Iraqi Journal of Statistical Sciences, Vol. 21, No. 1, 2024, Pp (76-89)

Iraqi Journal of Statistical Sciences

www.stats.mosuljournals.com

A Parametric Regression Model Using Power Chris-Jerry Distribution With Application To Censored Data

Christiana I. Ezeilo, Iheanyi S. Onyeagu and Uzoma E. Umeh

Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

Article information

Article history:
Received December 31 ,2023
Accepted February 25, 2024
Available online June 1, 2024

Keywords:
Corruption Perception Index, Gross Domestic Product, Real Gross Domestic Product, Nigeria Economy, Transparency International, Treasury Single Account

Correspondence:
Christiana I. Ezeilo
ifenyinwaeze@gmail.com

Abstract

The interdependency of various areas of Statistics is gaining good attention in the literature. This is possible through innovations such as the log-transformation of distribution to a parametric regression model hence integrating contemporary probability distribution models with the classical regression method. The new model is essentially preferred due to its applicability in wider scenarios. In this article, the base distribution is the Power Chris-Jerry distribution. The reparametrized regression model equivalent was carefully derived with the maximum estimation procedure under censored sample also considered. A simulation study of the log-power Chris-Jerry regression model was carried out with measures of performance presented. The COVID-19 patient lifetime censored data was deployed to justify the motivation for developing the new model and twelve competing models were used to compare the proposed regression. The results show that the proposed model is indeed better and preferred to its competitors.

DOI 10.33899/IQJOSS.2024.183249

Introduction

A good number of authors have investigated parametric regression model from the probability scenario using various approaches. Cordeiro, Biazatti, and Santana (2023) introduced a four-parameter Weibull extended Weibull (WEW) distribution that presents greater flexibility and can model data with bathtub-shape and unimodal failure rate taking the Extended Weibull PDF as baseline to form the new distribution. The new support for the new distribution is $x > 0$, with properties such as quantile function, kurtosis, skewness, moments were discussed. Estimation of the parameters was done using method of maximum likelihood. Carlo monte simulation study was carried out to show the new distribution WEW has consistent MLE's with lower AIC, BIC and Global Deviance GD from generated data. Also regression model $y_i = V_i^T \gamma + \delta z_i$ was constructed on Log-Weilbull Extended Weibull (LWEW) distribution for $Y = \log X$ when X has WEW pdf where z_i has LWEW Pdf, $\gamma = (\gamma_1, \gamma_2 ..., \gamma_p)^T$ is the vector coefficients and $v_i = (v_{i1}, v_{i2}, ... v_{ip})^T$ is the vector covariates for the ith response y_i which models the location parameter $\mu_i = v_i^T \gamma$. Two data set was used to show the applicability and superiority of the proposed model over other existing ones compared.

Biazatti, Cordeiro, Rodrigues, Ortega and De Santana (2022) introduced Weibull-Beta Prime (WBP) distribution from Beta Prime BP distribution due to wide use of the BP and to provide better fit to complex real data. Some structural properties of the new distribution such as quantile function, linear representation and moment were obtained. Method of maximum likelihood estimation was used for parameter estimation. The simulation study carried out shows that all estimators improve as n increases. Furthermore, the WBP regression model was constructed for censored samples. Since censored samples are commonly considered as systematic component for the shape parameter α. Considering systematic component $\alpha_i = \exp(v_i^T \lambda)$ where $v_i^T = (v_{i1}, v_{i2}, ... v_{ip})$ is the vector of covariates and $\lambda = (\lambda_1, ..., \lambda_p)^T$ is the vector of unknown parameters. Real data set were used to show the importance and superiority of the proposed model when compared with some known competing models. Rodrigues, Ortega, Cordeiro and Vila (2022) proposed Odd Log-Logistic Weibull (OLLW) regression model for censored data to identify factors that increase the risk of death of hospitalized patients diagnosed with Covid'19.

The properties of the distribution; mode, stochastic representation, closure under changes of scale and of power, identifiability, tail behavior and moment were discussed. OLLW regression model was defined by two systematic component for α_i and λ_i for(i=1,..,n) follows $g_1(\lambda_i) = \eta_{i1} = X_{i1}\beta_1$ and $g_2(\alpha_i) = \eta_{i2} = X_{i2}\beta_2$ where $\beta_j = (\beta j_0, ..., \beta_{jp})$ $(j = 1,2)$ are vector length $(p_j + 1)$ for unknown coefficients functionally independent, p_j is the number of explanatory variables related to the jth parameter, η_{ij} are the linear predictors and $X_{ij} = (v_{ij1}, ..., v_{ijp_j})$ are observation on p_1 and p_2 known repressors. From the data set, older age, asthma, diabetes obesity and chronic neurological diseases were identified as risk factors associated with death of diagnosed Covid'19 patients in the city of Campinas, Brazil.

Segovia, Gomez and Gallardo (2021) introduced an Extension of Power Maxwell (EPW) distribution called Exponentiated Power Maxwell and proposed Re-parameterized Exponentiated Power Maxwell (REPM) distribution. The properties of the new distribution which includes, moment, quantile function, median were discussed. They also introduced regression framework for applying the model to any quantile of the distribution, where quantile of such variables is related to a set of covariates say $X_i^T = (x_{i1}, x_{i2}, ..., x_{ip})$ through the logarithmic link $\log \mu_{i(p)} = X_i^T \tau(p)$ i=1,2,…n where $\tau(p) = (\tau_0(p), ..., \tau_p(p))$ are the regression coefficient. The maximum likelihood estimation for REPM regression model under classical approach was discussed and simulation study to assess the performance of the ML estimators for REPM regression model was conducted. Two real data set used to show the applicability of the model and was compared with other proposed in literature shows the value of AIC and K-S of REPM lower than the compared ones.

Wongrin and Bodhisuwan (2017) used the generalized linear model to create a new linear regression model called Generalized Poisson Linear model which was based on a Generalized Poisson-Lindley (GPLi) distribution for seven parameters. The conditional distribution $f\left(y_i / \ x_i^T\right)$ for $GPLi(\theta,\alpha,\beta,k,\eta,\phi,\delta)$, the GPL mean, the probability mass function $Y/X_i^T \sim \text{GPLi}(\theta,\alpha,\beta,k,\eta,\phi,\delta)$ were discussed. the maximum likelihood estimation for the model parameter estimation was derived. The applicability of the new model was seen in the analyzing a real dataset on the corona virus-infected patients.

Reis (2023) proposed a new distribution called Pezeta distribution that has support on the interval $(0,1)$. It was obtained after transforming the random variable $Y = \dfrac{1}{1+\omega}$ with exponential distribution $f(\omega,\lambda) = \lambda e^{\lambda\omega}$. Its properties such as; mode, moment, quantile function, random number generation, proof of exponential family, MLE and MLE bias correction when n is small were discussed. Also, regression model was introduced for the dependent variable with support at $(0,1)$. The model has a regression structure on the median of the distribution $\eta_i = g(\tau_i) = X_i^{\tau}\beta$ where $\beta = (\beta_1,...,\beta_k)^T$ is the k-vector of unknown parameters $x_i = (x_{i1},...,x_{ik})^T$ is the vector of k explanatory variables $(k < n)$ which are assumed fixed and known η_i is the linear predictor. Simulation study was conducted to show the performance of the MLE's for the proposed regression model $In\left(\dfrac{\tau_i}{1-\tau_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + ...$ Pezeta regression model was compared with the unit Lindley UL regression model using a dataset. Discriminate between the two regression models was assessed using AIC, BIC and Hannan Information Criterion (HQIC). The new model presents the smallest value of these statistics which shows its superiority over the compared one.

Badmus, Akinyemi and Onyeka-Ubaka (2021) introduced a location-scale regression model based on the logarithm of an extended Raleigh Lomax distribution which has the ability to model survival data than classical regression model called Log-Beta Rayleigh Lomax (LBRL) regression model. They presented two important classes of the distribution, firstly Beta Rayleigh Lomax (BRL)distribution using Logit Beta function. Secondly, Log –Beta Rayleigh Lomax LBRL distribution. Hazard function, reliability function, moment, moment generating function linear combination and other properties of the new distribution were derived. Based on the LBRL distribution, a linear regression model linking the response variables y_i explanatory variables x_i is defined as $y_i = X_i^T\beta + \sigma z_i, i =,2,..,n$ where the random error z_i has LBRL density function with $\beta = (\beta_1,...,\beta_p)^T$ the unknown parameters and $X_i^T = (x_1,...,x_p)$ the explanatory variable vector modeling the linear predictor $\mu_i = X_i^T\beta$. The linear predictor vector

$\mu = (\mu_1, ..., \mu_n)^T$ of LBRL regression model is written as $\mu = X\beta$ where $X = (x_1, ..., x_n)$ is the known model matrix. The MLE was used for parameter estimation. Applicability of the new model was shown using breast cancer dataset referring to time spent (t) and explanatory variables age (x_1), occupation (x_2), martial status (x_3), event statues (x_4) and type of treatment (x_5) for n=623 observations. Fitting the above dataset with proposed model and 5 other existing regression model and using model selection criteria AIC, BIC and CAIC the proposed model outperformed the compared ones.

Eliwa, Attun, Alhussian, Ahmed, Salah, Ahamed and El-Morshedy (2021) deployed the Odd Lindley-G family Oli-G to proposed a new generalization of Half Logistic HL distribution with only one parameter called Odd Lindley Half Logistic (OLiHL) distribution. The statistical properties of the new distribution such as raw and central moment, incomplete moment, moment generating functions, quantile function were discussed. the estimation method used are MLE, LS, Weighted Least Square and Cramer-Von Mises. Simulation study shows the relative performance of the used estimation methods. A log-location-scale regression model called log-OLiHL regression model was introduced based on the $Y = \log(X)$ transformation and a suitable re-parameterization on the baseline distribution OLiHL considering $y_i = X_i^T \beta + \sigma z_i$ where the response variable y_i has the Log-OLiHL density, the covariates are linked to location of y_i with identity link function $\mu_i = X_i^T \beta$ where $X = (x_1, x_2, ..., x_p)$ is the model matrix consists of the observation and the independent variables and $\beta = (\beta_0, \beta_1, ..., \beta_k)$ is the unknown regression coefficients. Two datasets were considered to show the flexibility of the OLiHL distributions against the several one-parameter competitive model and it showed better modeling ability.

Nasiru, Abubakari, Chesneau (2022) proposed the Bounded Truncated Cauchy Power Exponential (BTCPE) distribution for modeling dataset on the unit interval

Relevant properties of the BTCPE distribution which includes the distribution of inequalities, quantile function, moment, moment generating function and order statistics were discussed. The bivariate extension of the new model was shown. The parameter estimation method used are; MLE, OLS, WLS, Cramer-Von Mises, Percentile estimation, Anderson-Darling method and maximum and minimum product spacing method. Simulation study was conducted to compare the estimation methods using bias, RMSE of the estimates. Using 3 dataset, application of the BTCPE distribution was illustrated and its performance was compared to other competitive distributions defined in the unit interval based on AIC, BIC criterion. They also define BTCPE quantile regression as $g(\rho_i) = Z_i'\theta$ where $\theta = (\theta_0, \theta_1, ..., \theta_p)'$ is the vector of unknown parameter, ρ_i is the ith quantile parameter and $Z_i = (1, z_{i1}, z_{i2}, ..., z_{ip})$ are the known ith vector covariates. The log-likelihood for estimating the parameter of the BTCPE quantile regression was given. Monte Carlo simulation was carried out to examine the performance of the ML estimates of the parameter of the model using Absolute Bias and RMSEs. It shows that the regression parameter are consistent.

Nwankwo, Nwankwo and Obulezi (2023) proposed a new three exponentiated Power Akash (EPA) distribution. The properties of the distribution such as moment, r^{th} incomplete moment were discussed. Maximum likelihood estimation was used for the parameter estimation. Letting $Y = \log(X)$ where $X \sim \mathrm{EPA}(c, \alpha, \theta)$ and defining $\alpha = \frac{1}{\sigma}$ and $\theta = e^{-\frac{\mu}{\sigma}}$, the log-Exponentiated Power Akash (LEPA) density for $y \in R$ was derived and a parametric regression model for response variable y_i and vector of explanatory variables $V_i = (v_{i1}, v_{i2}, ..., v_{ip})'$ constructed as $y_i = V'\beta + \sigma z_i$ for i=1,2,…,n where $\mu_i = v'\beta$, and $\beta = (\beta_1, ..., \beta_p)'$ is the vector of unknown regression coefficient and z is the random error, also likelihood of β was derived. Using a Covid'19 censored data, the applicability and performance of the new distribution when compared with other competitive ones in the literature were shown using Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Hannan-Quinn Information Criterion (HQIC) measure criteria.

Methods: Log-Transformed Regression from Power Chris-Jerry Distribution

Ezeilo, Umeh, Osuagwu and Onyekwere (2023) introduced the Power Chris-Jerry (PCJ) distribution with PDF given as

$$g(x) = \frac{\alpha\theta^2}{\theta+2}(1+\theta x^{2\alpha})x^{\alpha-1}e^{-\theta x^{\alpha}} \tag{1}$$

and CDF

$$G(x) = 1 - \left(1 + \frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta+2}\right)e^{-\theta x^{\alpha}} \tag{2}$$

Where $\alpha > 0$ and $\theta > 0$ are the shape and scale parameters respectively. Essentially, the goal here is to create a new reparameterized regression model using log-transformation of the PCJ distribution. The baseline distribution being the one-parameter Chris-Jerry distribution proposed by Onyekwere and Obulezi (2022) and its extensions namely Oramulu et al. (2023a, 2023b), Chukwuma et al. (2024), Chinedu et al. (2023a, 2023b), Etaga et al. (2023a, 2023b, 2023c), Tolba et al. (2023), Musa et al. (2023a, 2023b), Anabike et al. (2023), Obulezi et al. (2023a, 2023b, 2023c, 2023d, 2023e), Oha et al. (2024), Onyekwere et al. (2022), Nwankwo B. C. et al. (2023), and Nwankwo, M. P. et al. (2023).

Let $Y = \log(X)$ where $X \sim \mathrm{PCJ}(\alpha, \theta)$ defined in equation (2.1). Assume $\alpha = \frac{1}{\sigma}$ and $\theta = e^{-\frac{\mu}{\sigma}}$, the log-PCJ density for $y \in R$ using $f(y;\, \sigma, \mu) = g(y)\left|\frac{dx}{dy}\right|$ is

$$f(y;\sigma,\mu) = \frac{e^{\frac{y-2\mu}{\sigma}}}{\sigma\left(e^{-\frac{\mu}{\sigma}}+2\right)}\left(1+e^{\frac{2y-\mu}{\sigma}}\right)e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}} \tag{3}$$

Where $\sigma > 0$ and $\mu \in \Re$. If $X \sim \mathrm{PCJ}(\alpha, \theta)$, the $Y = \log X \sim \mathrm{LPCJ}(\sigma, \mu)$. Similarly, we deploy the same technique in equation (2.3) to derive the survival and density function of $Z = \dfrac{Y - \mu}{\sigma}$ which are

$$s(Z; \sigma, \mu) = \left(1 + \frac{e^{z}\left(e^{z} + 2\right)}{e^{-\frac{\mu}{\sigma}} + 2}\right) e^{-e^{z}} \tag{4}$$

$$f(z; \sigma, \mu) = \frac{\omega(z) e^{-e^{z}}}{\sigma\left(e^{\frac{\sigma z - \sigma}{\sigma}} + 2\right)}; z \in \Re \tag{5}$$

Where $\omega(z) = e^{\frac{\sigma z - \mu}{\sigma}}\left(1 + e^{\frac{\mu + \sigma z}{\sigma}}\right)$

Using equation (2.5), we construct a parametric regression model for the response variable Y_i and a vector of explanatory variables $v_i' = \left(v_{i1}, \ldots, v_{ip}\right)$ as

$$y_i = v'\beta + \sigma z_i \qquad \text{i = 1, 2, …, n} \tag{6}$$

Where $\mu_i = v'\beta$, $\beta = \left(\beta_1, \ldots, \beta_p\right)$ is the vector of unknown regression coefficients and z is the random error with density in equation (2.6), define the survival and density function of Y_i / v' as

$$s(y/v') = \left(1 + \frac{e^{z_i}\left(e^{z_i} + 2\right)}{e^{-\frac{\mu_i}{\sigma}} + 2}\right) e^{-z_i} \tag{7}$$

and

$$f(y/v') = \frac{\omega(z_i) e^{-e^{z_i}}}{\sigma\left(e^{-\frac{\mu_i}{\sigma}} + 2\right)} \tag{8}$$

Where $\omega(z_i) = e^{\frac{\sigma z_i - \mu_i}{\sigma}}\left(1 + e^{\frac{\mu_i + \sigma z_i}{\sigma}}\right)$ and $z_i = \dfrac{y_i - \mu_i}{\sigma}$

2.1 Maximum Likelihood Estimation of Log-PCJ Parameters under Censored Sample

To estimate the parameters in equation (2.6) for right censored data, we defined y_i and C_i as the lifetime and non-information censoring time (assuming independence) and $y_i = \min(y_i, C_i)$. Then, the log-likelihood function for $\xi = (\sigma, \beta^T)$ is

$$l((\xi))=\sum_{i\in F}\log\left(e^{-\frac{\mu_i}{\sigma}}+2\right)+\sum_{i\in F}\log[\omega(z_i)]-\sum_{i\in F}e^{z_i}+\sum_{i\in C}\left\{1+\frac{e^{z_i}\left(e^{z_i}+2\right)}{e^{-\frac{\mu_i}{\sigma}}+2}\right\}e^{-z_i} \qquad (9)$$

Where F and C are the sets of uncensored and censored observations respectively and d is the number of failures. The MLE $\hat{\xi}$ of the unknown parameter vector can be obtained by maximizing equation (9).

3.0 Simulation Study

The simulation conditions deployed by Ferreira and Cordeiro (2023) are used in this article due to the compatibility of the two distributions. For the LPCJ distribution under different scenarios, the accuracy of the MLEs is examined. For 1000 repetitions, the acceptance and rejection method is adopted to generate random samples of sizes n=50, 100, 300, and 600 from the LPCJ distribution. The Average estimates (AEs) of the parameters, Biases, and mean squared error (MSEs) are calculated. The algorithm for generating random samples uses the acceptance-rejection method. Note that LPCJ means Log-PCJ.

Table 1: Simulation Measures from the Log-PCJ Regression Model

		Initial parameter values (9 3.5)			(10 0.5)			(7.0 5.0)		
n	ξ	AE	BIAS	MSE	AE	BIAS	MSE	AE	BIAS	MSE
50	θ	10.6914	1.6914	9.7092	12.1308	2.1308	14.7143	7.7550	0.7550	3.3489
	α	4.0445	0.5445	0.4968	0.5762	0.0762	0.0098	5.7678	0.7678	1.0205
100	θ	9.8019	0.8019	2.9971	11.1301	1.1300	4.7523	7.2193	0.2193	1.0605
	α	3.9427	0.4427	0.2925	0.5622	0.0622	0.0058	5.6215	0.6215	0.5974
300	θ	9.2463	0.2463	0.7356	10.4710	0.4710	1.1512	6.8341	0.1659	0.3155
	α	3.8617	0.3617	0.1615	0.5515	0.0515	0.0033	5.4979	0.4979	0.3180
600	θ	9.0443	0.0443	0.3075	10.2460	0.2460	0.4795	6.6972	0.3028	0.2240
	α	3.8325	0.3325	0.1252	0.5477	0.0477	0.0026	5.4462	0.4462	0.2341

The statistics in Table 1, indicate that the AEs converge to the true parameters and that the biases and MSEs tend to zero when n increases, which proves the consistency of the LPCJ estimators. Overall, the simulation results suggest that larger sample sizes and the appropriate choice of ξ are crucial for accurate parameter estimation of the LPCJ distribution.

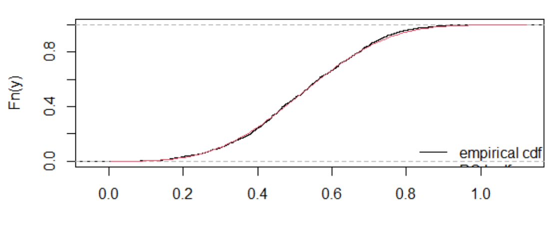


Fig. 1: Empirical cdf and estimated cdf for the scenario (9, 3.5)

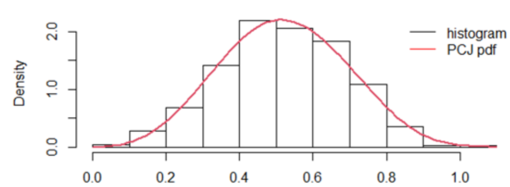


Fig 2: Estimated PDF and histogram for generated samples using samples using the scenario (9, 3.5).

Figures 1 and 2 reveal the approximation of the acceptance-rejection method. The estimated PDF and CDF of the PCJ distribution are very close to the histogram and empirical CDF of the generated samples, indicating a good performance of the method.

4.0 Application to COVID-19 Censored Data

The dataset comprises the lifetime (in days) of 322 individuals diagnosed with COVID-19 through RT-PCR screening in Campinas, Brazil. These data were previously studied by Ferreira and Cordeiro (2023), and Nwankwo, Nwankwo and Obulezi (2023). The response variable y_i represents the time elapsed from the onset of symptoms until death due to COVID-19 (failure). Ferreira and Cordeiro (2023), observed that about 66.45% of the observations are censored. The variables considered, for $(i = 1, \dots, 322)$ include δ_i: censoring indicator; 0 for censored and 1 for observed lifetime; v_{i1}: age (in years), and v_{i2} : diabetes mellitus $1 =$ yes, $0 =$ no or uninformed. The suggested regression model for these COVID-19 data is written as

$$y_i = \beta_0 + \beta_1 v_{i1} + \beta_2 v_{i2} + \delta z_i; \;\; i = 1, \dots, 322 \tag{10}$$

where $z_i \sim$ the PDF in equation (2.8).

The Power Lomax (PLO) distribution by Rady, El-Houssainy, Hassanein and Elhaddad (2016), Power Zeghdoudi (PZe) distribution (new), Power Suja (PSuj) distribution (new), exponentiated Power Ishita (EPI) by Ferreira and Cordeiro (2023), Power Ishita (PI) by Shukla and Shanker (2018), exponentiated Weibull (EWe) by Pal, Ali and Woo (2006), Power Rama (PR) by Abebe, Tesfay, Eyob and Shanker (2019), exponentiated Frechet (EF) by Nadarajah and Kotz (2003), Power Lindley (PLi) by Ghitany, Al-Mutairi, Balakrishnan and Al-Enezi (2013), exponentiated Power Akash (EPA) by Nwankwo, Nwankwo and Obulezi (2023), exponentiated Power Lindley (EPLi) by Ashour and Eltehiwy (2015) are used to compare with the proposed Power Chris-Jerry (PCJ) distribution. Note, that the log- of each distribution is derived following the procedure in section 2 to obtain LPLO, LPZe, LPSuj, LEPI, LPI, LEWe, LPR, LEF, LPLi, LEPA and LEPLi respectively.

The result from Table 3 shows that the explanatory variables age and diabetes mellitus are significant at the 5% significance level. Note that in table 3, c is the exponentiated parameter for some of the fitted distributions. The negative signs of β_1 and β_2 mean that older individuals or those with diabetes tend to have shorter failure times. This result is in agreement with that obtained from earlier study by Ferreira and Cordeiro (2023. From Table 2, the LPCJ regression has the lowest criterion values hence confirming that the LPCJ model provides a better fit for the COVID-19 data than the competing regression model using Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Hannan-Quinn Information Criterion (HQIC) measure criteria.

Table 2: Metrics of Model Performance for the Log-transformed Regression Models using the COVID-19 Censored data

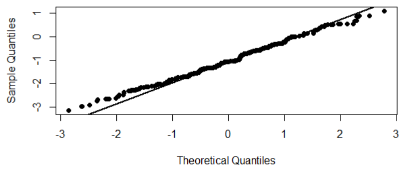
Distributions	AIC	CAIC	BIC	HQIC
LPCJ	428.7153	428.7153	443.5469	434.4763
LPLo	431.8449	432.2016	450.7177	439.3795

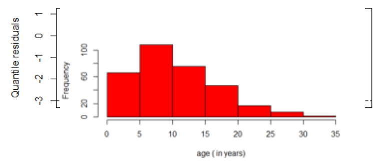
LPZe	432.6857	432.9524	447.7839	438.7134
LPSuj	436.1518	436.4185	451.25	442.1795
LEPI	428.9597	429.3164	447.8325	436.4943
LPI	433.5943	433.8609	448.6925	439.6219
LEWe	429.8493	430.2016	448.722	437.3839
LPR	641.4597	641.7264	656.5579	647.4874
LEF	441.9085	442.2652	460.7812	449.4431
LPLi	430.2784	430.6351	449.1512	437.8130
LEPA	433.6238	433.9805	452.4966	441.1584
LEPLi	430.2784	430.6351	449.1512	437.8130

Table 3: Parameter Estimates of the Log-transformed Regression Models using COVID-19 censored data

Distribution	c	σ	β_0	β_1	β_2
LPCJ	1	0.97559 (0.0586)	3.2564 (0.2660)	-0.0168 (0.0035) $<1.8687\times10^{-6}$	-0.2680 (0.2660) $<1.7881\times10^{-2}$
LPLo	0.3408 (4.0688)	0.5442 (0.0439)	-5.4568 (0.5656)	0.0193 (0.0041) $<3.1327\times10^{-6}$	0.2963 (0.1261) $<1.9895\times10^{-2}$
LPZe	1	1.1404 (0.0772)	3.2899 (0.2955)	-0.0192 (0.0039) $<1.01956\times10^{-6}$	-0.2866 (0.1269) $<2.4626\times10^{-2}$
LPSuj	1	1.5333 (0.0848)	1.9531 (0.2880)	-0.0185 (0.0036) $<4.4779\times10^{-7}$	-0.2656 (0.1237) $<3.2508\times10^{-2}$
LEPI	0.2070 (0.0887)	0.3971 (0.1461)	4.4367 (0.2365)	-0.0181 (0.0038) $<2.8899\times10^{-6}$	-0.2439 (0.1162) $<3.6568\times10^{-2}$
LPI	1	1.1926 (0.0774)	3.2401 (0.3021)	-0.0197 (0.0039) $<7.5442\times10^{-7}$	-0.2848 (0.1297) $<2.8741\times10^{-2}$
LEW	0.8878 (0.7995)	0.5106 (0.3515)	4.5518 (0.3585)	-0.0182 (0.0035) $<1.8687\times10^{-6}$	-0.2803 (0.1266) $<2.7473\times10^{-2}$

LPR	1	1.3356 (0.0866)	3.5766 (0.3052)	-0.0273 (0.0040) $< 6.2066\times10^{-11}$	-0.4021 (0.1308) $< 2.2848\times10^{-3}$





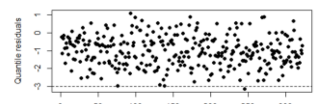


Fig 4.3: QQ plot Fig. 4.4: Histogram Fig. 4.5: Quantile residual plot

Figures 4.3, 4.4 and 4.5 are the non-parametric plots of the data set. The potential of the log-PCJ regression modeling skewed data is one interesting feature explored using the COVID-19 data. The properties of this distribution were derived and the log-transformation has been used to create a parametric regression model called the Power Chris-Jerry distribution regression model. The maximum likelihood estimation aided the estimation process for uncensored samples while the procedure for the estimation of the unknown parameters when data is censored was also shown. Essentially, the censored COVID-19 data set with the age of patients and diabetic mellitus index was deployed to justify the importance of the distribution. Furthermore, the distribution was fitted to the data on infant mortality rate (below age 5 years) reported for some countries by the World Health Organization in 2021. The distribution performs pretty well in both instances of application.

Conclusion

A great deal of effort has been made to investigate regression modeling using Power Chris-Jerry distribution in this article. The log-transformed Power Chris-Jerry regression model was developed and its survival function, and other functional forms were studied. The model parameters were estimated using maximum likelihood method under censored sample consequent upon the nature of the data deployed. The log-transformed Power Chris-Jerry regression model (LPCJ) designed in this paper was compared with those of the Log-Power Lomax (LPLO) distribution, Log-Power Zeghdoudi (LPZe) distribution, Log-Power Suja (LPSuj) distribution, Log-Exponentiated Power Ishita (LEPI), Log-Power Ishita (LPI), Log-Exponentiated Weibull (LEWe), Log-Power Rama (LPR), Log-Exponentiated Frechet (LEF), Log-Power Lindley (LPLi), Log-Exponentiated Power Akash (LEPA), Log-Exponentiated Power Lindley (LEPLi) distributions. The LPCJ outperformed the rest of the models and proves to fit better any skewed data.

Acknowledgment

The authors are very grateful to the anonymous reviewers and the editors for their comments which improved the quality of this article.

Conflict of interest

The author declare that there is no conflict of interest.

References

1. Abebe, B., Tesfay, M., Eyob, T., & Shanker, R. (2019). A two-parameter power Rama distribution with properties and applications. *Biometrics and Biostatistics International Journal*, *8*(1), 6-11.
2. Aidi, K., Al-Omari, A. I., & Alsultan, R. (2022). The Power Zeghdoudi Distribution: Properties, Estimation, and Applications to Real Right-Censored Data. Applied Sciences, 12(23), 12081.
3. Anabike, I. C., Igbokwe, C. P., Onyekwere, C. K., & Obulezi, O. J. (2023). Inference on the parameters of Zubair-Exponential distribution with application to survival times of Guinea Pigs. Journal of Advances in Mathematics and Computer Science, 38(7), 12-35.
4. Ashour, S. K., & Eltehiwy, M. A. (2015). Exponentiated power Lindley distribution. Journal of advanced research, 6(6), 895-905.
5. Badmus, N. I., Akinyem, M. I., & Onyeka-Ubaka, J. N. (2021). A Log-Beta Rayleigh Lomax Regression Model. Afrika Statistika, 16(4), 2993-3007.
6. Biazatti, E. C., Cordeiro, G. M., Rodrigues, G. M., Ortega, E. M., & de Santana, L. H. (2022). A Weibull-beta prime distribution to model COVID-19 data with the presence of covariates and censored data. Stats, 5(4), 1159-1173.
7. Chinedu, E.Q.; Chukwudum, Q.C.; Alsadat, N.; Obulezi, O.J.; Almetwally, E.M.; Tolba, A.H. New Lifetime Distribution with Applications to Single Acceptance Sampling Plan and Scenarios of Increasing Hazard Rates. Symmetry **2023a**, 15, 1881. https://doi.org/10.3390/sym15101881
8. Chinedu, E. Q., Asogwa, E. C., Sunday, B. T., Onyeizu, N. M., & Obulezi, J. O. (2023b). Unraveling Emotions: Contemporary Approaches in Sentiment Analysis. J Sen Net Data Comm, 3(1), 223-230.
9. Chukwuma, P. O., Harrison, E. O., Ibeakuzie, P., Anabike, I. C., & Obulezi, O. J. (2024). A new reduced quantile function for generating families of distributions. Ann Math Phys, 7(1), 001-015.
10. Cordeiro, G. M., Biazatti, E. C., & de Santana, L. H. (2023). A New Extended Weibull Distribution with Application to Influenza and Hepatitis Data. Stats, 6(2), 657-673.
11. Eliwa, M. S., Altun, E., Alhussain, Z. A., Ahmed, E. A., Salah, M. M., Ahmed, H. H., & El-Morshedy, M. (2021). A new one-parameter lifetime distribution and its regression model with applications. Plos one, 16(2), e0246969.
12. Etaga, H. O., Nwankwo, M. P., Oramulu, D. O., & Obulezi, O. J. (2023a). An Improved XShanker Distribution with Applications to Rainfall and Vinyl Chloride Data. Sch J Eng Tech, 9, 212-224.
13. Etaga, H. O., Onyekwere, C. K., Lydia, O. I., Nwankwo, M. P., Oramulu, D. O., & Obulezi, O. J. (2023b). Estimation of the Xrama Distribution Parameter Under Complete and Progressive Type-II Censored Schemes. Sch J Phys Math Stat, 10, 203-219.
14. Etaga, H. O., Nwankwo, M. P., Oramulu, D. O., Anabike, I. C., & Obulezi, O. J. (2023c). The Double XLindley Distribution: Properties and Applications. Sch J Phys Math Stat, 10, 192-202.

15. Ezeilo, C. I., Umeh, E. U., Osuagwu, D. C., & Onyekwere, C. K. (2023). Exploring the impact of factors affecting the lifespan of HIVs/AIDS patient's Survival: An investigation using advanced statistical techniques. Open Journal of Statistics, 13(4), 595-619.
16. Ferreira, A. A., & Cordeiro, G. M. (2023). The exponentiated power Ishita distribution: Properties, simulations, regression, and applications. Chilean Journal of Statistics (ChJS), 14(1).
17. Gallardo, D. I., Gómez, Y. M., & Segovia, F. A. (2021). Exponentiated power Maxwell distribution with quantile regression and applications. SORT-Statistics and Operations Research Transactions, 45(2), 181-200.
18. Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., & Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. Computational Statistics & Data Analysis, 64, 20-33.
19. Nadarajah, S., & Kotz, S. (2003). The exponentiated Fréchet distribution. Interstat Electronic Journal, 14, 01-07.
20. Nasiru, S., Abubakari, A. G., & Chesneau, C. (2022). New Lifetime Distribution for Modeling Data on the Unit Interval: Properties, Applications and Quantile Regression. Mathematical and Computational Applications, 27(6), 105.
21. Nwankwo, B. C., Orjiakoh, J. N., Nwankwo, M. P., Chukwu, E. I. M. I., & Obulezi, O. J. (2024). A New Distribution for Modeling both Blood Cancer Data and Median Effective Dose (ED50) of Artemether-Lumefantrine against P. falciparum. Earthline Journal of Mathematical Sciences, 14(1), 41-62.
22. Nwankwo, M. P., Nwankwo, B. C., & Obulezi, O. J. (2023). The Exponentiated Power Akash Distribution: Properties, Regression, and Applications to Infant Mortality Rate and COVID-19 Patients' Life Cycle. Annals of Biostatistics and Biometric Applications, 5(4), 1-12.
23. Musa, A., Onyeagu, S. I., & Obulezi, O. J. (2023). Comparative Study Based on Simulation of Some Methods of Classical Estimation of the Parameters of Exponentiated Lindley–Logarithmic Distribution. Asian Journal of Probability and Statistics, 22(4), 14-30.
24. Musa, A., Onyeagu, S. I., & Obulezi, O. J. (2023). Exponentiated Power Lindley-Logarithmic distribution and its applications. Asian Research Journal of Mathematics, 19(8), 47-60.
25. Obulezi, O. J. and Anabike, I. C. and Okoye, G.C. and Igbokwe, C. P. and Etaga, H. O. and Onyekwere, C. K. (2023a), "The Kumaraswamy Chris-Jerry Distribution and its Applications". In: Journal of Xidian University 17(6), 575– 591.
26. Obulezi, O. J., Obiorah, U. E., Anabike, I. C., & Igbokwe, C. P. (2023b). The Exponentiated Power Chris-Jerry Distribution: Properties, Regression, Simulation and Applications to Infant Mortality Rate and Lifetime of COVID-19 Patients. TWIST, 18(4), 328-337. https://twistjournal.net/twist/article/view/125
27. Obulezi, O. J., Chinedu, E. Q., Oramulu, D. O., Etaga, H. O., Onyeizu, N. M., & Ejike, C. O. (2023c). Machine Learning Models for Predicting Transportation Costs Inflated by Fuel Subsidy Removal Policy in Nigeria. Int. Res. J. Mod. Eng. Technol. Sci, 5, 1053-1070.
28. Obulezi, O., Igbokwe, C. P., & Anabike, I. C. (2023d). Single acceptance sampling plan based on truncated life tests for zubair-exponential distribution. Earthline Journal of Mathematical Sciences, 13(1), 165-181.

29. Obulezi, O. J., Nwankwo, B. C., Nwankwo, M. P., Anabike, I. C., Igbokwe, C. P., & Igwebudu, C. N. (2023e). Modeling Mortality Rate of COVID-19 Patients in the United Kingdom, Canada, and the Survival Rate of COVID-19 Patients in Spain. Journal of Xidian University, 17(11), 520-538.
30. Oha, F. C., Etaga, H. O., Ibeakuzie, P. O., Anabike, I. C., & Obulezi, O. J. (2024). Power XShanker Distribution: Properties, Estimation, and Applications. Eng OA, 2(1), 01-20.
31. Onyekwere, C. K., and Obulezi, O. J. (2022). Chris-Jerry distribution and its applications. Asian Journal of Probability and Statistics, 20(1), 16-30.
32. Onyekwere, C. K., Okoro, C. N., Obulezi, O. J., Udofia, E. M., & Anabike, I. C. (2022). Modification of Shanker distribution using quadratic rank transmutation map. Journal of Xidian University, 16(8), 179-198.
33. Oramulu, Dorathy O. and Igbokwe, Chinyere P. and Anabike, Ifeanyi C. and Etaga, Harrison O. and Obulezi, Okechukwu J. (2023a) Simulation Study of the Bayesian and Non-Bayesian Estimation of a new Lifetime Distribution Parameters with Increasing Hazard Rate. Asian Research Journal of Mathematics, 19 (9). pp. 183-211. ISSN 2456-477X
34. Oramulu, D. O., Etaga, H. O., Onuorah, A. J., & Obulezi, O. J. (2023b). A New Member in the Lindley Class of Distributions with Flexible Applications. Sch J Phys Math Stat, 7, 148-159.
35. Pal, M., Ali, M. M., & Woo, J. (2006). Exponentiated weibull distribution. Statistica, 66(2), 139-147.
36. Rady, E. H. A., Hassanein, W. A., & Elhaddad, T. A. (2016). The power Lomax distribution with an application to bladder cancer data. SpringerPlus, 5, 1-22.
37. Reis LDR. The Pezeta regression model: an alternative to unit Lindley regression model. Biom Biostat Int J. 2023;12(4):107–112.DOI: 10.15406/bbij.2023.12.00393
38. Rodrigues, G. M., Vila, R., Ortega, E. M., Cordeiro, G. M., & Serra, V. (2022). New results and regression model for the exponentiated odd log-logistic Weibull family of distributions with applications. arXiv preprint arXiv:2203.14189.
39. Shukla, K. K., & Shanker, R. (2018). Power Ishita distribution and its application to model lifetime data. Statistics in Transition new series, 19(1).
40. Tolba, A. H., Onyekwere, C. K., El-Saeed, A. R., Alsadat, N., Alohali, H., & Obulezi, O. J. (2023). A New Distribution for Modeling Data with Increasing Hazard Rate: A Case of COVID-19 Pandemic and Vinyl Chloride Data. Sustainability, 15(17), 12782.
41. Wongrin, W., & Bodhisuwan, W. (2017). Generalized Poisson–Lindley linear model for count data. Journal of Applied Statistics, 44(15), 2659-2671.

نموذج الانحدار معلمي باستخدام السلطة توزيع كريس جيري مع تطبيق لبيانات رقابة

كريستيانا إيزيلو وإيهاني س. أونييغو وأوزوما إي أوميه

قسم الإحصاء ، كلية العلوم الفيزيائية ، جامعة نامدي أزيكيوي ، أوكا ، نيجيري

الخلاصة: يكتسب الترابط بين مختلف مجالات الإحصاء اهتماما جيدا في الأدبيات. هذا ممكن من خلال الابتكارات مثل التحول اللوغاريتمي للتوزيع إلى نموذج الانحدار البارامتري ومن ثم دمج نماذج التوزيع الاحتمالي المعاصرة مع طريقة الانحدار الكلاسيكية. يفضل النموذج الجديد بشكل أساسي بسبب قابليته للتطبيق في سيناريوهات أوسع. في هذه المقالة ، التوزيع الأساسي هو توزيع الطاقة كريس جيري. تم اشتقاق مكافئ نموذج الانحدار المعاد تصنيعه بعناية مع مراعاة إجراء التقدير الأقصى تحت العينة الخاضعة للرقابة أيضا. تم إجراء دراسة محاكاة لنموذج انحدار كريس جيري بقوة السجل مع تقديم مقاييس الأداء. تم نشر البيانات الخاضعة للرقابة على مدى حياة مريض كوفيد 19− لتبرير الدافع لتطوير النموذج الجديد وتم استخدام اثني عشر نموذجا متنافسا لمقارنة الانحدار المقترح. تظهر النتائج أن النموذج المقترح أفضل بالفعل ويفضل على منافسيه .

الكلمات المفتاحية: مؤشر إدراك الفساد ، الناتج المحلي الإجمالي ، الناتج المحلي الإجمالي الحقيقي ، اقتصاد نيجيريا ، منظمة الشفافية الدولية ، حساب الخزانة الموحد.