



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2023, 2023:18

<https://doi.org/10.28919/cmbn/6144>

ISSN: 2052-2541

USING BETA REGRESSION MODELING IN MEDICAL SCIENCES: A COMPARATIVE STUDY

MOHAMED R. ABONAZEL^{1,*}, HADIR A. SAID¹, ELSAYED TAG-ELDIN², SUZAN ABDEL-RAHMAN³,
IBRAHIM G. KHATTAB⁴

¹Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo
University, Giza, Egypt

²Electrical Engineering Department, Faculty of Engineering & Technology, Future University in Egypt, New Cairo,
Egypt

³Department of Biostatistics and Demography, Faculty of Graduate Studies for Statistical Research, Cairo
University, Giza, Egypt

⁴Department of Statistics, Mathematics, and Insurance, Faculty of Business, Alexandria University, Alexandria,
Egypt

Copyright © 2023 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Beta regression (BR) models provide an adequate approach for modeling continuous outcomes of limited intervals (0, 1). The BR model assumes that the dependent variable follows a beta distribution and that its mean is affiliated to a set of exploratory variables through a linear predictor known as coefficients and link function. The BR model also includes a dispersion parameter. This paper describes the BR model along with its properties. Furthermore, the comparison between different link functions of the BR model is conducted through a medical real-life application

*Corresponding author

E-mail address: mabonazel@cu.edu.eg

Received May 27, 2021

to laparoscopic surgical operations aiming to widen congenital obstruction in the connection between the kidney and ureter, a condition called ureteropelvic junction obstruction (UJO).

Keywords: beta regression; maximum likelihood estimation; link functions; linear regression; ordinary least square; ureteropelvic junction obstruction.

2020 AMS Subject Classification: 62J12, 62J20, 62P10.

1. INTRODUCTION

The BR is a type of generalized linear model (GLM) as it is a member of the exponential family of distributions. The GLM constructed assuming a dependent variable is marginally distributed following a beta distribution which is referred to as the BR model. Beta regression was formally introduced in political science and has many applications in medical sciences. It is a suitable candidate to traditional linear regression when the dependent variable follows a beta distribution rather than a normal distribution. The beta distribution can be parameterized by its mean and variance like the normal distribution. However, unlike the normal distribution, the variance of a beta distribution is a function of its mean and a ‘precision’ parameter, which is a scale measure of how tightly the observed data is clustered.

The usual practice used to transform the data so that the transformed response, say \tilde{y} , assumes values in the real line and then applies a standard linear regression analysis. A commonly used transformation is the logit function: $\tilde{y} = \log\left(\frac{y}{1-y}\right)$, where \log is the natural logarithm. However, this approach has some disadvantages, such as:

- 1- The regression parameters are interpretable in terms of the mean of \tilde{y} , and not in terms of the mean of y (given Jensen’s inequality).
- 2- Regressions involving data from the unit interval such as rates and proportions which are typically heteroskedastic: display more variation around the mean and less variation as we approach the lower and upper limits of the standard unit interval.
- 3- The distributions of rates and proportions are typically asymmetric, and thus Gaussian-based approximations for interval estimation and hypothesis testing can be quite inaccurate in small

samples.

Ferrari and Cribari [1] proposed a regression model for continuous variates which assumes values in the standard unit interval, rates, and proportions or concentration indices. Since the model assumes that the response is beta distributed, they called their model the BR model. In their model, the regression parameters are interpretable in terms of the mean of y (the variable of interest) and the model is natural [2].

Let y a random variable following a beta distribution: $y \sim \text{beta}(p, q)$; where p, q are shape parameters and $p, q > 0$. The probability density function (PDF) of y is given as follows:

$$f(y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad p, q > 0, \quad 0 < y < 1, \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function. The mean and variance of y are $E(y) = \frac{p}{p+q}$,

$$\text{Var}(y) = \frac{pq}{(p+q)^2(p+q+1)}.$$

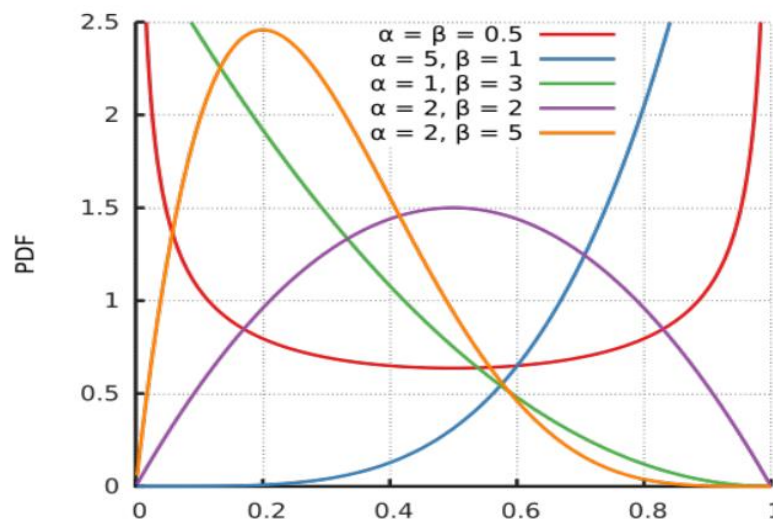


Figure 1: Beta density function with different combinations of parameters α and β .

Source: Wikipedia site.

Figure 1 shows that the beta distribution is highly flexible and able to accommodate the varying severity of skewness, heteroskedastic, and asymmetries (which make normalizing transformations impossible) [3]. However, the previous application of beta distribution did not involve the situations that the response variable can be modeled as a function of exogenous variables until the BR model proposed by [1].

2. LITERATURE REVIEW

Various beta regression models have been proposed as a function of location and precision parameters [4–6]. Paolino[4] demonstrated that maximum likelihood estimation of proportions data using the beta distribution provides more accurate and precise results than OLS approach. Kieschnick and McCullough [5] identified different specifications for variates observed on [0,1] intervals and recommended that regression models based on the beta distribution should be used to model for these data. Smithson and Verkuilen [6] presented maximum likelihood regression models assuming that the response variable is conditionally beta distributed. Location and variances were modeled using continuous and categorical variables considering the heteroscedasticity problem. Cribari and Vasconcellos [7] analyzed the finite-sample behavior of three bias corrections to the maximum likelihood estimators of Beta distribution parameters. While Vasconcellos and Cribari [8] proposed a new class of regression models for beta-distributed response variables considering that the beta distribution's parameters are related to the regression parameters and covariates. They also highlighted the bias of the maximum likelihood estimator in small samples.

Simas et al. [9] extended the beta regression model proposed by [1] allowing the regression structure to be nonlinear. They defined bias-corrected estimators by derived formulas for maximum-likelihood estimators with second-order biases. Cepeda [10] proposed joint mean and variance beta regression models and applied the Bayesian estimation method. Bayer et al. [11] proposed the beta regression control chart (BRCC). Ferrari and Pinheiro [12] developed two modified likelihood ratio tests to test restrictions on the beta regression. Bayer and Cribari [13] proposed three Bartlett corrected likelihood ratio tests for fixed dispersion of the beta regression. Also, Bayesian estimation of the BR model was proposed by [14]–[16]

Many studies have used the BR model in medical applications. Swearingen et al. [17] modeled ischemic stroke lesions using beta regression because of their highly skewed distribution. Cepeda et al.[10] modeled the meteorological data assuming a beta distribution and both the mean and precision parameters are being modeled. Yellareddygar et al. [18] proposed a BR model for

predicting the development of pink rot in potato tubers during storage and compared the BR model with the linear model on real data. Aktaş and Unlu [19] applied the BR model and describes its properties on well-being index data in Turkey . Gayawan et al. [20] adopted the beta regression model to examine covariate effects on the child mortality index in Nigeria.

Another group of studies proposed and developed the inflated BR model [21]–[25]. For the selection criteria model, Bayer and Cribari [11] proposed a model selection criterion for beta regression with varying dispersion, focusing on the selection of covariates for both mean and dispersion sub-models. Espinheira et al. [26] proposed also model selection criteria considering the residuals, leverage, and influential points both to systematic linear and nonlinear components.

3. METHODOLOGY

3.1 Beta Regression Model

The BR model introduced by Ferrari and Cribari [1] considered the precision parameter ϕ to be constant across all observations. Nevertheless, assuming a constant ϕ could lead to substantial loss in efficiency of the estimators [27]. In BR with varying dispersion, the precision parameter is assumed to be variable throughout the observations and modeled by covariates, unknown parameters, and one link function. The BR model assumes that the response variable follows beta distribution, which is a family of continuous probability distributions strictly defined on the interval $(0, 1)$ with two shape parameters (namely α and β). Those two positive shape parameters control the shape of the distribution in one-unit interval. Ferrari and Cribari [1] defined a regression structure for beta distributed responses that differs from (1). Let $\mu = \frac{p}{p+q}$, $\phi = p + q$, $p = \mu\phi$, and $q = (1 - \mu)\phi$. Ferrari and Cribari [1] proposed a different new parameterization, the beta density in (1) can be written as:

$$f(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (2)$$

where $y \sim (\mu, \phi)$, $0 < \mu < 1$, and $\phi > 0$. Then $E(y) = \mu$ and $Var(y) = \frac{\mu(1-\mu)}{1+\phi}$, where ϕ is known as a precision parameter and ϕ^{-1} is the dispersion parameter. Let $\{y_1, \dots, y_n\}$ be the

response variables that are independent of each other and each $y_i (i = 1, \dots, n)$ follows the beta density with μ and ϕ , the model is obtained by assuming that the mean of y_i can be written as

$$g(\mu_i) = \eta_i = f_1(x_i^T; \beta) = x_i^T \beta, \quad (3)$$

$$\mu_i = g^{-1}(\eta_i), \quad (4)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ is a vector of unknown regression parameters ($\beta \in \mathbb{R}^k$) and $x_i^T = (x_{1i}, \dots, x_{ki})$ are observations on k covariates ($k < n$) or (independent variables or covariates) and η_i a linear predictor, $\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$, in case of the model has an intercept. Finally, $g(\mu_i)$ is a strictly monotonic and twice differentiable link function that maps $(0, 1)$ into \mathbb{R} , Note that the variance of y_i is a function of μ_i and consequently of the covariate values. Hence, non-constant response variances are naturally accommodated into the model.

We can define a dispersion sub model to account for possibly varying dispersion:

$$h(\phi_i) = f_2(z_i^T; \theta) = z_i^T \theta, \quad (5)$$

where $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ is a vector of unknown regression parameters, and $z_i^T = (z_{1i}, \dots, z_{qi})$ are observations on q covariates ($q < n - k$). Note that under our parametrization, one can use for the dispersion sub-model the same link functions that are typically used for the mean sub-model namely: logit, probit, clog log, and log-log. For details on these links (see Table 1).

3.2 Link Functions and Estimation

Different link functions can be considered for the BR model; as in other generalized linear models [28]–[30]. When the parameters of interest are in the continuous interval $(0, 1)$, possibilities include the symmetric and asymmetric link functions [31], Box-Cox transformation link function [32], Gosset link function [33], Pregibon link function, and generalized logit function [34]. Canterle and Bayer [35] proposed the BR model with parameter-indexed link functions in the mean and dispersion sub models. A smooth and invertible linearizing link function $g(\cdot)$ is another element of the generalization, which transforms the expectation of the response variable, $g(\mu_i) = E(y_i)$ to the linear predictor:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad (6)$$

The inverse link $g^{-1}(\cdot)$ is also called the mean function commonly employed link function and their inverse are shown in Table1. The identity link simply returns its argument unaltered, $\eta_i = g(\mu_i)$, and $\mu_i = g^{-1}(\eta_i)$.

Table 1: Link functions and their inverse

Link function	Formula: $g(\mu_i) = \eta_i$	Inverse: $\mu_i = g^{-1}(\eta_i)$
Logit	$\eta_i = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$	$\mu_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$
Probit	$\eta_i = \phi^{-1}(\mu_i)$	$\mu_i = \phi(\eta_i)$
Complementary log-log (Clog-log)	$\eta_i = \log[-\log(1 - \mu_i)]$	$\mu_i = 1 - \exp[-\exp(\eta_i)]$
Log-log	$\eta_i = -\log[-\log(\mu_i)]$	$\mu_i = \exp[-\exp(-\eta_i)]$

Some processes require that we model the scale parameter as a function of covariates as shown in table 2.

Table 2: The definitions of the scale-link functions

Name	Function
Identity	$h(\phi_i) = \phi_i$
Log	$h(\phi_i) = \log(\phi_i)$
Square root	$h(\phi_i) = \sqrt{\phi_i}$

Note that the variance of y is a function of μ which renders the regression model based on this parameterization naturally heteroskedastic. In particular

$$\text{Var}(y_i) = \frac{\mu_i(1-\mu_i)}{1+\phi} = \frac{g^{-1}(x_i^T\beta)[1-g^{-1}(x_i^T\beta)]}{1+\phi}. \tag{7}$$

The log-likelihood function for the BR model is:

$$\begin{aligned} \ell(\beta, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi) = & \log\Gamma(\phi) - \log\Gamma(\mu_i\phi) - \log\Gamma((1 - \mu_i)\phi) + (\mu_i\phi - 1)\log y_i + \\ & ((1 - \mu_i)\phi - 1)\log(1 - y_i). \end{aligned} \tag{8}$$

It is possible to show that this BR model is a regular model since all regularity conditions are described. Furthermore, with an invertible reparameterization, one can guarantee that the ML estimations are unique [36]-[37]. A particularly useful link function is the logit link, in which case we can write $\mu_i = \frac{x_i^T\beta}{1+x_i^T\beta}$.

3.3 Model Selection Criteria

Model comparison can be made via information criteria such as Akaike information criterion (AIC)

introduced by [36] or the Bayesian information criterion (BIC). AIC is a well-known flaw of being dependent on sample size; it tends to capitalize on chance and thus favors more complex models when the sample size is large. For this reason, many prefer the BIC [6].

$$\text{AIC} = -2 \log(L_{fit}) + 2k; \text{BIC} = -2 \log(L_{fit}) + k \log(n), \quad (9)$$

where $\log(L_{fit})$ is the log-likelihood of whatever model was fitted, k is the number of parameters estimated, and n is the number of observations.

To assess the goodness of the model indices to assess the predictive capacity of the generalized regression model. These pseudo R^2 indices have been developed that are intended as analogs of R^2 as used in ordinary least squares (OLS) estimator. One such index introduced a penalized version of the BR pseudo-R-squared used in [1]. A simple candidate is a proportional reduction of error (PRE) statistic based on log-likelihoods:

$$\text{Pseudo } R^2 = 1 - \left[\frac{\log(L_{null})}{\log(L_{fit})} \right], \quad (10)$$

where $\log(L_{null})$ is the log-likelihood of the null model as defined earlier, and $\log(L_{fit})$ is the log-likelihood of whatever model was fitted.

4. REAL-LIFE APPLICATION

The dataset used in this application obtained from Abdelmohsen et al. [38]. Data is collected from a master's thesis in pediatric surgery conducted at Al-Azhar University in Egypt, which recorded the pre-and post-operative data for children aged from one year to twelve years since they were suffering from a congenital obstruction in the connection between the kidney and the ureter, a condition called ureteropelvic junction obstruction. The number of observations is twenty patients divided into two groups (ten patients in each group) to compare two minimally invasive techniques using surgical laparoscopy to widen this obstruction.

4.1 Descriptive Statistics

STATA software version 14.2 is used to analyze the data. Tables 3 and 4 provides labels and descriptive statistics for the study variables.

Table 3: Description of variables

Variable	Description
Dependent variable	
Post-Split Function (PO.SF)	Proportion percent after Surgery.
Independent variables	
Groups (GRO)	A dummy variable (Laparoscopic Group = 0, Laparoscopic- Assisted Group = 1).
Preoperative Split Function (PRE.SF)	Proportion percent before Surgery.
GIT Recovery (GIT)	Recovery of intestines after surgery in hours.
Presentation (PRES)	Pain Complaint (Incidental = 0, Prenatal = 1, Abdominal Pain=2).
HOS	Staying in hospital after surgery in days.
UTI	Dummy variable (Yes = 0, No = 1) of urinal inflammation.
PAIN	Measures of pain in 3 days after surgery.

Table 4: Descriptive statistics and different mean test

Variable	Mean	Max.	Min.	SD	CV	Jarque-Bera test		t-test	
						Statistic	P-value	t-Statistic	P-value
PO.SF	0.509	0.780	0.200	0.158	0.310	0.508	0.776	-0.207	0.839
PRE.SF	0.255	0.360	0.100	0.070	0.275	1.413	0.493	0.192	0.850
GIT	9.100	14.000	5.000	3.024	0.332	1.898	0.387	0.000	1.000
HOS	9.850	15.000	7.000	2.084	0.212	1.833	0.400	2.561	0.020
PAIN	5.300	8.000	4.000	1.342	0.253	1.273	0.529	-1.756	0.096

Note: SD: Standard deviation, CV: coefficient of variation, Min.: minimum, Max.: maximum

Table 4 indicates that all variables have a small variation as CV values for all variables is less than one. This is an indication that the data does not contain outliers. Jarque-Bera test is used to test normality. Since the p-values of all variables are more than 0.05, indicating that all variables are

normally distributed. So, t-test is used to compare between two groups. There are no significant differences in PO.SF, PRE.SF, GTI, and Pain variables between the two group. While there is a significant difference between the two groups for HOS variable.

4.2 Diagnostic Tests

To test the multicollinearity problem in the model, we used the correlation matrix and variance inflation factor (VIF). As shown in Table 5, the model has no multicollinearity problem, because all correlation coefficients less than 0.8, and VIF values for all variables less than 5 [6,35,39-40]. Some authors have introduced new estimators to reduce the impact of the multicollinearity problem in the BR model [30,39-41].

Table 5: Correlation matrix and variance inflation factor (VIF)

Variable	PO.SF	PRE.SF	GIT	PRES	HOS	UTI	PAIN
PO.SF	1.0000						
PRE.SF	0.0487	1.0000					
GIT	-0.1676	0.0001	1.0000				
PRES	0.2931	-0.5443	0.0739	1.0000			
HOS	-0.2848	-0.5168	0.3198	0.3751	1.0000		
UTI	0.0441	0.1400	0.3943	-0.1524	0.1068	1.0000	
PAIN	-0.1784	0.3824	-0.2672	-0.1249	-0.1148	-0.0107	1.0000
VIF	-----	1.69	1.55	1.71	1.76	1.64	1.41

To test the heteroskedasticity problem, we used the Breusch-pagan-Cook-Weisberg (BPCW) test. There is no heteroskedasticity problem (P-value = 0.8635 > 0.05).

4.3 Beta Regression Models

Table 6 compares between the BR model with different link functions (Logit, Probit, Clog-log, and Log-log link functions) and the linear regression model with the OLS estimation method to choose the best model for this data. Table 6 indicates that four BR models are significant, because the p-value of LR χ^2 test is less than 0.05. Moreover, the coefficients of all variables of the four BR models are significant except PAIN variable. This may be due to complications after surgery

as well. The dummy variable of GRO is positive and significant in all BR models, concluding that the laparoscopic-assisted surgery was the favorable one in this case. Unlike the linear regression model (OLS), it was found that the p-value of F-statistic for the model is $0.1683 > 0.05$ indicating that the model is non-significant, and all variables are non-significant. The BR model is the best model for this medical application.

Table 6: Estimated Regression Models

Variable	Linear regression	Beta regression			
	OLS	Logit	Probit	Clog-log	Log-log
Constant	0.474	-2.868***	-1.808***	-2.579***	-1.457**
GRO	0.138	0.718***	0.449***	0.509***	0.540**
PRE.SF	0.973	3.963***	2.487***	2.859***	2.798**
GIT	-0.017	-0.048*	-0.031*	-0.042**	-0.034
PRES	0.156	0.941***	0.577***	0.637***	0.682***
HOS	-0.019	-0.088**	-0.053**	-0.048**	-0.065*
UTI	0.013	1.275***	0.798***	0.928***	0.8733***
PAIN	-0.032	0.047	0.033	0.052	0.0036
Scale					
PRE.SF	-----	-89.606***	-90.013**	-90.968**	-78.877**
UTI	-----	10.199***	10.326***	10.10***	9.489**
PRES	-----	2.0503**	2.048**	1.996*	1.888*
Constant	-----	16.011***	16.017***	16.531***	13.895**
F-statistic	1.77				
LR χ^2	-----	32.75***	32.53***	33.56***	29.39**

Notes: ***, **, and * indicate statistical significance at the 0.001, 0.01 and 0.05 level, respectively.

To prove that the link function is the best of the model, we used goodness criteria to select the best model as shown in table 7. Table 7 compares the BR model with four link functions (Logit, Probit, Clog-log, Log-log). We found that clog-log has the smallest values of BIC and AIC, and the largest value of log-likelihood and Pseudo R^2 . Thus, we concluded that the BR model with clog-log is the best model to analyze our data.

Table 7: Goodness' criteria of beta models

Link Function	Log-likelihood	Pseudo R^2	AIC	BIC
Logit	25.791	0.635	-27.582	-15.634
Probit	25.681	0.633	-26.834	-14.885
Clog-log	26.198	0.641	-28.397	-16.448
Log-log	24.111	0.609	-24.224	-12.275

5. CONCLUSION

Beta regression assumes that the response variable is beta-distributed. The BR model has mean and dispersion parameters assuming that the mean is related to a set of independent variables through a linear predictor with link function and unknown coefficients. In this paper, we studied the BR model with different link functions as Logit, Probit, Clog-log, and Log-log and applied in the medical field. We utilized the selection criteria to compare between the different BR models and choose the best model for analyzing the data. We used a real data in medical research to compare two minimally invasive techniques using surgical laparoscopy and laparoscopy-assisted techniques for widening the ureter used. The results indicated that the linear regression model is not suitable for this application, unlike the BR model with clog-log which provided the best results.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

- [1] S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, *J. Appl. Stat.* 31 (2004), 799–815. <https://doi.org/10.1080/0266476042000214501>.
- [2] F. Cribari-Neto, A. Zeileis, Beta regression in R, *J. Stat. Softw.* 34 (2010), 1-24. <https://doi.org/10.18637/jss.v034.i02>.
- [3] P.K. Andersen, R.D. Gill, Cox's regression model for counting processes: a large sample study, 10 (1982), 1100-

1120. <https://www.jstor.org/stable/2240714>.
- [4] P. Paolino, Maximum likelihood estimation of models with beta-distributed dependent variables, *Polit. Anal.* 9 (2001), 325–346. <https://doi.org/10.1093/oxfordjournals.pan.a004873>.
- [5] R. Kieschnick, B.D. McCullough, Regression analysis of variates observed on (0, 1): percentages, proportions and fractions, *Stat. Model.* 3 (2003), 193–213. <https://doi.org/10.1191/1471082x03st053oa>.
- [6] M. Smithson, J. Verkuilen, A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, *Psychol. Methods.* 11 (2006), 54–71. <https://doi.org/10.1037/1082-989x.11.1.54>.
- [7] F. Cribari-Neto, K.L.P. Vasconcellos, Nearly unbiased maximum likelihood estimation for the beta distribution, *J. Stat. Comput. Simul.* 72 (2002), 107–118. <https://doi.org/10.1080/00949650212144>.
- [8] K.L.P. Vasconcellos, F. Cribari-Neto, Improved maximum likelihood estimation in a new class of beta regression models, *Brazilian J. Probab. Stat.* 19 (2005), 13–31. <https://www.jstor.org/stable/43601052>.
- [9] A.B. Simas, W. Barreto-Souza, A.V. Rocha, Improved estimators for a general class of beta regression models, *Comput. Stat. Data Anal.* 54 (2010), 348–366. <https://doi.org/10.1016/j.csda.2009.08.017>.
- [10] E. Cepeda-Cuervo, M.G. Andrade, J.A. Achcar, “Beta meteorological time series: application to air humidity data, Departamento de Estadística, pp. 1–14, 2013.
- [11] F.M. Bayer, C.M. Tondolo, F.M. Müller, Beta regression control chart for monitoring fractions and proportions, *Computers Ind. Eng.* 119 (2018), 416–426. <https://doi.org/10.1016/j.cie.2018.04.006>.
- [12] S.L.P. Ferrari, E.C. Pinheiro, Improved likelihood inference in beta regression, *J. Stat. Comput. Simul.* 81 (2011), 431–443. <https://doi.org/10.1080/00949650903389993>.
- [13] F.M. Bayer, F. Cribari-Neto, Bartlett corrections in beta regression models, *J. Stat. Plan. Inference.* 143 (2013), 531–547. <https://doi.org/10.1016/j.jspi.2012.08.018>.
- [14] J. Buckley, Estimation of models with beta-distributed dependent variables: a replication and extension of paolino’s study, *Polit. Anal.* 11 (2003), 204–205. <https://doi.org/10.1093/pan/mpg010>.
- [15] A.J. Branscum, W.O. Johnson, M.C. Thurmond, Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses, *Aust. N. Z. J. Stat.* 49 (2007), 287–301. <https://doi.org/10.1111/j.1467-842x.2007.00481.x>.
- [16] E. Cepeda-Cuervo, V. Núñez-Antón, Spatial double generalized beta regression models, *J. Educ. Behav. Stat.* 38 (2013), 604–628. <https://doi.org/10.3102/1076998613499779>.

- [17] C.J. Swearingen, B.C. Tilley, R.J. Adams, et al. Application of beta regression to analyze ischemic stroke volume in NINDS rt-PA clinical trials, *Neuroepidemiology*. 37 (2011), 73–82. <https://doi.org/10.1159/000330375>.
- [18] S.K.R. Yellareddygar, J.S. Pasche, R.J. Taylor, et al. Beta regression model for predicting the development of pink rot in potato tubers during storage, *Plant Dis*. 100 (2016), 1118–1124. <https://doi.org/10.1094/pdis-06-15-0696-re>.
- [19] S. Aktaş, H. Unlu, Beta regression for the indicator values of well-being index for provinces in Turkey, *J. Eng. Technol. Appl. Sci.* 2 (2017), 101–111. <https://doi.org/10.30931/jetas.321165>.
- [20] E. Gayawan, O.D. Fasusi, D. Bandyopadhyay, Structured additive distributional zero augmented beta regression modeling of mortality in Nigeria, *Spatial Stat.* 35 (2020), 100415. <https://doi.org/10.1016/j.spasta.2020.100415>.
- [21] R. Ospina, S.L.P. Ferrari, Inflated beta distributions, *Stat. Papers*. 51 (2008), 111–126. <https://doi.org/10.1007/s00362-008-0125-4>.
- [22] R. Ospina, S.L.P. Ferrari, A general class of zero-or-one inflated beta regression models, *Comput. Stat. Data Anal.* 56 (2012), 1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>.
- [23] T.L. Pereira, F. Cribari-Neto, Detecting model misspecification in inflated beta regressions, *Commun. Stat. - Simul. Comput.* 43 (2013), 631–656. <https://doi.org/10.1080/03610918.2012.712183>.
- [24] F. Liu, E.C. Eugenio, A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression, *Stat. Methods Med. Res.* 27 (2016) 1024–1044. <https://doi.org/10.1177/0962280216650699>.
- [25] P. Li, Efficient MCMC estimation of inflated beta regression models, *Comput. Stat.* 33 (2017), 127–158. <https://doi.org/10.1007/s00180-017-0747-x>.
- [26] P. Espinheira, L. da Silva, A. Silva, et al. Model selection criteria on beta regression for machine learning, *Mach. Learn. Knowl. Extr.* 1 (2019), 427–449. <https://doi.org/10.3390/make1010026>.
- [27] F.M. Bayer, F. Cribari-Neto, Model selection criteria in beta regression with varying dispersion, *Commun. Stat. - Simul. Comput.* 46 (2016), 729–746. <https://doi.org/10.1080/03610918.2014.977918>.
- [28] F.A. Awwad, K.A. Odeniyi, I. Dawoud, et al. New Two-Parameter Estimators For The Logistic Regression Model With Multicollinearity, *WSEAS Trans. Math.* 21 (2022), 403–414. <https://doi.org/10.37394/23206.2022.21.48>.
- [29] M.N. Akram, M.R. Abonazel, M. Amin, et al. A new Stein estimator for the zero - inflated negative binomial regression model, *Concurrency Comput.* 34 (2022), e7045. <https://doi.org/10.1002/cpe.7045>.

- [30] Z.Y. Algamal, M.R. Abonazel, F.A. Awwad, et al. Modified jackknife ridge estimator for the Conway-Maxwell-Poisson model, *Sci. Afr.* 19 (2023), e01543. <https://doi.org/10.1016/j.sciaf.2023.e01543>.
- [31] F.J. Aranda-Ordaz, On two families of transformations to additivity for binary response data, *Biometrika.* 68 (1981), 357–363. <https://doi.org/10.1093/biomet/68.2.357>.
- [32] V.M. Guerrero, R.A. Johnson, Use of the Box-Cox transformation with binary response models, *Biometrika.* 69 (1982), 309–314. <https://doi.org/10.1093/biomet/69.2.309>.
- [33] R. Koenker, J. Yoon, Parametric links for binary choice models: A Fisherian–Bayesian colloquy, *J. Econometrics.* 152 (2009), 120–130. <https://doi.org/10.1016/j.jeconom.2009.01.009>.
- [34] E.A. Ramalho, J.J.S. Ramalho, J.M.R. Murteira, Alternative estimating and testing empirical strategies for fractional regression models, *J. Econ. Surveys.* 25 (2011), 19–68. <https://doi.org/10.1111/j.1467-6419.2009.00602.x>.
- [35] D.R. Canterle, F.M. Bayer, Variable dispersion beta regressions with parametric link functions, *Stat. Papers.* 60 (2017), 1541–1567. <https://doi.org/10.1007/s00362-017-0885-9>.
- [36] M. Dishon, G.H. Weiss, Small sample comparison of estimation methods for the beta distribution, *J. Stat. Comput. Simul.* 11 (1980), 1–11. <https://doi.org/10.1080/00949658008810385>.
- [37] R. Ospina, F. Cribari-Neto, K.L.P. Vasconcellos, Improved point and interval estimation for a beta regression model, *Comput. Stat. Data Anal.* 51 (2006), 960–981. <https://doi.org/10.1016/j.csda.2005.10.002>.
- [38] A. Abdelmohsen, R. Shalaby, H. Galal, et al. Transperitoneal laparoscopic pyeloplasty versus laparoscopic assisted pyeloplasty for the treatment of ureteropelvic junction obstruction in children, *Egypt. J. Hosp. Med.* 80 (2020), 1046–1053. <https://doi.org/10.21608/ejhm.2020.107572>.
- [39] M.R. Abonazel, I. Dawoud, F.A. Awwad, et al. New estimators for the probit regression model with multicollinearity, *Sci. Afr.* 19 (2023), e01565. <https://doi.org/10.1016/j.sciaf.2023.e01565>.
- [40] M.R. Abonazel, A.A. Saber, F.A. Awwad, Kibria–Lukman estimator for the Conway–Maxwell Poisson regression model: Simulation and applications, *Sci. Afr.* 19 (2023), e01553. <https://doi.org/10.1016/j.sciaf.2023.e01553>.
- [41] M.R. Abonazel, I.M. Taha, Beta ridge regression estimators: simulation and application, *Commun. Stat. - Simul. Comput.* (2021). <https://doi.org/10.1080/03610918.2021.1960373>.