Poisson Regression Modeling of Diarrhea Events in Pasuruan Regency with Maximum Likelihood Estimates and Generalized Method Moment

Abstract

Context: Diarrhea characterized by a frequency increased of defecation more than 3 times/day accompanied by changes in consistency (becoming liquid). The causes of diarrhea can be divided into 2 parts, which are direct causes and indirect causes that can facilitate or accelerate the occurrence of diarrhea, including bacteria, nutritional conditions, hygiene and sanitation, social culture such as population density, economic status, low birth weight, and immunization. **Aims:** The purpose of this study to examine the factors that influence the incidence of diarrhea. **Methods:** This research used secondary data, the prevalence of diarrhea and risk factors in Pasuruan Regency Health Center. Poisson regression approach with maximum likelihood estimator (MLE) estimation and Generalized Method Moment (GMM) used in this study. **Results:** The results showed that GMM estimation method in the Poisson regression model gave better performance in terms of significance parameters compared to the MLE method. **Conclusions:** Factors affecting the increase of diarrhea occurrences in area with an estimated MLE Percentage of non-exclusive breastfeeding and Percentage of normal nutritional status. Whereas the GMM estimation is the percentage of non-exclusive breastfeeding, the percentage of low birth weight, the percentage of population density, the percentage of smokers among family members in the house, the percentage of incomplete immunizations, the percentage of under-five years old children less than 2, the percentage of normal nutritional status, and the percentage of middle class socioeconomic status.

Keywords: *Diarrhea, likelihood functions, regression analysis*

Bambang W. Otok,
Dhian S. Rachmawati[1],
Purhadi,
Santi W. Purnami,
Mahpolah Mahpolah[2]

Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, [1]Sekolah Tinggi Ilmu Kesehatan Hang Tuah Surabaya, [2]Health Polytechnic Banjarmasin, Ministry of Health, Banjarmasin, Indonesia

Introduction

Health problems in Indonesia are very complex, ranging from infectious diseases and non-communicable diseases. Based on the media or mode of transmission, infectious diseases can be divided into three, namely through the airborne, foodborne, waterborne, and others transmission, and through vectors. Examples of infectious diseases are Acute Respiratory Infection (ARI), diarrhea, malaria, HIV/AIDS, and others. While non-communicable diseases are diseases that cannot be transmitted from one person to another. Examples of non-communicable diseases are asthma, hypertension, coronary heart disease, kidney disease, and others. Many factors that influence health problems include poor environmental conditions, lack of public awareness for healthy living, and lack of health services available in the community.

Diarrhea is a one form of bowel disorders. A person will be indicated with diarrhea if the person defecated more than three times a day with watery feces, can be accompanied by blood or mucus. According to the results of the Health Research in 2013 by Indonesian Ministry of Health,[1] the Prevalence Period of diarrhea is 3.5 percent. The five provinces with the highest incidence of diarrhea are Aceh (10.2 percent), Papua (9.6 percent), DKI Jakarta (8.9 percent), South Sulawesi (8.1 percent), and Banten (8 percent). East Java is among the 13 provinces with the highest incidence of diarrhea, which is 6.6 percent. Based on the health profile of Pasuruan Regency in 2015 and the number of patient visits to the Puskesmas, diarrhea was in the eighth position out of the top 10 diseases. Infants are an age group that is very susceptible to diarrhea. When observed from the occupational group that suffers from diarrhea, farmers/fishermen/laborers have the highest proportion compared to other occupation. ORS and zinc can be used as first aid to treat diarrhea. ORS is used for rehydration for the people that suffers from

Address for correspondence:
Dr. Bambang W. Otok, Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur, 60111 Indonesia.
E-mail: dr.otok.bw@gmail.com; bambang_wo@statistika.its.ac.id

diarrhea. Zinc is used to increase the availability of zinc in the body which helps the healing process of diarrhea. World Health Organization (WHO) stated that diarrhea kills 2 million children each year.[2] According to Sukersa (2001), diarrhea outbreaks in Indonesia are the second largest causes of death for children and third largest causes of death for infants and fifth largest causes of death for all ages, around 162 thousand children die every year or around 460 children every day. Every child in Indonesia experiences 1.6–2 episodes of diarrhea per year (www. piogama.ugm.ac.id). Survey of Diarrhea morbidity and knowledge, attitudes, and behaviors carried out by the Indonesian Ministry of Health in 2000 found that diarrhea morbidity for all ages in East Java was 283 per 1,000 population, while episodes in infants were 1.3 times per year.

There are several factors that influence the occurrence of diarrhea in infants. Maki, *et al*., with Mann Whitney analysis, states there are differences in the incidence of diarrhea.[3] First, the majority of diarrhea occurrences in 6- to 12-month-old infants who are exclusively breastfed have a history of non-recurring diarrhea. Second, infants aged 6–12 months who get formula milk that have a history of recurrent diarrhea. Aisyah, *et al*. (2013) using the Chi-Square hypothesis test results obtained there are differences in the incidence of diarrhea in 0–6 months old infants who are exclusively breastfed and those who are breastfed with early supplementary food.[4]

The poor quality of exclusive breastfeeding in Indonesia, the limited food supply at the household and the limited access to health care services caused 5 million children to suffer from malnutrition.[5] Risk factors that can cause the occurrence of diarrhea are internal factors consisting of age of approximately 2 months, LBW, male, nutritional status, vitamin A deficiency, too early supplementary feeding. While the external factors are exclusive breastfeeding, immunization, air pollution (smoking habits among family members), ssresidential density, inadequate ventilation, and socioeconomic.[1] The frequency of diarrhea occurrences in East Java, especially in Pasuruan Regency is suspected to experience underdispersion or overdispersion. Next apply poisson regression using GMM in the case of the number of diarrhea in infants in Pasuruan Regency. The results obtained are expected to help health centers in Pasuruan Regency to determine policies so that cases of diarrhea in infants decrease.

Subjects and Methods

The research data is secondary data, namely the prevalence of diarrhea and risk factors in the area of Pasuruan Regency Health Center. The conceptual framework for the occurrence of diarrhea is as follows.

Based on Figure 1, the method that used in this study is Poisson regression with GMM approach. Poisson regression

is used to analyze discrete type response variables, where the observational response (Y) is in the form of counted objects which are the functions of a number of certain characteristics (x). The probability of Y "number of events" with Poisson distribution, as follow[6,7,8].

$$P(Y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \quad (y = 0,1,2,\dots) \tag{2.1}$$

where μ is the average number of events with Poisson distribution. The parameter μ is very dependent on a particular unit or period of time, distance, area, volume and so on. The Poisson distribution is used to model the events that are relatively rare during certain time intervals.

Poisson regression model, the log function used is ln (μ_i) = η_i, so that the relationship function is presented in equations (2.2) and equation (2.3)

$$\ln \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \tag{2.2}$$

$$\mu_i = \exp(x_i^T \beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) \tag{2.3}$$

where i is the observation unit i = 1,2,…, n.

The GMM method is an estimation extension from method of moment. Population moment represent the information that will be used. The GMM method takes the estimation concept from the method of moment, where if in the method of moment, the number of population moment is equal to the number of parameters to be estimated, while for the GMM method the number of population moment is greater or equal to the number of parameters to be estimated. The population moment is the function of the parameter model and data that has been determined so that the expected value of the function is zero in the actual value of the parameter. The steps in applying logistic regression analysis with GMM estimation are as follows:[10,11,12,13]

a. Identifying research data
b. Conduct descriptive analysis of response variables and predictor variables from research data
c. Use and modify GMM algorithms in R software based on the package
d. Interpretation of parameter coefficients through values $\mu_i = \exp(x_i^T \beta)$.
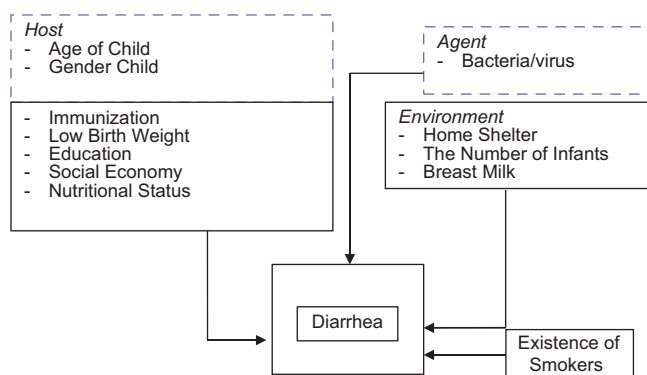

Figure 1: Conceptual framework with L. Green Theory Application (Notoatmodjo, 2010)[14]

Results

The response variable used in this study is the number of diarrhea prevalence at Pasuruan Regency in 2017. The number of diarrhea cases follows the Poisson distribution, this is indicated through the Kolmogorov Smirnov test with Asymp values. Sig. (2 tailed) = 0.329 which is greater than α = 5%, so it fails to reject Ho, which means the data follows the Poisson distribution. The results of estimation of MLE and GMM in the Poisson regression model are processed through package R, as follows:

Table 1 shows with the MLE method that the variables X2, X3, X4, X6, X9 are significant because the Sig value is smaller than α = 10%, whereas with the GMM method, all variables are statistically significant, except X7. Models with small Goodness of Fit values are said to be good models, then the GMM method is better than the MLE method. The model obtained from the GMM estimation results are as follows:

$$\log(\mu_i) = 1.838 + 0.367X_1 + 0.132X_2 - 0.072X_3 - 0.183X_4$$
$$- 0.236X_5 - 0.037X_6 + 0.021X_7 + 0.066X_8 + 0.091X_9$$

then,

$$\hat{\mu}_i = \exp(1.838 + 0.367X_1 + 0.132X_2 - 0.072X_3 - 0.183X_4$$
$$- 0.236X_5 - 0.037X_6 + 0.021X_7 + 0.066X_8 + 0.091X_9)$$

Discussion

Percentage of non-exclusive breastfeeding (X1)

Every time there is an increase in the percentage of non-exclusive breastfeeding (X1), the average value of diarrhea prevalence will increase by 1.443 times with provision of all other variables are constant. This is in accordance with the research by Hegar *et al*.[15] which

states the adhesion (adhesion) of Bifidobacterium bifidum in the digestive mucosa of healthy infants is greater than that of infants with allergies. This situation supports the relationship between the incidence of allergic diseases and the composition of Bifidobacteria in the digestive tract. According to Juffrie,[16] for infant who are exclusively breastfed, the practical definition of diarrhea is an increase in the frequency of bowel movements or feces with watery consistency, according to the mother, that is abnormal or unusual.

Percentage of low birth weight (X2)

Every time there is an increase in the percentage of low birth weight (X2), the average value of diarrhea prevalence will increase by 1.141 times with the provision of all other variables are constant. This is also in line with Widodo's[17] study which states that things that can affect the incidence of pneumonia in children are LBW. Low birth weight (LBW) show a tendency to be more susceptible to infectious diseases than babies with normal birth weight and it is a cause of high rates of infant mortality.

Percentage of Number of children under five less than 2 (X3)

Every time there is an increase in the percentage of the number of children (Under 5 years old) less than 2 (X3), the average value of diarrhea prevalence will decrease by 0.931 times with the provision that all other variables are constant. This is in accordance with the research of Susanti *et al*.,[18] which states that the number of children under five years old and the incidence of diarrhea depends on the origin of the disease, which is contracted or not.

Table 1: Significance test of the parameter coefficient of the poisson regression model in the diarrhea occurrence using MLE and GMM

Variables	Poisson Regressiom									
	MLE					GMM				
	Estimate	Std. Error	Wald	Sig.	μ_i	Estimate	Std. Error	Wald	Sig.	μ_i
Constants	0.175	0.478	2.738	0.012	1.191	1.838	0.343	5.356	0.000	6.285
Percentage of non-exclusive breastfeeding (X1)	-0.128	0.140	-1.097	0.215	0.880	0.367	0.150	2.447	0.023	1.443
Low birth weight percentage (X2)	0.020	0.063	3.160	0.004	1.020	0.132	0.041	-3.191	0.004	1.141
Percentage of number of children (under 5 years old) less than 2 (X3)	-0.017	0.034	-1.950	0.062	0.983	-0.072	0.033	2.183	0.040	0.931
Percentage of normal nutritional status (X4)	0.024	0.046	1.937	0.064	1.024	-0.183	0.065	-2.831	0.010	0.833
Percentage of mother's education level (High School or above) (X5)	-0.060	0.095	-1.572	0.116	0.941	-0.236	0.106	2.221	0.037	0.790
Percentage of middle socioeconomic level (Provincial Minimum Wage (UMP) or above) (X6)	-0.014	0.027	1.943	0.063	0.986	-0.037	0.018	-2.025	0.054	0.964
Percentage of occupancy density (X7)	0.022	0.035	1.562	0.118	1.023	0.021	0.026	0.811	0.283	1.021
Percentage of smokers among family members (X8)	-0.024	0.031	-1.272	0.175	0.976	0.066	0.030	2.177	0.041	1.068
Percentage of incomplete immunization (X9)	0.023	0.047	2.041	0.053	1.023	0.091	0.037	-2.419	0.025	1.095

Percentage of normal nutritional status (X4)

Every time there is an increase in the percentage of normal nutritional status (X4), the average value of diarrhea prevalence will decrease by 0.833 times with provision that all other variables are constant.

Percentage of mother's education level (High School or above) (X5)

Every time there is an increase in the percentage of mother's education level (High School or above) (X5), the average value of diarrhea prevalence will decrease by 0.790 times with provision that all other variables are constant. This shows that the level of education is closely related to socioeconomic conditions, and also related to parental knowledge. Lack of knowledge causes some cases of Acute Respiratory Infection (ARI) to be undetected by the parents and not treated properly.

Percentage of middle socioeconomic level (Provincial Minimum Wage (UMP) or above) (X6)

Every time there is an increase in middle socioeconomic level (Provincial Minimum Wage (UMP) or above) (X6), the average value of diarrhea prevalence will decrease by 0.964 times with provision that all other variables are constant. Children with low socioeconomic level are more likely to have diarrhea than children with high socioeconomic level. This can be explained that socioeconomic status influences education and other factors such as nutrition, environment and acceptance of health services. Parents from high socioeconomic levels are better able to provide healthy food, vitamins and supplements that can help improve family health status. Children who come from families with low socioeconomic level have a greater risk of experiencing diarrhea.

Percentage of smokers among family members in the house (X8)

Every time there is an increase in the percentage of smokers of family members (X8), the average value of diarrhea prevalence will increase by 1.068 with provision that all other variables are constant. This shows that one of the causes of children suffers from diarrhea is because some family members are smoker, so children are exposed to cigarette smoke which causes ARI and indirectly affects diarrhea. The high prevalence of passive smokers in toddlers and young age is caused because they still live at home with parents or siblings who smoke at home.

Percentage of incomplete immunization (X9)

Every time there is an increase in the percentage of incomplete immunization (X9), the average value of diarrhea prevalence will increase by 1.095 times with provision that all other variables are constant. Children with incomplete immunization status are more likely to experience it than children with complete immunization status. This is related to Measles, Pertussis, and several other diseases can increase the risk of Diarrhea and aggravate Diarrhea itself, but actually can be prevented. Measles, pertussis, and diphtheria together can cause 15–25% of all deaths related to diarrhea.

References

1. Dinas Kesehatan Pemerintah Kabupaten Pasuruan. Profil Kesehatan Kabupaten Pasuruan Tahun 2015. Pasuruan. 2015.
2. WHO. Penanganan ISPA Pada Anak di Rumah Sakit Kecil Negara Berkembang. Jakarta: Penerbit Buku Kedokteran EGC; 2003.
3. Maki F, Umboh A, Ismanto AY. Perbedaan Pemberian Asi Eksklusif dan Susu Formula Terhadap Kejadian Diare Pada Bayi Usia 6 12 bulan di Wilayah Kerja Puskesmas Ranotana Weru. eJournal Keperawatan (eKp) 2017;5. Available from : https://media.neliti.com/media/publications/112546-ID-perbedaan-pemberian-asi-eksklusif-dan-su.pdf [Last accessed on 2020 Jan 02].
4. Afriani B. Peranan petugas kesehatan dan ketersediaan sarana air bersih dengan kejadian diare. Aisyah: Jurnal Ilmu Kesehatan 2017;2:117-22.
5. Chausse P. Computing generalized method of moments and generalized empirical likelihood with R. J Stat Softw 2010;34:1-35.
6. Agresti A. Categorical Data Analysis. 2nd ed. New York: John Wiley and Sons; 2002.
7. Cameron AC, Trivedi PK. Regression Analysis of Count Data. Cambridge: CambCambridge University Press; 1998.
8. Famoye F, Bae S, Wulu JT, Bartolucci AA, Singh KP. On the generalized poisson regression model with an application to accident data. J Data Sci 2004;2:287-95.
9. Hall AR. Generalized Method of Moments: Advanced Text in Econometrics. New York: Oxford University Press Inc; 2005.
10. MacKinnon JG. Notes on the Generalized Method of Moments. Economics: 851 854. USA: Oxford University Press Inc; 2009.
11. Mahpolah, Suharto, Wibowo A, Otok BW. Generalized method moment logistic regression model on the prevalence of acute respiratory tract infections. J Appl Environ Biol Sci 2017;7:92-6.
12. Rasyidin, MT, Otok BW, latra IN, Estimation of Generalized Method of Moment in Logistic Regression Model. Prosiding Seminar Nasional Matematika, Universitas Jember; 2014:167-74.
13. Purwanengsi SE, Kuntoro, Windu P, Bambang WO. The modeling level of birth weight using a maximum likelihood estimation and generalized method of moment. TextRoad Publication. J Appl Environ Biol Sci, 2016;6:125-8.
14. Notoatmodjo S. Metodologi Penelitian Kesehatan. Cetakan ke 3. 2005.

15. Hegar, dkk. BedahASI. Jakarta; Ikatan Dokter Anak Indonesia; 2008. hlm 65.

16. Juffrie. Buku Ajar Gastroenterologi Hepatologi. Cetakan ke 1. Jakarta: Badan Penerbit IDAI; 2010. hlm 88.

17. Widodo Y. Pertumbuhan bayi usia 0 – 4 bulan yang mendapat ASI eksklusif dan makanan pendamping ASI (penelitian lanjutan) Badan Litbang Kesehatan. 2005.

18. Susanti WE, Novrikasari, Sunarsih E. Determinan kejadian diare pada anak balita di Indonesia (analisis lanjut data SDKI 2012). Jurnal Ilmu Kesehatan Masyarakat 2016;7:64-72.