mathematics

MDPI

Article

A Bimodal Exponential Regression Model for Analyzing Dengue Fever Case Rates in the Federal District of Brazil

Nicollas S. S. da Costa [1,*], Maria do Carmo Soares de Lima [2] and Gauss Moutinho Cordeiro [2]

[1] Coordenadoria Estratégica de Dados de Pessoal, Decanato de Gestão de Pessoas, Universidade de Brasília, Campus Darcy Ribeiro, Brasília 70910-900, Brazil
[2] Departamento de Estatística, Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Cidade Universitária, Recife 52070-040, Brazil
* Correspondence: nicollascosta@unb.br

Abstract: Dengue fever remains a significant epidemiological challenge globally, particularly in Brazil, where recurring outbreaks strain healthcare systems. Traditional statistical models often struggle to accurately capture the complexities of dengue case distributions, especially when data exhibit bimodal patterns. This study introduces a novel bimodal regression model based on the log-generalized odd log-logistic exponential distribution, offering enhanced flexibility and precision for analyzing epidemiological data. By effectively addressing multimodal distributions, the proposed model overcomes the limitations of unimodal models, making it well suited for public health applications. Through regression analysis of dengue case data from the Federal District of Brazil during the epidemiological weeks of 2022, the model demonstrates its capacity to improve the fit of the disease rate. The model's parameters are estimated using maximum likelihood estimation, and Monte Carlo simulations validate their accuracy. Additionally, local influence measures and residual analysis ensure the proposed model's goodness-of-fit. While this innovative regression model offers substantial advantages, its effectiveness depends on the availability of high-quality data, and further validation is necessary to confirm its applicability across diverse diseases and regions with varying epidemiological characteristics.

Keywords: dengue fever; epidemiological data; generalized odd log-logistic family; maximum likelihood; regression model; simulation

MSC: 62E99; 62J99; 62P10

check for updates

Citation: da Costa, N.S.S.; Lima, M.d.C.S.d.; Cordeiro, G.M. A Bimodal Exponential Regression Model for Analyzing Dengue Fever Case Rates in the Federal District of Brazil. *Mathematics* **2024**, *12*, 3386. https://doi.org/10.3390/math12213386

Academic Editor: Chang-Xing Ma

Received: 8 August 2024
Revised: 13 September 2024
Accepted: 18 September 2024
Published: 29 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Dengue fever presents a major challenge to epidemiology worldwide, especially in Brazil, where recurring epidemics occur in certain endemic regions. The rapid, uncontrolled growth of metropolitan populations and the lack of awareness and preventive measures are factors that contribute to the spread of the disease and place a significant burden on the healthcare system. More than 100 countries face dengue fever as an endemic issue, affecting millions of people globally each year. One of the most critical factors driving the current global spread of the virus is the rapidly accelerating climate change. Lately, ref. [1] proposed a novel class of discrete distributions and examined eight different datasets illustrating mortality, infection, and medication statistics. The results demonstrated its superiority over typical discrete modeling options in terms of model fit and flexibility, particularly in handling heavy-tailed datasets. In this context, numerous applications of generalized extreme value (GEV) distributions can be found in epidemiological studies. Ref. [2] presented a review of the relationship between dengue fever and meteorological parameters, along with a meta-analysis investigating the impact of ambient temperature and precipitation. In [3], dengue case counts during outbreaks in Thailand were modeled using extreme value theory (EVT). A zero-inflated GEV regression model is applied to

the Vietnam dengue data in [4], estimating the infection risk for individuals based on covariates such as age and weight. In a related study, ref. [5] presented a GEV approach to investigate the frequency and intensity of extreme novel epidemics, including those similar to COVID-19. Extreme value statistics have also been used to predict severe influenza epidemics in real time [6]. Additionally, ref. [7] explored the extreme correlation between infectious disease outbreaks and crude oil futures. Finally, ref. [8] estimated the disease burden of dengue in endemic regions to analyze the influence of socioeconomic factors. It is recommended to allocate more resources to areas with high population expansion and urbanization. Furthermore, ref. [9] explored the characteristics and studied the temporal–spatial distribution of overseas imported dengue fever cases in outbreak provinces of China. Ref. [10] conducted a systematic review on myocardial manifestations related to dengue fever cases. In [11], dengue fever cases in the Brazilian state of Alagoas were modeled monthly using the GEV distribution. The findings underscore the importance of ongoing monitoring and assistance in this area. Therefore, this study focuses on dengue fever cases during the epidemiological weeks of 2022 in the Federal District of Brazil. Regression analysis is applied to study extreme events (epidemiological events that affect health centers and the economy), and the maximum likelihood method is used to estimate the parameters. The accuracy of the estimators is confirmed through Monte Carlo simulations. Local influence measures and residual analyses are employed to validate the goodness-of-fit of the proposed model.

The study introduces a new bimodal regression model based on the generalized odd log-logistic exponential (GOLLE) distribution. The broad flexibility of the generalized odd log-logistic (GOLL-G) class, which allows for modeling both its tails and skewness, combined with the exponential distribution's closed mathematical form, makes this novel regression model highly relevant for applications in many fields, including the analysis of extreme events. Traditional models often fail to capture environmental and socioeconomic variations, especially when the data is bimodal. Consequently, this model is particularly innovative, as it provides a more accurate fit for data that deviate from the assumptions of unimodal models. It is specifically designed to analyze epidemiological data with multimodal patterns in case rates and holds potential for use in public health contexts. Given this, the novel LGOLLE regression model has the potential to enhance accuracy in estimating disease case rates, leading to more timely public health decisions for prevention or treatment. However, while the model offers many advantages, it may require high-quality data, which could be a limitation in some cases. Additionally, its applicability to other diseases or regions with different epidemiological factors would require further validation.

The remaining work is organized as follows: Section 2 presents the GOLLE distribution, as discussed in [12,13]. It also explores the linear representation and its mathematical properties. The maximum likelihood estimate (MLE) is explained, along with a Monte Carlo simulation analysis demonstrating the estimators' consistency. Section 3 introduces the new LGOLLE distribution, employs maximum likelihood estimation, and conducts simulations to evaluate the MLEs' consistency. A novel regression model based on the LGOLLE distribution is proposed, incorporating location parametrization. Simulations are performed to investigate the behavior of the estimates, and the model's fit is evaluated using global influence measures and residual analysis. Section 4 applies the novel regression model to epidemiological data, presenting findings and comments. Finally, Section 5 offers concluding remarks.

2. Materials and Methods

Recently, the development of new distributions based on well-known ones has aimed to better capture the underlying distribution of the data, leading to more precise estimates of key quantities of interest.

The GOLL-G family, introduced by [14], is a versatile class of continuous distributions that can effectively model various types of data. The study demonstrated the advantages of this family, highlighting its flexibility in fitting skewed, bimodal, and asymmetric data

sets, as well as its ability to capture a wide range of hazard function shapes. Ref. [15] developed a novel bimodal normal regression model based on the GOLLN distribution to assess patients' survival time in the intensive care unit due to COVID-19 in a Brazilian hospital. Additionally, Ref. [16] investigated factors influencing county-level COVID-19 vaccination rates in Texas, United States, using the GOLLL regression model.

This class of distributions is based on the transformer-transformer (T-X) generator defined by [17]. Consider a baseline cumulative distribution function (cdf) $G(x) = G(x;\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ denotes an unknown parameter vector. The GOLL-G cdf is defined by integrating the log-logistic density function, namely

$$F(y) = \int_0^{\frac{G(x)^\theta}{1-G(x)^\theta}} \frac{\alpha w^{\alpha-1}}{(1-w)^2} dw = \frac{G(x)^{\alpha\theta}}{G(x)^{\alpha\theta} + [1-G(x)^\theta]^\alpha}, \tag{1}$$

where $\alpha > 0$ and $\theta > 0$ are two extra shape parameters.

The probability density function (pdf) corresponding to (1) can be expressed as

$$f(y) = \frac{\alpha\,\theta\,g(x)\,G(x)^{\alpha\theta-1}\,[1-G(x)^\theta]^{\alpha-1}}{\{G(x)^{\alpha\theta} + [1-G(x)^\theta]^\alpha\}^2}, \tag{2}$$

where $g(x) = g(x;\xi)$ is the baseline pdf.

These equations define key characteristics of the GOLL-G family, allowing it to effectively model a wide range of data types. The extra parameters α and θ are crucial in shaping the distribution. Table 1 reports several sub-models of Equation (1).

Table 1. Sub-models of the GOLL-G family of distributions.

α	θ	**Sub-Model**
-	1	Generalized log-logistic family [18]
1	-	Proportional reversed hazard rate family [19]
1	1	Baseline

The cdf and the pdf of the GOLLE distribution are defined in [12,13], respectively, inserting the cdf and the pdf of the exponential distribution, $G(y;\lambda)$, into Equations (1) and (2), as follows (for $y > 0$)

$$F(y;\alpha,\theta,\lambda) = \frac{\left(1-\mathrm{e}^{-\lambda y}\right)^{\alpha\theta}}{\left(1-\mathrm{e}^{-\lambda y}\right)^{\alpha\theta} + \left[1-\left(1-\mathrm{e}^{-\lambda y}\right)^\theta\right]^\alpha} \tag{3}$$

and

$$f(y;\alpha,\theta,\lambda) = \frac{\alpha\,\theta\,\lambda\,\mathrm{e}^{-\lambda y}\left(1-\mathrm{e}^{-\lambda y}\right)^{\alpha\theta-1}\left[1-\left(1-\mathrm{e}^{-\lambda y}\right)^\theta\right]^{\alpha-1}}{\left\{\left(1-\mathrm{e}^{-\lambda y}\right)^{\alpha\theta} + \left[1-\left(1-\mathrm{e}^{-\lambda y}\right)^\theta\right]^\alpha\right\}^2}, \tag{4}$$

where $\alpha, \theta, \lambda > 0$.

The corresponding hrf ($\tau(y) = f(y)/[1-F(y)]$) is easily determined as

$$\tau(y;\alpha,\theta,\lambda) = \frac{\alpha\theta\left(\lambda\mathrm{e}^{-\lambda x}\right)\left(1-\mathrm{e}^{-\lambda x}\right)^{\alpha\theta-1}}{\left[1-\left(1-\mathrm{e}^{-\lambda x}\right)^\theta\right]\left\{\left(1-\mathrm{e}^{-\lambda x}\right)^{\alpha\theta} + \left[1-\left(1-\mathrm{e}^{-\lambda x}\right)^\theta\right]^\alpha\right\}}. \tag{5}$$

In addition, Equations (3) and (4) do not involve complex mathematical functions, unlike the gamma and beta distributions. In Table 2, the sub-models obtained from Equation (4) are presented. Their ability to handle data fitting across a wide range of distributions demonstrates their versatility and applicability.

Table 2. Sub-models of the GOLLE distribution.

α	θ	Sub-Model
-	1	Odd log-logistic exponential (OLLE) distribution, see [18]
1	-	Exponentiated-exponential (EE) distribution, see [20]
1	1	Exponential (E) distribution

Figures 1 and 2 show plots of the Y's hrfs and histograms for selected parameters. One of the most notable properties of the GOLLE distribution is its ability to generate a wide range of hazard shapes, in contrast to the exponential hrf's constant behavior over time. Figure 1 illustrates many forms, including the inverse J-shape, increasing–decreasing, decreasing–increasing and bathtub. Figure 2 demonstrates that the model is effective for modeling non-normal data sets with diverse histogram patterns (e.g., asymmetric, heavy-tailed, multimodal).

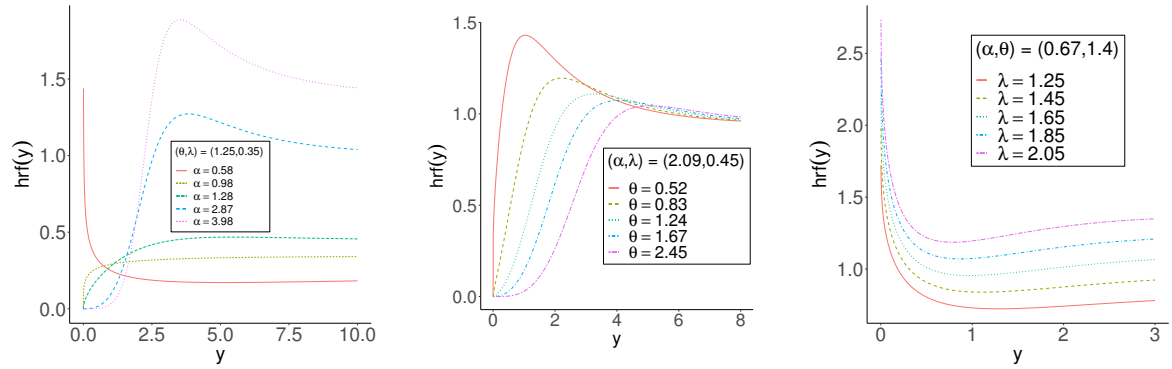


Figure 1. (**a**–**c**) GOLLE hrf for selected values.

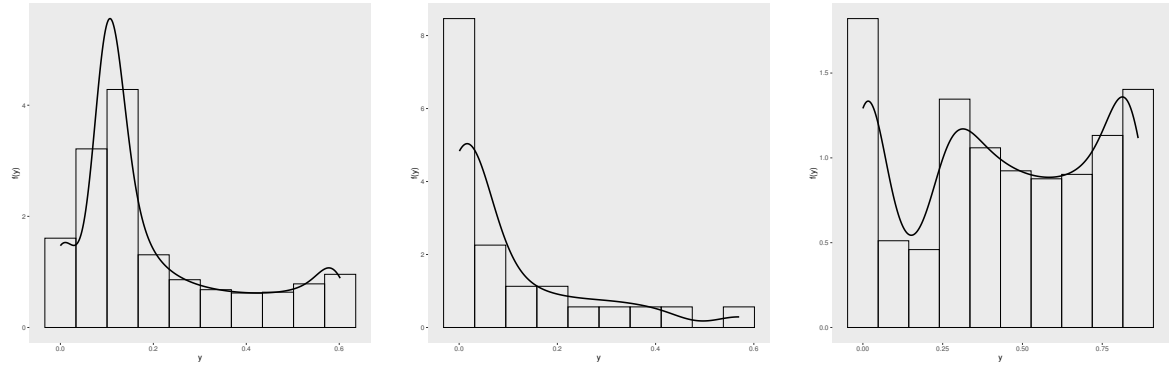


Figure 2. GOLLE histogram. (**a**) GOLLE (0.15, 73, 2.50). (**b**) GOLLE (0.22, 1.13, 7.50). (**c**) GOLLE (0.07, 120.13, 3.50).

2.1. Main Properties

This Section reviews the linear representation of the GOLLE distribution's density function, including the quantile function (qf), moments, and the moment generating function (mgf), as shown in [13].

2.1.1. Representation

Definition 1. *The GOLLE density (4) can be represented linearly using exponential densities as*

$$f(y;\alpha,\theta,\lambda^*) = \sum_{k,m=0}^{\infty} h_{k,m}\, g(y;\lambda^*), \tag{6}$$

where $g(\cdot)$ is the exponential density with a shared parameter $\lambda^* = \lambda^*(\lambda, m) = \lambda(m+1)$ and $h_{k,m}$ is defined below

$$h_{k,m} = \frac{(-1)^m(k+1)\binom{k}{m}}{(m+1)} b_k.$$

2.1.2. Quantile Function

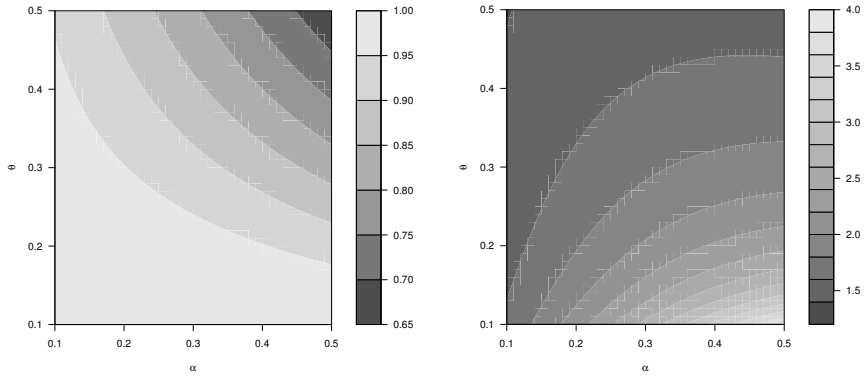
Definition 2. *The GOLLE qf of Y, used to simulate the density, is expressed by*

$$Q(u) = -\frac{1}{\lambda}\log[1 - \varepsilon_{\alpha.\theta}(u)], \tag{7}$$

where

$$\varepsilon_{\alpha,\theta}(u) = \left[\frac{\left(\frac{u}{1-u}\right)^{1/\alpha}}{1+\left(\frac{u}{1-u}\right)^{1/\alpha}}\right]^{1/\theta}.$$

Figure 3 displays Galton's skewness and Moors' kurtosis for different α and θ, with $\lambda = 1.58$. These plots illustrate that as the parameter α increases, the distribution becomes more right-skewed and leptokurtic, eventually reaching a minimum value.



(**a**) (**b**)

Figure 3. GOLLE distribution. (**a**) Galton's skewness. (**b**) Moors' kurtosis.

2.1.3. Moments

Definition 3. *The nth moment of the GOLLE distribution is defined by*

$$\mu'_n = E(Y^n) = \sum_{k,m=0}^{\infty} \frac{1}{\lambda^*} h_{k,m} = \sum_{k,m=0}^{\infty} \frac{(-1)^m(k+1)\binom{k}{m}}{(m+1)\lambda^*} b_k.$$

2.1.4. Generating Function

Definition 4. *The mgf of the GOLLE density can be expressed as*

$$M_Y(t) = \sum_{k,m=0}^{\infty} \frac{\lambda^*}{\lambda^* - t} h_{k,m} = \sum_{k,m=0}^{\infty} \frac{(-1)^m(k+1)\lambda^*}{\lambda^* - t} b_k, \quad t < \lambda^*.$$

2.1.5. Estimation

The MLEs of the GOLLE parameters vector $\boldsymbol{\psi} = (\alpha, \theta, \lambda)^\top$ are calculated from a complete sample $y_1, \ldots, y_n$ of the Equation (4) by maximizing the log-likelihood function

$$\begin{aligned}\mathcal{L}_n(\boldsymbol{\psi}) = {} & n\log(\alpha\theta\lambda) - \lambda\sum_{i=1}^{n} y_i + (\alpha\theta - 1)\sum_{i=1}^{n}\log(1 - \mathrm{e}^{-\lambda y_i}) + (\alpha - 1)\sum_{i=1}^{n}\log\left[1 - (1 - \mathrm{e}^{-\lambda y_i})^\theta\right] \\ & - 2\sum_{i=1}^{n}\log\left\{(1 - \mathrm{e}^{-\lambda y_i})^{\alpha\theta} + \left[1 - (1 - \mathrm{e}^{-\lambda y_i})^\theta\right]^\alpha\right\}.\end{aligned} \tag{8}$$

Let's consider

$$A_i(\lambda) = A_i = 1 - e^{\lambda y_i}.$$

Therefore, the elements of the score vector can be formulated as follows

$$\begin{aligned} U_\alpha &= \frac{n}{\alpha} + \theta \sum_{i=1}^{n} \log(A_i) + \sum_{i=1}^{n} \log\left(1 - {A_i}^\theta\right) \\ &- 2\sum_{i=1}^{n} \frac{\theta \log(A_i) {A_i}^{\alpha\theta} + (1 - {A_i}^\theta)^\alpha \log(1 - {A_i}^\theta)}{{A_i}^{\alpha\theta} + (1 - {A_i}^\theta)^\alpha}, \end{aligned}$$

$$\begin{aligned} U_\theta &= \frac{n}{\theta} + \alpha \sum_{i=1}^{n} \log(A_i) - (\alpha - 1) \sum_{i=1}^{n} \frac{{A_i}^\theta \log(A_i)}{1 - {A_i}^\theta} \\ &+ \sum_{i=1}^{n} \frac{\alpha {A_i}^{\alpha\theta} \log(A_i) + (1 - {A_i}^\alpha)^\theta \log(1 - {A_i}^\alpha)}{{A_i}^{\alpha\theta} + (1 - {A_i}^\alpha)^\theta}, \end{aligned}$$

and

$$\begin{aligned} U_\lambda &= \frac{n}{\lambda} - \sum_{i=1}^{n} y_i + (\alpha\theta - 1) \sum_{i=1}^{n} \frac{(1 - A_i)}{A_i} - \theta(\alpha - 1) \sum_{i=1}^{n} \frac{(1 - A_i) A_i^{\theta-1}}{1 - A_i^\theta} \\ &- 2\alpha\theta \sum_{i=1}^{n} \frac{(1 - A_i)[A_i^{\alpha\theta-1} - A_i^{\theta-1}(1 - A_i^\theta)^{\alpha-1}]}{A_i^{\alpha\theta} + (1 - A_i^\theta)^\alpha}. \end{aligned}$$

Using a Newton–Raphson type method and setting the score equations $U_\alpha = U_\theta = U_\lambda = 0$, the MLEs can be calculated. Due to the complexity of the equations for the GOLLE model, analytical solutions are generally not feasible. Therefore, numerical optimization techniques are employed to solve these equations. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, implemented through the optim function in [21] (version 4.4.1), is used to maximize Equation (8). This method is well-suited for maximizing complex likelihood functions due to its balance between computational efficiency and robustness.

2.1.6. Simulation Study

In two scenarios, based on the settings depicted in Figures 1c and 2b, Monte Carlo simulations were generated by 1000 samples of varying sizes from the GOLLE distribution are utilized to assess the accuracy of MLEs. For each sample size, $n = \{50, 150, 300, 500, 750, 1000\}$, the average estimates (AEs), absolute biases (ABs), and root mean square errors (RMSEs) are computed, for each $(\alpha, \theta, \lambda)$ (Tables 3 and 4).

Table 3. Simulations results for GOLLE distribution for scenario 1.

Scenario 1—GOLLE (0.67, 1.40, 1.25)									
Par	n = 50			n = 150			n = 300		
	AE	AB	RMSE	AE	AB	RMSE	AE	AB	RMSE
α	0.817	0.147	0.750	0.742	0.072	0.381	0.724	0.054	0.292
θ	2.609	1.209	3.157	1.756	0.356	1.194	1.557	0.157	0.728
λ	1.729	0.479	1.364	1.393	0.143	0.713	1.307	0.057	0.508
Par	n = 500			n = 750			n = 1000		
α	0.707	0.037	0.212	0.676	0.006	0.155	0.681	0.011	0.135
θ	1.477	0.077	0.537	1.496	0.096	0.433	1.452	0.052	0.351
λ	1.269	0.019	0.392	1.301	0.051	0.317	1.273	0.023	0.267

Table 4. Simulations results for GOLLE distribution for scenario 2.

Scenario 2—GOLLE (0.22, 1.13, 7.50)									
Par		n = 50			n = 150			n = 300	
α	0.261	0.041	0.185	0.240	0.020	0.104	0.231	0.011	0.063
θ	1.413	0.283	0.877	1.212	0.082	0.465	1.151	0.021	0.308
λ	9.167	1.667	5.438	7.954	0.454	2.990	7.655	0.155	2.206
Par		n = 500			n = 750			n = 1000	
α	0.223	0.003	0.044	0.224	0.004	0.040	0.222	0.002	0.030
θ	1.155	0.025	0.223	1.141	0.011	0.197	1.141	0.011	0.156
λ	7.625	0.125	1.448	7.577	0.077	1.214	7.589	0.089	1.008

As predicted by the consistency requirement, the results in Tables 3 and 4 indicate that AEs approximate the real values and ABs and RMSE approach zero as n increases. It is notable that for scenario 1, all the estimates obtained when $n = 50$ were overestimated, while for scenario 2, the parameters θ and λ were overestimated. This shows the sensitivity of the model's parameters to some values, but in general, as the sample size increases, convergence towards the true values is achieved.

3. The Proposal LGOLLE Distribution

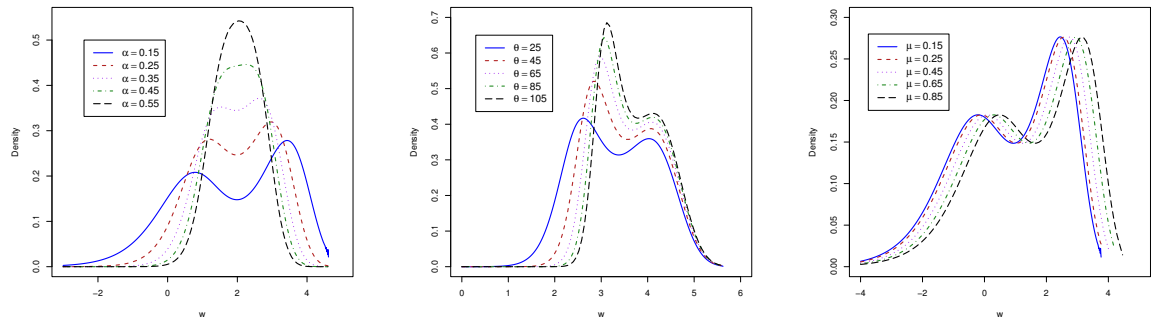
The LGOLLE distribution is proposed to address particular problems that arise when studying epidemiological data, especially in cases like dengue disease, where the data contains bimodal distributions. Traditional models usually use a unimodal distribution, which may not accurately reflect the complexity of the actual data.

Let Y be a random variable (rv) following the GOLLE density function (4). Define $W = \log(Y)$. Setting $\lambda = -e^{\mu}$, the density function of W is expressed as (for $y \in \mathbb{R}$)

$$f(y;\alpha,\theta,\mu) = \frac{\alpha\,\theta\,e^{[(y-\mu)-e^{y-\mu}]}\left[1-e^{-e^{y-\mu}}\right]^{\alpha\theta-1}\left[1-\left(1-e^{-e^{y-\mu}}\right)^{\theta}\right]^{\alpha-1}}{\left\{\left(1-e^{-e^{y-\mu}}\right)^{\alpha\theta}+\left[1-\left(1-e^{-e^{y-\mu}}\right)^{\theta}\right]^{\alpha}\right\}^{2}}, \tag{9}$$

where $\mu \in \mathbb{R}$.

The Equation (9) represents the LGOLLE distribution, denoted as $W \sim \text{LGOLLE}(y;\alpha,\theta,\mu)$, where μ is the location parameter. Thus, if $Y \sim \text{GOLLE}(y;\alpha,\theta)$ then $W = \log(Y) \sim \text{LGOLLE}(y;\alpha,\theta,\mu)$. The density function in Figure 4 is displayed for selected values of the parameters α, θ, and μ. These plots illustrate the versatility of the new distribution, demonstrating its skewed and bimodal shapes.



(**a**) (**b**) (**c**)

Figure 4. Plots of the LGOLLE density function for selected values. (**a**) For $\theta = 10.5$ and $\mu = 1$. (**b**) For $\alpha = 0.25$ and $\mu = 2$. (**c**) For $\alpha = 0.17$ and $\theta = 7.42$.

3.1. The New LGOLLE Regression Model

The novel regression model has wide potential in public health, notably, for epidemiologists, policymakers, and data scientists. The model's capacity to handle bimodal distributions enables more accurate predictions in complex epidemiological scenarios, making it a viable option for applications beyond dengue fever. While the recently developed regression model is relevant for diseases with bimodal distributions, it is essential to expand its application to more conditions or diseases for further validation.

Thus, let be $W \sim \text{LGOLLE}(\alpha, \theta, \mu)$ the density function, thus the rv $Z = Y - \mu$ is given by

$$f(z;\alpha,\theta,\mu) = \frac{\alpha\,\theta\,\mathrm{e}^{[z-\mathrm{e}^z]}\left[1-\mathrm{e}^{-\mathrm{e}^z}\right]^{\alpha\theta-1}\left[1-\left(1-\mathrm{e}^{-\mathrm{e}^z}\right)^{\theta}\right]^{\alpha-1}}{\left\{(1-\mathrm{e}^{-\mathrm{e}^z})^{\alpha\theta}+\left[1-(1-\mathrm{e}^{-\mathrm{e}^z})^{\theta}\right]^{\alpha}\right\}^2}, \tag{10}$$

and the rv $Z \sim \text{LGOLLE}(\alpha, \theta, 0)$.

In order to introduce a new regression structure in the class of models (10), the parameter μ_i is assumed to vary across observations through a regression structure expressed as

$$y_i = \mu_i + z_i, \quad i = 1, \ldots, n, \tag{11}$$

where the random error z_i has density function (10) and μ_i is parameterized by

$$\mu_i = \mu_i(\boldsymbol{x_i}^{\top}\boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\beta, \ldots, \beta_p)^{\top}$ is the parameter vector of dimension p associated with the explanatory variables $\boldsymbol{x_i}^{\top} = (x_{i1}, \ldots, x_{ip})$ for the location parameter.

3.2. Estimation

Except for μ_i, the components of the score vector for U_α and U_θ are the same as those obtained from the Equations presented in Section 2.1.5. The score component of the location parameter μ_i is defined to add the regression part as follows

$$\begin{aligned} U_{\mu_i} &= \sum_{i=1}^{\infty}\frac{\partial_{\boldsymbol{\beta}} f(z_i;\mu_i)}{f(z_i;\mu_i)} + (\alpha\theta-1)\sum_{i=1}^{\infty}\frac{\partial_{\boldsymbol{\beta}} F(z_i;\mu_i)}{F(z_i;\mu_i)} + \theta(1-\alpha)\sum_{i=1}^{\infty}\frac{\partial_{\boldsymbol{\beta}} F(z_i;\mu_i)GFz_i;\mu_i)^{\theta-1}}{1-F(z_i;\mu_i)^{\theta}} \\ &- 2\sum_{i=1}^{\infty}\partial_{\boldsymbol{\beta}} F(z_i;\mu_i)\frac{F(z_i;\mu_i)^{\alpha\theta-1}-F(z_i;\mu_i)^{\theta-1}[1-F(z_i;\mu_i)^{\theta}]^{\alpha-1}}{F(z_i;\mu_i)^{\alpha\theta}+[1-F(z_i;\mu_i)^{\theta}]^{\alpha}}, \end{aligned}$$

where $F(\cdot)$ and $f(\cdot)$ are the cdf and pdf of the exponential distribution, respectively, $\partial_{\boldsymbol{\beta}} f(z_i;\mu_i) = \partial_{\mu_i} f(z_i;\mu_i)\partial_{\boldsymbol{\beta}}\mu_i(z_i;\boldsymbol{\beta})$ and $\partial_{\boldsymbol{\beta}} F(z_i;\mu_i) = \partial_{\mu_i} F(z_i;\mu_i)\partial_{\boldsymbol{\beta}}\mu_i(z_i;\boldsymbol{\beta})$ denotes the derivatives of the parameter μ_i using the chain rule.

The MLE $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ of the regression model is calculated either by setting the score equations $U_\alpha = U_\theta = U_{\mu_i} = 0$ or using the optim routine in [21] (version 4.4.1).

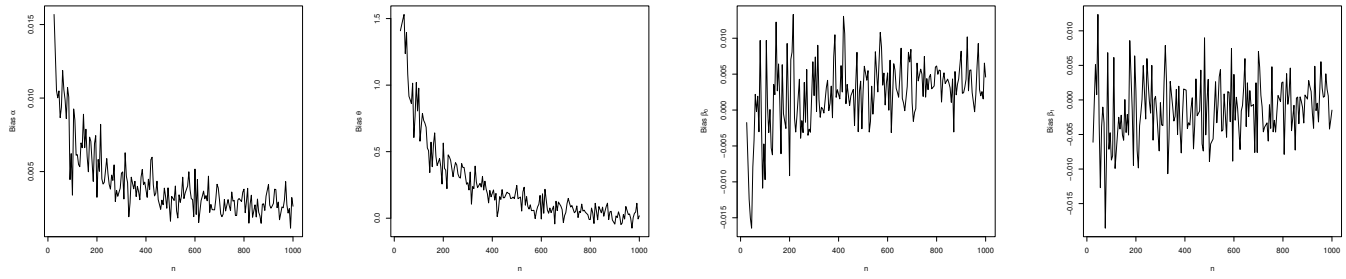
3.3. Regression Simulation Study

To show the accuracy of the MLEs for $\alpha = 0.25$, $\theta = 9.75$, $\beta_0 = 0.89$, and $\beta_1 = 1.15$, 1000 samples of size $n = \{25, \ldots, 1000\}$ were generated. The study is based on the following measurements: biases, mean square errors (MSEs), and average lengths (ALs). The measures are (for $\epsilon = \alpha, \theta, \lambda$)

$$Bias_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon), \quad MSE_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon)^2 \quad \text{and} \quad AL_\epsilon(n) = \frac{3.919928}{N}\sum_{i=1}^{N} s_{\hat{\epsilon}_i}.$$

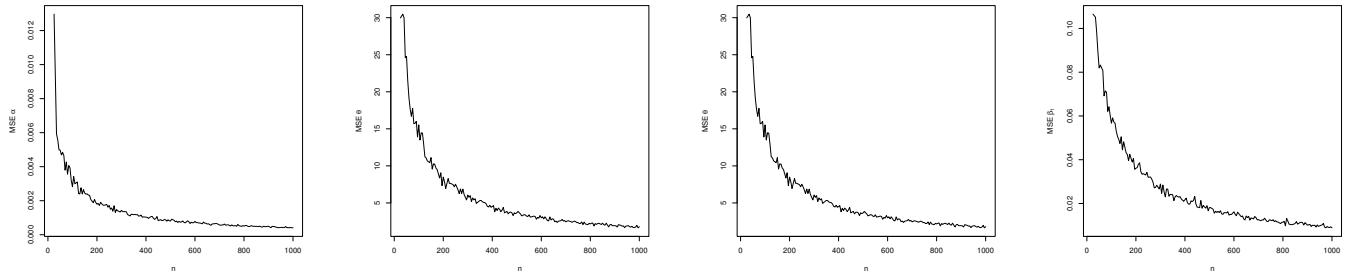
Figures 5–7 illustrate the relationship between these measures and n. As the sample size increases, biases, MSEs, and ALs tend toward zero. However, the estimate of β_0 consis-

tently exhibits an overestimation bias, while the estimate of β_1 demonstrates oscillatory behavior. These findings suggest potential optimization challenges for certain parameter values and sample sizes. Nonetheless, the biases ultimately converge to zero, confirming the consistency of the MLEs.



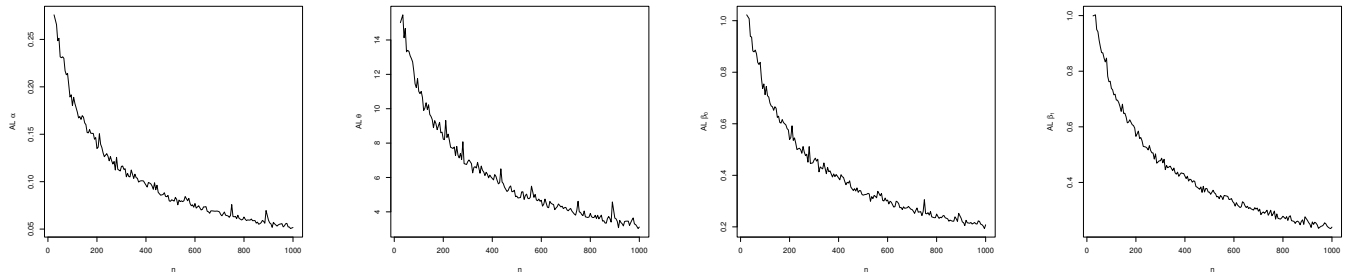
(**a**) (**b**) (**c**) (**d**)

Figure 5. (**a**–**d**) Biases versus sample size from LGOLLE regression model.



(**a**) (**b**) (**c**) (**d**)

Figure 6. (**a**–**d**) MSEs versus sample size from LGOLLE regression model.



(**a**) (**b**) (**c**) (**d**)

Figure 7. (**a**–**d**) ALs versus sample size from LGOLLE regression model.

3.4. Model Checking

Several approaches to analyzing outliers have been documented in the literature, including [22–24]. Outlier detection methods such as observation exclusion are used to identify influential observations in the proposed regression model.

In this context, for the proposed systematic component, the exclusion of observations follows

$$\mu_l = \mu(\boldsymbol{x_l}^\top \boldsymbol{\beta_j}), \quad l = 1, \ldots, n, \quad l \neq i. \tag{12}$$

For investigating the influential observations, the generalized Cook's distance is given by

$$GCD_i = \left(\hat{\boldsymbol{\psi}}_{(i)} - \hat{\boldsymbol{\psi}}\right)^\top [\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\psi}})] \left(\hat{\boldsymbol{\psi}}_{(i)} - \hat{\boldsymbol{\psi}}\right), \tag{13}$$

and the likelihood distance, as

$$LD_i = 2\Big\{\mathcal{L}(\hat{\boldsymbol{\psi}}) - \mathcal{L}\Big(\hat{\boldsymbol{\psi}}_{(i)}\Big)\Big\}, \tag{14}$$

where the subscript i denotes the observation deleted from the dataset and $\ddot{\boldsymbol{L}}(\boldsymbol{\psi})$ is the observed information matrix.

Moreover, the objective of residual analysis is to identify trends or characteristics in the residuals that may affect the model's validity. Therefore, the deviance residuals are commonly used to assess the goodness-of-fit of regression models [25]. It follows that deviance residuals are given by

$$r_{D_i} = \operatorname{sgn}(\hat{r}_{M_i})\big\{-2[\hat{r}_{M_i} + \log(1-\hat{r}_{M_i})]\big\}^{1/2}, \tag{15}$$

where

$$\hat{r}_{M_i} = 1 + \log\left\{1 - \frac{\left[1-\mathrm{e}^{-\mathrm{e}^{y-\hat{\mu}}}\right]^{\hat{\alpha}\hat{\theta}}}{\left[1-\mathrm{e}^{-\mathrm{e}^{y-\hat{\mu}}}\right]^{\hat{\alpha}\hat{\theta}} + \left[1-\left(1-\mathrm{e}^{-\mathrm{e}^{y-\hat{\mu}}}\right)^{\hat{\theta}}\right]^{\hat{\alpha}}}\right\}, \tag{16}$$

are the martingale residuals and the sign function $\operatorname{sign}(\cdot)$ is the signal function with a value $+1$ if the argument is positive and -1 if the argument is negative.

Ref. [26] proposed the construction of envelopes to support the analysis of the residuals, with normal probability plots. To construct these envelopes confidence bands are simulated, and if the model fits well, the majority of the points will be randomly distributed within these bands.

4. Application: Dengue Fever Cases Data

To demonstrate the potential of the GOLLE distribution, Table 5 illustrates several alternative distributions generated by well-known models, in addition to the nested model.

Table 5. Competitive distributions compared to the GOLLE distribution.

Distribution	Reference
Kumaraswamy-Fréchet (KwFr)	[27]
Kumaraswamy-Exponential (KwE)	[28]
Gamma-Fréchet (GFr)	[29]
Gamma-Exponentital (GE)	[30]
Beta-Exponentital (BE)	[31]
Fréchet (Fr)	[32]

The distributions are presented (for $x > 0$), respectively, as

$$F_{\mathrm{KwFr}}(x) = \{1 - [F_{\mathrm{Fr}}(x)]^a\}^b,$$

$$F_{\mathrm{KwE}}(x) = \{1 - [G(x)]^a\}^b,$$

$$F_{\mathrm{GFr}}(x) = \frac{\gamma\{a, -\log[1-F_{\mathrm{Fr}}(x)]/b\}}{\Gamma(a)},$$

$$F_{\mathrm{GE}}(x) = \frac{\gamma\{a, -\log[1-G(x)]/b\}}{\Gamma(a)},$$

$$F_{\mathrm{BE}}(x) = I_{G(x)}(a,b) = \frac{1}{B(a,b)}\int_0^{G(x)} w^{a-1}(1-w)^{b-1}dw$$

and

$$F_{\mathrm{Fr}}(x) = \mathrm{e}^{[-(x-a)^{-b}]},$$

where all of the parameters are positive, $\gamma(\cdot)$ is the incomplete gamma function and $G(x)$ and $F_{\mathrm{Fr}}(x)$ represent the exponential and Fréchet distributions, respectively. The goodness.fit function of AdequacyModel package (version 2.0.0) (see [33]) computes the MLEs (with standard errors (SEs) in parentheses) for all fitted models using the BFGS approach.

4.1. Descriptive of the Data

The data set was extracted from the Health Problem and Notification Information System (SINAN) (https://datasus.saude.gov.br/acesso-a-informacao/doencas-e-agravos-de-notificacao-de-2007-em-diante-sinan/, accessed on 2 July 2024). SINAN is a repository of patient notifications including a wide range of diseases, injuries and public health incidents listed as nationally mandatory for reporting. Notably, this includes more than 40 diseases (dengue fever, chikungunya fever, pandemic influenza, etc.). The data comprise of notifications related to dengue fever cases reported within the Federal District, Brazil, spanning all 49 epidemiological weeks (observations) throughout the year 2022: 689, 1205, 938, 1121, 1523, 1469, 1508, 1999, 2468, 2827, 3196, 3651, 3550, 4142, 4118, 4178, 3853, 2700, 6726, 2183, 2581, 1616, 1126, 898, 548, 752, 622, 415, 309, 291, 396, 476, 411, 360, 500, 402, 418, 313, 385, 475, 406, 277, 323, 433, 505, 574, 465, 1.682.

The focus of the study is on the variables below:

- y_i: total dengue fever cases of a epidemiological week (DG) (response variable);
- m_{ij}: month (levels: 0—January to 11—December). Thus, for $i = 1, \ldots, 49$ and $j = 0, \ldots, 11$, dummy variables.

The proposed model provides both advantages and disadvantages compared counting models. The exponential distribution used as a baseline has several important characteristics, including applicability in some epidemiological contexts, memorylessness, a good fit featuring heavy tails, flexibility and a simple density function form. Several concerns may arise, such as a lack of flexibility for trend modeling, violations of the independence assumption, and limitations with inflated zero data. Nonetheless, the model captures the significance of the exploratory variables as well as extreme events involving dengue fever patients in the temporal scenario.

Table 6 presents descriptive statistics. The number of dengue fever cases varied from very low (277) to high (6726). The standard deviation of 1445.35 indicates increased variability in dengue fever cases over time. The distribution is skewed to the right (1.509), suggesting that there are more extreme values at the higher end of the scale, and the kurtosis indicates heavier tails (4.998).

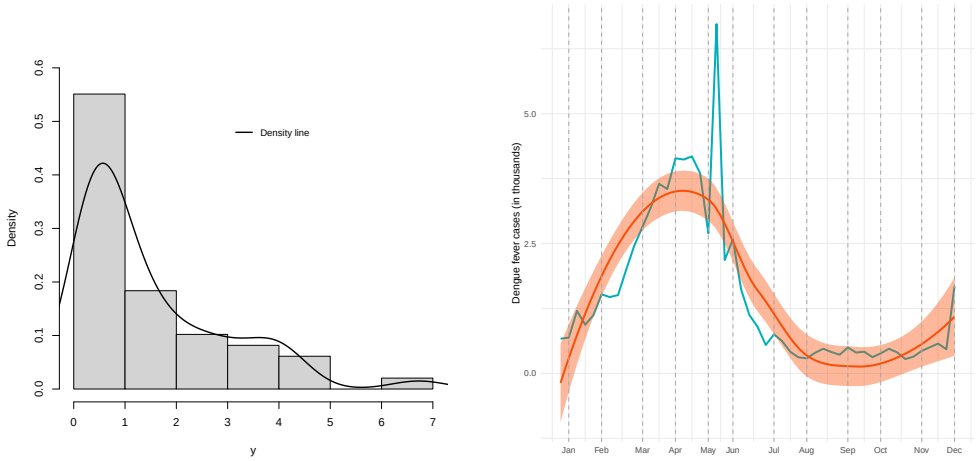
Table 6. Descriptive statistics of dengue fever cases data.

Variable	Min.	Max.	Mean	Median	SD	Skewness	Kurtosis
DG	277	6726	1483	752	1445.35	1.509	4.998

Figure 8 shows the histogram and time series of the data. Figure 8a indicates heavy tail behavior, consistent with extreme event data. Figure 8b shows anomalous activity between May and June. This represents the highest number of cases since 1998, (https://www.saude.df.gov.br/informes-dengue-chikungunya-zika-febre-amarela, accessed on 2 July 2024), demonstrating the atypical behavior of the observations, which deviate significantly from the historical average, indicating an unusual outbreak, or, in epidemiology, an extreme event of dengue fever that can have an impact on both the health system and the economy. In addition, the plot shows an increase in tendency between February and June, when the disease is most likely to develop in the Federal District and a decrease in tendency between July and December, which is the drought period.

Despite being a continental country with varying disease patterns, Brazil's Midwest region has the highest incidence of dengue fever, as reported by the Arbovirus Monitoring Panel of the Ministry of Health (https://www.gov.br/saude/pt-br/assuntos/saude-de-a-

a-z/a/aedes-aegypti/monitoramento-das-arboviroses, accessed on 2 July 2024). This is the region where the study data were collected.



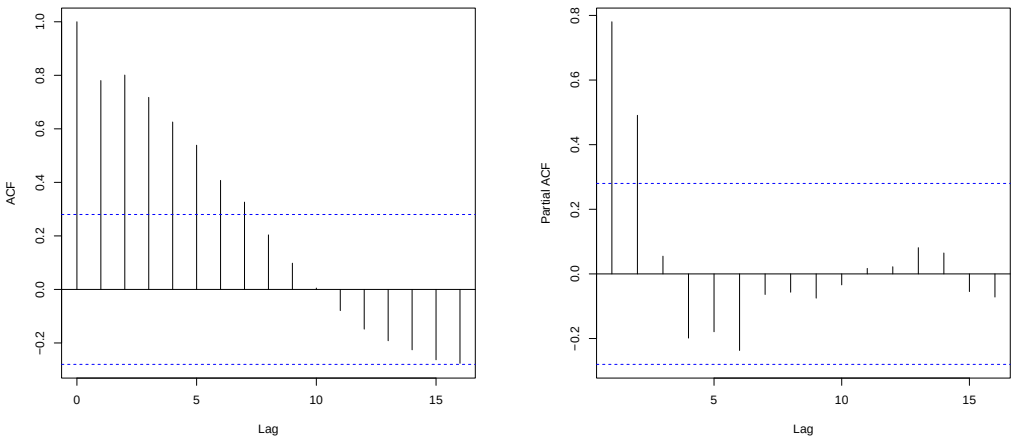
(**a**) (**b**)

Figure 8. Dengue fever cases data. (**a**) Histogram and empirical density. (**b**) Variation across months with trend smoothed line in shadow light red.

4.2. Findings from GOLLE Distribution

The analysis of time series data requires a detailed inspection, which includes identifying relationships between the observations. Failure to account for these associations may result in an inadequate model that neglects temporal dependence, potentially leading to incorrect forecasts and interpretations. To identify serial correlation, it is essential to analyze the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Figure 9 shows the ACF and PACF for an autoregressive integrated moving average (ARIMA) model with one lag in the differenced series, specifically, ARIMA(1,1,0).

Although the dataset shows a correlation, indicating a dependency between variables, the model can still be used if this assumption is relaxed, especially given the small sample size and the implementation of diagnostic and residual analyses.



(**a**) (**b**)

Figure 9. Dengue fever cases data. (**a**) ACF. (**b**) PACF.

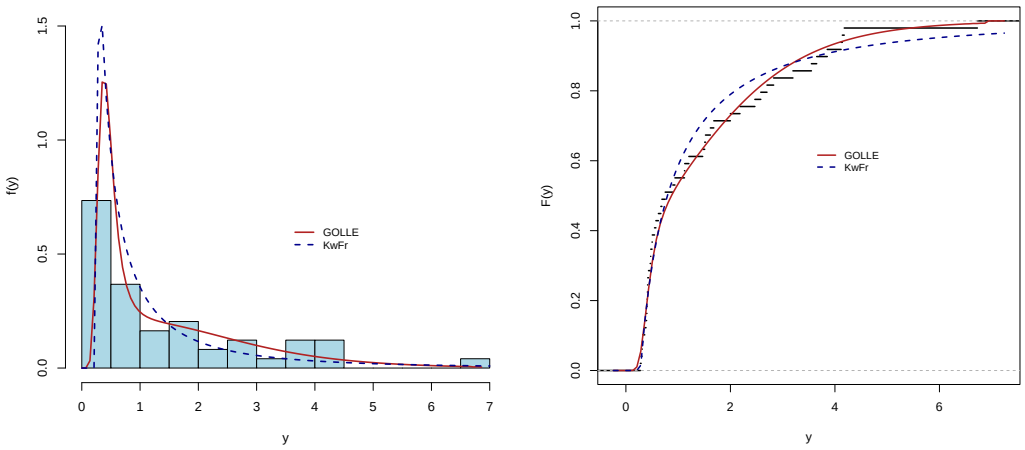
Table 7 presents the results of the fitted distributions to the data and shows that the GOLLE distribution is the best fit. Figure 10a,b show histograms and plots of estimated densities, as well as empirical cdf estimated ones. The Fréchet distribution is commonly

used to model extreme occurrences and these findings suggest that the KwFr distribution (the second-best model) is competitive with the presented model.

Table 7. Findings from the fitted models of dengue fever cases data.

Model	Parameters				W*	A*	KS
GOLLE † **$(\alpha, \theta, \lambda)$**	**0.154**	**76.500**	**5.402**		**0.060**	**0.400**	**0.077**
	(0.018)	**(0.019)**	**(0.003)**				**(0.914)**
OLLE(α, λ)	1.180	1	0.634		0.318	1.929	0.160
	(0.142)	(-)	(0.086)				(0.145)
EE(θ, λ)	1	1.391	0.830		0.313	1.898	0.175
	(-)	(0.284)	(0.147)				(0.088)
E(λ)	1	1	0.674		0.316	1.913	0.170
	(-)	(-)	(0.096)				(0.103)
KwFr(β, γ, a, b)	3.851	51.070	0.172	0.271	0.087	0.559	0.097
	(1.409)	(71.389)	(0.060)	(0.008)			(0.705)
KwE(β, γ, λ)	4.500	0.151	5.402		0.242	1.482	0.205
	(0.005)	(0.022)	(0.003)				(0.028)
GFr(β, a, b)	0.465	0.777	0.225		0.128	0.830	0.120
	(0.082)	(0.142)	(0.039)				(0.443)
BE(β, γ, λ)	3.027	0.150	5.402		0.253	1.548	0.197
	(1.054)	(0.023)	(0.003)				(0.038)
GE(β, λ)	1.323	0.892			0.317	1.917	0.173
	(0.241)	(0.197)					(0.096)
Fr(a, b)	1.791	-0.281			0.235	1.449	0.173
	(0.285)	(0.076)					(0.094)

† Best fit model in bold.



(**a**) (**b**)

Figure 10. Fitted models of dengue fever cases data. (**a**) Histogram and estimated pdfs. (**b**) Empirical and estimated cdfs.

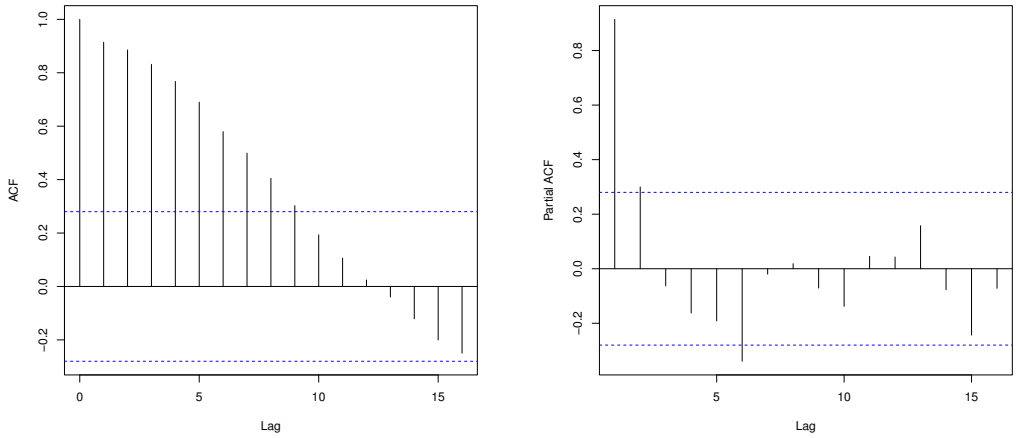
Likelihood-ratio (LR) tests were used to compare the GOLLE distribution with its nested models. Table 8 shows that adding more parameters significantly influences the accuracy of modeling the existing data.

Table 8. LR tests of the GOLLE distribution.

Models	Statistic w	p-Value
GOLLE vs. E	29.657	<0.0001
GOLLE vs. EE	27.143	<0.0001
GOLLE vs. OLLE	27.937	<0.0001

4.3. Findings from LGOLLE Regression Model

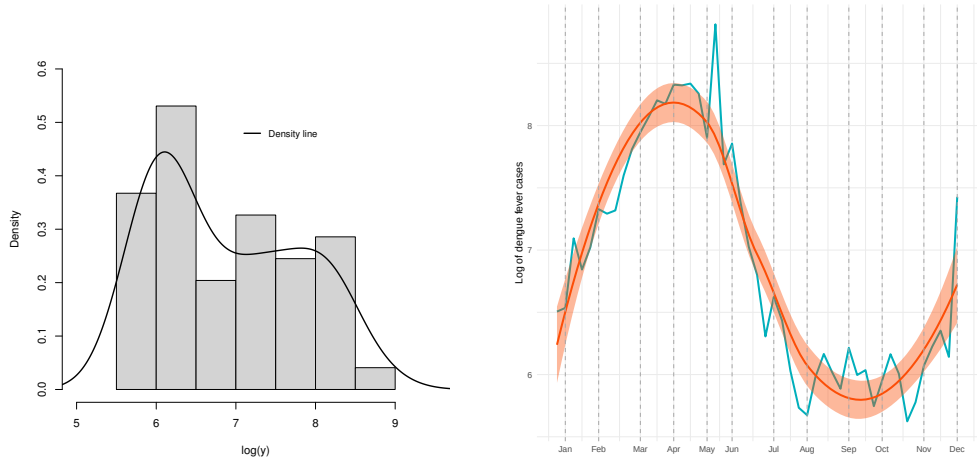
To assess the LGOLLE model, Figure 11 displays the ACF and PACF plots for the differenced series ARIMA(0,1,0), which indicates a random walk process. Therefore, the proposed regression model is suitable for the current data set.



(**a**) (**b**)

Figure 11. Log-dengue fever cases data. (**a**) ACF. (**b**) PACF.

A histogram and time series of the log data are displayed in Figure 12. Figure 12a shows a bimodal distribution, with extreme occurrences observed between May and June. Figure 12b illustrates a clear trend over time.



(**a**) (**b**)

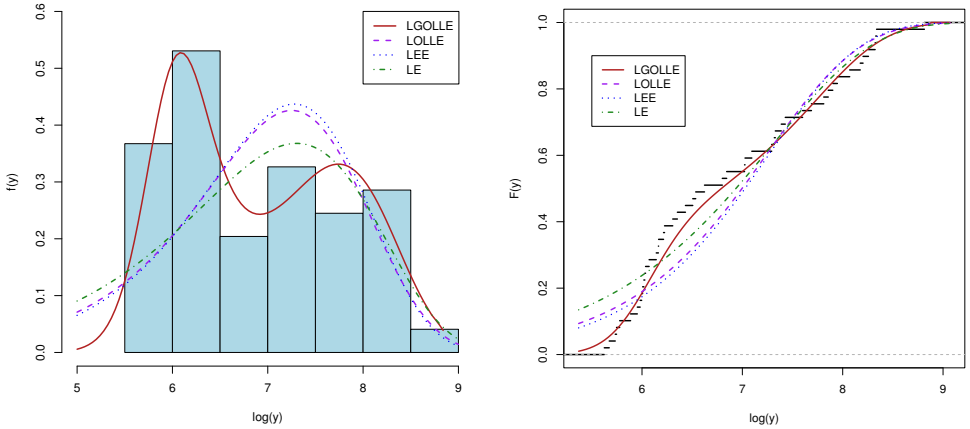
Figure 12. Log dengue fever cases data. (**a**) Histogram and empirical density. (**b**) Variation across months with trend smoothed line in shadow light red.

The LGOLLE distribution is the best fit, compared to the nested models log odd log-logistic exponential (LOLLE), log exponentiated exponential (LEE) and log exponential (LE), as demonstrated by Table 9, which displays the results of the fitted nested distributions to the log data. In addition to histograms and the estimated density functions in Figure 13a,b displays the empirical cdf and estimated ones.

Table 9. Findings from the fitted models of log-dengue fever cases data.

Model	Parameters			W*	A*	KS
LGOLLE † (α, θ, μ)	**0.1517**	**78.7499**	**5.2352**	**0.054**	**0.3664**	**0.0862**
	(0.0178)	**(0.0274)**	**(0.0034)**			**(0.8293)**
LOLLE(α, μ)	1.1798	1	7.3631	0.3184	1.9289	0.1601
	(0.1423)	(-)	(0.1353)			(0.1453)
LEE(θ, μ)	1	1.3911	7.0935	0.3132	1.8975	0.1750
	(-)	(0.2839)	(0.1769)			(0.0879)
LE(μ)	1	1	7.3019	0.3161	1.9131	0.1704
	(-)	(-)	(0.1429)			(0.1032)

† Best fit model in bold.



(**a**) (**b**)

Figure 13. Fitted models of log-dengue fever cases data. (**a**) Histogram and estimated pdfs. (**b**) Empirical and estimated cdfs.

LR tests are employed to compare the nested models of the LGOLLE distribution. Table 10 shows that adding extra parameters has a considerable influence on the modeling performance with the given data.

Table 10. LR tests of the LGOLLE distribution.

Models	Statistic w	p-Value
LGOLLE vs LE	29.650	<0.0001
LGOLLE vs LEE	27.136	<0.0001
LGOLLE vs LOLLE	27.930	<0.0001

The new LGOLLE regression model for the dengue fever cases data in Federal District, Brazil can be expressed as follows

$$y_i = \mu_i + z_i,$$

where $z_1, z_2, \ldots, z_{49}$ are independent rv with density function (10) and the systematic structure is defined by

$$\mu_i = \beta_0 + \sum_{j=1}^{11} \beta_j m_j.$$

The results (MLEs, SEs, Confidence Intervals (CIs), and p-values) of the fitted LGOLLE regression model are presented in Table 11. Therefore, using a significance level of 5%, the months in the temporal scenario are significant and can be used to model the location.

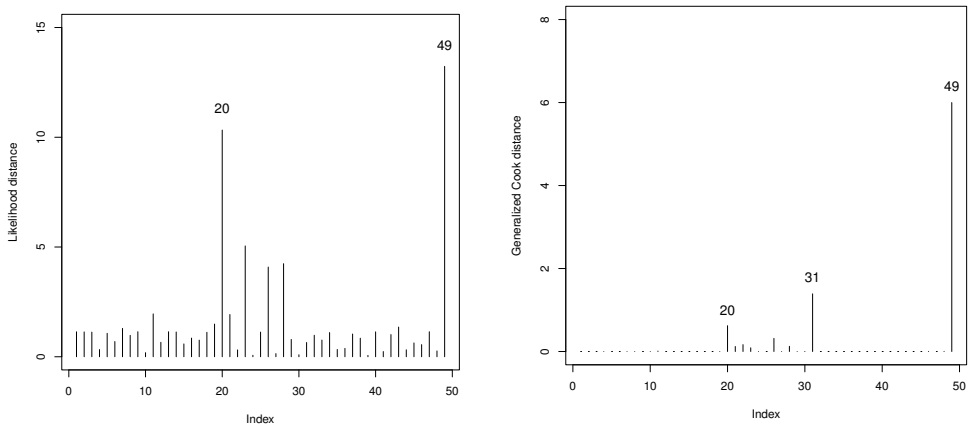
Table 11. Fitted LGOLLE regression model of dengue fever cases data.

Parameter	Estimate	SE	CI (95%)	p-Value
β_1 (Fev)	**0.5511**	**0.1836**	**(0.1913; 0.9109)**	**0.0049**
β_2 (Mar)	**1.0827**	**0.1909**	**(0.7085; 1.4569)**	**<0.0001**
β_3 (Apr)	**1.6788**	**0.1764**	**(1.3331; 2.0245)**	**<0.0001**
β_4 (May)	**1.7125**	**0.1822**	**(1.3554; 2.0696)**	**<0.0001**
β_5 (Jun)	**1.4393**	**0.2262**	**(0.9960; 1.8826)**	**<0.0001**
β_6 (Jul)	0.3091	0.1935	(-0.0702; 0.6884)	0.1192
β_7 (Ago)	**−0.5881**	**0.2006**	**(−0.9813; −0.1949)**	**0.0059**
β_8 (Sep)	**−0.4623**	**0.1723**	**(−0.8000; −0.1246)**	**0.0110**
β_9 (Oct)	**−0.5778**	**0.1776**	**(−0.9259; −0.2297)**	**0.0025**
β_{10} (Nov)	**−0.5749**	**0.1802**	**(−0.9281; −0.2217)**	**0.0030**
β_{11} (Dec)	−0.1612	0.1928	(−0.5391; 0.2167)	0.4088
α	47.1443	6.4074	(34.5861; 59.7026)	-
θ	0.0751	0.0054	(0.0645; 0.0857)	-

Significant parameters in bold.

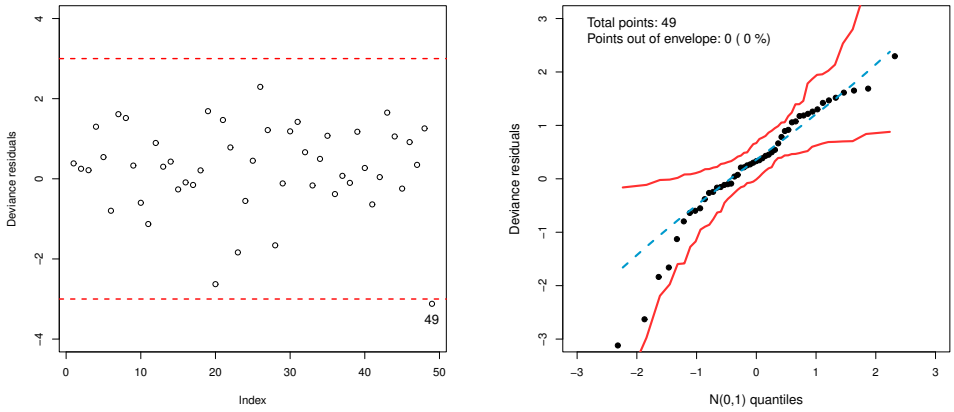
Figure 14 presents two potentially significant observations based on the LD and GCD measures. It is worth noting that the 20th and 49th observations, correspond to the beginning of June, and the last epidemiological week. The first matches the record (https://www.metropoles.com/distrito-federal/boletim-revela-aumento-dos-casos-de-dengue-em-todas-as-regioes-do-df, accessed on 2 July 2024) of dengue fever cases in the Federal District. The second relates to the end of year vacation/recess period, which results in a backlog of alerts due to a lack of healthcare professionals available to notify cases and input data into the system.

Nonetheless, Figure 15 indicates that the residual deviations behave randomly across the range and remain within the simulated envelope, implying that the observations have minimal impact on the regression model.



(**a**) (**b**)

Figure 14. The LGOLLE regression model. (**a**) LD. (**b**) GCD.



(**a**) (**b**)

Figure 15. The LGOLLE regression model. (**a**) Deviance residual index. (**b**) Simulated envelope.

4.4. Discussion

The findings indicate that the LGOLLE regression model is suitable for explaining the weekly dengue fever cases in the Federal District. Table 11 presents parameter estimates for the LGOLLE regression model, which becomes (for $i = 1, \ldots, 49$)

$$\begin{aligned}\hat{\mu}_i = {} & 15.7495 + 0.5511\, m_{i1} + 1.0827\, m_{i2} + 1.6788\, m_{i3} + 1.7125\, m_{i4} \\ & + 1.4393\, m_{i5} - 0.5881\, m_{i7} - 0.4623\, m_{i8} - 0.5778\, m_{i9} - 0.5749\, m_{i10}.\end{aligned} \tag{17}$$

The following discussion examines the systematic structure, using January as the month of reference.

Interpretations for $\hat{\mu}$

- Except for the covariates m_6 and m_{11}, referring to the months of July and December, all other covariates are significant at a 5% level of significance. This indicates that there is a difference in dengue fever cases registered in the Federal District between the other months and January. The months of July and December are probably not significant due to their behavior being similar to the reference month;
- The months of February to June have positive estimates, which is significant, showing an increase in comparison to January. This can be seen in Figure 12b, which shows an extreme event occurring between May and June in the data for that time window;
- The months of August to November have negative values, indicating a decline in dengue fever cases compared to January. During this period, the Federal District experiences a drought, which corroborates the study's findings (https://portal.inmet.gov.br/uploads/notastecnicas/Estado-do-clima-no-Brasil-em-2022-OFICIAL.pdf, accessed on 2 July 2024).

5. Conclusions

The paper defines the generalized odd log-logistic exponential (GOLLE) distribution (see [12,13]) and introduces a new bimodal regression model based on this distribution, incorporating a location-systematic structure, to investigate weekly dengue fever cases in the Federal District for 2022. The paper reviews some mathematical properties, estimates the parameters using the maximum likelihood method, and evaluates the consistency criterion through Monte Carlo simulations. The consistency of the MLEs for the regression model is assessed using various simulation measures. Additionally, global influence measures and residual analysis are conducted to examine the fit of the new model.

Some important findings are presented. Aside from the months of July and December, the remaining months are significant at the 5% significance level. From February to June, positive estimates are observed, suggesting a positive impact on dengue fever cases during

this period. This aligns with the climatological effects that increase cases in these months. Between August and November, a drought occurs, supporting the negative estimates during this period, indicating a negative effect on weekly dengue fever cases.

The epidemiology dataset showed that the novel regression model is more flexible than other nested and well-established models. Therefore, the proposed model enhances the understanding of dengue fever cases in the Federal District and accounts for an extreme event that occurred during the study period. To generalize these findings, further validation across diverse datasets and regions is necessary. Future research should explore the model's applicability to other diseases or epidemiological settings.

Author Contributions: Conceptualization, N.S.S.d.C.; methodology, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; software, N.S.S.d.C.; validation, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; formal analysis, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; investigation, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; data curation, N.S.S.d.C.; writing—original draft preparation, N.S.S.d.C. and M.d.C.S.d.L.; writing–review and editing, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; visualization, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; supervision, N.S.S.d.C., M.d.C.S.d.L. and G.M.C.; project administration, N.S.S.d.C. and M.d.C.S.d.L. All authors have read and agreed to the current version of the manuscript.

Funding: This research is awaiting to external funding.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

A^*	Anderson Darling
ACF	autocorrelation function
AE	average estimate
AL	average estimate length
ARIMA	autoregressive integrated moving average model
BFr	beta-Fréchet
cdf	cumulative distribution function
CI	confidence interval
COVID-19	corona virus disease 2019
DG	dengue fever cases
E	exponential distribution
EE	exponentiated exponential distribution
EVT	extreme value theory
Fr	Fréchet
GCD	generalized Cook distance
GEV	generalized extreme value
GE	gamma-exponential distribution
GFr	gamma-Fréchet distribution
GOLLE	generalized odd log-logistic exponential distribution
GOLL-G	generalized odd log-logistic distribution
hrf	hazard rate function
KS	Kolmogorov-Sminorv
KwE	Kumaraswamy exponential distribution
KwFr	Kumaraswamy Fréchet distribution
LD	loglikelihood distance
LE	log exponential distribution
LGOLLE	log generalized odd log-logistic exponential distribution
LLE	log exponentiated exponential distribution

LOLLE	log odd log-logistic exponential distribution
LR	likelihood ratio
mgf	moment generation function
MLE	maximum likelihood estimate
MSE	mean squared error
OLLE	odd log-logistic exponential distribution
PACF	partial autocorrelation function
pdf	probability distribution function
qf	quantile function
RMSE	root mean squared error
SE	standard error
SINAN	sistema de informação de agravos de notificação
T-X	transformer-transformer generator
W*	Cramér-von Misses

References

1. Joseph, J.; Gillariose, J. A novel discrete Slash family of distributions with application to epidemiology informatics data. *Int. J. Data Sci. Anal.* **2024**, 1–17. [CrossRef]
2. Li, Y.; Dou, Q.; Lu, Y.; Xiang, H.; Yu, X.; Liu, S. Effects of ambient temperature and precipitation on the risk of dengue fever: A systematic review and updated meta-analysis. *Environ. Res.* **2020**, *191*, 110043. [CrossRef]
3. Lim, J.T.; Dickens, B.S.L.; Cook, A.R. Modelling the epidemic extremities of dengue transmissions in Thailand. *Epidemics* **2020**, *33*, 100402. [CrossRef]
4. Diop, A.; Deme, E.H.; Diop, A. zero-inflated generalized extreme value regression model for binary response data and application in health study. *J. Stat. Comput. Simul.* **2023**, *93*, 1–24. [CrossRef]
5. Marani, M.; Katul, G.G.; Pan, W.K.; Parolari, A.J. Intensity and frequency of extreme novel epidemics. *Proc. Natl. Acad. Sci. USA* **2021**, *35*, e2105482118. [CrossRef]
6. Thomas, M.; RootzÉn, H. Real-time prediction of severe influenza epidemics using extreme value statistics. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2022**, *71*, 376–394. [CrossRef]
7. Lin, H.; Zhang, Z. Extreme co-movements between infectious disease events and crude oil futures prices: From extreme value analysis perspective. *Energy Econ.* **2022**, *110*, 106054. [CrossRef]
8. Tian, N.; Zheng, J.-X.; Guo, Z.-Y.; Li, L.-H.; Xia, S.; Lv, S.; Zhou, X.-N. Dengue incidence trends and its burden in major endemic regions from 1990 to 2019. *Trop. Med. Infect. Dis.* **2022**, *7*, 180. [CrossRef]
9. Lun, X.; Wang, Y.; Zhao, C.; Wu, H.; Zhu, C.; Ma, D.; Xu, M.; Wang, J.; Liu, Q.; Xu, L.; et al. Epidemiological characteristics and temporal-spatial analysis of overseas imported dengue fever cases in outbreak provinces of China, 2005–2019. *Infect. Dis. Poverty* **2022**, *11*, 12. [CrossRef]
10. Sandeep, M.; Padhi, B.K.; Yella, S.S.T.; Sruthi, K.G.; Venkatesan, R.G.; Sasanka, K.S.; Krishna, B.S.; Satapathy, P.; Mohanty, A.; Al-Tawfiq, J.A.; et al. Myocarditis manifestations in dengue cases: A systematic review and meta-analysis. *J. Infect. Public Health* **2023**, *16*, 1761–1768. [CrossRef]
11. de Oliveira-Júnior, J.F.; Souza, A.; Abreu, M.C.; Nunes, R.S.C.; Nascimento, L.S.; Silva, S.D.; Correia Filho, W.L.F.; Silva, E.B. Modeling of dengue by cluster analysis and probability distribution functions in the state of Alagoas in Brazilian. *Braz. Arch. Biol. Technol.* **2023**, *66*, e23220086. [CrossRef]
12. Qoshja, A.; Muça, M. A new modified generalized odd log-logistic distribution with three parameters. *Math. Theory Model.* **2018**, *8*, 2224–5804. Available online: https://www.researchgate.net/publication/331483356_A_NEW_MODIFIED_GENERALIZED_ODD_LOG-LOGISTIC_DISTRIBUTION_WITH_THREE_PARAMETERS (accessed on 28 June 2024).
13. Afify, A.Z.; Suzuki, A.K.; Zhang, C.; Nassar, M. On three-parameter exponential distribution: Properties, Bayesian and non-Bayesian estimation based on complete and censored samples. *Commun. Stat.-Simul. Comput.* **2021**, *50*, 3799–3819. [CrossRef]
14. Cordeiro, G.M.; Alizadeh, M.; Ozel, G.; Hosseini, B.; Ortega, E.M.M.; Altun, E. The generalized odd log-logistic family of distributions: Properties, regression models and applications. *J. Stat. Comput. Simul.* **2017**, *87*, 908–932. [CrossRef]
15. da Costa, N.S.S.; Cordeiro, G.M. A new normal regression with medical applications. *Appl. Math. Inf. Sci.* **2023**, *17*, 309–322. [CrossRef]
16. da Costa, N.S.S.; do Carmo, M.d.C.S.; Cordeiro, G.M. Analyzing county-level COVID-19 vaccination rates in Texas: A new Lindley regression model. *COVID* **2023**, *3*, 1761–1780. [CrossRef]
17. Alzaatreh, A.; Lee, C.; Famoye, F. A new method for generating families of continuous distributions. *Metron* **2013**, *71*, 63–79. [CrossRef]
18. Gleaton, J.U.; Lynch, J.D. Properties of generalized log-logistic families of lifetime distributions. *J. Probab. Stat. Sci.* **2006**, *4*, 51–64. Available online: https://www.researchgate.net/publication/283595537_Properties_of_generalized_log-logistic_families_of_lifetime_distributions (accessed on 28 June 2024).
19. Gupta, R.C.; Gupta, R.D. Proportional reversed hazard rate model and its applications. *J. Stat. Plan. Inference* **2007**, *137*, 3525–3536. [CrossRef]

20. Gupta, R.C.; Gupta, R.D. Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biom. J. J. Math. Methods Biosci.* **2001**, *43*, 117–130. [CrossRef]
21. R Core Team. *R Core Team: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024.
22. Cox, D.R.; Snell, E.J. A general definition of residuals. *J. R. Stat. Soc. Ser. B (Methodol.)* **1968**, *30*, 248–275. [CrossRef]
23. Cook, R.D.; Weisberg, S. *Residuals and Influence in Regression*; Chapman & Hall: London, UK, 1982.
24. Ortega, E.M.M.; Paula, G.A.; Bolfarine, H. Deviance residuals in generalized log-gamma regression models with censored observations. *J. Stat. Comput. Simul.* **2008**, *78*, 747–764. [CrossRef]
25. Silva, G.O.; Ortega, E.M.M.; Paula, G.A. Residuals for log-Burr XII regression models in survival analysis. *J. Appl. Stat.* **2011**, *38*, 1435–1445. [CrossRef]
26. Atkinson, A.C. *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*; Clarendon Press: Oxford, UK, 1987. [CrossRef]
27. Mead, M.E.A. A note on Kumaraswamy Fréchet distribution. *Australia* **2014**, *8*, 294–300. Available online: http://www.ajbasweb.com/old/ajbas/2014/September/294-300.pdf (accessed on 28 June 2024).
28. Adepoju, K.A.; Chukwu, O.I. Maximum likelihood estimation of the Kumaraswamy exponential distribution with applications. *J. Mod. Appl. Stat. Methods* **2015**, *14*, 208–214. [CrossRef]
29. Silva, R.; Andrade, T.; Maciel, D.; Campos, R.; Cordeiro, G. A new lifetime model: The gamma extended Frechet distribution. *J. Stat. Theory Appl.* **2013**, *12*, 39. [CrossRef]
30. Kudriavtsev, A.A. On the representation of gamma-exponential and generalized negative binomial distributions. *Inform. Its Appl.* **2019**, *13*, 76–80. [CrossRef]
31. Nadarajah, S.; Kotz, S. The beta exponential distribution. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 689–697. [CrossRef]
32. Fréchet, M. Sur La Loi de Probabilité de L'écart Maximum. *Annales de La Societe Polonaise de Mathematique*. Available online: https://cir.nii.ac.jp/crid/1572261550191409280 (accessed on 28 June 2024).
33. Marinho, P.R.D.; Silva, R.B.; Bourguignon, M.; Cordeiro, G.M.; Nadarajah, S. AdequacyModel: An R package for probability distributions and general purpose optimization. *PLoS ONE* **2019**, *14*, e0221487. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.