Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/170213

A New Normal Regression with Medical Applications

Nicollas S. S da Costa * *and Gauss M. Cordeiro*

Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brazil

Received: 2 Dec. 2022, Revised: 22 Jan. 2023, Accepted: 22 Feb. 2023
Published online: 1 Mar. 2023

Abstract: We define the generalized odd log-logistic normal regression with a dispersion systematic component. We obtain a linear representation, some of its properties, and maximum likelihood estimates. Furthermore, we carry out several simulations for different schemes to evaluate the accuracy of the estimators. The robustness of the new regression model is proved by modeling COVID-19 data. The proposed model explains COVID-19 ICU survival times of the patients in a Brazilian hospital.

Keywords: Generalized odd log-logistic distribution, Maximum likelihood, Normal distribution, Regression model, Simulation

1 Introduction

In late 2019, first identified in Wuhan, China, a disease to that point unknown to researchers emerged, coronavirus 2019 (COVID-19) which caused a global pandemic outbreak. Early research conducted, [1] and [2], identified that elderly people had elevate hospitalizations and death outcome. The main possible symptoms are described, and it is continuous update by CDC[1]. According to information updated on December 16, 2021 from Worldometers[2], for COVID-19 worldwide, over 272 million cases were reported, exceeding 5,3 millions that passed away. In light of this, several studies addressing aspects about the virus and the pandemic have been published, ranging from screening for possible infected [3], investigations into the disease in relation to demographic characteristics and comorbidities [4], [5] and [6], to ways of diagnosing [7] and possible treatments [8].

The study by [9] evaluated the interventions made, to date, by countries in combating the pandemic. The performance of twenty nations were ranked according to eight criteria and the scores obtained were considered in evaluating the effort to mitigate the pandemic. Countries such as New Zealand, Australia, South Korea, and Norway obtained the best scores, by adopting pandemic containment and control measures that ranged from mass population testing to the containment of individuals in regions with curfews and lockdown. Brazil ranked 17th, which can be showed by the article [10] in which it elaborates general statistics about the sad scenario of COVID-19 in Brazil.

Due to the large spectrum of clinical manifestations that the coronavirus can trigger, from asymptomatic or momentary diseases given by the symptoms, to more severe pictures, we still have a range of parameters that vary according to age, sex, comorbidities, habits, among others, that make it difficult for researchers to interpret and identify [11]. Recently, several studies have been published looking at both the time to recovery or death considering variables such as gender and age. For example, the study of 5,769 patients in Israel regarding recovery time under the effects of age and sex assessed by [12], in addition to the analysis of [13], which presented studies on the variables sex, region, reasons for infection, age on the rates of recovered cases and deaths.

Other works in the area are related to risk factors associated with patient mortality or survival. The study by [14] explored some characteristics of critically ill diabetes patients in Mexico and associated to their survival times. [15] presented a survival analysis of death predictors associated with COVID-19, and [16] related factors that can affect patient outcomes. For application purposes, we use real data from hospitalization in the intensive care unit (ICU) of a hospital in São Paulo state, Brazil. The survival times of the patients are displayed in Figure 1 which shows an asymmetric shape of the data.

Recently, new generalized models have been developed to expand parent distributions and produce very flexible forms to fitting real data. The authors [17]

[1] https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html
[2] https://www.worldometers.info/coronavirus/

* Corresponding author e-mail: nicollas.stefan@gmail.com

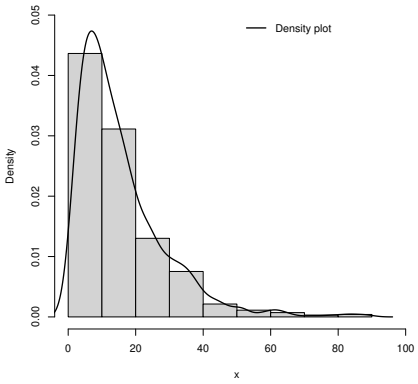


Fig. 1: Histogram and empirical density of COVID-19 data.

introduced the *generalized odd log-logistic-G* (GOLL-G) family which includes two well-known cases: the generalized log-logistic family and the proportional reversed hazard rate family. More recently, [18] introduced two regressions for bimodal data using the odd log-logistic family. [19] introduced the GOLL flexible Weibull distribution, while [20] defined the GOLL exponential distribution based on samples with complete and censored data, and applications in reparable systems. Based on the GOLL Maxwell distribution, [21] presented a distribution to model data with excess zeros and a parametric regression to estimate nonlinear effects. Its usefulness is illustrated by an experimental design evaluated in a sugarcane field. The GOLL Maxwell regression was also applied to real engineering data [22].

Based on this, we define the *generalized odd log-logistic normal* (GOLLN) distribution which has some advantages (bimodal, asymmetric and data-tailed heavy modeling) over other competing distributions. The new distribution has the baseline normal, the exponentiated normal (EN), and the odd log-logistic normal (OLLN) models as particular cases. We derive a linear representation and present useful properties. In addition, maximum likelihood estimates (MLEs) of the parameters are found. Overall, we note many phenomenons or cases where there are exists associated variables, one call explanatory and another response. For example, in different fields, such as medical, economics, engineering, psychology, etc, we investigate the effects of these variables from regression models.

Further, we provide a regression model based on the new distribution and some global influence measures, and a detailed residual analysis.

The paper is structured as follows. Section 2 introduces the GOLLN distribution. Section 3 presents a linear representation, and some of its mathematical properties. Some simulations in Section 4 examine the accuracy of the MLEs. A new GOLLN regression with a dispersion systematic structure is constructed in Section 5. Section 6 provides diagnostic measures and simulated envelopes for the new regression. Some simulations in Section 7 also reveal the efficiency of the parameter estimates. A COVID-19 data set in Section 8 confirms that the new model performs better than some other models. Some valuable findings are also presented. Section 9 ends with some comments.

2 The GOLLN definition

The cumulative distribution function (cdf) and probability density function (pdf) of the normal with mean $\mu \in \mathbb{R}$ and dispersion $\sigma > 0$ are (for $y \in \mathbb{R}$)

$$G(y) = \Phi(z) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \tag{1}$$

and

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma}\, \text{e}^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\,\phi(\text{z}), \tag{2}$$

respectively, where $\text{erf}(\cdot)$ is the error function, $z = z(y) = (y-\mu)/\sigma$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal, respectively.

Consider a baseline cdf $G(x;\xi)$ with unknown parameter vector ξ. [23] characterized the cdf of the OLL-G family as

$$F(x) = \frac{G(x;\xi)^{\alpha}}{G(x;\xi)^{\alpha} + [1 - G(x;\xi)]^{\alpha}}, \quad x \in \mathbb{R}, \tag{3}$$

where $\alpha > 0$. Based on the transformer-transformer (T-X) generator [24], [17] defined the cdf of the GOLL-G family by

$$F(x) = \frac{G(x;\xi)^{\alpha\theta}}{G(x;\xi)^{\alpha\theta} + [1 - G(x;\xi)^{\theta}]^{\alpha}}, \tag{4}$$

where $\alpha > 0$ and $\theta > 0$.

Hence, Equation (4) contains three sub-models reported in Table 1.

Table 1: Some special models.

α	θ	Reduced model
-	1	Generalized log-logistic family [23]
1	-	Proportional reversed hazard rate family [25]
1	1	Baseline

Let $X \sim$ GOLL-G(α, θ, ξ) be a random variable (rv) with cdf (4). Differentiating Equation (4), the pdf of X reduces to

$$f(x) = \frac{\alpha\theta g(x;\xi) G(x;\xi)^{\alpha\theta-1}[1 - G(x;\xi)^{\theta}]^{\alpha-1}}{\{G(x;\xi)^{\alpha\theta} + [1 - G(x;\xi)^{\theta}]^{\alpha}\}^2}, \tag{5}$$

where $g(x;\xi)$ is the parent pdf. Due to great flexibility, the density function (5) is widely used in many areas. If the baseline distribution has closed-form, the generated distribution can be more mathematically tractable.

The hazard rate function (hrf) of X is

$$h(x) = \frac{\alpha\theta g(x;\xi)G(x,\xi)^{\alpha\theta-1}}{[1-G(x,\xi)^{\theta}]\{G(x,\xi)^{\alpha\theta}+[1-G(x,\xi)^{\theta}]^{\alpha}\}}. \quad (6)$$

The four-parameter GOLLN cdf follows by substituting (1) into Equation (4)

$$F(x) = \frac{\Phi(z)^{\alpha\theta}}{\Phi(z)^{\alpha\theta}+[1-\Phi(z)^{\theta}]^{\alpha}}. \quad (7)$$

By inserting (2) into Equation (5), the corresponding pdf is

$$f(x) = \frac{\alpha\theta\phi(z)\Phi(z)^{\alpha\theta-1}[1-\Phi(z)^{\theta}]^{\alpha-1}}{\sigma\{\Phi(z)^{\alpha\theta}+[1-\Phi(z)^{\theta}]^{\alpha}\}^2}. \quad (8)$$

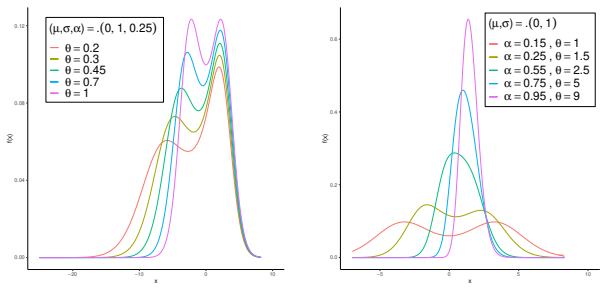
Substituting (1) and (2) into Equation (6), the hrf follows as

$$h(x) = \frac{\alpha\theta\phi(z)\Phi(z)^{\alpha\theta-1}}{\sigma[1-\Phi(z)^{\theta}]\{\Phi(z)^{\alpha\theta}+[1-\Phi(z)^{\theta}]^{\alpha}\}}. \quad (9)$$

The GOLLN model (denoted from now on by X) contains three special distributions:

(i) for $\theta = 1$, OLLN distribution, see [26];
(ii) for $\alpha = 1$, EN distribution;
(iii) for $\theta = \sigma = 1$, normal distribution.

Figures 2 and 3 display the pdf and hrf of X. Note that one of the main features of the GOLLN distribution is that its hrf can have many shapes (bathtub, unimodal, among others).

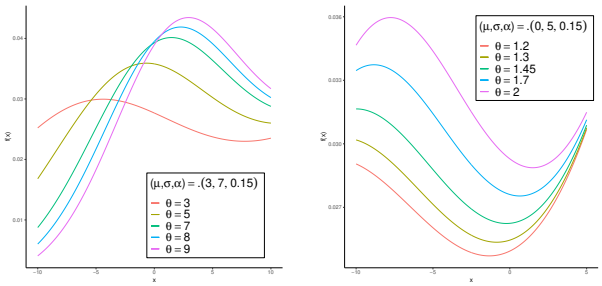


(a) **(b)**

Fig. 2: GOLLN pdf. (a) Selected values for θ, fixed α, μ, σ. (b) Selected values for α and θ, fixed μ, σ

By inverting (7), the quantile function (qf) of X, say $x = Q(u) = F^{-1}(u)$, is

$$x = Q(u) = Q_N\left(\left[\frac{\left(\frac{u}{\bar{u}}\right)^{1/\alpha}}{1+\left(\frac{u}{\bar{u}}\right)^{1/\alpha}}\right]^{1/\theta}\right), \quad (10)$$



(a) **(b)**

Fig. 3: GOLLN hrf. (a) Selected values for θ, fixed α, μ, σ. (b) Selected values for θ, fixed α, μ, σ.

where $\bar{u} = 1-u$, and $Q_N(u) = \mu + \sigma\Phi^{-1}(u)$ is the normal qf.

Equation (10) is practical for simulations. Hence, $X = Q(U)$ has the GOLLN distribution if U $\sim$ U$(0,1)$.

3 Properties of the GOLLN model

We determine a linear representation for the GOLLN density in terms of EN densities, whose properties in [27] can be used to find its mathematical properties.

The structural properties of the GOLL-G family can be obtained from those of the *exponentiated*-G (Exp-G) class, see [17]. Formally, for more than fifty baselines G, several authors studied the characteristics of the Exp-G class, e.g., [28] for Exp-Log-Normal, [29] for Exp-Gamma, and [30] for Exp-Gumbel, among others.

So, the pdf (8) can follow from equations in [17] as

$$f(x) = \sum_{k=0}^{\infty} b_k\, h_{k+1}(x), \quad (11)$$

where $h_{k+1} = (k+1)\,\sigma^{-1}\,\phi\,(z)\,\Phi^{k}\,(z)$ is the EN density with power parameter $(k+1)$ (for $k \geq 0$), and

$$b_k = \frac{\alpha\theta}{k+1}\sum_{i,j=0}^{\infty}\sum_{l=k}^{\infty}(-1)^{j+k+l}\binom{-2}{i}\binom{l}{k}\times \binom{-(i+1)\alpha}{j}\binom{(i+1)\alpha\theta+j\theta-1}{l}.$$

Equation (11) is the main result of this section. Thus, using well-established properties for the EN distribution in [27], we derive the GOLLN mathematical properties more simply and accurately.

3.1 Moments

If $Z = (X-\mu)/\sigma$, then the rv Z has the GOLLN$(\alpha,\theta,0,1)$ distribution. The moments of X can

be easily found from the moments of Z by $E(X^n) = E[(\mu + \sigma Z)^n] = \sum_{r=0}^{n} \binom{n}{r} \mu^{n-r}\sigma^r E(Z^r)$. Like so, the standard representation for the GOLLN distribution is preferred used.

The nth moment (μ'_n) of Z is calculated from [27]

$$\mu'_n = E(Z^n) = \sum_{k,i=0}^{\infty} (k+1)(2\pi)^{i/2}\, b_k\, c_{n,i} \sum_{r=0}^{i} \frac{(-1)^r 2^{-r}\binom{i}{r}}{(k+i+1-r)}, \tag{12}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are obtained from the recurrence equation $c_{n,i} = (ia_0)^{-1} \sum_{m=1}^{i}[m(n+1)-i]\, a_m\, c_{n,i-m}$, $a_i = 0$ for $i = 0, 2, 4, \ldots$ and $a_i = b_{(i-1)/2}$ for $i = 1, 3, \ldots$, $c_{n,0} = a_0^n$, and b_l is determined recursively from

$$b_{l+1} = \frac{1}{2(2l+3)} \sum_{s=0}^{l} \frac{(2s+1)(2l-2s+1)\, b_l\, b_{l-s}}{(s+1)(2s+1)}.$$

A second representation for μ'_n follows from [27] as

$$\mu'_n = \sum_{k=0}^{\infty} (k+1)\, b_k\, \tau_{n,r}, \tag{13}$$

where $\tau_{n,r} = E[Z^n \Phi^k(z)]$ is the (n,r)th probability weighted moment (PWM) (for n and r integers) represented by the Lauricella type A function as

$$\tau_{n,r} = 2^{n/2}\pi^{-(r+1/2)} \sum_{\substack{p=0 \\ (n+r-p)\ even}}^{r} 2^{-p}\, \pi^p \binom{r}{p} \Gamma\left(\frac{n+r-p+1}{2}\right) \times F_A^{(r-p)}\left(\frac{n+r-p+1}{2}; \frac{1}{2}, \ldots, \frac{1}{2}; \frac{3}{2}, \ldots, \frac{3}{2}; -1, \ldots, -1\right), \tag{14}$$

for $n + r - p$ even and $\tau_{n,r}$ vanishes for $n + r - p$ odd.

3.2 Generating function

The generating function (gf) of Z, i.e., $M(-t) = E(\mathrm{e}^{-tZ})$, can be expressed with auxiliary properties in [27] from (11) as

$$M(-t) = \frac{1}{\sqrt{2\pi}} \sum_{k,j=0}^{\infty} (k+1)\, b_k\, c_{k,j}\, J_j(t), \tag{15}$$

where $J_j(t) = (-1)^j \sqrt{2\pi}\, \frac{\partial^j}{\partial t^j}(\mathrm{e}^{t^2/2})$.

4 Estimation

Consider observed values $x_1, \cdots, x_n$ from the GOLLN$(\alpha, \theta, \mu, \sigma)$ distribution. The log-likelihood function for $\psi = (\alpha, \theta, \mu, \sigma)^\top$ is defined as

$$\begin{aligned} l(\psi) =& n\log(\alpha\theta) - n\log(\sigma) + \sum_{i=1}^{n} \log\left[\phi\left(z_i\right)\right] + \\ & (\alpha\theta - 1)\sum_{i=1}^{n} \log\left[\Phi\left(z_i\right)\right] + \\ & (\alpha - 1)\sum_{i=1}^{n} \log\left\{\left[1 - \Phi\left(z_i\right)^\theta\right]\right\} - \\ & 2\sum_{i=1}^{n} \log\left\{\Phi\left(z_i\right)^{\alpha\theta} + \left[1 - \Phi\left(z_i\right)^\theta\right]^\alpha\right\}, \end{aligned} \tag{16}$$

where $z_i = (x_i - \mu)/\sigma$.

The MLE $\hat{\psi}$ can be found by maximizing (16) in R software (optim funtion). First, we fitted the reduced model for $\alpha = \theta = 1$ to get initial values for baseline parameters.

By differentiating (16) for each parameter, we can determine four nonlinear equations for $\hat{\psi}$. However, they cannot be solved analytically but any Newton-Raphson type algorithm or the numerical BGFS procedure in R software can solve them numerically.

4.1 Simulation study

The estimate properties are examened by the measures: bias, mean square error (MSE), estimated average length (AL), and coverage probability (CP). One thousand samples of size $n = 50, 55, \ldots, 750$ are drawn from the GOLLN distribution using Equation (10) with true parameters: $\alpha = 0.80$, $\theta = 0.50$, $\mu = 0.65$, and $\sigma = 1.25$. For each sample, the MLEs and their standard errors (SEs) are calculated, and the biases and MSEs can be expressed as

$$\widehat{Bias}_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon), \quad \widehat{MSE}_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon)^2, \tag{17}$$

whereas the ALs and CPs are given by

$$AL_\epsilon(n) = \frac{3.919928}{N}\sum_{i=1}^{N} s_{\hat{\epsilon}_i}, \tag{18}$$

$$CP_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N} I(\hat{\epsilon}_i - 1.95996 \cdot s_{\hat{\epsilon}_i}, \hat{\epsilon}_i + 1.95996 \cdot s_{\hat{\epsilon}_i}), \tag{19}$$

for $\epsilon = \alpha, \theta, \mu\, \sigma$.

Figures 4-6 display these measures versus the sample size n. Based on these plots, the biases approach zero when n increases (as expected). Similar results are highlighted for the MSEs. In addition, the ALs decrease, and the CPs converge to the nominal level 0.95 when n

increases. These results demonstrate the consistency of the MLEs.

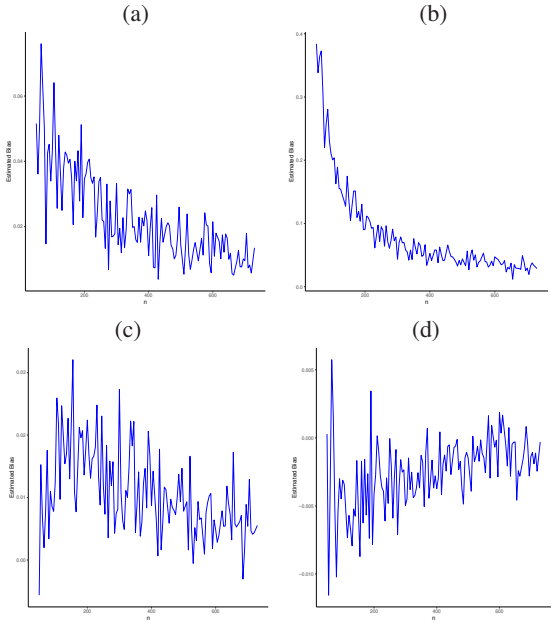


Fig. 4: Biases versus sample size. (a) $\hat{\alpha}$. (b) $\hat{\theta}$. (c) $\hat{\mu}$. (d) $\hat{\sigma}$.

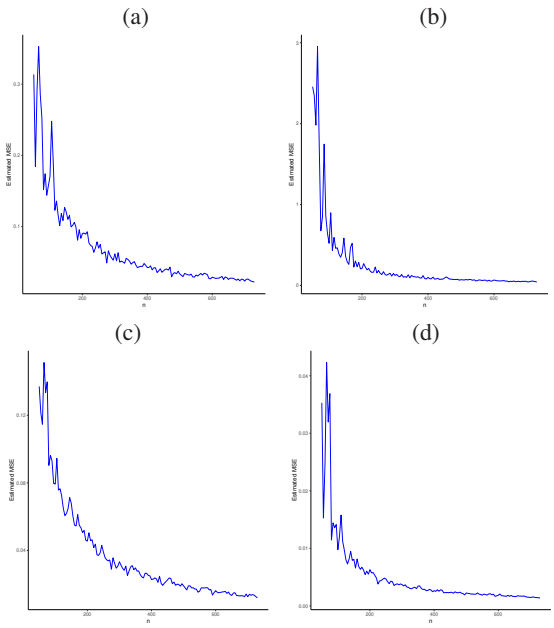


Fig. 5: MSEs versus sample size. (a) $\hat{\alpha}$. (b) $\hat{\theta}$. (c) $\hat{\mu}$. (d) $\hat{\sigma}$.

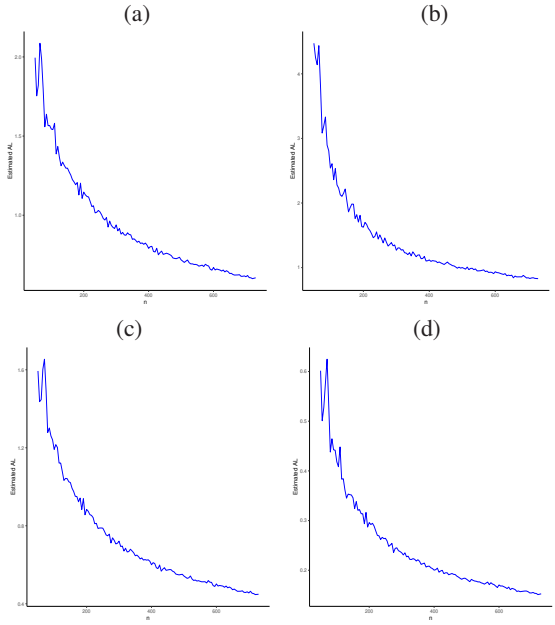


Fig. 6: ALs versus sample size. (a) $\hat{\alpha}$. (b) $\hat{\theta}$. (c) $\hat{\mu}$. (d) $\hat{\sigma}$.

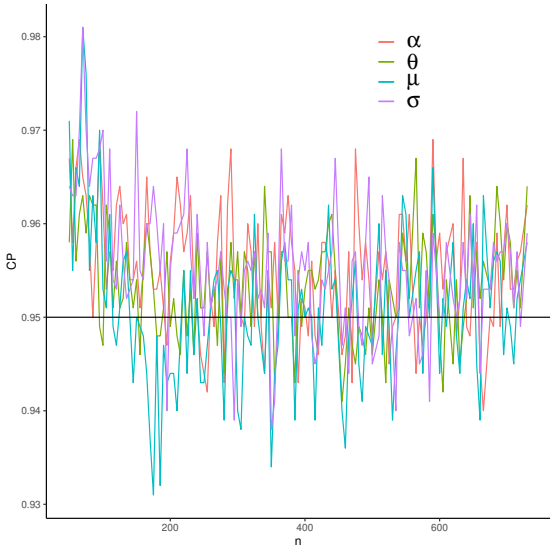


Fig. 7: CPs versus sample size.

5 The GOLLN regression model

5.1 Definition

Regression models for survival analysis are commonly used in medical sciences and engineering in different ways. Recently, generalized regressions have been the focus of several papers, which have become an invaluable

source in applied statistics, both in the context of uncensored and censored data.

For example, [19] proposed a regression based on the GOLL flexible Weibull distribution, [21] introduced the GOLLMaxwell regression applied to microbiology, and [18] presented two GOLL-log-normal regressions (one parametric and other with an additive non-parametric part). In this section, we define a GOLLN regression which provides interesting results for modeling censored survival data as an affordable alternative to location-scale regressions.

For defining a lifetime regression, we specify the distribution of a lifetime given a vector $\boldsymbol{x} = (x_1, \cdots, x_p)^\top$ of covariates, and a function of the parameters which depend on $\boldsymbol{x}$. We adopt the systematic component for the dispersion σ of the density function (8) via the log-linear link function to allow variation across the observations ($i = 1, \ldots, n$) such that

$$g(\sigma_i) = \exp(\boldsymbol{x}^\top \boldsymbol{\beta}), \tag{20}$$

where $g(\cdot)$ is a log-linear link function (twice continuously differentiable), and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$ is the parameter vector of dimension p.

If $X \sim$ GOLLN$(\alpha, \theta, \mu, \sigma)$, the density function of $Z = (X - \mu)/\sigma$ has the form

$$f(z) = \frac{\alpha\theta\phi(z)\Phi(z)^{\alpha\theta-1}[1-\Phi(z)^\theta]^{\alpha-1}}{\sigma\{\Phi(z)^{\alpha\theta} + [1-\Phi(z)^\theta]^\alpha\}^2}. \tag{21}$$

We define the GOLLN regression by $X_i = \mu + \sigma_i Z_i$, where the random component Z_i has density (21), and the dispersion parameter σ_i varies with the observations under the systematic component (20).

5.2 Estimation

Consider $(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)$ from unrelated observations (sample size n). The response y_i is defined by $min\{t_i, c_i\}$, where t_i and c_i are the survival and censored times, and let F and C denote the associated sets, respectively.

For the parameter vector $\psi = (\alpha, \theta, \mu, \boldsymbol{\beta}^\top)^\top$ of the regression model given by (20) and (21), the total log-likelihood can be expressed as $l(\psi) = \sum_{i \in F} l_i(\psi) + \sum_{i \in C} l_i^{(c)}(\psi)$, where $l_i(\psi) = \log[f(y_i)]$ and $l_i^{(c)}(\psi) = \log[S(y_i)]$, and $f(y_i)$ and $S(y_i)$ are the pdf and survival function of Y_i, respectively. Thus, $l(\psi)$ can be expressed as

$$\begin{aligned} l(\psi) =& n\log(\alpha\theta) - \sum_{i\in F}\log(\sigma_i) + \sum_{i\in F}\log\left[\phi\left(z_i\right)\right] + \\ & (\alpha\theta - 1)\sum_{i\in F}\log\left[\Phi\left(z_i\right)\right] + \\ & (\alpha - 1)\sum_{i\in F}\log\left\{\left[1-\Phi\left(z_i\right)^\theta\right]\right\} - \\ & 2\sum_{i\in F}\log\left\{\Phi\left(z_i\right)^{\alpha\theta} + \left[1-\Phi\left(z_i\right)^\theta\right]^\alpha\right\} + \\ & \sum_{i\in C}\log\left\{1 - \frac{\Phi\left(z_i\right)^{\alpha\theta}}{\Phi\left(z_i\right)^{\alpha\theta} + \left[1-\Phi\left(z_i\right)^\theta\right]^\alpha}\right\}, \end{aligned} \tag{22}$$

where $z_i = (x_i - \mu)/\sigma_i$ is the transformed varible.

The MLE of ψ can be computed by maximizing Equation (22) using R software (optim function). Fitting the reduced regression with $\alpha = \theta = 1$ provides the initial values for $\boldsymbol{\beta}$ and μ.

6 Diagnostic and residual analysis

The residual analysis aims to investigate features that compromise the validity of the model, i.e., analysis of inherent characteristics in the data. When checking for outliers, for example, several approaches are reported by [31], [32] and [33]. Hence, we adopt diagnostic measures based on the exclusion of observations to find out influential observations in the proposed regression. The systematic component based on the exclusion of observations follows from Equation (20)

$$g(\sigma_l) = \exp(\boldsymbol{x}_{\boldsymbol{l}}^\top \boldsymbol{\beta}), \quad l = 1, \ldots, n, \quad l \neq i. \tag{23}$$

Henceforth, the subscript (i) denotes the observation deleted from the dataset. Thus, $l_{(i)}(\psi)$ is the log-likelihood function for ψ from (23) which is maximized by $\hat{\psi}_{(i)}$. In this context, the difference between $\hat{\psi}_{(i)}$, and $\hat{\psi}$ show the influence of the ith case on the MLE of ψ and more consideration should be paid to this observation.

For investigating the influential observations, we define as a first measure the generalized Cook's distance by

$$GD_i = (\hat{\psi}_{(i)} - \hat{\psi})^\top [\ddot{\boldsymbol{L}}(\hat{\psi})](\hat{\psi}_{(i)} - \hat{\psi}). \tag{24}$$

A second usual measure is the likelihood distance

$$LD_i = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(i)})\}. \tag{25}$$

According to [34], despite censoring time, the deviance residuals have been commonly used to assess the goodness-of-fit of regression models. Further, they can be adopted for the GOLLN regression to study the model assumptions or presence of outliers.

It follows that, for censored data, the deviance residuals for the GOLLN regression are given by

$$r_{D_i} = \text{sgn}(\hat{r}_{M_i})\{-2[\hat{r}_{M_i} + \delta_i \log(\delta_i - \hat{r}_{M_i})]\}^{1/2}, \quad (26)$$

where

$$\hat{r}_{M_i} = \delta_i + \log\left\{1 - \frac{\hat{\alpha}\hat{\theta}\phi(\hat{z}_i)^{\hat{\alpha}\hat{\theta}-1}[1-\phi(\hat{z}_i)^{\hat{\theta}}]^{\hat{\alpha}-1}}{\hat{\sigma}_i\{\phi(\hat{z}_i)^{\hat{\alpha}\hat{\theta}} + [1-\phi(\hat{z}_i)^{\hat{\theta}}]^{\hat{\alpha}}\}^2}\right\} \quad (27)$$

are the martingale residuals, δ_i is the censoring indicator, sign$(\cdot)$ is the signal function, and $\hat{z}_i = (y_i - \hat{\mu})/\hat{\sigma}_i$.

If the regression model has a good fit, the martingale and deviance residuals should exhibit a random pattern around zero. Further, [35] proposed the construction of envelopes to support the analysis of the residuals with normal probability (NP) plots. Confidence bands are simulated for these envelopes, and if the model gives a good fit, most of the points will lie randomly inside. The construction of these confidence bands can be calculated using the following steps:

(i)Calculate the r_{D_i}'s for the considered model;
(ii)Using the fitted model, the response variable is simulated (k samples);
(iii)Calculate the deviance residuals for each fitted model to the sample (for $j = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, n$);
(iv)Sort the n residuals to have $r_{D_{(i)j}}$'s for each group;
(v)Calculate descriptive statistics (mean, minimum and maximum) of arrange the residuals for each i;
(vi)Plot the residuals r_{D_i} versus the expected percentile of the standard normal, and obtain the descriptive statistics.

7 Simulation study

The properties of the MLEs of the GOLLN regression model are investigated. One-thousand simulations are done using R software. We generate survival times t_i, and observations x_i from a uniform U(0,1) (for $i = 1, \ldots, n$). The MLEs are calculated for each replication of size n from the GOLLN$(\alpha, \theta, \mu, \sigma_i)$ regression.

We consider three scenarios for sample sizes $n = 100, 350, 850$, and censoring percentages approximately equal to 0% (scenario 1), 10% (scenario 2) and 30% (scenario 3). For three scenarios, the parameters are fixed as: $\beta_0 = -2.45$, $\beta_1 = 0.35$, $\alpha = 0.80$, $\theta = 0.50$, and $\mu = 0.65$. The generation process is given below:

(i)Generate observations as $x_i \sim U(0, 1)$;
(ii)Generate censored observations as $c \sim U(0, \nu)$, where ν is the proportion of censoring data;
(iii)Generate $z_i \sim$ GOLLN$(\alpha, \theta, 0, 1)$ from Equation (21);
(iv)Set $t_i = \mu + \exp(\beta_0 + \beta_1 x_i) z_i$;
(v)Let $y_i = min(t_i, c)$;
(vi)Define a vector δ (length n) that gives 1 if $y_i \leq c$, and 0 otherwise.

For each sample, the MLEs and SEs are found. For the three scenarios, Tables 2-4[3] reveal that the AEs tend to the true values. The biases and MSEs converge to zero (as expected), and the ALs decrease when n increases. Further, the CPs approach to the nominal value when n increases.

Table 2: Simulation results for scenario 1.

		0% censoring				
n	ψ	AE	Bias	MSE	AL	C(ψ)
100	β_0	-2.5821	-0.1321	0.3281	2.0380	0.944
	β_1	0.3588	0.0088	0.0654	0.8946	0.919
	α	0.9304	0.1304	1.0906	2.9130	0.970
	θ	0.8531	0.3531	2.6830	3.5146	0.968
	μ	0.6515	0.0015	0.0211	0.4524	0.902
350	β_0	-2.4856	-0.0356	0.0604	0.9143	0.941
	β_1	0.3484	-0.0016	0.0150	0.4818	0.961
	α	0.8171	0.0171	0.0558	0.8614	0.944
	θ	0.5415	0.0415	0.0720	0.9967	0.949
	μ	0.6533	0.0033	0.0030	0.2134	0.945
850	β_0	-2.4519	-0.0019	0.0234	0.5717	0.936
	β_1	0.3464	-0.0036	0.0061	0.3080	0.956
	α	0.8157	0.0157	0.0193	0.5297	0.944
	θ	0.5180	0.0180	0.0273	0.6142	0.945
	μ	0.6514	0.0014	0.0012	0.1326	0.943

Table 3: Simulation results for scenario 2.

		10% censoring				
n	ψ	AEs	Biases	MSEs	ALs	C(ψ)
100	β_0	-2.5827	-0.1327	0.4106	2.2411	0.938
	β_1	0.3474	-0.0026	0.0618	0.9216	0.932
	α	0.9924	0.1924	1.7933	3.4200	0.965
	θ	0.8724	0.3724	4.5549	4.2369	0.974
	μ	0.6557	0.0057	0.0267	0.5114	0.930
350	β_0	-2.4907	-0.0407	0.0617	0.9831	0.959
	β_1	0.3493	-0.0007	0.0163	0.5021	0.952
	α	0.8171	0.0171	0.0597	0.9000	0.943
	θ	0.5441	0.0441	0.0839	1.0455	0.935
	μ	0.6544	0.0044	0.0039	0.2259	0.942
850	β_0	-2.4680	-0.0180	0.0229	0.6003	0.951
	β_1	0.3511	0.0011	0.0067	0.3190	0.948
	α	0.7997	-0.0003	0.0192	0.5421	0.955
	θ	0.5203	0.0203	0.0265	0.6379	0.954
	μ	0.6501	0.0001	0.0012	0.1351	0.953

Due to the analytical difficulty in the mathematical treatment of the GOLLN distribution, to verify the regularity conditions, a simulation was done to check the

[3] The measures are denoted in Section 4.1, where AE means average estimate.

Table 4: Simulation results for scenario 3.

		30% censoring				
n	ψ	AEs	Biases	MSEs	ALs	C(ψ)
100	β_0	-2.5860	-0.1360	0.5443	2.7517	0.942
	β_1	0.3589	0.0089	0.0907	1.0283	0.896
	α	1.0929	0.2929	2.3628	5.3010	0.961
	θ	0.9407	0.4407	3.7158	4.4725	0.961
	μ	0.6535	0.0035	0.0288	0.5636	0.932
350	β_0	-2.5063	-0.0563	0.1028	1.2078	0.951
	β_1	0.3534	0.0034	0.0206	0.5531	0.944
	α	0.8272	0.0272	0.1014	1.1043	0.949
	θ	0.5671	0.0671	0.2029	1.2692	0.947
	μ	0.6537	0.0037	0.0061	0.2588	0.945
850	β_0	-2.4723	-0.0223	0.0342	0.7242	0.952
	β_1	0.3544	0.0044	0.0079	0.3534	0.953
	α	0.8003	0.0003	0.0250	0.6227	0.954
	θ	0.5286	0.0286	0.0368	0.7354	0.953
	μ	0.6491	-0.0009	0.0014	0.1506	0.942

accuracy of the MLEs. Likewise, for each fitted scheme, we get the deviance residuals r'_{D_i}s to analyze the empirical distribution (ED), and the normal probability plots.

Figures 8-10 display the normal probability plots for the deviance residuals. According to [36], these plots play the role of assessing departures from the normal assumption of the residuals. For all scenarios and n large, the ED of the deviance residuals is in conformity with the model assumptions. Also, Figures 11-13 reveal convergence to the true parameter values, and a normality shape.

8 Application

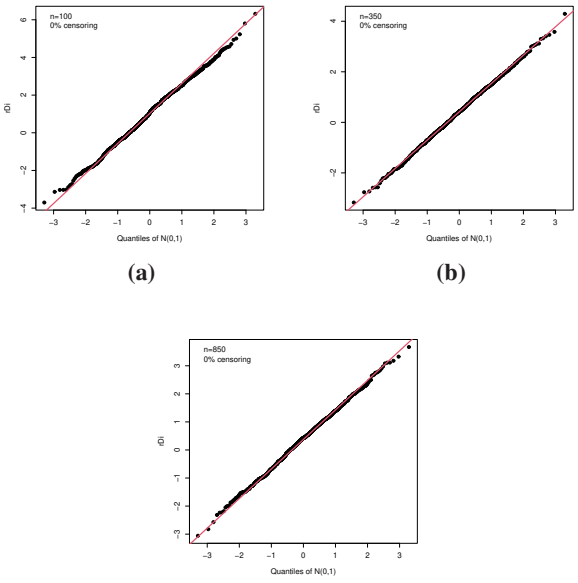
We perform an application of the proposed regression to COVID-19 lifetime data. We fit the normal, EN, OLLN and GOLLN distributions, and other non-nested models: transmuted exponentiated generalized normal (TEGNormal) [37] and [38], transmuted normal (TNormal) [37], beta-normal (BN) [39] and Kumaraswamy-normal (KwN) [40] distributions.

We determine the MLEs of the parameters and their SEs. In addition, we calculate the measures: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Cramér von-Mises (W^*), Anderson Darling (A^*) and Kolmogorov-Smirnov (KS) statistics from the fitted models using the *AdequacyModel* package [41] in R software. The best model gives smaller values of the good-of-fit (GoF) measures.

8.1 COVID-19 data

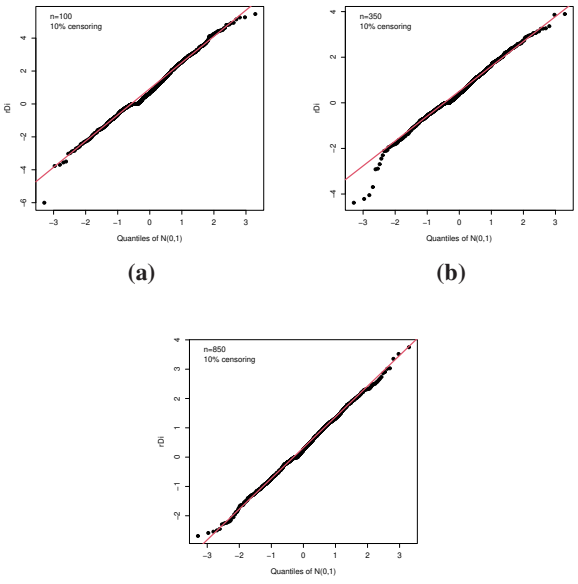
The dataset refers to 983 COVID-19 patients in the ICU obtained from SIVEP-Gripe[4].

[4] https://opendatasus.saude.gov.br/gl/dataset/bd-srag-2020



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

Fig. 8: NP plots for the r_{D_i}'s under scenario 1. (a) $n = 100$. (b) $n = 350$. (c) $n = 850$.



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

Fig. 9: NP plots for the r_{D_i}'s under scenario 2. (a) $n = 100$. (b) $n = 350$. (c) $n = 850$.

SIVEP-Gripe was created in 2012 by the Ministry of Health and has helped to oversee information about

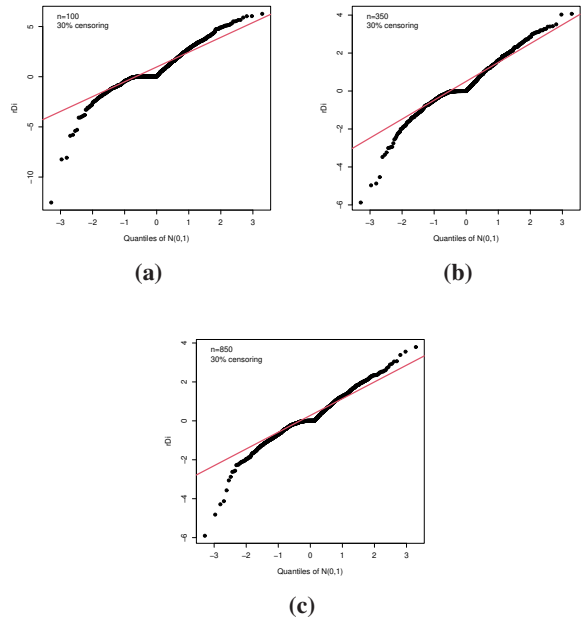


Fig. 10: NP plots for the r_{D_i}'s under scenario 3. (a) $n = 100$. (b) $n = 350$. (c) $n = 850$.

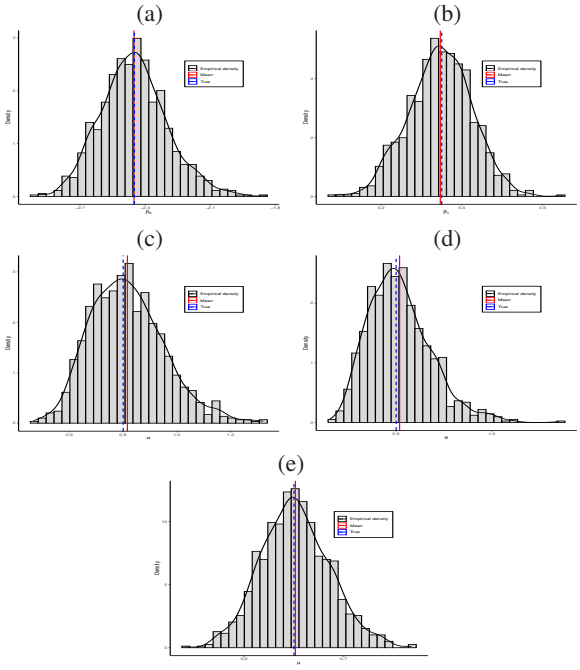


Fig. 11: Plots of the EDs for $n = 850$ in scenario 1. (a) $\hat{\beta}_0$. (b) $\hat{\beta}_1$. (c) $\hat{\alpha}$. (d) $\hat{\theta}$. (e) $\hat{\mu}$.

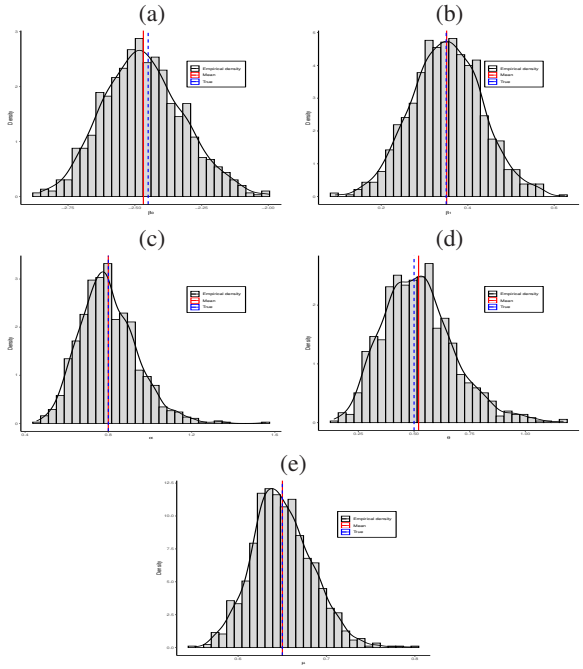


Fig. 12: Plots of the EDs for $n = 850$ in scenario 2. (a) $\hat{\beta}_0$. (b) $\hat{\beta}_1$. (c) $\hat{\alpha}$. (d) $\hat{\theta}$. (e) $\hat{\mu}$.

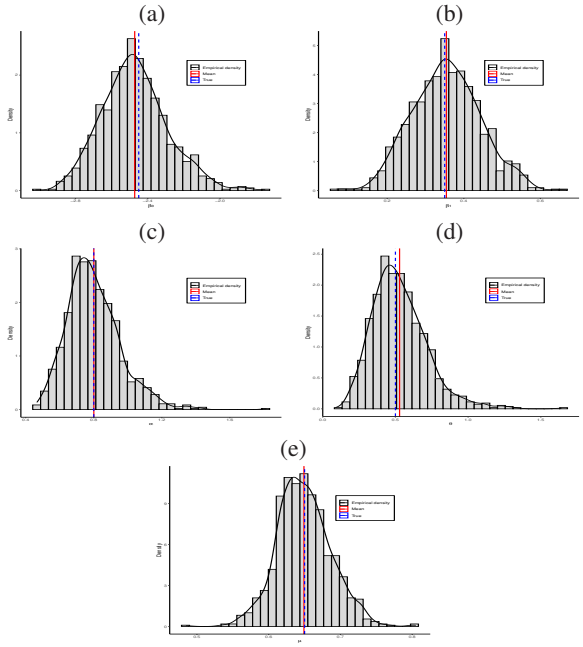


Fig. 13: Plots of the EDs for $n = 850$ in scenario 3. (a) $\hat{\beta}_0$. (b) $\hat{\beta}_1$. (c) $\hat{\alpha}$. (d) $\hat{\theta}$. (e) $\hat{\mu}$.

COVID-19 hospitalizations, admissions and deaths in Brazil. For this propose, it consolidates data from admissions of COVID-19 patients in both public and

private hospitals, as those are mandatory notifications throughout the Brazilian territory.

The registry[5] includes, for each patient, demographic issues, self-reported symptoms, comorbidities, ICU admission, ventilatory support, test results, hospital admission and discharge dates, hospitalization outcome, among several others. [42] presented a retrospective analysis of the COVID-19 cases in Brazil, thus illustrating demographic and descriptive characteristics to investigate the effect of coronavirus on health care resources employed to fight the pandemic, as well as on in-hospital mortality.

The period analysis is defined from the 8th epidemiological week (16th January 2020) to the 53th epidemiological week (2nd January 2021) at the Hospital de Base de São José do Rio Preto (state of São Paulo). All positive RT-PCR (RT-qPCR) and SARS-CoV-2 records of the patients admitted in the ICU of the hospital are included in the research.

The patients survival time is specified as the number of days admitted to ICU, while the censoring indicator refers to the hospitalization discharge before the 53th week of the study. We focus on the explanatory variables below:

– y_i: time in ICU (days);
– δ_i censoring indicator ($0 =$ censoring, $1 =$ uncensoring);
– x_{i1} is gender ($0 =$ male, $1 =$ female);
– x_{i2} is the age group ($0 = <65$ years, $1 = \geq 65$);
– x_{i3} is chronic cardiovascular disease ($0 =$ no, $1 =$ yes);
– x_{i4} is chronic hematological disease ($0 =$ no, $1 =$ yes);
– x_{i5} is chronic liver disease ($0 =$ no, $1 =$ yes);
– x_{i6} is diabetes mellitus ($0 =$ no, $1 =$ yes);
– x_{i7} is chronic neurological disease ($0 =$ no, $1 =$ yes);
– x_{i8} is other chronic pneumatopathy ($0 =$ no, $1 =$ yes);
– x_{i9} is immunodeficiency ($0 =$ no, $1 =$ yes);
– x_{i10} is chronic kidney disease ($0 =$ no, $1 =$ yes);
– x_{i11} is obesity ($0 =$ no, $1 =$ yes).

Table 5 reports a descriptive summary of ICU times for COVID-19 patients. The positive values for the skewness and kurtosis are confirmed in Figure 1.

Table 5: Descriptive statistics for the COVID-19 patients[6].

min	q1	median	mean	q3	max	sd	skewness	kurtosis
1.00	7.00	12.00	15.66	21.00	89.00	13.03	1.89	5.17

[6] q1 and q3 are the first and thrid quantile
sd means the standard deviation

Tables 6 and 7 report the MLEs and GoF measures for the fitted distributions to the current data, and reveal that the GOLLN distribution is the most suitable model.

Im fact, the GOLL model is competitive to the other models. The inclusion of the extra parameters (α and θ) is confirmed by three likelihood ratio (LR) tests in Table 8 which indicate that the extra parameters are highly significant for modeling these data.

The histogram and plots of the three best fitted densities for the GOLLN, BN, and KwN models in Figure 14(a) confirm that the estimated GOLLN distribution provides the best model. Further, the estimated cdfs of these models in Figure 14(b) reveal that the new distribution gives the best fit.

Table 6: Findings from the fitted models to COVID-19 data.

Model	μ	σ	α	θ	
GOLLN	**-14.587**	**12.533**	**0.391**	**73.636**	
	(0.035)	**(0.010)**	**(0.011)**	**(0.310)**	
OLLN	13.800	111.713	10.595	1	
	(0.362)	(42.421)	(4.056)	(-)	
EN	-39.837	24.266	1	53.697	
	(3.038)	(0.772)	(-)	(9.493)	
Normal	15.664	13.026	1	1	
	(0.415)	(0.294)	(-)	(-)	
Model	μ	σ	δ	λ	ν
TEGNormal	0.063	30.031	0.546	2.333	14.047
	(5.681)	(5.969)	(0.069)	(0.823)	(6.767)
GNormal	2.984	10.384	1.345	2.134	
	(0.020)	(0.012)	(0.055)	(0.104)	
TNormal	20.480	13.542	0.691		
	(0.573)	(0.344)	(0.053)		
Model	μ	σ	δ	λ	
BN	-22.116	13.409	32.803	0.350	
	(0.027)	(0.011)	(2.109)	(0.013)	
KwN	-14.981	12.583	16.434	0.360	
	(0.278)	(0.010)	(0.057)	(0.013)	

Table 7: GoF statistics[7].

Model	AIC	BIC	W*	A*	KS
GOLLN	**7425.687**	**7445.250**	**1.311**	**8.597**	**0.119** (<**0.0001**)
OLLN	7695.684	7710.356	4.643	27.721	0.126 (<0.0001)
EN	7552.998	7567.670	3.022	18.572	0.108 (<0.0001)
Normal	7840.381	7850.163	6.446	37.692	0.143 (<0.0001)
Model	AIC	BIC	W*	A*	KS
TEGNormal	7592.503	7616.956	3.380	20.583	0.116 (<0.0001)
GNormal	7765.378	7784.940	5.474	32.393	0.143 (<0.0001)
TNormal	7762.333	7777.005	5.507	32.400	0.124 (<0.0001)
Model	AIC	BIC	W*	A*	KS
BN	7462.971	7482.533	1.874	11.926	0.109 (<0.0001)
KwN	7494.852	7514.414	2.252	14.150	0.112 (<0.0001)

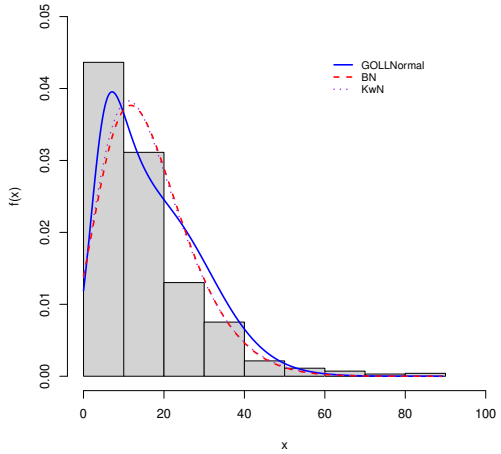
[7] The descriptions of GoF are denoted in Section 8

[5] https://opendatasus.saude.gov.br/dataset/srag-2020/resource/9f6ba348-0033-49b1-abbe-719a0ffbeb28

Table 8: LR tests for COVID-19 data.

Models	Statistic w	p-value
GOLLN vs OLLN	271.997	< 0.0001
GOLLN vs EN	129.311	< 0.0001
GOLLN vs Normal	418.694	< 0.0001

(a)



(b)

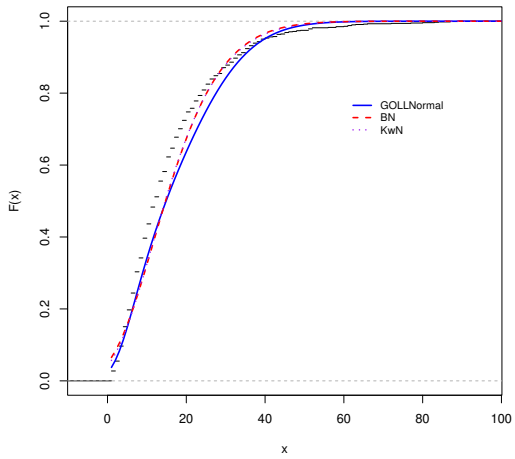


Fig. 14: (a) Estimated pdfs for three models. (b) Estimated cdfs for three models.

Next, we consider the systematic structures below:

$$\sigma_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \beta_{10} x_{i10} + \beta_{11} x_{i11}),$$

for $i = 1, \ldots, 983$.

Table 9 contain the results (MLEs, SEs and p-value) for the fitted GOLLN regression to the COVID-19 data.

Table 9: Results from GOLLN regression[8].

Parameter	MLE	SE	p-value
β_0	2.9349	0.1136	< 0.0001
β_1	0.2685	0.0736	0.0002
β_2	-0.2153	0.0756	0.0045
β_3	- 0.4852	0.1052	$<$0.0001
β_4	- 0.7283	0.2653	0.0062
β_5	- 0.4531	0.1845	0.0142
β_6	- 0.1568	0.0706	0.0266
β_7	- 0.5007	0.0926	$<$0.0001
β_8	- 0.3482	0.1136	0.0022
β_9	- 0.2545	0.1150	0.0271
β_{10}	- 0.3313	0.0975	0.0007
β_{11}	0.3590	0.0762	$<$0.0001

[8] The descriptions of β are in subsection 8.1

Some conclusions are drawn from Table 9:

- All explanatory variables are significant (5% level). So, the groups within each covariate, are not statistically equal;
- β_1 represents the variable gender with a p-value of 0.0002, and its estimate is positive. Thus, the female patients has survival time in ICU larger than male patients;
- β_2 is the age group and its extremely significant (p-value 0.0045). The estimate is negative, and then patients with age $\geq$ 65 remain more time in ICU, which corroborates with the research in [43];
- From β_3 to β_{10} all covariates are significant, and has negative estimate. Studies from [44] and [45] analyze the association between comorbidities for the hospitalization and understand related predictors of ICU admission;
- Despite the estimate of β_{11} is positive, previus studies ([46] and [47]) indicated that obesity increases mortality and admission in COVID-19 patients.

8.2 Model checking

To detect possible influence points in the GOLLN regression, we illustrate in Figure 15 the GD and LD measures. These plots reveal some possible influential observations, but we can verify that there is a small amount of observations, which do not impact the fitted model.

Finally, Figure 16 show the deviance residuals with simulated envelope. These plot display that they lie inside the bands randomly, thus confirming a good fit of the GOLLN regression model.

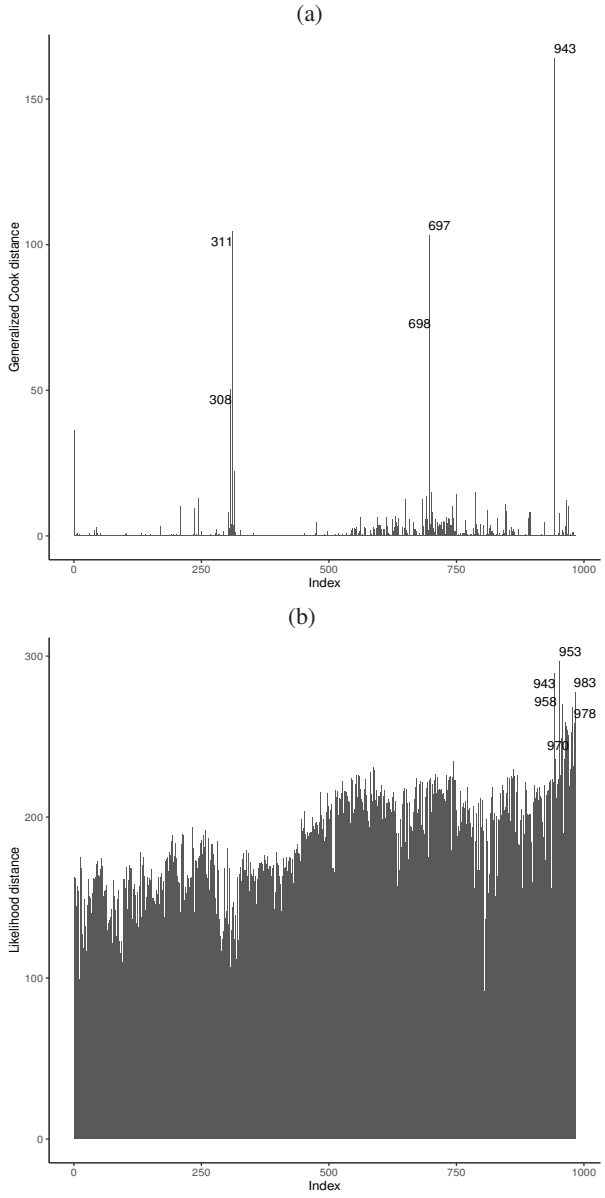


Fig. 15: (a) GD for the fitted GOLLN regression. (b) LD for the fitted GOLLN regression.

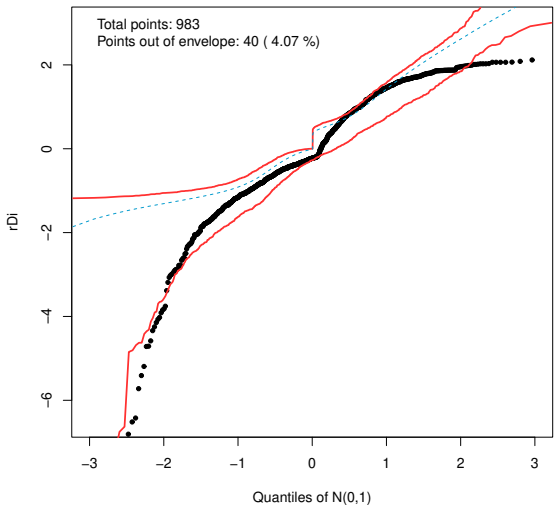


Fig. 16: NP plots for the r_{D_i}'s with envelope from the fitted GOLLN regression to COVID-19 data.

9 Concluding Remarks

We defined a four-parameter model called the generalized odd log-logistic normal distribution (GOLLN). We determined a linear combination for its density, obtained some mathematical properties, maximum likelihood estimates, and carried out several simulations for different scenarios to evaluate the consistency of the estimates. We constructed a GOLLN regression, and checked the accuracy of the estimators by Monte Carlo simulations. The normality assumptions were tested by evaluating the normal probability plots and the empirical distribution of the variance residuals. A diagnostic analysis is addressed to test global influential observations and residuals analysis to evaluate model fit through simulated envelopes. The usefulness of the introduced regression was discussed by an application for modeling survival times in a ICU of the COVID-19 patients. The proposed model was the best to explain the current data. Further, some helpful findings are presented to COVID-19 data.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] S. Richardson, J.S. Hirsch, M. Narasimhan, J.M. Crawford, T. McGinn, K.W. Davidson and the Northwell COVID-19 Research Consortium. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area, *JAMA*, **323**, 2052-2059, (2020).

[2] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen and B. Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, *The Lancet*, **395**, 1054-1062, (2020).

[3] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya,

B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, Sheffield COVID-19 Genomics Group, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire and D.C. Montefiori. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, *Cell*, **182**, 812-827, (2020).

[4] M. Alkhouli, A. Nanjundappa, F. Annie, M.C. Bates and D. L. Bhatt. Sex differences in COVID-19 case fatality rate: insights from a multinational registry, *Mayo Clinic Proceedings*, **95**, 1613–1620, (2020).

[5] J. Hewitt, B. Carter, A. Vilches-Moraga, T. Quinn, P. Braude, A. Verduri, L. Pearce, M. Stechman, R. Short, A. Price, J. Collins, E. Bruce, A. Einarsson, F. Rickard, E. Mitchell, M. Holloway, J. Hesford, F. Barlow-Pay, E. Clini and G. Guaraldi. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study, *The Lancet Public Health*, **5**, 444-451, (2020).

[6] J.M. Jin, P. Bai, W. He, F. Wu, X.F. Liu, D.M. Han, S. Liu, and J.K. Yang. Gender differences in patients with COVID-19: focus on severity and mortality, *Frontiers in Public Health*, **8**, (2020).

[7] J.Y. Kim, P.G. Choe, Y. Oh, K.J. Oh, J. Kim, S.J. Park, J.H. Park, H.K. Na and M.D. Oh. The first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures, *Journal of Korean Medical Science*, **35**, (2020).

[8] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C. B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak and H.C. Lane. Remdesivir for the treatment of COVID-19 — final report, *New England Journal of Medicine*, **383**, 1813-1826, (2020).

[9] D. Neogi. Performance appraisal of select nations in mitigation of COVID-19 pandemic using entropy based TOPSIS method, *Ciência & Saúde Coletiva*, **26**, 1419-1428, (2021).

[10] P.R.D. Marinho, G.M. Cordeiro, H.F.C. Coelho and S.C.S. Brandão. COVID-19 in Brazil: a sad scenario, *Cytokine Growth Factor Rev.*, **58**, 51-54, (2020).

[11] O.P. Mehta, P. Bhandari, A. Raut, S.E.O. Kacimi and N.T. Huy. Coronavirus disease (COVID-19): comprehensive review of clinical presentation, *Frontiers in Public Health*, **8**, (2021).

[12] I. Voinsky, G. Baristaite and D. Gurwitz. Effects of age and sex on recovery from COVID-19: analysis of 5769 Israeli patients, *The Journal of Infection*, **81**, 102–103, (2020).

[13] N. AL-Rousan and H. AL-Najjar. Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases, *Journal of Medical Virology*, **92**, 1603-1608, (2020).

[14] Y. Yan, Y. Yang, F. Wang, H. Ren, S. Zhang, X. Shi, X. Yu and K. Dong. Clinical characteristics and outcomes of patients with severe COVID-19 with diabetes, *BMJ Open Diabetes Res. Care*, **8**, (2020).

[15] G. Salinas-Escudero, M.F. Carrillo-Vega, V. Granados-García, S. Martínez-Valverde, F. Toledano-Toledano and J. Garduño-Espinosa. A survival analysis of COVID-19 in the Mexican population, *BMC Public Health*, **20**, 1-8, (2020).

[16] G. Sousa, T. Garces, V. Cestari, R. Florêncio, T. Moreira and M. Pereira. Mortality and survival of COVID-19, *Epidemiology and Infection*, **148**, (2020).

[17] G.M. Cordeiro, M. Alizadeh, G. Ozel, E.M.M. Ortega and E. Altun. The generalized odd log-logistic family of distributions: properties, regression models and applications, *Journal of Statistical Computation and Simulation*, **87**, 908-932, (2016).

[18] J.C.S. Vasconcelos, G.M. Cordeiro, E.M.M. Ortega and M.A.M. Biaggioni. The parametric and additive partial linear regressions based on the generalized odd log-logistic log-normal distribution, *Communications in Statistics - Theory and Methods*, **51**, 3480-3507, (2020).

[19] F. Prataviera, E.M.M. Ortega, G.M. Cordeiro, R.R. Pescim and B.A.W. Verssani. A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems, *Reliability Engineering & System Safety*, **176**, 13-26, (2018).

[20] A.Z. Afify, A.K. Suzuki, C. Zhang and M. Nassar. On three-parameter exponential distribution: properties, Bayesian and non-Bayesian estimation based on complete and censored samples, *Communications in Statistics - Simulation and Computation*, **50**, 3799-3819, (2021).

[21] F. Prataviera, A.M.M. Silva, E.J.B.N. Cardoso, G.M. Cordeiro and E.M.M. Ortega. A novel generalized odd log-logistic Maxwell-based regression with application to microbiology, *Applied Mathematical Modelling*, **93**, 148-164, (2021).

[22] F. Prataviera, E.M.M. Ortega and G.M. Cordeiro. A new bimodal Maxwell regression model with engineering applications, *Applied Mathematics & Information Sciences*, **14**, 817-831, (2020).

[23] J.U. Gleaton and J. Lynch, Properties of generalized log-logistic families of lifetime distributions, *Journal of Probability and Statistical Science*, **4**, 51-64, (2006).

[24] A. Alzaatreh, C. Lee and F. Famoye. A new method for generating families of continuous distributions, *METRON*, **71**, 63-79, (2013).

[25] R.C Gupta and R.D. Gupta. Proportional reversed hazard rate model and its applications, *Journal of Statistical Planning and Inference*, **137**, 3525-3536, (2007).

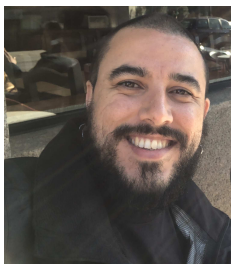
[26] A.S. Braga, G.M. Cordeiro, E.M.M. Ortega and J.N da Cruz. The odd log–logistic normal distribution: theory and applications in analysis of experiments, *Journal of Statistical Theory and Practice*, **10**, 311-335, (2016).

[27] G.M. Cordeiro, R.J. Cintra, L.C. Rego and E.M.M. Ortega. The McDonald normal distribution, *Pakistan Journal of Statistics and Operation Research*, **8**, 301-329, (2012).

[28] D.T. Shirke and C.S. Kakade. On exponentiated log-normal distribution, *International Journal of Agricultural and Statistical Science*, **2**, 319-326, (2006).

[29] S. Nadarajah and A.K. Gupta. The exponentiated gamma distribution with application to drought data, *Calcutta Statist. Assoc. Bull.*, **59**, 29-54, (2007).

[30] S. Nadarajah. The exponentiated Gumbel distribution with climate application, *Environmetrics*, **17**, 13-23, (2005).
[31] D.R. Cox and E.J. Snell. A general definition of residuals, *Journal of the Royal Statistical Society B*, **30**, 248-275, (1968).
[32] R.D. Cook and S. Weisberg. *Residuals and influence in regression*, Chapman and Hall, New York, (1982).
[33] E.M.M. Ortega, G.A. Paula and H. Bolfarine. Deviance residuals in generalized log-gamma regression models with censored observations, *Journal of Statistical Computation and Simulation*, **78**, 747-764, (2008).
[34] G.O. Silva, E.M.M. Ortega and G.A. Paula. Residuals for log-Burr XII regression models in survival analysis, *Journal of Applied Statistics*, **38**, 1435-1445, (2011).
[35] A.C. Atkinson. *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*, Oxford University Press, New York, (1985).
[36] S. Weisberg. *Applied linear regression*, John Wiley and Sons, New Jersey, (2005).
[37] W.T. Shaw and I.R.C. Buckley. The alchemy of probability distributions: beyond Gram-Charlier expansions and a skew-kurtotic-normal distribution from a rank transmutation map, *Statistical Finance*, (2007).
[38] G.M. Cordeiro, E.M.M. Ortega and D. Cunha. The exponentiated generalized class of distributions, *Journal of Data Science*, **11**, 1-27, (2013).
[39] N. Eugene, C. Lee and F. Famoye. Beta-normal distribution and its applications, *Communications in Statistics - Theory and Methods*, **31**, 497-512, (2002).
[40] G.M. Cordeiro and M. Castro. A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, **81**, 883-898, (2011).
[41] P.R.D. Marinho, M. Bourguignon and C.R.B. Dias. AdequacyModel: adequacy of probabilistic models and general purpose optimization, *PLOS ONE*, **14**, (2016).
[42] O.T. Ranzani, L.S.L. Bastos, J.G.M. Gelli, J.F. Marchesi, F. Baião, S. Hamacher and F.A. Bozza. Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data, *The Lancet Respiratory Medicine*, **9**, 407-418, (2021).
[43] K.C. Mascarello, A.C.B. C. Vieira, A.S.S. Souza, W.D. Marcarini, V.G. Barauna and E.L.N. Maciel.COVID-19 hospitalization and death and relationship with social determinants of health and morbidities in Espírito Santo State, Brazil: a cross-sectional study, *Epidemiologia e Serviços de Saúde : Revista do Sistema Unico de Saude do Brasil*, **30**, (2021).
[44] G. Iaccarino, G. Grassi, C. Borghi, S. Carugo, F. Fallo, C. Ferri, C. Giannattasio, D. Grassi, C. Letizia, C. Mancusi, P. Minuz, S. Perlini, G. Pucci, D. Rizzoni, M. Salvetti, R. Sarzani, L. Sechi, F. Veglio, M. Volpe, M.L. Muiesan and SARS-RAS Investigators. Gender differences in predictors of intensive care units admission among COVID-19 patients: The results of the SARS-RAS study of the Italian Society of Hypertension, *PLoS ONE*, **10**, e0237297, (2020).
[45] D.A.M. Meijs, B.C.T. van Bussel, B. Stessel, J. Mehagnoul-Schipper, A. Hana, C.I.E. Scheeren, S.A.E. Peters, W.N.K.A. van Mook, I.C.C. van der Horst, G. Marx, D. Mesotten, C. Ghossein-Doha and CoDaP Investigators. Better COVID-19 Intensive Care Unit survival in females, independent of age, disease severity, comorbidities, and treatment, *Scientific Reports*, **12**, (2021).
[46] L. Sjögren, E. Stenberg, M. Thuccani, J. Martikainen, C. Rylander, V. Wallenius, T. Olbers, and J.M. Kindblom. Impact of obesity on intensive care outcomes in patients with COVID-19 in Sweden-A cohort study, *PLoS One*, **16**, (2021).
[47] V.B. Paravidino, T.H. Leite, M.F.F Mediano, F. Mediano, R. Sichieri, G.A. Silva, V. Cravo, A. Balduino, E. Salgueiro, B.A.M.P. Besen, R.C. Moreira, C.E. Brandão, D.C.K. Gomes, C.A.G. Assemany and P. Cougo. Association between obesity and COVID-19 mortality and length of stay in intensive care unit patients in Brazil: a retrospective cohort study, *Scientific Reports*, **12**, (2022).



Nicollas Costa is PhD student in Statistic from Federal University of Pernambuco, UFPE, Brazil (2020). His research interests are parametric and semiparametric regression models, distribution theory and survival analysis.

Gauss Cordeiro is a Brazilian statistician who has made significant contributions to the theory of statistical inference, mainly through asymptotic theory and applied probability. He received his PhD in Statistics in 1982 from Imperial College, London. Currently, Cordeiro is a Class A researcher of the Brazilian Research Council-CNPq, Full Professor of Federal University of Pernambuco (Brazil) and Member of the Post- Graduate Program in Statistics at the same University. He has published more than 430 research articles in international scientific journals with referee practice (2020) and supervised more than 65 MSc dissertations and DSc theses.