Journal of Statistics: Advances in Theory and Applications
Volume 16, Number 1, 2016, Pages 45-65
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jsata_7100121695

AN APPLICATION OF THE GENERALIZED LINEAR MODEL FOR THE GEOMETRIC DISTRIBUTION

FARZANA JAHAN, BORHAN SIDDIKA and M. ATAHARUL ISLAM

Department of Applied Statistics
East West University
Aftabnagar
Dhaka 1212
Bangladesh
e-mail: kakon18@yahoo.com
mataharul@yahoo.com

Abstract

Application of the generalized linear models (GLMs) in real life problems are well established and has extensive use. However, the GLM for the geometric distribution is not explored yet. The present study consists of the derivation of the GLM for the geometric distribution, estimation of parameters, and test procedures. An application is made to Bangladesh Demographic and Health Survey 2011 data to find the significant factors associated with the first occurrence of infant death in terms of birth order. Two different generalized linear models are fitted, one using the natural link function and the other one using the log link function. At the end, the results of both models are compared. It is found that the model fitted using log link function has lower Akaike's information criteria (AIC) and deviance than the model fitted using the natural link function, that means the GLM for the geometric distribution using log link function provides better result.

2010 Mathematics Subject Classification: 62E20, 62F03, 62F10, 62J12, 62P25.
Keywords and phrases: geometric distribution, generalized linear model, infant mortality, birth order.
Received July 23, 2016

1. Introduction

The development of the GLMs to obtain maximum likelihood estimates of parameters with observations which were distributed according to some exponential family had been established [30]. In this paper, authors discussed the estimation of parameters and the approach to compare among models with different number of independent variables. They had given examples using four common population distributions such as: normal, binomial, Poisson, and gamma.

Later on, there had been many works on GLMs. These contained studies on general aspect of the GLM and its application on different real life situations [19], [40], [43], [26], [11] etc. There had been works based on the GLM for specific distributions also. These works included both theories and application on different type of data from binomial, Poisson, and exponential distribution [15], [16], [24], [21], [43], [7] etc. Though, there are other distributions those belong to exponential family of distribution, still all of them are not explored in detail, for example, the geometric distribution. There had been several works on characterization of the geometric distribution [3], [8], [33], [10] and generalizations of the geometric distribution [9], [41]. But none of these studies attempted to derive the GLM for the geometric distribution. The form of the link function and the deviance of the geometric distribution were stated but no application had been shown [28].

The present study is an attempt to derive the GLM for the geometric distribution to show in detail the estimation and test procedures, to discuss the possible challenges faced when applying to real life data. An application of this approach is made to Bangladesh Demographic and Health Survey (BDHS) 2011 data using suitable dependent and independent variables.

2. Methods

2.1. Geometric distribution

The geometric distribution refers to the probability of the number of times needed to do something until getting a desired result. If the probability of success on each trial is p, then the probability that y_i failures are needed to get one success is

$$f(y_i, \theta) = pq^{y_i}, \tag{1}$$

where $y_i = 0, 1, 2, \ldots$ and $p + q = 1$. Equation (1) is the probability mass function of the geometric distribution with mean and variance, respectively, $E(Y) = \mu = \frac{q}{p}$ and $\mathrm{Var}(Y) = \frac{q}{p^2} = \mu(\mu - 1)$.

There is another form of the geometric mass function where number of trials needed to get the 1^{st} success is modelled as

$$f(y_i, \theta) = pq^{y_i - 1}, \tag{2}$$

where $y_i = 1, 2, 3, \ldots$ and $p + q = 1$. The mean and variance of Equation (2) are, respectively, $E(Y) = \mu = \frac{1}{p}$ and $\mathrm{Var}(Y) = \frac{q}{p^2} = \mu(\mu - 1)$.

The geometric distribution is the only discrete distribution which has lack of memory property. It is a discrete analogue of the exponential distribution [29], [2]. Also, the geometric distribution can be obtained from negative binomial distribution when the number of success, $r = 1$ [22].

2.2. GLM for the Geometric distribution

2.2.1. Use of natural link function

Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be n independent random variables, $Y_i =$ number of failures needed to get the 1^{st} success or number of events before the 1^{st} occurrence of the event of interest. We can show that Equation (2) follows exponential family of distribution with natural parameter $b(\theta) = \ln(1 - p)$, which yields the link function of the GLM as

$$g(\mu_i) = \ln\left(\frac{\mu_i - 1}{\mu_i}\right), \tag{3}$$

where $\mu_i = \frac{1}{1 - e^{\eta_i}}$, $\eta_i = X_i^T \beta$, X and β are vectors of covariates and parameters, respectively. Then score function and observed information matrix are $U_j = \sum_{i=1}^{n} X_{ij}(y_i - \mu_i)$ and $I_{ij} = \sum_{i=1}^{n} X_{ij} X_{ik} \quad (\mu_i(\mu_i - 1))$, respectively. The estimating equation using Newton-Raphson iterative procedure is

$$b^{(m)} = b^{(m-1)} + [I^{(m-1)}]^{-1} U^{(m-1)}, \tag{4}$$

where $m = 1, 2, 3, \ldots$ and $b^{(m)}$ is the vector of estimates of the parameters $\beta_1, \ldots, \beta_P$ at the m-th iteration. $[I^{(m-1)}]^{-1}$ is the inverse of observed information matrix and $U^{(m-1)}$ is the score vector all evaluated at $b^{(m-1)}$ [5].

Using the log likelihood of the geometric distribution we can get the deviance as

$$D = 2\left[\sum_{i=1}^{n} \left\{ (y_i - 1) \ln\left(\frac{y_i - 1}{\hat{\mu}_i - 1}\right) + y_i \ln\left(\frac{\hat{\mu}_i}{y_i}\right) \right\}\right], \tag{5}$$

which follows chi-square distribution with $(n - p)$ degrees of freedom.

2.2.2. Use of log link function

The log link function can also be used to fit the GLM for the geometric distribution [13], [20]. For the geometric random variable Y_i, log link function can also be used. The log-link function is $\ln \mu_i = e^{X_i^T \beta}$, that is,

$$\mu_i = e^{X_i^T \beta} = e^{\eta_i}. \tag{6}$$

The score function and observed information matrix are $U_j = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\mu_i - 1} X_{ij}$ and $I_{ij} = \sum_{i=1}^{n} \frac{\mu_i}{(\mu_i - 1)} x_{ij} x_{ik}$, respectively.

The estimating equation using Newton-Raphson iterative procedure and deviance will be same as for the GLM using natural link function for the geometric distribution.

2.3. Data

For the application of the GLM for the geometric distribution, we use the data collected in the Bangladesh Demographic and Health Survey conducted in 2011. Details of the reproductive history of women were collected using the individual women's questionnaire together with background information. The sample of the BDHS 2011 is nationally representative and covers the entire population in the country.

Sample was collected by two-stage stratified sample design. In the first stage, 600 enumeration areas (EA) were selected with probability proportional to the EA size, with 207 clusters in urban areas and 393 in rural areas. A complete household listing operation was then carried out in all the selected EAs to provide a sampling frame for the second-stage selection of households. In the second stage of sampling, a systematic sample of 30 households on average was selected per EA to provide statistically reliable estimates of key demographic and health variables for the country as a whole, for urban and rural areas separately, and for each of the seven divisions. With this design, the survey selected 17,141 residential households, which were expected to result in completed interviews with 17,842 ever-married women.

For our study, we used individual record of women data. Here women of reproductive age were interviewed for the information of their birth history. In BDHS 2011 data were given from last to first birth. For our study, we need the data for women for whom at least one child died at infant ages. We had to consider the order of the birth of that child as our response variable is the occurrence of 1^{st} infant death to woman in terms of birth order, that is, $Y_i = i,$ where i denotes the first infant death at birth order i. For our study, we selected data form 3469 women whose at least one child faced infant death at any order. The information about the covariates was selected for these women only.

3. Application of the GLM for the Geometric Distribution

Many studies on infant mortality showed the association of significant factors in past [32], [39], [6] etc. Among those some studies showed a significant association between birth order and infant mortality [25], [18], [38] etc. But none of the study considered the response variable as occurrence of 1^{st} infant death to woman in terms of birth order. Since there had been no work applying the generalized linear model for the geometric distribution in past. So this study is an attempt to find the significant factors associated with the response variable by using the GLM for the geometric distribution.

The response variable using BDHS 2011 dataset is the occurrence of 1^{st} infant death to woman in terms of birth order follows the geometric distribution. That is, $Y_i = i,$ where i denotes the first infant death at birth order i.

Since in this study, we consider women for whom there is a positive response for infant mortality of child, we use Equation (2) as the probability mass function (pmf) of the geometric distribution. The corresponding link function (Equation (3)), estimates (Equation (4)), and deviance (Equation (5)) are used for further analysis.

In conventional way, we see log link function is frequently used for the GLM of negative binomial distribution (e.g., in built in function of 'glm.nb' under library 'MASS' of R-Programming Language [36], [42]). As the geometric distribution can be obtained from the negative binomial distribution, for modelling the response variable of interest in present study, log link function can be used with its corresponding estimates (Equation (6)) and deviance. The use of log link function in application of the GLM for the negative binomial distribution had been found in most cases [13], [20]. So, in this present study, the results using log-link function and the natural link function from the likelihood function of the geometric distribution will be compared. This comparison will tell us whether use of log link function should be preferred to the natural link function while fitting the GLM of the geometric distribution.

To select the covariates for the model application, the past studies such as [18], [25], [38] etc. are reviewed and the covariates that are found in the data consistent with the literature are taken in the model. The selected covariates are mother's education, wealth index, type of place of residence, mother's involvement with NGO, sex of child, birth plurality, and age of mother at birth. The indicator variables are introduced as follows, mother's education ($X_1 = 1$ for primary education, 0, otherwise), ($X_2 = 1$ for secondary education, 0, otherwise), ($X_3 = 1$ for higher education, 0, otherwise), wealth index ($X_4 = 1$ for poor class, 0, otherwise), ($X_5 = 1$ for middle class, 0, otherwise), type of place of residence ($X_6 = 1$ for urban mother, 0, rural mother), mother's involvement with NGO ($X_7 = 1$ for involvement, 0, not involvement), sex of child ($X_8 = 1$ for male child, 0, female child), birth plurality ($X_9 = 1$ for single birth, 0 for multiple birth), and age of mother at birth ($X_{10} = 1$ for mothers' age below 18 years, 0, otherwise).

Including the above indicator variables, the models that are fitted in this study can be expressed as

Model 1. Using natural link function:

$$E(Y_i) = \frac{1}{1 - e^{\eta i}}.$$

Model 2. Using log link function:

$$E(Y_i) = e^{\eta i},$$

where, $Y_i = i,$ where i denotes the first infant death at birth order i.

4. Results and Discussion

4.1. Univariate analysis

Table 1.1. Proportion of 1^{st} infant death in terms of birth order with respect to total births at each order

Birth order	Proportion of 1^{st} infant death (No of women = 3469)
1	$0.11 (N_1 = 1832)$
2	$0.07 (N_2 = 856)$
3	$0.05 (N_3 = 416)$
4	$0.05 (N_4 = 211)$
5	$0.04 (N_5 = 101)$
6	$0.02 (N_6 = 28)$
7	$0.03 (N_7 = 18)$
8	$0.02 (N_8 = 7)$

It is observed in the Table 1.1 using BDHS 2011 data, there have been occurrences of 1^{st} infant death (according to birth order) up to 8^{th} child. Among these, highest infant death is observed for 1^{st} child, and subsequent frequencies are decreasing with the increasing birth order. So, the distribution of Y ($Y_i = i,$ where i denotes the first infant death at birth order i) is the geometric.

Table 1.2. Frequency and percentage distribution of women who had experienced infant deaths according to some specified characteristics

Variable	Frequency (n = 3469)	Percentage
Mother's education		
No education	1628	46.9
Primary	1181	34.0
Secondary	589	17.0
Higher	71	2.0
Wealth index		
Poor	1643	47.4
Middle class	676	19.5
Rich	1150	33.2
Involvement with NGO		
Yes	192	5.5
No	3277	94.5
Sex of child		
Male	1851	53.4
Female	1618	46.6
Type of place of residence		
Rural	2467	71.1
Urban	1002	28.9
Birth plurality		
Single	3332	96.1
Plural	137	3.9
Age of mothers at birth		
Below 18	1200	65.4
Above 18	2269	36.4

Data Source: Bangladesh Demographic Health Survey data 2011.

The distribution of women experiencing 1^{st} occurrence of infant death in terms of birth order according to the pre-specified covariates are displayed in the Table 1.2.

4.2. Bivariate analysis

The bivariate association between the response variable and covariates are measured and the cross tabulation is shown in Table 2.1. In Table 2.1, the explanatory variables which have significant association with the response variable 1^{st} infant death according to birth order are: mother's highest education level, sex of child, and age of mother at birth. Socio-economic status of the mother in terms of wealth index, NGO involvement of mothers, place of residence, and birth plurality exhibit non-significant bivariate relationship with the response variable.

Table 2.1. Percentage distribution of covariates in terms of 1st infant death according to birth order

1st Infant death according to birth order (***n*** **= 3469**)	1	2	3	4	5	6	7	8
Mother's education***								
No education	51.7	23.7	12.3	6.7	3.7	0.7	0.7	0.4
Primary	50.8	25.8	11.6	7.1	2.9	1.2	0.5	0.1
Secondary	57.7	25.5	12.2	3.1	1.0	0.5	0.0	0.0
Higher	70.4	21.1	8.5	0.0	0.0	0.0	0.0	0.0
Wealth index								
Poor	54.0	23.5	11.9	6.0	2.9	0.7	0.6	0.3
Middle class	50.6	27.2	11.2	6.2	3.1	1.0	0.4	0.1
Rich	52.3	24.9	12.5	6.1	2.8	0.9	0.4	0.1
Involvement with NGO								
Yes	50.5	29.2	12.0	3.6	2.6	0.5	1.6	0.0
No	52.9	24.4	12.0	6.2	2.9	0.8	0.5	0.2
Sex of child**								
Male	54.5	24.0	12.0	5.5	2.9	0.8	0.2	0.1
Female	50.9	25.5	11.9	6.7	2.9	0.8	0.9	0.4
Type of place of residence								
Rural	52.3	24.9	11.8	6.0	3.2	0.9	0.6	0.3
Urban	54.0	24.2	12.5	6.4	2.2	0.5	0.3	0.0
Birth plurality								
Single	53.9	24.2	11.7	5.9	2.7	0.8	0.5	0.2
Plural	27.0	35.0	19.0	10.2	7.3	0.7	0.7	0.0
Age***								
Above 18	35.0	30.9	18.0	9.3	4.5	1.2	0.8	0.3
Below 18	86.4	12.9	0.7	0.0	0.0	0.0	0.0	0.0

Data Source: Bangladesh Demographic Health Survey data 2011, Significance codes: 0.01 '***' 0.05 '**' 0.1 '*'.

It is well established that infant mortality is reduced with the increase in mother's educational level [38], [25], [32] etc. But our results of bivariate analysis between mother's education and 1st infant death according to birth order differ from these literatures. In this study, the occurrence of 1st infant death at birth order 1 is the highest for the mothers who have higher education, and then it decreases for the mothers with secondary education. The occurrence of 1st infant death among mothers with no education and primary education are very close and lower compared with mothers with higher education. This unusual pattern can be found because of the nature of data. This data is cross sectional and it contains information of women of different age. Some women might have given birth to their 1st child 20 years ago, some might have only 1 or 2 years ago. 50 women gave birth to their 1st child at the age of 11, 12, 13 (very early age), 1134 women gave birth at age of 20-25, 43 gave birth even at the age of 35-46. In this analysis, women of all age cohort and all generation are taken altogether without considering their different reproductive patterns, socio-demographic conditions of their times. As the women are from different generation, their lifestyle, impact of education in real life are very much different from each other. So when analyzing education as a factor to determine the 1st occurrence of infant death, it cannot provide desired output, because so many other factors are operating in this regard. It is also mentioned in literature that shifts in the reproductive pattern (as measured by birth interval, birth order, and maternal age) cannot explain the relationship between education and child mortality [14], [4]. Same conditions are present in the present study; as a result, the bivariate relationship between the response variable and mother's education cannot be explained. Zeria [45] found that women's average educational level in their community exerts a greater influence on infant survival than the mother's educational level. The women for whom the information is available in present study data (BDHS 2011), they belong to different groups in terms of their age, generation, socio-demographic conditions. As a result, their individual educational level's impact on the 1^{st} occurrence of infant mortality in terms of birth order cannot be meaningfully interpreted.

In the present study, mother's age at birth is categorized as below 18 years and 18 years to above years. For mothers with age below 18 years at the 1^{st} birth, have higher rate of occurrence of infant mortality at the 1^{st} birth. This is obvious because of the young age of mothers, the children who are born often be malnutritious, or underweight or suffer from various diseases, which leads to infant mortality [12], [17], [1], [34] etc. In most of the cases, young mother belong to low income groups, so they suffer from lack of prenatal care which increases the risk of infant mortality [23]. But the rate of 1^{st} infant death in 2^{nd} order decreases for the mothers age below 18 years age, as after the 1^{st} birth. They may get some knowledge and experience about children's health care, as a result infant mortality at 2^{nd} order birth decreases for these women. But for the women aged above 18, the occurrences of 1^{st} infant death is higher than the below 18 age group for birth order 2 and so on. This can happen because of women's increasing age at birth, the risk of infant mortality increases. So for the older aged women, the occurrence of 1^{st} infant death in birth order 2 or higher is more than the women of below 18 age group.

Sex of the child has significant impact on the 1^{st} occurrence of infant death in terms of birth order. It is seen in Table 2.1 that the 1^{st} occurrence of infant mortality in birth order 1 is higher for male children than female children. This is well known that infant mortality is higher in males than females because male children are biologically weaker and more susceptible to diseases and premature death than female, so there is high infant mortality among male and higher childhood mortality among females [31], [35], [37] etc. For birth order 2, it is observed that 1st occurrence of infant mortality for females is higher than the male. For birth order 1, only the biological facts work, in most of the cases, but for 2nd and higher order birth, son preferences play a vital role. As a result, female infants cannot access all the health care related facilities that male infants can, this erodes their biological advantages over male infants [37]. And as a result infant mortality increases for female children for 2nd and higher order birth.

4.3. Multivariate analysis

The result after applying the GLM for the geometric distribution to the data taking the pre-specified covariates (using natural link function and log link function) are shown in Table 3.1 (Result of Model 1: using natural link function) and Table 3.2 (Result of Model 2: using log link function).

Table 3.1. Parameter estimates for the GLM for the geometric distribution using natural link function (Model 1)

Variable Name	Coefficient	Standard error	Test statistic	*P* Value
Constant	− 0.4079521	0.0486898	− 8.3785946	0.0***
Wealth index (Ref: Poor)				
Middle	0.03322179	0.03247333	1.0230486	0.306
Rich	0.05037203	0.03121852	1.6135307	0.107
Place of residence (Ref: Rural)				
Urban	− 0.02168074	0.0303259	− 0.714925	0.475
Education (Ref: No education)				
Primary	− 0.02089642	0.02657691	− 0.7862624	0.432
Secondary	− 0.22913375	0.04253333	− 5.3871577	7.16e-08***
Higher	− 0.70586858	0.16402738	− 4.3033581	1.68e-05***
Sex (Ref: Female)	− 0.04583765	0.02431173	− 1.8854127	0.059*
Plurality(Ref: Plural)	− 0.13011142	0.0453968	− 2.8660924	4.16e-03***
NGO (Ref: Not involved)	− 0.02155033	0.05366789	− 0.4015497	0.688
Below 18 (Ref: 18 and above)	− 1.49597496	0.07261925	− 20.6002546	0.0***
Deviance: 771.0389	AIC: 18713.35			

Significance codes: 0.01 '***' 0.05 '**' 0.1 '*'.

From the results in Table 3.1, we observe that, secondary and higher education of mother, sex of child, birth plurality, and mother's age at birth (below 18) have significant effect on the occurrence of 1^{st} infant death according to birth order using natural link function for the GLM for the geometric distribution.

From the results in Table 3.2, we observe that, rich socio-economic class, secondary and higher education of mother, sex of child, birth plurality, and mother's age at birth (below 18) have significant effect on the occurrence of 1^{st} infant death according to birth order using log link function for the GLM for the geometric distribution. So, one additional significant variable is found when the log link function is used.

For comparing the two models, deviance and Akaike's information criteria (AIC) for both models are calculated and shown in Tables 3.1 and 3.2. According to this two criteria, the GLM for the geometric using log link provides better result, as it has minimum deviance and AIC.

Table 3.2. Parameter estimates for the GLM for the geometric distribution using log link function (Model 2)

Variable Name	Coefficient	Standard error	Test statistic	*P* Value
Constant	1.0365	0.04501	23.03	< 2e-16 ***
Wealth index (Ref: Poor)				
Middle	20.03165	0.02363	1.339	0.1806
Rich	0.0449	0.02272	1.976	0.0482 **
Place of residence (Ref: Rural)				
Urban	– 0.01792	0.0212	– 0.845	0.3982
Education (Ref: No education)				
Primary	– 0.02206	0.01972	– 1.119	0.2634
Secondary	– 0.20043	0.02627	– 7.629	3.05e-14 ***
Higher	– 0.48748	0.06636	– 7.346	2.53e-13 ***
Sex (Ref: Female)	– 0.04221	0.01742	– 2.423	0.0154 **
Plurality (Ref: Plural)	– 0.17367	0.04291	– 4.048	5.29e-05 ***
NGO (Ref: Not involved)	– 0.02138	0.03821	– 0.56	0.5758
Below 18 (Ref: 18 and above)	– 0.68452	0.01925	– 35.557	< 2e-16 ***
Deviance: 740.83		AIC: 12668		

Significance codes: 0.01 '***' 0.05 '**' 0.1 '*'.

5. Conclusion

The geometric distribution is a very important probability distribution where 1^{st} occurrence of any event can be modelled. The generalized linear model (GLM) for the geometric distribution can contribute significantly to exhibit many important facts associated with the 1st occurrence of any event. But in the past no studies have been made to apply the GLM for the geometric distribution. So in this paper, an attempt is made to apply this important technique to find out the significant factors associated with the first occurrence of infant death according to birth order. A comparison of the GLM for the geometric distribution using natural and log link function is made in this study. And it is observed that the model using log link function (Model 2) is better fitted to the data than the model using natural link function (Model 1) in terms of minimum AIC and deviance value. So it may be suggested that for the GLM for the geometric random variable, log link function is better suited than the natural link function.

Acknowledgement

This work was supported by the [University Grants Commission of Bangladesh, World Bank] under grant [Higher Education Quality Enhancement Project (HEQEP) sub-project, (CP-3293)].

References

[1] A. B. Abdullah, C.-K. Chen, J. Buford et al., Investigating risk factors associated with teenage pregnancy and infant mortality in the United States, European International Journal of Science and Technology 2(6) (2013), 226-234.

[2] A. Alzaatreh, C. Lee and F. Famoye, On the discrete analogues of continuous distributions, Statistical Methodology 9(6) (2012), 589-603.

[3] B. C. Arnold, Two characterization of the geometric distribution, Journal of Applied Probability 17(2) (1980), 570-573.

[4] J. G. Cleland and J. K. V. Ginneken, Maternal education and child survival in developing countries: The search for pathways of influence, Social Science and Medicine 27(12) (1988), 1357-1368.

[5] A. J. Dobson and A. Barnett, An Introduction to Generalized Linear Models, 3rd Edition, Chapman & Hall, 2008.

[6] C. Dollfus, M. Patetta, E. Siegel and A. W. Cross, Infant mortality: A practical approach to the analysis of the leading causes of death and risk factors, Pediatrics 86(176) (1990), 176-183.

[7] O. Ebenezer, Maximum Likelihood Analysis for Bivariate Exponential Distribution, This paper is submitted to Georg-August-Universität Göttingen, Ghana, 2007.

[8] T. S. Ferguson, A characterization of the geometric distribution, The American Mathematical Monthly 72(3) (1965), 256-260.

[9] E. Gómez-Déniz, Another generalization of the geometric distribution, Test 19(2) (2010), 399-415.

doi:10.1007/s11749-009-0169-3

[10] O. E. Gultekin and I. Bairamov, A trivariate geometric distribution, Characterization and asymptotic distribution, Ege University Journal of Faculty of Science 37(1) (2013), 1-18.

[11] S. Heberman and A. E. Renshaw, Generalized linear models and actuarial science, Journal of the Royal Statistical Society, Series D (The Statistician) 45(4) (1996), 407-436.

[12] H. H. Hibbs, The mother and infant mortality, Publications of the American Statistical Association 15(113) (1916), 66-79.

[13] J. Hilbe, Log-Negative Binomial Regression as a Generalized Linear Model, The Selected Works of Joseph M. Hilbe, 1993. Available at:

http://works.bepress.com/joseph_hilbe/11

[14] J. N. Hobcraft, J. W. McDonald and S. O. Rustenin, Socio-economic factors in infant and child mortality: A cross-national comparison, Population Study (Camb) 38 (1984), 193-223.

[15] M. A. Islam, A. A. Alzaid, R. I. Chowdhury and K. S. Sultan, A generalized bivariate Bernoulli model with covariate dependence, Journal of Applied Statistics 40(5) (2013), 1064-1075.

[16] M. A. Islam, R. I. Chowdhury and L. Briollais, A bivariate binary model for testing dependence in outcomes, Bulletin of the Malaysian Mathematical Sciences Society 35(4) (2012), 845-858.

[17] M. S. Islam, M. M. Rahaman, K. M. S. Aziz et al., Infant mortality in rural Bangladesh: An analysis of causes during neonatal and postnatal period, International Center for Diarrhoeal Disease Research, Bangladesh, Scientific Report No. 44 (1981).

[18] J. L. Kiley, J. C. Kleinman and M. Kiley, Triplets and higher-order multiple births: Time trends and infant mortality, American Journal of Diseases of Children 146(7) (1992), 862-868.

[19] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73(1) (1986), 13-22.

[20] J. Mann, P. Larsen and J. Brinkley, Exploring the use of negative binomial regression modeling for pediatric peripheral intravenous catheterization, Journal of Medical Statistics and Informatics 2(Article 6) (2014).

http://dx.doi.org/10.7243/2053-7662-2-6

[21] A. W. Marshall and I. Olkin, A generalized bivariate exponential distribution, Journal of Applied Probability 4(2) (1967), 291-302.

[22] A. W. Marshall and I. Olkin, A family of bivariate distributions generated by the bivariate Bernoulli distribution, Journal of the American Statistical Association 80(390) (1985), 332-338.

[23] T. J. Mathews, M. F. Macdorman and V. Statistics, Infant mortality statistics from the 2010 period linked birth/infant death data set, National Vital Statistics Reports 62 (8) (2013), 1-25.

[24] A. Melliana, Y. Setyorini, H. Eko and E. Al, The comparison of generalized Poisson regression and negative binomial regression methods in overcoming over dispersion, International Journal of Scientific & Technology Research 2(8) (2013), 255-258.

[25] B. Modin, Birth order and mortality: A life-long follow-up of 14200 boys and girls born in early 20th century Sweden, Social Science & Medicine 54 (2002), 1051-1064.

[26] G. G. Moisen and T. C. Edwards, Use of generalized linear models and digital data in a forest inventory of northern Utah, Journal of Agricultural, Biological, and Environmental Statistics 4(4) (1999), 372-390.

[27] N. I. Mondal, K. Hossain and K. Ali, Factors influencing infant and child mortality: A case study of Rajshahi district, Bangladesh, Journal of Human Ecology 26(1) (2009), 31-39.

[28] M. Müller, Generalized Linear Models, In: J. Gentle, W. Härdle, Y. Mori (Eds): Handbook of Computational Statistics (Volume I), Concepts and Fundamentals, Springer, 2004.

[29] V. Nekoukhou, M. H. Alamatsaz, H. Bidram and A. H. Aghajani, Discrete beta-exponential distribution, Communications in Statistics - Theory and Methods 41(11) (2012).

[30] J. A. Nelder and R. W. M. Wedderburn, Generalized linear models, Journal of Royal Statistical Society: Series A (General) 135(3) (1972), 370-384.

[31] A. Pandey, M. K.. Choe, N. Y. Luther and D. Sahu, Infant and child mortality in India, National Family Health Survey Subject Reports Number 11 (1998).

[32] M. K. Pandey, Maternal health and child mortality in rural India, ASARC Working Paper (2009), 1-14.

[33] A. S. Paulson and V. R. R. Uppuluri, A characterization of the geometric distribution and a bivariate geometric distribution, The Indian Journal of Statistics, Series A 34(3) (1972), 297-300.

[34] M. G. Phipps, M. Sowers and S. M. Demonner, The risk for infant mortality among adolescent childbearing groups, Journal of Women's Health 11(10) (2002), 889-897.

[35] R. Pongou, Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins, Demography 50(2) (2013), 421-444.

[36] R Core Team R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013.

URL http://www.R-project.org/

[37] C. C. Sawyer, Child mortality estimation: Estimating sex differences in childhood mortality since the 1970s, PLOS Medicine 9(8) (2012).

doi:10.1371/journal.pmed.1001287

[38] R. Singh and V. Tripathi, Maternal factors contributing to under-five mortality at birth order 1 to 5 in India: A comprehensive multivariate study, Springer Plus 2(284) (2013).

doi:10.1186/2193-1801-2-284

[39] G. C. S. Smith, J. P. Pell and R. Dobbie, Birth order, gestational age, and risk of delivery related prenatal death in twins: Retrospective cohort study, BMJ 325(1004) (2002).

[40] G. K. Smyth, Generalized linear models with varying dispersion, Journal of Royal Statistical Society Series B (Methodological) 51(1) (1989), 47-60.

[41] R. C. Tripathi, R. C. Gupta, T. J. White and B. R. A. M. C. Tripathi, Some generalizations of the geometric distribution, The Indian Journal of Statistics, Series B 49(3) (1987), 218-223.

[42] W. N. Venables and B. D. Ripley, Modern Applied Statistics with S, Fourth Edition, Springer, New York, 2002. ISBN 0-387-95457-0.

[43] P. Wang, M. L. Puterman, I. Cockburn and N. Le, Mixed poisson regression models with covariate dependent rates, Biometrics 52(2) (1996), 381-400.

[44] S. Zeger and M. R. Karim, Generalized linear models with random effects; a Gibbs sampling approach, Journal of the American Statistical Association 86(413) (1991), 79-86.

[45] A. Zeria, Preventive health strategies and infant survival in Zimbabwe, African Population Studies 11(1) (1996), 29-62.

■