes or values. Regression models provide insights into the nature and strength of relationships

between variables, helping us understand how changes in one variable affect another. They also support the decision-making process by offering quantitative guidance, allowing us to optimize strategies, allocate resources effectively, and identify influential factors within a system (see, [1, 2]). There are several methodologies of developing regression models depending on the applied studies such as quantile regression models (see, [3, 4, 5]), proportional hazards regression for survival modeling (see, [6, 7]), for some other regressions (see, [subsec. 4.3, 8]) among others. In this study we will focus on log-linear regression modeling.

There has been numerous parametric log-regression model in the literature, recently, [9] introduced log-generalized exponential-exponential regression model. Log-weighted exponential [10], log-extended exponential [11], log-exponentiated generalized Weibull exponential [12], log-new extended Weibull [13], log-new weighted Burr XII [14], log-sine alpha power-G [15], log-modified half-logistic [16], among others.

Log transformation is commonly applied in regression models especially those involving probabilistic distributions to stabilize variance, linear relationships, and ensure parameter constraints. When the response variable or parameters of a distribution are strictly positive, taking the natural logarithm helps convert nonlinear relationships into a linear form that can be modeled more easily. The log-transformed scale or rate often has a more stable variance and a more linear relationship with covariates. This reduces heteroscedasticity and increases interpretability of regression coefficients. When the response variable is highly skewed as frequently seen in lifetime, count, or reliability data, the log transformation makes the data more symmetric, which improves regression performance.

In this work, we construct a log-generalized exponential-Poisson regression model based on the generalized exponential-Poisson distribution [17] and discuss its important properties and data-generation process. We apply the maximum likelihood estimation method and use R 4.3.0 to estimate the parameters of the regression model and the coefficients of the linear predictor. Simulation results confirm the effectiveness of maximum likelihood estimation. Finally, we apply the proposed model to two experimental datasets, one on guinea-pig anesthesia response and the other on percentage of fat in human body, then compare its performance with other existing models using several model-selection criteria and residual analyses.

The paper is organized as follows: in Section 2, the log-generalized exponential-Poisson (LGEP) distribution is derived. In addition, some important properties along with some graphical characteristics of its density function are presented. In Section 3, maximum likelihood estimation, numerical simulation studies for the model, and discusses residual analysis are discussed. In Section 4, we present the application of the log-GEP regression model to experimental data. Section 6 provides the conclusion.

## 2. LGEP regression model derivation

In this section, we derived the LGEP distribution as well as the LGEP-regression model and several important properties. The generalized-exponential-Poisson distribution (GEP) was proposed by [17], and has cumulative distribution function CDF and probability density function (PDF) given by

$$F(x) = \frac{1 - e^{-\lambda(1 - e^{-\beta x})^\alpha}}{1 - e^{-\lambda}} \quad (1)$$

and

$$f(x) = \frac{\lambda\alpha\beta}{1 - e^{-\lambda}} e^{-\lambda} (1 - e^{-\beta x})^\alpha (1 - e^{-\beta x})^{\alpha-1} e^{-\beta x},$$

respectively, where  $x, \alpha, \beta > 0$ ,  $\lambda \in R - \{0\}$ . the quantile function of the GEP is given by

$$Q(u) = \frac{\ln \left\{ 1 - \left[ \frac{\ln \left[ 1 - (u(1-e^{-\lambda}))^{\frac{1}{\alpha}} \right]}{-\lambda} \right]^{\frac{1}{\alpha}} \right\}}{-\beta}, \quad u \in (0, 1).$$

### 2.1. LGEP and Its Properties

Let  $X$  be a random variable following the GEP. We define a logarithmic transformation and use the parameters  $\alpha$  and  $\lambda$  to derive the cumulative distribution function and probability density function of the LGEP. Let  $Y = \sigma \ln X$ , this implies that  $X = e^{\frac{Y}{\sigma}}$ ; let  $\beta = e^{-\frac{\mu}{\sigma}}$ , by applying this transformation in (1) we have the LGEP model CDF as

$$F(y) = \frac{1-e^{-\lambda \left( 1 - e^{-e^{-\frac{y-\mu}{\sigma}}} \right)^{\alpha}}}{1-e^{-\lambda}}, \tag{2}$$

and the corresponding PDF of the LGEP is derive from (2) as

$$f(y) = \frac{\alpha\lambda}{\sigma(1-e^{-\lambda})} e^{\frac{y-\mu}{\sigma}} e^{-e^{-\frac{y-\mu}{\sigma}}} \left( 1 - e^{-e^{-\frac{y-\mu}{\sigma}}} \right)^{\alpha-1} e^{-\lambda \left( 1 - e^{-e^{-\frac{y-\mu}{\sigma}}} \right)^{\alpha}} \tag{3}$$

Where  $-\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0, \alpha > 0, \lambda \in R - \{0\}$ . Figure 1 below presents the graphs of the LGEP for selected parameter values.

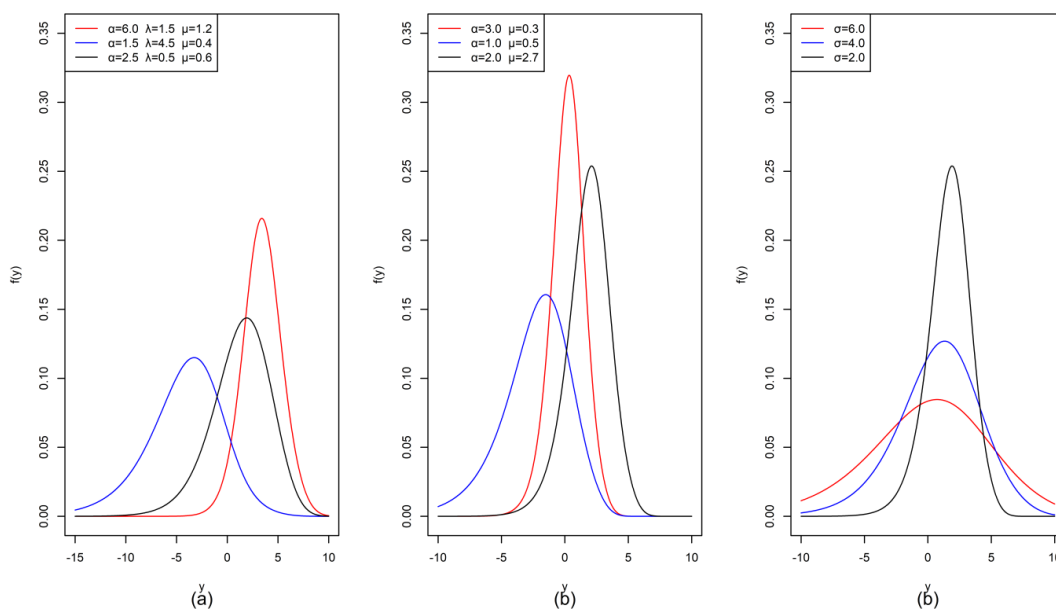


Figure 1. LGEP density plots for: (a)  $\sigma=3.9$ , (b)  $\lambda=2.5, \sigma=2$  (c)  $\sigma=2, \lambda=2.5, \mu=2.5$

The corresponding survival function  $S(y)$  can be obtained as

$$S(y) = 1 - F(y) = 1 - \frac{1 - e^{-\lambda \left( 1 - e^{-e^{-\frac{y-\mu}{\sigma}}} \right)^{\alpha}}}{1 - e^{-\lambda}}.$$

The quantile function is useful in both theoretical and applied research, and it is also employed for generating random data. The quantile function of the LGEP is derived as

$$Q(u) = \sigma \ln \left\{ -\ln \left\{ 1 - \left[ \frac{\ln [1 - u(1 - e^{-\lambda})]}{-\lambda} \right]^{\frac{1}{\alpha}} \right\} \right\} + \mu. \tag{4}$$

Table 1 below provides the algorithm for generating data from the LGEP distribution.

**Table 1.** The algorithm for LGEP random data generation

1. Select the parameters $\alpha, \lambda, \mu, \sigma$ ; set the number of sample size $n$
2. Let $u$ be from uniform distribution $u_i \sim U(0, 1)$ , for $i = 1, 2, \dots, n$
3. Sample the random variable using $y_i = Q(u_i, \alpha, \lambda, \sigma, \mu)$ , where $Q(\cdot)$ given in (4)

## 2.2. LGEP-regression model

Parametric models are type of statistical tool that can effectively fit data, thereby providing us with more accurate predictions. Using the LGEP we can define a flexible regression model as follows. Let  $\mathbf{x}_i^T = (1, x_{1i}, \dots, x_{pi})$  be the explanatory variables associated with the response variable  $Y_i$ . Based on LGEP distribution, a linear regression model can be constructed to link the dependent variable  $y_i$  with the explanatory variables  $x_i$ . Let

$$\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\varphi} + \boldsymbol{\sigma} z_i, \quad i = 1, \dots, n \tag{5}$$

The error term can be obtained from the density function of  $z_i$ , where  $\boldsymbol{\varphi} = (\varphi_0, \dots, \varphi_p)^T$  is an unknown parameter vector of the explanatory variable  $\mathbf{x}_i^T = (1, x_{1i}, \dots, x_{pi})$  associated with the location parameter  $\mu_i$ , and  $\sigma > 0, \alpha > 0, \lambda \in R - \{0\}$ . The parameter  $\sigma$  controls the scale, while  $\alpha$  and  $\lambda$  control the shape. Therefore, the location-parameter vector of the LGEP regression model can be expressed as the linear model  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\varphi}$  where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$  is a known model matrix. Substituting the regression model in equation (5) into the distribution function (2) of the LGEP distribution yields the corresponding distribution function as follows:

$$F(y) = \frac{1 - e^{-\lambda(1 - e^{-e^{-\frac{y - \mathbf{x}_i^T \boldsymbol{\varphi}}{\sigma}}})^\alpha}}{1 - e^{-\lambda}}, \tag{6}$$

The corresponding density function is given by

$$f(y) = \frac{\alpha \lambda}{\sigma(1 - e^{-\lambda})} e^{-\lambda(1 - e^{-e^{-\frac{y - \mathbf{x}_i^T \boldsymbol{\varphi}}{\sigma}}})^\alpha} (1 - e^{-e^{-\frac{y - \mathbf{x}_i^T \boldsymbol{\varphi}}{\sigma}}})^{\alpha-1} e^{-e^{-\frac{y - \mathbf{x}_i^T \boldsymbol{\varphi}}{\sigma}}} e^{\frac{y - \mathbf{x}_i^T \boldsymbol{\varphi}}{\sigma}} \tag{7}$$

Where  $-\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0, \alpha > 0, \lambda \in R - \{0\}$ , and  $\boldsymbol{\mu} = \mathbf{X}^T \boldsymbol{\varphi}$ . To propose the standard form of the LGEP regression model, we let a random variable  $Z = \frac{Y - \mu}{\sigma}$ , then the standardized PDF can be obtained as follows

$$f(z) = \frac{\lambda \alpha}{1 - e^{-\lambda}} e^z e^{-e^z} (1 - e^{-e^z})^{\alpha-1} e^{-\lambda(1 - e^{-e^z})^\alpha}$$

## 3. Parameter estimation, residual analysis and simulation studies

In this section, we discuss maximum likelihood estimation (MLE) for parameters of the LGEP regression model, with simulation assessment, additionally, we also discuss residual analysis

### 3.1. Maximum Likelihood estimation

Let  $Y_1, Y_2, \dots, Y_n$  be a random sample from the LGEP regression model, let  $\boldsymbol{\theta} = (\alpha, \lambda, \sigma, \boldsymbol{\varphi})^T$  be a vector of parameters of the model. The log likelihood function of the proposed regression model is given by

$$\begin{aligned} \log L(\theta) = & n \ln \alpha + n \ln \lambda - n \ln \sigma - n \ln(1 - e^{-\lambda}) + \sum_{i=1}^n -\lambda \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right)^\alpha \\ & + (\alpha - 1) \sum_{i=1}^n \ln \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right) - \sum_{i=1}^n e^{-\frac{y_i - X_i^T \varphi}{\sigma}} + \sum_{i=1}^n \frac{y_i - X_i^T \varphi}{\sigma}. \end{aligned}$$

The MLE can be obtained by maximizing, or by solving the following partial derivatives numerically using mathematical software (such as R or Matlab). Below, are the first-order partial derivatives of  $\log L(\theta)$  function with respect to the parameters respectively.

$$\frac{\partial \log L(\theta)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n -\lambda \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right)^\alpha \ln \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right) + \sum_{i=1}^n \ln \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right),$$

$$\frac{\partial \log L(\theta)}{\partial \lambda} = \frac{n}{\lambda} - \frac{ne^{-\lambda}}{1 - e^{-\lambda}} + \sum_{i=1}^n -\left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right)^\alpha,$$

$$\begin{aligned} \frac{\partial \log L(\theta)}{\partial \sigma} = & -\frac{n}{\sigma} + \sum_{i=1}^n -\lambda \alpha \left(1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right)^{\alpha-1} \left(-e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right) \left(-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}\right) \left(-\frac{y_i - X_i^T \varphi}{\sigma^2}\right) \\ & + \sum_{i=1}^n (\alpha - 1) \frac{1}{1 - e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}} \left(-e^{-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}}\right) \left(-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}\right) \left(-\frac{y_i - X_i^T \varphi}{\sigma^2}\right) + \sum_{i=1}^n \left(-e^{-\frac{y_i - X_i^T \varphi}{\sigma}}\right) \left(-\frac{y_i - X_i^T \varphi}{\sigma^2}\right) \\ & + \sum_{i=1}^n \left(-\frac{y_i - X_i^T \varphi}{\sigma^2}\right). \end{aligned}$$

In addition, we need  $\frac{\partial \log L(\theta)}{\partial \varphi_i}$ , where  $i = 0, 1, 2, \dots, p$ , is the first-order derivative of  $\varphi = (\varphi_0, \dots, \varphi_p)^T$

with respect to the vector of covariates parameters in order to obtain the MLEs. This optimization can be carried out in R through the *maxLik* package, which provides routines such as **BFGS** and **SANN** suitable for nonlinear likelihood maximization. Provided that the usual regularity requirements hold, particularly that the true parameter lies in the interior of the parameter space the estimator satisfies the asymptotic property:

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N_{p+3}(0, J(\theta)^{-1}),$$

Where  $N_{p+3}(0, J(\theta)^{-1})$  is a multivariate normal distribution and  $J(\theta)$  is the observed information matrix defined by the second derivatives of the log-likelihood function:  $J(\hat{\theta}) = [\partial^2 \log L(\theta) / \partial \theta \partial \theta^T]$ , an  $(p + 3) \times (p + 3)$  matrix evaluated at  $\theta$  can be used to construct this asymptotic approximation enables the construction of  $(100(1 - \varepsilon))\%$  confidence intervals for the components of  $\hat{\theta}$ .

### 3.2. Residual analysis

Residual analysis is a diagnostic technique used to judge how well a statistical model represents the data. It relies on examining the residuals the differences between observed values and the model's predictions to evaluate whether the underlying model assumptions are appropriate. Although model selection criteria can provide an overall measure of model performance, they often fail to capture localized discrepancies or structural issues within the model. The central idea behind residual analysis is to investigate whether the residuals behave according to the assumptions required by the model. Typically, these assumptions include independence, identical distribution, zero mean, and constant variance. When the residuals conform to these

properties, the model can be regarded as adequately specified. However, violations of these patterns indicate potential problems, suggesting the need to revisit the model formulation or reconsider the chosen parameters. Here, we discuss two important residual measures called Martingale residuals and adjusted deviance residuals

### 3.2.1. Martingale residuals

Martingale residuals serve as an important diagnostic measure for evaluating how well a survival model aligns with observed data. Unlike standard residuals which simply reflect the difference between observed and predicted responses martingale residuals also account for the subject's survival history up to the evaluation time. In essence, a martingale residual summarizes the deviation between an individual's observed survival outcome at a given moment and the expected outcome implied by the model, incorporating the individual's earlier survival pattern. Under a correctly specified model, the residuals should have mean zero and exhibit stable variance. Systematic departures from these characteristics suggest that the underlying model assumptions may be violated. The general form of a martingale residual is given by

$$r_{M_i} = 1 + \log[S(y_i)]$$

Consequently, the martingale residual for the LGEP framework takes the form

$$r_{M_i} = 1 + \log \left( 1 - \frac{1 - e^{-\lambda(1 - e^{-e^{-\frac{y_i - x_i^T \boldsymbol{\varphi}}{\sigma}})^{\alpha}}}}{1 - e^{-\lambda}} \right).$$

### 3.2.2. Adjusted Deviance Residuals

Adjusted deviance residuals provide another route for examining model fit, particularly when dealing with complex structures. Derived using bootstrap-based ideas, these residuals can be used to investigate how closely the model matches the data, to scrutinize key assumptions, and to detect potential outliers or influential cases. In the context of a Cox proportional hazards model that incorporates time-varying covariates, the deviance component residual can be expressed as

$$r_{D_i} = \text{sgn}(r_{M_i}) \{-2[r_{M_i} + \log(1 - r_{M_i})]\}.$$

### 3.3. Simulation studies

Simulation refers to the use of computer algorithms to generate random data in order to mimic the behavior of a given system or stochastic process. By conducting simulations, one can explore a wide range of scenarios and assess the impact of different parameter settings. In the context of MLE, simulation plays a crucial role. It allows us to repeatedly generate large numbers of datasets, thereby providing a basis for evaluating the accuracy, stability, and parameter sensitivity of the resulting estimators. This, in turn, helps identify estimation procedures that perform well and supports more reliable data analysis and prediction. In our study, we generate samples from the LGEP regression for various parameter configurations and sample sizes  $n=50, 100, 200$ . A Monte Carlo experiment with 1000 replications is carried out, and the estimators are assessed using the average bias and mean squared error (MSE). We consider the model where the covariates are drawn from a uniform distribution on the unit interval, the model is:

$$y_i = \varphi_0 + \varphi_1 x_i + \sigma z_i, \quad i = 1, 2, \dots, n, \quad ,$$

and the samples are generated according to equation (4). The bias and MSE are computed using the following formulas

$$\text{Bias} = \sum_{i=1}^n \frac{\hat{\theta}_i - \theta_i}{n}, \quad \text{and} \quad \text{MSE} = \sum_{i=1}^n \frac{(\hat{\theta}_i - \theta_i)^2}{n}.$$

Tables 2-4 report the results of the simulation study. From these tables, it is evident that the MLE performs very well under the LGEP setting: the mean squared errors of the estimators decrease as the sample size increases. This pattern indicates that, in practical applications, the maximum likelihood approach provides an effective procedure for estimating the parameters of the LGEP regression modeling.

**Table 2.** Numerical results for the simulation results -I

4. n	5. Parameter	6. Bias	7. MSE
8.	11. $\alpha = 1.5$	12. -0.9464	13. 0.8957
9.	14. $\lambda = 1.8$	15. -1.2860	16. 1.6539
10. 50	17. $\sigma = 0.5$	18. -0.4961	19. 0.2461
	20. $\varphi_0 = 1.0$	21. -0.9959	22. 0.9918
	23. $\varphi_1 = 1.5$	24. -1.4470	25. 2.0937
26.	29.	30. -0.9384	31. 0.8806
27.		32. -1.2847	33. 1.6505
28. 100		34. -0.4960	35. 0.2460
		36. -0.9812	37. 0.9629
		38. -1.4194	39. 2.0147
40.	43.	44. -0.8964	45. 0.8035
41.		46. -1.2763	47. 1.6290
42. 200		48. -0.4959	49. 0.2459
		50. -0.9859	51. 0.9608
		52. -1.3880	53. 1.9267
54.	57. $\alpha = 1.6$	58. -1.0468	59. 1.0958
55.	60. $\lambda = 1.6$	61. -1.0861	62. 1.1795
56. 50	63. $\sigma = 0.6$	64. -0.5961	65. 0.3553
	66. $\varphi_0 = 1.2$	67. -1.1351	68. 1.2884
	69. $\varphi_1 = 1.3$	70. -1.1668	71. 1.3614
72.	75.	76. -1.0388	77. 1.0791
73.		78. -1.0848	79. 1.1768
74. 100		80. -0.5959	81. 0.3551
		82. -1.1205	83. 1.2556
		84. -1.1494	85. 1.3211
86.	89.	90. -0.9968	91. 0.9937
87.		92. -1.0762	93. 1.1583
88. 200		94. -0.5958	95. 0.3550
		96. -1.1194	97. 1.2530
		98. -1.1283	99. 1.2730

**Table 4.** Numerical results for the simulation results -II

100. n	101. Parameter	102. Bias	103. MSE
104.	107. $\alpha = 1.4$	108. -0.8504	109. 0.7231
105.	110. $\lambda = 1.4$	111. -0.8855	112. 0.7841
106. 50	113. $\sigma = 0.4$	114. -0.3962	115. 0.1570
	116. $\varphi_0 = 1.4$	117. -1.3841	118. 1.9158
	119. $\varphi_1 = 1.1$	120. -1.0218	121. 1.0440
122.	125.	126. -0.8414	127. 0.7080

100. n	101. Parameter	102. Bias	103. MSE
123.		128. -0.8834	129. 0.7805
124. 100		130. -0.3961	131. 0.1569
		132. -1.3720	133. 1.8824
		134. -1.0056	135. 1.0112
136.	139.	140. -0.7930	141. 0.6289
137.		142. -0.8720	143. 0.7604
138. 200		144. -0.3960	145. 0.1568
		146. -1.3708	147. 1.8791
		148. -0.9843	149. 0.9688
150.	153. $\alpha = 1.6$	154. -1.0524	155. 1.1075
151.	156. $\lambda = 1.6$	157. -1.0801	158. 1.1797
152. 50	159. $\sigma = 0.6$	160. -0.5152	161. 0.3543
	162. $\varphi_0 = 1.6$	163. -1.5792	164. 2.4939
	165. $\varphi_1 = 1.4$	166. -1.3393	167. 1.7937
168.	171.	172. -1.0473	173. 1.0969
169.		174. -1.0852	175. 1.1778
170. 100		176. -0.5950	177. 0.3541
		178. -1.5775	179. 2.4886
		180. -1.3384	181. 1.7914
182.	185.	186. -1.0406	187. 1.0829
183.		188. -1.0680	189. 1.1406
184. 200		190. -0.5949	191. 0.3539
		192. -1.5716	193. 2.4698
		194. -1.3355	195. 1.7835

Table 5. Numerical results for the simulation results -III

196. n	197. Parameter	198. Bias	199. MSE
200.	203. $\alpha = 1.4$	204. -0.8490	205. 0.7208
201.	206. $\lambda = 1.4$	207. -0.8860	208. 0.7851
202. 50	209. $\sigma = 0.5$	210. -0.4952	211. 0.2459
	212. $\varphi_0 = 1.6$	213. -1.1886	214. 1.4128
	215. $\varphi_1 = 1.4$	216. -1.3435	217. 1.8049
218.	221.	222. -0.8358	223. 0.6986
219.		224. -0.8832	225. 0.7780
220. 100		226. -0.4958	227. 0.2458
		228. -1.1811	229. 1.3950
		230. -1.3296	231. 1.7678
232.	235.	236. -0.7964	237. 0.6342
233.		238. -0.8749	239. 0.7654
234. 200		240. -0.4957	241. 0.2457
		242. -1.1723	243. 1.3742
		244. -1.3045	245. 1.7018
246.	249. $\alpha = 1.5$	250. -0.9518	251. 0.9058

247.	252. $\lambda = 1.6$	253. -1.0857	254. 1.1787
248. 50	255. $\sigma = 0.4$	256. -0.3956	257. 0.1565
	258. $\varphi_0 = 1.4$	259. -1.3799	260. 1.9040
	261. $\varphi_1 = 1.6$	262. -1.5395	263. 2.3700
264.	267.	268. -0.9472	269. 0.8972
265.		270. -1.0848	271. 1.1769
266. 100		272. -0.3954	273. 0.1564
		274. -1.3779	275. 1.8987
		276. -1.5386	277. 2.3674
		282. -0.9393	283. 0.8823
278.	281.	284. -1.0673	285. 1.1392
279.		286. -0.3953	287. 0.1563
280. 200		288. -1.3732	289. 1.8856
		290. -1.5364	291. 2.3605

#### 4. Real data applications

In this section, we illustrate the practical usefulness of the LGEP regression model using a two real dataset. We assess the performance using some model selection criteria such as Akaike information criterion (AIC) defined by  $AIC = -2 \log L + 2k$ , Hannan–Quinn information criterion (HQIC) defined by  $HQIC = -2 \log L + 2k \ln(\ln(n))$ ,  $\log L$  is the log-likelihood value,  $k$  number of parameters, and  $n$  sample size. the competing models include log-exponential Poisson (LEP) regression model, log-generalize exponential (LGE) regression model, and log-generalize extended exponential (LGEE) regression model by [18].

##### 4.1. First data: Anesthetic response in guinea pigs

An experimental study on anesthetic response in guinea pigs reported in [19]. The original dataset contains 30 observations, including guinea pigs that exhibited no response as well as several clear outliers. For the present analysis, we focus on a cleaned subset of  $n = 20$  valid observations. The variables under study are:  $y_i$ , the sleep time of the guinea pigs (in minutes), and  $x_i$ , the administered dose of ketamine ( $i=1, 2, 3, \dots, 20$ ). We then fit the proposed LGEP regression model to these data, with the resulting fitted regression model given by:

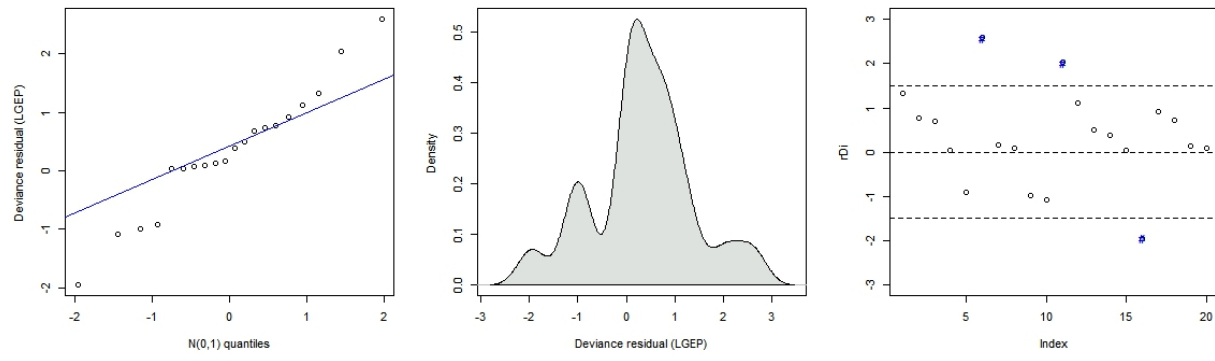
$$y_i = \varphi_0 + \varphi_1 x_i + \sigma Z_i,$$

Here, the response variable  $Y_i$  (conditional on the covariates) is modeled by an LGEP distribution. We implement model fitting in R version 4.3.0 . The resulting goodness-of-fit measures and parameter estimates are reported in the following Table 5.

**Table 5.** Numerical results of the MLEs, standard error (SE), p-values, log L, AIC, and HQC for the first data

Parameter	Log-GEP			Log-EP			Log-GE		
	Estimate	S.E.	P-value	Estimate	S.E.	P-value	Estimate	S.E.	P-value
$\alpha$	12.7747	18.3304	0.4859						
$\beta$							1.2273	0.2847	$7.26 \times 10^{-6}$
$\lambda$	2.7685	3.3239	0.4049	4.1694	2.4401	0.0875			
$\sigma$	1.0559	0.6306	0.0941	0.3099	0.0497	$4.63 \times 10^{-10}$	0.3006	0.6077	0.6209
$\varphi_0$	0.6189	1.1516	0.5910	2.0934	0.3305	$2.39 \times 10^{-10}$	0.4231	0.1203	0.0004
$\varphi_1$	0.4159	0.1134	0.0002	0.3786	0.1385	0.0063	35.3072	23.0040	0.1248

<b>Log L</b>	-8.3070	-9.6759
<b>AIC</b>	26.6140	27.3517
<b>HQIC</b>	27.5859	28.1292



**Figure 2.** LGEF residuals quantile-quantile (Q-Q) plot (left), residual density(center), and residual plot (right) for the first data

Table 5 reports the MLE, SEs, p-values, log-likelihoods, AIC and HQIC values for the three fitted models, the LGEF, LEP, and LGE regression models. Using a 5% significance level, the results show that the p-value for the LGE model is 0.1248, which exceeds 0.05, indicating that this model does not provide a statistically significant fit for the dataset. In contrast, the LGEF and LEP models yield p-values of 0.002 and 0.0063, respectively, both below the 5% threshold, demonstrating that these two models achieve statistically meaningful fits. Among the competing models, LGEF attains the smallest AIC and HQIC values, suggesting that it offers the best overall fit relative to LEP and LGE. This supports the use of the LGEF distribution for further modeling and analysis. The Table 5 also clearly indicates a positive relationship between ketamine dosage and the sleep duration of the guinea pigs. The estimated linear regression model is given by:

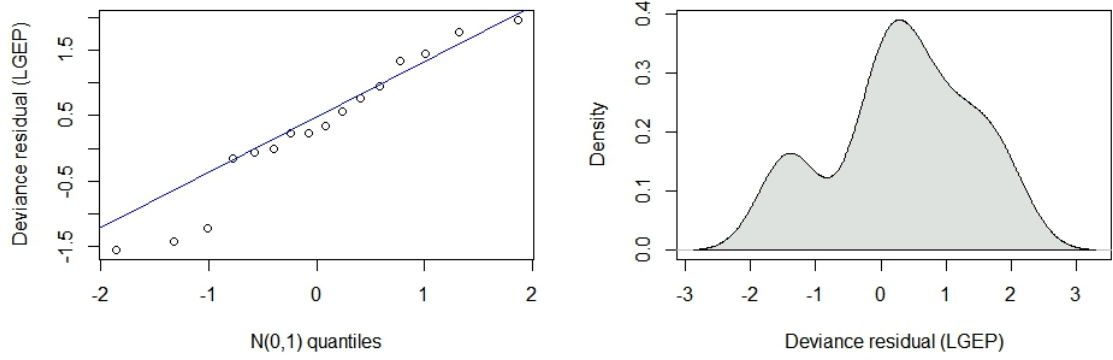
$$y_i = 0.6189 + 0.4159 x_i.$$

Referring to Figure 2, the left panel presents the Q-Q plot of LGEF regression residuals, the middle panel shows their density, and the right panel displays the residual plot. The residual plot reveals three points that clearly deviate from the general pattern and can be regarded as outliers. Consequently, these three observations were removed, and the LGEF model was refitted to evaluate the impact of the outliers on model performance. The new estimates of the LGEF regression are given in Table 6, and Figure 3 as follows.

**Table 6.** Updated estimates of the LGEF regression: MLEs, SE, p-values, logL, AIC and HQC for the first data

<b>Log-GEP</b>			
<b>Parameter</b>	<b>Estimate</b>	<b>S.E.</b>	<b>P-value</b>
<b><math>\alpha</math></b>	130.50	20.1470	$1.2 \times 10^{-9}$
<b><math>\lambda</math></b>	1.3780	2.4350	0.5710
<b><math>\sigma</math></b>	1.0160	0.2577	$8.1 \times 10^{-5}$
<b><math>\varphi_0</math></b>	0.0086	0.3605	0.9810
<b><math>\varphi_1</math></b>	0.3586	0.0692	$2.2 \times 10^{-7}$

<i>Log L</i>	2.2843
<i>AIC</i>	5.4313
<i>HQIC</i>	5.6291



**Figure 3.** Q-Q plot of residual (left) and density plot of residua for LGEP regression for the updated first data

The updated results, shown in Table 6, confirm that the removed observations were indeed anomalous. Also, the improvements clearly shown in the Figure 3. The values of Log L, AIC, and HQIC are all improved, and the overall parameter estimates remained largely unchanged after the removing the outliers. The newly estimated linear model is:

$$y_i^* = 0.0086 + 0.3586 x_i^*.$$

The interpretation is as follows: the estimated coefficient for the dosage variable  $x_i$  is 0.3586, indicating that changes in Ketamine dosage are associated with corresponding changes in sleep duration. Specifically, each one-unit increase in Ketamine dosage is expected to raise the predicted sleep time by approximately 0.3586 units.

#### 4.2. Second data: Human age and fatness

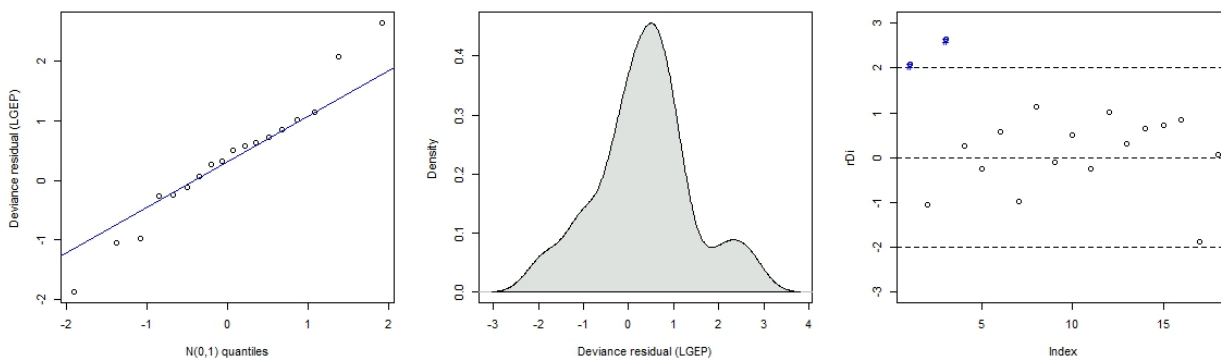
The dataset originates from research aimed at evaluating a novel technique for assessing human body composition, the data was originally from [20], and can be found in [21, 22]. It includes measurements of body fat percentage, age, and sex for 18 adults between 23 and 61 years old. The study involved “eighteen healthy adult participants, four younger men and fourteen women aged 25 to 60, none of whom had chronic illnesses, were on medication, or showed skeletal conditions suggestive of osteoporosis.” Body mass index (BMI) values used in the analysis were calculated from the height and weight information reported in [20]. The variables under study are:  $y_i$ , the body fat percentage; and the covariates are:  $x_{i1}$ , the age in years;  $x_{i2}$ , the gender;  $x_{i3}$ , the BMI in meters per kilogram-squared, for ( $i=1, 2, 3, \dots, 18$ ). For gender, we used for Female ( $F=1$ ), and Male ( $M=0$ ). We then fit the proposed LGEP regression model to these data, with the resulting fitted regression model given by:

$$y_i = \varphi_0 + \varphi_1 x_{i1} + \varphi_2 x_{i2} + \varphi_3 x_{i3} + \sigma z_i,$$

The resulting goodness-of-fit measures and parameter estimates are reported in the following Table 7.

**Table 7.** Numerical results of the MLEs, standard error (SE), p-values, log L, AIC, and HQC for the second data

Parameter	Log-GEP			Log-GEE		
	Estimate	S.E.	P-value	Estimate	S.E.	P-value
$\alpha$	0.3223	0.3932	0.4123			
$\beta$				1.4660	0.0518	$2.0 \times 10^{-16}$
$\lambda$	-3.2510	2.9314	0.2674			
$\sigma$	0.1215	0.0723	0.0929	0.0703	$1.3 \times 10^{-4}$	$2.0 \times 10^{-16}$
$\mathbf{b}$				0.1611	0.0413	$9.51 \times 10^{-5}$
$\varphi_0$	2.1618	0.3428	$2.8 \times 10^{-10}$	2.1420	$4.6 \times 10^{-5}$	$2.0 \times 10^{-16}$
$\varphi_1$	0.0083	0.0024	0.0007	0.0090	$1.05 \times 10^{-4}$	$2.0 \times 10^{-16}$
$\varphi_2$	0.2839	0.0798	0.0004	0.3160	0.3317	$2.0 \times 10^{-16}$
$\varphi_3$	0.0258	0.0119	0.0314	0.0271	$1.41 \times 10^{-4}$	$2.0 \times 10^{-16}$
Log L	5.6172			5.1022		
AIC	2.7657			3.7957		
HQIC	3.6251			4.6551		



**Figure 4.** LGEP residuals quantile-quantile (Q-Q) plot (left), residual density(center), and residual plot (right) for the second data

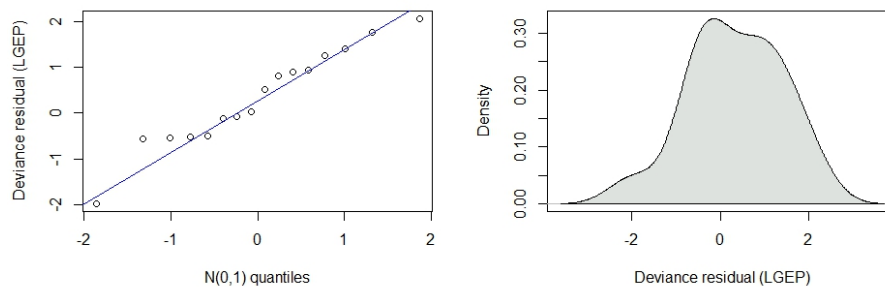
Table 7 summarizes the maximum likelihood estimates together with their standard errors, associated p-values, log-likelihoods, and the AIC and HQIC criteria for the two competing models, LGEP and LGEE. At the 5% significance threshold, both models yield p-values below 0.05, implying that each model can provide statistical representation of the data. However, the LGEP model achieves the lowest AIC and HQIC scores, indicating that it delivers a superior fit when compared with the LGEE alternative. These findings therefore justify selecting the LGEP distribution for subsequent inference and modeling. In addition, Table 7 reveals a clear positive significant association between all covariates and the subjects’ body fat percentages. The fitted linear model is expressed as follows:

$$y_i = 2.1618 + 0.0083 x_{i1} + 0.2839 x_{i2} + 0.0258 x_{i3}.$$

Figure 4 provides several diagnostic displays for the LGEP regression model: the Q–Q plot of the residuals appears on the left, the estimated residual density is shown in the central panel, and the residual scatter plot is displayed on the right. The residual plot highlights two observations that fall noticeably outside the general trend, suggesting the presence of outliers. These atypical data points were therefore excluded, and the LGEP model was re-estimated to assess how their removal influences the fitted results. The updated set of parameter estimates obtained after refitting the model is summarized in Table 8.

**Table 8.** Updated estimates of the LGEP regression: MLEs, SE, p-values, logL, AIC and HQC for second data

<i>Log-GEP</i>			
<i>Parameter</i>	Estimate	S.E.	P-value
$\alpha$	37.3691	37.4212	0.3179
$\lambda$	2.3032	4.3961	0.6003
$\sigma$	0.3746	0.1727	0.0302
$\varphi_0$	1.7079	0.3194	$8.91 \times 10^{-8}$
$\varphi_1$	0.0050	0.0032	0.1186
$\varphi_2$	0.2450	0.0882	0.0055
$\varphi_3$	0.0323	0.0090	0.0003
<i>Log L</i>	15.0238		
<i>AIC</i>	-16.0477		
<i>HQIC</i>	-15.7707		



**Figure 5.** LGEP residuals quantile-quantile (Q-Q) plot (left), residual density(center) for the updated second data

The revised estimates presented in Table 8 indicate that the excluded data points were genuine outliers. This conclusion is also visually supported by the improved diagnostic patterns in Figure 3. After removing the anomalous observations, notable enhancements are observed in the model’s fit. The log-likelihood increases from 5.6172 to 15.0238, while the AIC decreases from 2.7657 to -16.0477 and the HQIC from 3.6251 to -15.7707. Despite these substantial improvements in the fit statistics, the parameter estimates remain broadly consistent with those obtained before the outliers were removed. Moreover, at the 5% significance level, the coefficients for  $\varphi_2$  and  $\varphi_3$  continue to be statistically significant, whereas the coefficient for becomes  $\varphi_1$  non-significant (p-value = 0.1186), suggesting that age in years may not meaningfully influence body fat percentage in this dataset. The updated linear regression model is therefore:

$$y_i^* = 1.7079 + 0.2450 x_{i2}^* + 0.0323 x_{i3}^* .$$

The interpretation is as follows: Females have, on average, 0.2450 units higher predicted percentage body fat than males, holding BMI constant, and for every 1 unit increase in BMI, the predicted value of percentage body fat increases by 0.0323, assuming gender stays the same. Thus, model suggests that both gender and BMI positively influence the percentage body fat across the subjects.

## 5. Conclusions

In this work, we introduced the LGEP regression model and established several of its key structural properties, including its cumulative distribution function, density function, and survival function. The quantile function was also derived to facilitate random data generation. The LGEP family includes, as a special case, the

LEP distribution. To estimate the parameters of the LGEP regression model, we employed MLE. A simulation study was conducted to examine the performance of the MLEs, where MSE served as the evaluation metric. The results demonstrated that the MSE decreases as the sample size increases, confirming the desirable behavior of the estimators. To illustrate the practical relevance of the LGEP regression model, we analyzed data from a biomedical experiment on anesthetic responses in guinea pigs, and data on percentage fat in human body. The performance of the LGEP regression model was compared with that of the LEP, LGEE, and LGE regression models. Model assessment was based on several criteria, including significance on p-values, AIC, HQIC, and residual-based diagnostics analysis. The findings consistently indicated that the LGEP regression model achieved a superior fit and yielded more reliable estimates and data representation than the competing models. Furthermore, under the LGEP, in the first data the model reveals that the dosage of the anesthetic drug was shown to exert a positive effect on guinea pigs sleep duration once it reached an adequate level; also, for the second data, the model suggests that both gender and BMI positively influence the percentage body fat across the subjects. This highlights the flexibility and potential of the LGEP model in analyzing experimental and biomedical data.

We anticipate that the LGEP model will serve as a valuable tool in mathematical and applied statistical research. Future work may explore new regression formulations by incorporating different link functions, or extend the model using alternative estimation techniques such as Bayesian methods or least-squares approaches. Additionally, applying the LGEP model to a wider range of experimental datasets will further demonstrate its practical utility.

**Author contributions:** Conceptualization, K.D, M.M.; Methodology, K.D,R,L,M.M,J.X.; Writing-original draft, K.D,M.M.; Writing—review and editing, K.D, R.L,M.M,J.X.

**Funding Statement:** This work was funded by the Characteristic Innovation Project of Guangdong Province Ordinary University (2023KTSCX089), and Guangdong Provincial Education Science Planning Project (2024GXJK561).

**Acknowledgments:** The authors would like to express their appreciation to the Editor, and Reviewers for their time and comments, which significantly enhanced the quality of the paper.

**Conflict of interest:** The authors declare no conflict of interest.

## References

- [1] M. P. Allen, *Understanding regression analysis*, Springer Science & Business Media, 2004.
- [2] Fahrmeir, Ludwig, et al. "Regression models." *Regression: Models, methods and applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2022. 23-84.
- [3] Muhammad M, Sallam Salem Abdalla G, Faal A, Almetwally EM, Elgarhy M. A Flexible Approach to Quantile Regression Modeling With Unit Burr-XII-Poisson and Its Applications to Cancer, Chemotherapy, and Energy Data. *Engineering Reports*. 2025;7(8):e70312.
- [4] Muhammad M, Abba B, Xiao J, Alsadat N, Jamal F, Elgarhy M. A new three-parameter flexible unit distribution and its quantile regression model. *IEEE Access*. 2024;23.
- [5] Muhammad M. A new three-parameter model with support on a bounded domain: Properties and quantile regression model. *Journal of Computational Mathematics and Data Science*. 2023 Jan 1;6:100077.

- [6] Muse AH, Ngesa O, Mwalili S, Alshanbari HM, El-Bagoury AA. A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data. *Journal of Healthcare Engineering*. 2022;2022(1):2051642.
- [7] Sadiq, I. A., Kajuru, J. Y., Doguwa, S. I., Yahaya, S. S., Gambo, Y. Y., Hephzibah, A. A., ... Bello, A. (2025). Survival analysis in advanced lung cancer using Weibull survival regression model: estimation, interpretation, and clinical application. *Journal of Statistical Sciences and Computational Intelligence*, 1(2), 106–123.
- [8] Panitanarak, U., Ishaq, A. I., Usman, A., Sadiq, I. A., & Mohammed, A. S. (2025). The modified sine distribution and machine learning models for enhancing crude oil production prediction. *Journal of Statistical Sciences and Computational Intelligence*, 1(1), 29–45
- [9] Muhammad, M., Abba, B., Muhammad, I., Bakouch, H. S., & Xiao, J. (2025). A versatile family of distributions: Log-linear regression model and applications to real data. *Kuwait Journal of Science*, 52(2), 100385.
- [10] Altun, E. (2021). The log-weighted exponential regression model: alternative to the beta regression model. *Communications in Statistics-Theory and Methods*, 50(10), 2306-2321.
- [11] Arasan, J., & Midi, H. (2024). Modified outlier diagnostics for the extended exponential regression model with interval and right-censored data. *Communications in Statistics - Theory and Methods*, 54(7), 1991–2004.
- [12] Klakattawi, H.S. A Novel Exponentiated Generalized Weibull Exponential Distribution: Properties, Estimation, and Regression Model. *Axioms* 2025, 14, 706.
- [13] Dawlah, Alsulami, and Amani S. Alghamdi. "A New Extended Weibull Distribution: Estimation Methods and Applications in Engineering, Physics, and Medicine." *Mathematics* 13.20 (2025): 3262.
- [14] Anafo, Abdulzeid Yen, Ocloo, Selasi Kwaku, Nasiru, Suleman, New Weighted Burr XII Distribution: Statistical Properties, Applications, and Regression, *International Journal of Mathematics and Mathematical Sciences*, 2024(14), 4098771.
- [15] Alghamdi, A.S.; ALoufi, S.F.; Baharith, L.A. The Sine Alpha Power-G Family of Distributions: Characterizations, Regression Modeling, and Applications. *Symmetry* 2025, 17, 468.
- [16] E. Hussam, R. A. Aldallal, L. Prasad Sapkota, and A. M. Gemeay, " A Modified Half-Logistic Distribution With Regression Analysis," *Engineering Reports* 7( 11), 2025: e70487
- [17] Ristić MM, Nadarajah S. A new lifetime distribution. *Journal of Statistical Computation and Simulation* 2012. DOI:10.1080/00949655.2012.697163
- [18] Renwang Liao, Kaihong Dong, Mustapha Muhammad, Jinsen Xiao. Log-generalized extended exponential (LGEE) regression model and its application to dielectrics breakdown strength data. *Advances and Applications in Statistics*.2023 (Accepted manuscript)
- [19] Bailey R, Summe J, Homer L, et al. A Model for Analysis of the Anesthetic Response. *Biometrics*, 1978. 223-32.
- [20] R. B. Mazess, W. W. Pepler, and M. Gibbons. Total body composition by dualphoton (<sup>153</sup>Gd) absorptiometry. *American Journal of Clinical Nutrition*, 1984. 40, 834–839.
- [21] D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1994) *A Handbook of Small Data Sets*, London: Chapman and Hall. Dataset 17.
- [22] Dunn PK, Smyth GK. GLMsData: generalized linear model data sets. R package version. 2018.



**Disclaimer/Publisher's Note:** The views, opinions, and content expressed in all articles are solely those of the respective author(s) and contributor(s) and do not necessarily reflect those of the JSSCI, its editors, or the publisher. JSSCI and its editorial team assume no responsibility for any harm or damage resulting from the use of information, methods, or products mentioned in the publication.