Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:74
https://doi.org/10.28919/cmbn/7374
ISSN: 2052-2541

A NEW REGRESSION MODEL FOR POISSON LINDLEY DISTRIBUTION WITH APPLICATION

ABDUL HADI N. EBRAHEIM, SALAH M. MOHAMED, KHADEEJAH ABDULLAH MUAYW*

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research,

Cairo University, Cairo, Egypt

Abstract: The main goal of this study is to propose a new regression model using re-parameterization of the Poisson-Lindley distribution for seven parameters. The utility of real-world data is used to assess the accuracy of estimating algorithms. The suggested model is compared to well-known regression models for count data modelling, such as Poisson, on a real data set to demonstrate its utility. While fitting two real data sets, the (GPL7) linear model will be compared to the Poisson for seven parameters and the (GPL4) linear model will be compared to the Poisson for four parameters. The GPL7 linear model was found to be capable of fitting over-dispersed count data and to have the maximum log-likelihood, according to the results.

Keywords: count data; over-dispersion; link functions; Poisson-Lindley distribution; Poisson regression model; maximum likelihood estimation.

2010 AMS Subject Classification: 93A30, 62J05.

---

*Corresponding author
E-mail address: kdojh_abdallh@yahoo.com
Received March 21, 2022

1. INTRODUCTION

The Poisson regression model [1,2,21] is a count data model in which the number of occurrences of a defined experiment is random and the conditional mean of occurrence equals the conditional variance (equi-dispersion).

However, challenged in real-world data from several trials contradicts the assumption that the conditional mean is equal to the conditional variance. When the conditional variance and mean are not equal, there are two scenarios. When the conditional variance is less than the conditional mean, under-dispersion occurs. In the case of over-dispersion, the conditional variance exceeds the conditional mean [24,27]. According to Palmisano's research and Karlis and Xekalaki's research, the over-dispersion situation has been widely investigated [9, 23].

Mover, there are two techniques to examine count data with over-dispersion. To begin with, there is a large body of work on alternative discrete mixed distributions that may handle various levels of over dispersion. The Poisson-gamma or negative binomial (NB) distribution [15], and the Poisson–Lindley distribution [18], are two new mixed Poisson distributions. As one method of analyzing count data with over-dispersion, these have been used to model count data with over-dispersion. In addition, the use of new mixed Poisson distributions with auxiliary variables has been proposed to provide more appropriate models for forecasting the behavior of the count response variable than the predictions from the Poisson regression model [2,15].

There are many new mixed Poisson regression models that can be used. For example, the NB regression model [15], generalized Waring regression model [20], Poisson-normal or Hermite regression model [8], Poisson-Inverse Gaussian regression model [19], and hyper-Poisson regression model [3] have been developed in the context of generalized linear models.

In 2014, Zamani et al. published a Poisson-weighted exponential (P-WE) distribution using a regression model. [13]. Wongrin and Winai (2016) established a novel linear regression model for count data called generalized-Poisson Lindley (GPL) [29]. By re-parameterizing Poisson quasi-Lindley, Altun presented an alternative regression model for modelling over-dispersed count data sets in 2019 [4]. Karim and Mountainlike at the modified Poisson-Lindley linear model [16]. The generalized Poisson-Lindley linear model is an option to modelling over dispersed count data when claim frequency data is generated from populations with a generalized Poisson-Lindley distribution. Its regression model and MLE were also created using the generalized linear model.

So, the purpose of this study is to use the generalized linear model to create a new linear regression model called the GPL linear model, which is based on a generalized Poisson–Lindley (GPL7) distribution for seven parameters. Mahmoudi and Zakerzadeh introduced the GPL distribution in 2010 [11]. It's a mixed Poisson distribution [21], which gives a flexible model for count data with over-dispersion by combining the Poisson distribution with a generalized Lindley (GL) distribution. In addition, the MLE is employed to estimate the model's parameters. Its results are also compared to certain classic count data regression models. The GPL7 linear model is a more accurate way to explain the relationship between count data and a set of covariates more accurately.

In addition to this paper proposes a novel regression model. The research is structured as follows: Section 2 shows a mixed Poisson-Lindely distribution for seven parameters. A new linear regression model for count data has been constructed based on the GPL7 release, as detailed in Section 3. In addition, Section 4 shows how to infer statistical parameter estimators. Apply the GPL7 linear model to real-world data sets, then analyze the data summary and model performance in Section 5. Finally, various conclusions are illustrated in Section 6.

2. Generalized Poisson Lindley Distribution for Seven Parameters

The analysis for count data will conducted by using the Poisson distribution, which is considered the basic distribution for it. If a random variable Y is distributed as the Poisson with parameter λ, its probability mass function (pmf) is

$$\mathrm{p}(\mathrm{y},\lambda)=\frac{e^{-\lambda}\lambda^{y}}{y!};y=0,1,2,\ldots\ldots\ for\ \lambda>0 \tag{1}$$

Consequently

$$E\left(y\right)=Var\left(y\right)=\lambda$$

the pdf of the seven Parameter Lindley Distribution (SPL) distribution, is defined as

$$g(\lambda,\theta,\alpha,\beta,k,\eta,\phi,\sigma)=\frac{\theta^2 k}{\eta^\sigma+\theta^\varnothing k}\left[\frac{k\,\theta^{\varnothing-1}(\theta\lambda)^{\alpha-1}}{\Gamma(\alpha)}+\frac{\eta^\sigma(\theta\lambda)^{\beta-1}}{\theta\Gamma(\beta)}\right]e^{-\theta\lambda},\lambda>0 \qquad (2)$$

We note that $g\left(\lambda;\theta,\alpha,\beta,k,\eta,\varnothing,\sigma\right)$ incorporates seven parameters namely.

$\theta>0,\alpha>0,\beta>0,k\geq 0,\varnothing>0,\sigma>0 \, and \;\; \eta\geq 0,$

Then

$$E\left(\lambda\right)=\frac{\theta^2}{\eta^\sigma+\theta^\phi k}\left(\alpha k\,\theta^{\phi-3}+\beta\eta^\sigma\theta^{\beta-2}\right)$$

$$\text{var}\left(\lambda\right)=\sigma^2=\frac{1}{\eta^\sigma+\theta^\Phi k}\left[k\theta^{\phi-2}\alpha\left(\alpha+1\right)+\eta^\sigma\theta^{\beta-2}\beta\left(\beta+1\right)\right]-\left[\frac{\alpha k\theta^{\phi-1}+\eta^\sigma\beta\theta^{\beta-2}}{\eta^\sigma+\theta^\phi k}\right]^2$$

Hence, The GPL7 distribution with parameters $\alpha,\beta,k,\eta,\emptyset,\sigma$ and θ.The marginal pmf GPL7 $\left(\lambda;\theta,\alpha,\beta,k,\eta,\varnothing,\sigma\right)$, and its probability density function (pdf) is of

Y ~ GPL7 $g\left(\lambda;\theta,\alpha,\beta,k,\eta,\varnothing,\sigma\right)$ is

$$f\left(y\right)=\frac{\theta^2 k}{\eta^\sigma+\theta^\phi k}\left[\frac{k\theta^{\alpha+\phi-2}\Gamma\left(\alpha+y\right)}{y!\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}}+\frac{\eta^\sigma\theta^{\beta-1}\Gamma\left(\beta+y\right)}{\theta\Gamma\left(\beta\right)y!\left(\theta+1\right)^{\beta+y}})\right] \qquad (3)$$

$$y=0,1,2........y\sim GPL7$$

Figures (1), (2), (3), and (4) show some of the possible shapes of the GPL7 pdf for various values of the parameters θ, α, β, k, ∅, σ and η selected from the ranges indicated in Equation (3).

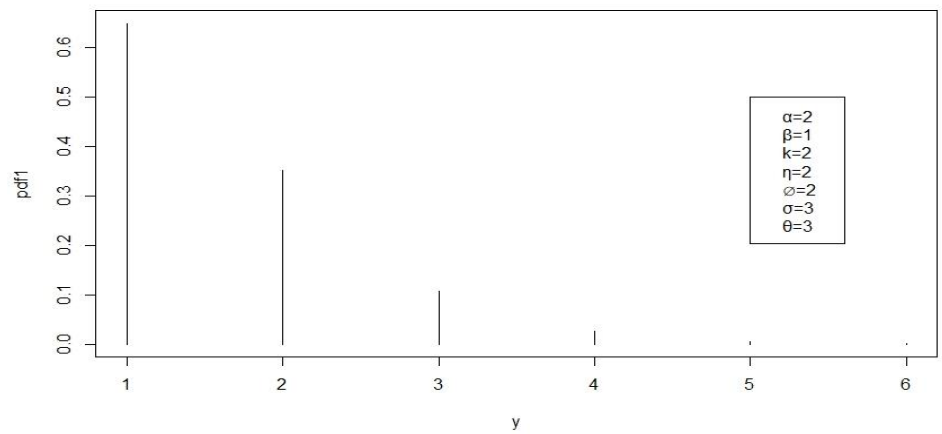

Figure 1: Different Shapes of the pdf for the GPL7

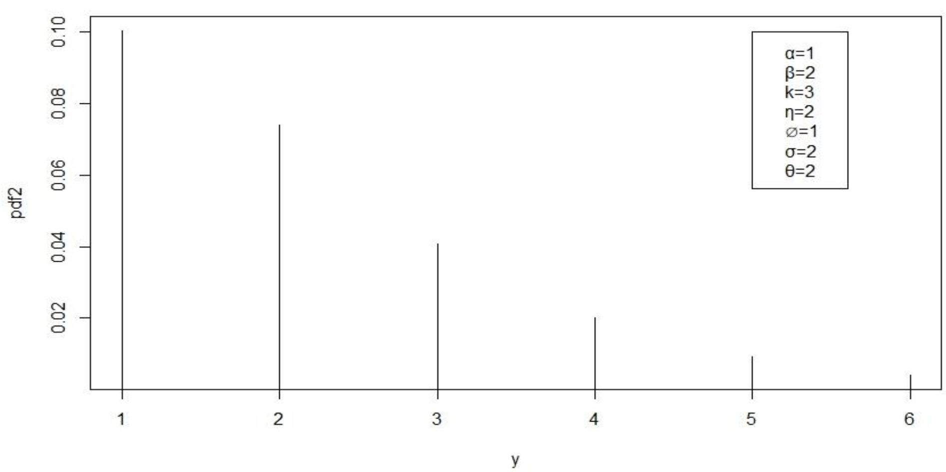

Figure 2: Different Shapes of the pdf for the GPL7

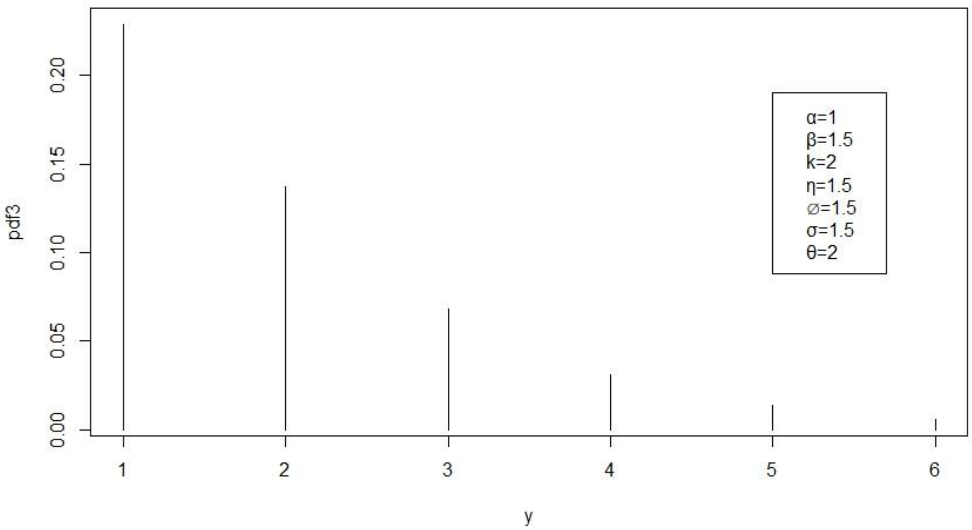

Figure 3: Different Shapes of the pdf for the GPL7

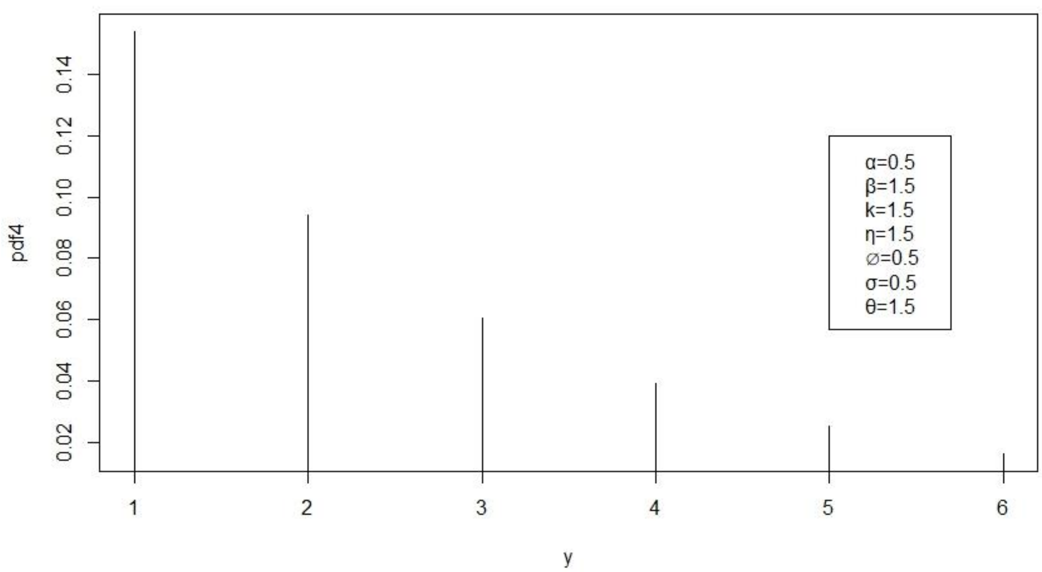

Figure 4: Different Shapes of the pdf for the GPL7

3. GENERALIZED POISSON-LINDLEY LINEAR MODEL FOR SEVEN PARAMETERS (GPL7).

In count data, the GPL distribution is a good alternative to the standard Poisson and NB distributions [9]. Covariates, on the other hand, were not used to explain the GPL variable. The GPL linear model is developed by generalizing the GPL7 distribution, which is the current statistical distribution. The extended linear model is an extension of the classic linear regression model where the continuous assumption of the response variable is broken, and the response can be a count variable. As a result, the link function needs to be considered. It must be a monotonic, invertible, and differentiable function that maps from $\boldsymbol{X\beta} \in \mathbb{R}^{n}$ to the response variable's mean (E(Y) >0) [1,2,24].

The log-linearity of the mean is considered the link function in the GPL linear model.

The vector-value link function is defined as η = g(μ), where $\eta_i = g(\mu_i) = \log(\mu_i) = x_i^T \beta$ and x_i^T is the ith row of a n × (k + 1) design matrix, X.

As a result, $E(Y_i) = \exp\left(\chi^T \beta\right)$, implying that the mean of the response variable is equal to the exponential of the linear predictor [9,11].

Proposition 3.1: $Y|x_i^T$ be a response variable with x_i^T as a covariates. Then, theconditional distribution of Y_i for x_i^T as follows distribution $GPL7(\theta, \alpha, \beta, k, \eta, \varnothing, \sigma)$ with seven parameters and mean greater than zero.

$Y \mid x_i^T \sim GPL7\left(\theta, \alpha, \beta, k, \eta, \varnothing, \sigma\right)$ is pmf can be written as**:**

$$f(y_i \mid x_i^T) = \left(\frac{\theta^{\phi+2} \exp\left(\chi^T \beta\right) - \alpha\theta^{\phi+1}}{\eta^{\sigma}.\Gamma\left(\chi+1\right).\left(\beta\theta^{\phi+\beta} - \alpha\theta^{\phi-1}\right)} \right) \times$$

$$\left[\frac{\eta^{\sigma}\theta^{\beta-1}\Gamma\left(y+\beta\right)\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left(\theta^{\phi}\exp\left(\chi^T\beta\right) - \alpha\theta^{\phi-1}\right) + \theta\Gamma\left(\beta\right)\left(\theta+1\right)^{y+\beta}}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right) - \alpha\theta^{\phi-1}\right]} \right. \qquad (4)$$

$$\left. \times \quad \frac{\left[\left(\beta\eta^{\sigma}\theta^{\alpha+\beta+\phi-2} - \eta^{\sigma}\theta^{\alpha+\phi-2}\exp\left(\chi^T\beta\right)\right)\Gamma\left(\alpha+y\right)\right]}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right) - \alpha\theta^{\phi-1}\right]} \right]$$

Proof:

If $Y|x_i^T$ has a pmf as in Equation (1) and the pdf of λ_i is proposed in Equation (2), then $Y\ |x_i^T \sim GPL(\theta,\alpha,\beta,k,\eta,\varnothing,\sigma)$ has a marginal pmf as in Equation (3). Furthermore, the GPL distribution's mean is

$$E(Y_i|x_i^T)=E\left[E\left[(Y_i|x_i^T)\,|\,\lambda_i\right]\right]$$

$$\mathrm{E}(x)=\frac{\theta^2}{\eta^\sigma+\theta^\phi k}\left(\alpha k\theta^{\phi-3}+\beta\eta^\sigma\theta^{\beta-2}\right)$$

$$=\mu_i$$
$$=\exp\left(\chi^T\beta\right)$$

By parameterizing, the mean of the GPL distribution is achieved

$$k=\frac{\beta\eta^\sigma\theta^\beta-\mu\eta^\sigma}{\mu\theta^\phi-\alpha\theta^{\phi-1}}$$

As a result, the pmf of $Y\ |x_i^T \sim GPL(\theta,\alpha,\beta,k,\eta,\varnothing,\sigma)$ can be described as a linear model with a log-link function by substituting $k=\dfrac{\beta\eta^\sigma\theta^\beta-\mu\eta^\sigma}{\mu\theta^\phi-\alpha\theta^{\phi-1}}$ into Equation (3) as

$$f\left(y_i|x_i^T\right)=\frac{\theta^2}{\Gamma(y+1)(\eta^\sigma+\theta^\phi k)}\left[\frac{K\theta^{\alpha+\phi-2}}{\Gamma(\alpha)(\theta+1)^{\alpha+y}}+\frac{\eta^\sigma\theta^{\beta-1}\Gamma(y+\beta)}{\theta\Gamma(\beta)(\theta+1)^{\beta+y}}\right]$$

Then, we obtained that

$$f\left(y_i|x_i^T\right)=\frac{\theta^2}{\Gamma(y+1)\eta^\sigma+\theta^\phi\left(\dfrac{\beta\eta^\sigma\theta^\beta-\mu\eta^\sigma}{\mu\theta^\phi-\alpha\theta^{\phi-1}}\right)}\times\left[\frac{\theta^{\alpha+\phi-2}}{\Gamma(\alpha)(\theta+1)^{\alpha+y}}\times\frac{\left(\beta\eta^\sigma\theta^\beta-\mu\eta^\sigma\right)}{\mu\theta^\phi-\alpha\theta^{\phi-1}}+\frac{\eta^\sigma\theta^{\beta-1}\Gamma(y+\beta)}{\theta\Gamma(\beta)(\theta+1)^{\beta+y}}\right]$$

Let

$$u=\frac{\mu\theta^{\phi+2}-\alpha\theta^{\phi+1}}{\eta^\sigma.\Gamma(y+1)\left[\beta\theta^{\phi+\beta}-\alpha\theta^{\phi-1}\right]}\qquad w=\frac{\eta^\sigma\theta^{\beta-1}\Gamma(y+\beta)}{\theta\Gamma(\beta)(\theta+1)^{\beta+y}}$$

$$v=\frac{\left[\beta\eta^\sigma\theta^{\alpha+\beta+\phi-2}-\mu\eta^\sigma\theta^{\alpha+\phi-2}\right]\Gamma(\alpha+y)}{\Gamma(\alpha)(\theta+1)^{\alpha+y}\left(\mu\theta^\phi-\alpha\theta^{\phi-1}\right)}$$

by substituting $\mu_i = \exp\left(\chi^T \beta\right)$ in u and v, then it can write $f\left(y_i | x_i^T\right) = u\left[v + w\right]$ as

$$f(y_i \mid x_i^T) = \left(\frac{\theta^{\phi+2}\exp\left(\chi^T\beta\right) - \alpha\theta^{\phi+1}}{\eta^{\sigma}.\Gamma\left(y+1\right).\left(\beta\theta^{\phi+\beta} - \alpha\theta^{\phi-1}\right)}\right) \times$$

$$\frac{\eta^{\sigma}\theta^{\beta-1}\Gamma\left(y+\beta\right)\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left(\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right)+\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{y+\beta}}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right]}$$

$$\times \quad \frac{\left[\left(\beta\eta^{\sigma}\theta^{\alpha+\beta+\phi-2} - \eta^{\sigma}\theta^{\alpha+\phi-2}\exp\left(\chi^T\beta\right)\right)\Gamma\left(\alpha+y\right)\right]}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right]}$$

4. MODEL ESTIMATION

In this section, model parameter estimation will be derived**.**

- **Maximum likelihood estimation**

Maximizing the log-likelihood function of parameters, L, called the MLE, was used to estimate the regression coefficients (β) and the distribution parameter (θ).

Let $\mathbf{\Theta} = (\beta^{\mathrm{T}}, \boldsymbol{\theta})^T$ be a vector of the parameters. Then log-likelihood function is

$$\ell\left(\mathbf{\Theta}\right) = \prod_{i=1}^{n} f\left(y_i | x_i^T\right) = \prod_{i=1}^{n} u\left[v+w\right]$$

$$L\left(\mathbf{\Theta}\right) = \sum_{i=1}^{n} \ln u + \sum_{i=1}^{n} \ln\left(v+w\right)$$

$$L\left(\mathbf{\Theta}\right) = \sum_{i=1}^{n} \ln\left(\frac{\mu\theta^{\phi+2} - \alpha\theta^{\phi+1}}{\eta^{\sigma}.\Gamma\left(y+1\right)\left[\beta\theta^{\phi+\beta} - \alpha\theta^{\phi-1}\right]}\right) +$$

$$\sum_{i=1}^{n} \ln\left[\frac{\eta^{\sigma}\theta^{\beta-1}\Gamma\left(y+\beta\right)\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left(\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right)+\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{y+\beta}}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right]}\right.$$

$$\left.\times \quad \frac{\left[\left(\beta\eta^{\sigma}\theta^{\alpha+\beta+\phi-2} - \eta^{\sigma}\theta^{\alpha+\phi-2}\exp\left(\chi^T\beta\right)\right)\Gamma\left(\alpha+y\right)\right]}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^T\beta\right)-\alpha\theta^{\phi-1}\right]}\right]$$

$$Ln\ u = Ln\left[\frac{\theta^{\phi+2}\exp\left(\chi^{T}\beta\right)-\alpha\theta^{\phi+1}}{\eta^{\sigma}.\Gamma\left(\chi+1\right).\left(\beta\theta^{\phi+\beta}-\alpha\theta^{\phi-1}\right)}\right]$$

$$= Ln\left(\theta^{\phi+2}\exp\left(\chi^{T}\beta\right)-\alpha\theta^{\phi+1}\right)-\mathrm{Ln}\ \eta^{\sigma}.-\mathrm{Ln}\ \Gamma\left(y+1\right)-Ln\left[\beta\theta^{\phi+\beta}-\alpha\theta^{\phi-1}\right]$$

$$\sum_{i=1}^{n}\ln u = \sum_{i=1}^{n}\left[Ln\left(\exp\left(\chi^{T}\beta\right)\theta^{\phi+2}-\alpha\theta^{\phi+1}\right)\right]-n\sigma\ \mathrm{Ln}\ \eta-\sum_{i=1}^{n}\mathrm{Ln}\ \Gamma\left(y+1\right)-n\ Ln\left[\beta\theta^{\phi+\beta}-\alpha\theta^{\phi-1}\right]$$

And,

$$Ln\left(v+w\right) = Ln\left[\frac{\eta^{\sigma}\theta^{\beta-1}\Gamma\left(y+\beta\right)\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left(\theta^{\phi}\exp\left(\chi^{T}\beta\right)-\alpha\theta^{\phi-1}\right)+\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{y+\beta}}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^{T}\beta\right)-\alpha\theta^{\phi-1}\right]}\right.$$

$$\left.\times\quad\frac{\left[\left(\beta\eta^{\sigma}\theta^{\alpha+\beta+\phi-2}-\eta^{\sigma}\theta^{\alpha+\phi-2}\exp\left(\chi^{T}\beta\right)\right)\Gamma\left(\alpha+y\right)\right]}{\theta\Gamma\left(\beta\right)\left(\theta+1\right)^{\beta+y}\Gamma\left(\alpha\right)\left(\theta+1\right)^{\alpha+y}\left[\theta^{\phi}\exp\left(\chi^{T}\beta\right)-\alpha\theta^{\phi-1}\right]}\right]$$

By differentiating L(Θ) with respect to each parameter $\theta, \alpha, \beta, k, \eta, \varnothing, \sigma$ and the score functions of the parameters, it will estimate all parameters.

$$\frac{\delta L(\Theta)}{\delta\theta},\frac{\delta L(\Theta)}{\delta\alpha},\frac{\delta L(\Theta)}{\delta\beta},\frac{\delta L(\Theta)}{\delta\mu},\frac{\delta L(\Theta)}{\delta\eta},\frac{\delta L(\Theta)}{\delta\phi},\frac{\delta L(\Theta)}{\delta\sigma}$$

Hence, setting these expressions to zero and solving the nonlinear resultant equations simultaneously produces the maximum likelihood estimators, which are obtained by using a numerical approach or a direct numerical search for the maximum of the log-likelihood surface.

5. APPLICATION TO REAL DATA

The data used in this study focuses on the relationship between the patient's stay in the hospital as a dependent variable and each of the following variables as independent variables: the procedures followed inside the hospital as the first variable, the patient's age as the second variable, and the patient's gender as the third variable. Where this data was obtained in the year 2020 from the

Damradash General Hospital in the Arab Republic of Egypt for a group of Corona virus-infected patients, where the total number of observations obtained was 1570, and each case recorded the length of stay in the hospital, as well as the type of infection. The following table shows some statistical measures that were calculated for each of the independent and dependent variables.

Table 1: Data summary for a group of Corona patients at Demerdash General Hospital (1570)

variables	mean	max	min	var	Rang
stay	15.24	30	10	2.13	20
age	35.35	80	12	3.21	68
procedures	4.2	9	2	2.1	7

Extrapolating Table (1). We should highlight that the average length of stay in the hospital during Corona's treatment was 15.24 days. The longest hospital stay was 30 days, and the shortest was 10 days, with a variance of A.23 for the variable. The second factor is age, with a median age of 35. The oldest person is 80 years old, and the youngest is 12 years old. The third variable is the length of time the hospital spends performing procedures. The average procedure lasts four days. Procedure 2 has a minimum duration of two days and a maximum duration of ten days.

Table 2: Model performance of corona virus patients in Demerdash General Hospital

Models	log-likelihood	AIC	BIC
Models(7p)	-12.345	24.125	24.0123
Models(4p)	-10.234	25.124	26.145

Through the second table, the value of the Maximum likelihood of regression for the distribution with seven parameters was less than the distribution with four parameters, indicating that the regression model for the distribution with seven parameters is better than the regression model for the distribution with four parameters, as shown in the second table. In the regression model with seven parameters, the AIC and BIC values were also lower.

Table 3: Modeling results for the patients in Demerdash General Hospital

	Passion Lindley (7)		Passion Lindley (4)	
Covariates	Estimate	p-Value	Estimate	p-Value
intercept	1.423 (0.0231)	<0.001	1.4532 (0.0312)	<0.001
Doctors' absence days	0.842 (0.0182)	<0.001	0.864 (0.0294)	<0.001
age	-0.1132 (0.0172)	<0.001	0.2132 (0.0264)	<0.001
procedures	0.1214 (0.0182)	<0.001	0.1245 (0.0283)	<0.001

By extrapolating Table No. 2, the regression model can be written as follows

$$\hat{\mu} = \exp\left(1.423 + 0.842 \text{ Doctors' absence days} - 0.1132 \text{ age} + 0.1214 \text{ procedures}\right)$$

The number of days doctors were absent from the hospital is the most influential element on the patient's stay in the hospital, according to the previous regression model, where the effect of this independent variable on the dependent variable was (0.842) and statistically significant. In terms of the protocols followed in the hospital, everything was in line. Because they were in second place and their effect was on the dependent variable, they had no effect on the length of stay (0.1214) Finally, the dependent variable is affected by the age variable in the opposite way.

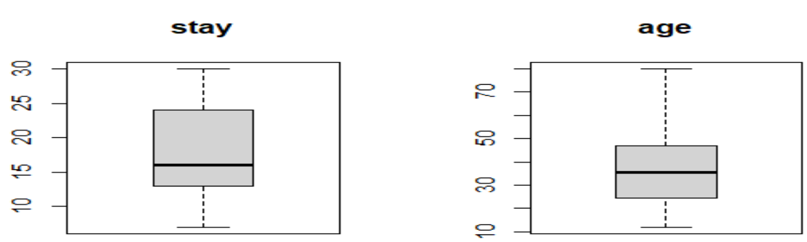

Figure 5: show independent variable and independent variables

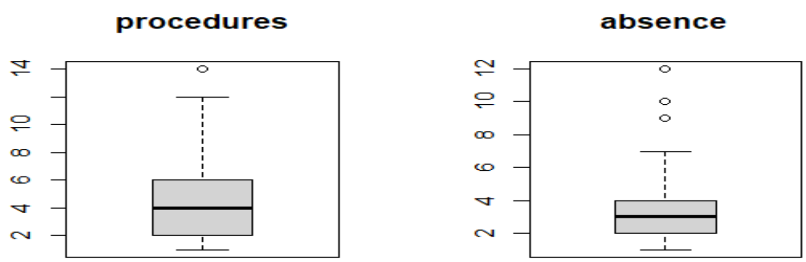

Figure 6: show independent variable and independent variables

6. CONCLUSIONS

The aim of this study is to propose a new regression model using re-parameterization of the Poisson-Lindley distribution for seven parameters. The proposed model was considered for analyzing a real dataset on the Corona virus-infected patients. The main conclusion one can make from the fitted model is that the proposed model able to discover which factors had the most impact on the length of stay in the hospital by using data from Corona needs. As a result, the new model of regression can be applied to many medical sectors because of its ability to discover which factors have the most impact on the length of stay in the hospital.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] A.C. Cameron, P.K. Trivedi, Regression analysis of count data, Cambridge University Press, Cambridge, (1989).

[2] A.C. Cameron, P.K. Trivedi, Regression analysis of count data (Vol. 53), Cambridge University Press, Cambridge, (2013).

[3] A.J. Sáez-Castillo, A. Conde-Sánchez, A hyper-Poisson regression model for overdispersed and underdispersed count data, Comput. Stat. Data Anal. 61 (2013), 148–157. https://doi.org/10.1016/j.csda.2012.12.009.

[4] E. Altun, A new model for over-dispersed count data: Poisson quasi-Lindley regression model, Math. Sci. 13

(2019), 241–247. https://doi.org/10.1007/s40096-019-0293-5.

[5] L. Cheng, S.R. Geedipally, D. Lord, The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data, Safe. Sci. 54 (2013), 38–42. https://doi.org/10.1016/j.ssci.2012.11.002.

[6] D. Karlis, E. Xekalaki, Mixed poisson distributions, Int. Stat. Rev. 73 (2005), 35-58. https://www.jstor.org/stable/25472639.

[7] D. Lord, S.R. Geedipally, The negative binomial–Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros, Accident Anal. Prevent. 43 (2011) 1738–1742. https://doi.org/10.1016/j.aap.2011.04.004.

[8] D.E. Giles, Hermite regression analysis of multi-modal count data. Econ. Bull. 30 (2010), 2936-2945.

[9] E. Mahmoudi, H. Zakerzadeh, Generalized Poisson–Lindley distribution, Commun. Stat. – Theory Methods. 39 (2010), 1785–1798. https://doi.org/10.1080/03610920902898514.

[10] E. Ohlsson, B. Johansson, Non-life insurance pricing with generalized linear models (Vol. 174). Springer, Berlin, (2010).

[11] E.G. Déniz, A new discrete distribution: properties and applications in medical care, J. Appl. Stat. 40 (2013), 2760–2770. https://doi.org/10.1080/02664763.2013.827161.

[12] H. Zakerzadeh, A. Dolati, Generalized Lindley distribution, J. Math. Ext. 3 (2009), 13-25.

[13] H. Zamani, N. Ismail, P. Faroughi, Poisson-weighted exponential univariate version and regression model with applications, J. Math. Stat. 10 (2014), 148-154.

[14] H. Fawzy, E.H.A. Rady, A.M.A. Fattah, Comparison between support vector machines and K-nearest neighbor for time series forecasting, J. Math. Comput. Sci. 10 (2020), 2342-2359. https://doi.org/10.28919/jmcs/4884.

[15] J.M. Hilbe, Negative binomial regression, Cambridge University Press, Cambridge, (2011).

[16] M. Karim, A.K. Mutaqin, Modeling claim frequency in Indonesia auto insurance using generalized Poisson-Lindley linear model, J. Mat. Stat. Komput. 16 (2020), 428-439. https://doi.org/10.20956/jmsk.v16i3.9315.

[17] E.H.A. Rady, H. Fawzy, A.M.A. Fattah, Time series forecasting using tree based methods, J. Stat. Appl. Probab. 10 (2021), 229-244.

[18] M. Sankaran, 275. note: The discrete Poisson-Lindley distribution, Biometrics, 26 (1970), 145-149. https://doi.org/10.2307/2529053.

[19] M.M. Shoukri, M.H. Asyali, R. VanDorp, D. Kelton, The Poisson inverse Gaussian regression model in the analysis of clustered counts data, J. Data Sci. 2 (2004), 17-32.

[20] J. Rodríguez-Avi, A. Conde-Sánchez, A.J. Sáez-Castillo, et al. A generalized Waring regression model for count data, Comput. Stat. Data Anal. 53 (2009) 3717–3725. https://doi.org/10.1016/j.csda.2009.03.013.

[21] N.L. Johnson, S. Kotz, A.W. Kemp, Univariate discrete distributions, John Wiley & Sons, Hoboken, (2005).

[22] P. Deb, P.K. Trivedi, Demand for medical care by the elderly: A finite mixture approach, J. Appl. Econ. 12 (1997), 313–336. https://doi.org/10.1002/(sici)1099-1255(199705)12:3<313::aid-jae440>3.0.co;2-g.

[23] P. McCullagh, J.A. Nelder, Generalized linear models, Chapman and Hall, London, (1989).

[24] R.C. Team, R: A languageand environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www. R-project. org. (2014).

[25] R. Varadhan, Numerical optimization in R: Beyond optim, J. Stat. Soft. 60 (2014), 1-3. https://doi.org/10.18637/jss.v060.i01.

[26] H. Fawzy, E.H.A. Rady, A.M.A. Fattah, Forecasting time series using a hybrid ARIMA-ANN methodology, J. Appl. Probab. Stat. 16 (2021), 95-106.

[27] W. Wongrin, W. Bodhisuwan, Generalized Poisson–Lindley linear model for count data, J. Appl. Stat. 44 (2016), 2659–2671. https://doi.org/10.1080/02664763.2016.1260095.