View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Lumbung Pustaka UNY (UNY Repository)

M – 26

NEGATIVE BINOMIAL REGRESSION METHODS TO ANALYZE FACTORS AFFECTING CHILD MORTALITY RATES IN WEST JAVA

Hadiyatul Fitriyah[1*], Anang Kurnia [2], Farit M. Afendi[3]
[1,2,3] *Department of Statistics, FMIPA, Bogor Agricultural University, Indonesia*
[*] *Corresponding email address: myhadiyah88@gmail.com*

Abstract

Data on the number of child mortality cases are discrete data (count) which are usually analyzed with Poisson regression. The characteristics of the Poisson regression mean and variance must be the same, whereas in fact the count data is often becoming variance greater than the mean, which is often referred to overdispersion. To deal with the problem over dispersion, modelling can be done with Negative Binomial Regression because it does not require the mean value equal to the value of variance. Model Negative Binomial produces Deviance/Degree Freely value of 1.6347 and Pearson Chi-Square of 1.4569. This value goes to 1, its means that overdispersion problem was sloved.
Key Word: child mortality rates; Negative Binomial Regression, overdisversion

INTRODUCTION

The result of SDKI 2012 (Survei Demografi dan Kesehatan Indonesia) about Child Mortality Rates (CMR) show that CMR in West Java is 38/1000 life birth. In other side National's CMR is 40/1000. MDGs targeting CMR in West Java become 32 mortlity in 2015. So according SDKI 2012, West Java hasn't reah out MDGs targets. There are many factors are affecting CMR, such as healt facility, poverty, education, nutrition and many other. There are many distric or city in West Java haven't gave fundamental healt service to society, healt promoting, and society empowerment in healt.

Data on the number of child mortality cases are discrete data (count) which are usually analyzed with Poisson regression. The characteristics of the Poisson regression mean and variance must be the same, whereas in fact the count data is often becoming variance greater than the mean, which is often referred to overdispersion. Theare are many method to overcoming overdisversion.

The studies used in this study was to negative binomial regression, also determine the factors on the numbe CMR in West Java. This research is carried out to obtain the best model by Negative Binomial Regression. The result is also expected to provide additional information about any factors that significantly influence the Child Mortality Rates.

LITERATURE REVIEWS

2.1 Child Mortality Rates

Child is both infant and children 0 until before 5 years old. In general called 0 - 4 years old. Child Mortality Rates is he number of chid mortality 0 - 4 years old in one year and regency/1000 child. Formulation of CMR is :

$$CMR = \frac{Y0 - < 5\ years\ old}{\Sigma\ \text{child}\ 0 - 4\ \text{years old}}\ x\ 1000$$

with:
CMR = Child Mortality Rates
$Y0$ - <5 years old = Number child mortality berusia (0-4) years old.
Σ child (0-4) years old = Number child berusia (0-4) years old.

2.2 Poisson Regression

Poisson regression is a nonlinear regression analysis of the Poisson distribution, where the analysis is highly suitable for use in analyzing discrete data (count). Poisson regression model is a Generalized Linear Model (GLM) is the Poisson distribution is assumed response data Poisson regression is said to contain over dispersion if the variance is greater than the value of the mean value. Over dispersion has the same impact with the assumption that if the offense discrete data occurred over dispersion but still used Poisson regression, the parameter estimates of the regression coefficients remain consistent but not efficient. This has an impact on the value of the standard error to under estimate, so that the conclusions become invalid. Over dispersion phenomenon can be written as $Var(Y) > E(Y)\ or\ E(Y) = \tau\, Var(Y)$.

2.3 Negative Binomial Regression (Nbr)

In addition to GPR, over dispersion handling on Poisson regression can also be performed by using negative binomial models approach. In the negative binomial regression, if θ goes to zero then variable (Yi) goes to μi so will converge towards the negative binomial Poisson [3], [4]. Negative binomial regression models have the same form as the Poisson regression model Negative binomial regression parameter estimation performed by using MLE method. Log-likelihood equation for the negative binomial is.

$$\ln L\,(\theta;\beta) = \sum_{i=1}^{n}\left\{\left(\sum_{i=0}^{y_i}\ln\,(y_j+\theta)\right)\ln y_i\,!\,(y_i+\theta)\ln\left(1+\theta\exp\left(x_i^T\beta\right)\right)+y_i ln\theta + y_i x_i^T\beta\right\}$$

2.4 Akaike Information Criterion

AIC is one of the criteria in determining the best model as follows.

$$AIC = -2\ln\,L\left(\hat{\theta}\right)+2k$$

Where $L\left(\hat{\theta}\right)$is *likelihoo valued*, and *k* the number of parameters. The best model is the model that has the smallest AIC value.

RESEARCH METHODOLOGY

3.1 Data

The data used in this research is data from PODES (Potensi Desa) 2011, West Java in Figure 2011 and SESENAS (Survei Ekonomi dan Sosial Nasional) 2011. Respon variabel is number child mortality in West Java. Predictor variabel is present child with mall nutrition (PBGB), child haven't vaksin (TDK_IMUNISASI), present of puskesmas (RPUS), healt officer (RTKS), present household's poverty (RKFM), present of house not habitable (PRTLH), the last birth asistens by non medis (PTKBNM), present sanitation not worthy (PRTSTL), average of school duration (RLS).

3.2 Analysis Stage

To get the best modelling, some measures are used as follows:

1. Explorating data
2. Getting the best model for the Poisson regression modelling of the number of child mortality cases.
3. Detecting the presence of over dispersion on the data by looking at the value of Pearson Chi-squares and Deviance divided by degrees freely.
4. Getting the best model for the negative binomial regression modelling of the number of child mortality case is to: a) Estimate the parameters of the negative binomial regression model, b) Test the significance of the negative binomial regression model parameters

simultaneously and partially, c) Calculate the AIC values of the negative binomial regression model.

5. Conclusion and discussion.

RESEARCH RESULT AND ANALYSIS

4.1 The Modelling Of Number Child Mortality Cases Using Poisson Regression

The data number child mortality cases are a data count, where this type of data follows the Poisson distribution. The modelling with Poisson regression analysis is conducted to determine the factors that influence the number of child mortality. MLE method is used to obtain the estimation of the Poisson regression model parameters as shown in Table 1 and the resulting AIC value of 4638.7315.

Tabel 1. Poisson Regression Model Parameter Estimation

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.0669	0.0297	-2.1252	-2.0087	4837.49	<.0001
PBGB	1	0.1093	0.0083	0.0931	0.1254	175.04	<.0001
TDK_IMUNISASI	1	-0.0032	0.0011	-0.0054	-0.0011	8.95	0.0028
RPUS	1	0.0705	0.0013	0.0680	0.0729	3130.97	<.0001
RTKS	1	0.0053	0.0001	0.0050	0.0055	1391.80	<.0001
RKFM	1	-0.0042	0.0003	-0.0047	-0.0037	237.07	<.0001
PRTLH	1	-0.0002	0.0005	-0.0013	0.0008	0.17	0.6829
PTKBNM	1	-0.0017	0.0004	-0.0025	-0.0009	16.83	<.0001
PRTSTL	1	-0.0077	0.0002	-0.0081	-0.0073	1420.32	<.0001
RLS	1	-0.1818	0.0026	-0.1870	-0.1766	4730.09	<.0001

Table 1 shows that the 5% significance level can be seen that the p-value of all the parameters is smaller than 0.05. so that the parameters intersept, present child with mall nutrition (PBGB), child haven't vaksin (TDK_IMUNISASI), present of puskesmas (RPUS), healt officer (RTKS), present household's poverty (RKFM), the last birth asistens by non medis (PTKBNM), present sanitation not worthy (PRTSTL) and average of school duration (RLS). have significant effect on the model. Then Poisson regression models generated by.

$$\hat{\mu} = \exp(-2.0669 + 0.1093X1 - 0.0032X2 + 0.0705X3 + 0.0053X4 - 0.0042X5 - 0.0017X7 - 0.0077X8 - 0.1818X9)$$

Predictor variables that influence the number of child cases in West Java is present child with mall nutrition (PBGB), child haven't vaksin (TDK_IMUNISASI), present of puskesmas (RPUS), healt officer (RTKS), present household's poverty (RKFM), the last birth asistens by non medis (PTKBNM), present sanitation not worthy (PRTSTL) and average of school duration (RLS).

4.2 Overdispersion Test

In the Poisson regression analysis using, there is assumptions contained equidispersion that the mean value of the variance must be met. However, this assumption is rarely fulfilled so that it appears the case over dispersion. For detecting the over dispersion, can be seen from the value of deviance / df or Pearson / df. If the value of deviance /df or Pearson / db is greater than 1, it can be said to be a case of over dispersion whereas if it is less than 1 then there under dispersion.

Table 2. Value of Deviance and Pearson Poisson Regression

Criterion	DF	Value	Value/DF
Deviance	16	4349.3274	271.8330
Scaled Deviance	16	4349.3274	271.8330
Pearson Chi-Square	16	4142.4590	258.9037
Scaled Pearson X2	16	4142.4590	258.9037

Table 3 shows that the value of deviance / df and Pearson chi-square/df greater than 1 so it can be concluded on the Poisson regression models the number of numbee child mortaity cases in West parameter estimates. Overdispersion presence will also have consequences on the value of the standard error estimator for the smaller (underestimate) which subsequently can lead to errors in inference for the parameters. To overcome this, the modelling is done using the Negative Binomial regression, where this methods can accommodate the dispersion parameter.

4.3 The Modelling Of Number Of Cervical Cancer Cases Using Negative Binomial Regression (Nbr)

Each way to handle overdispersion on the Poisson regression model aside of using negative binomial regression model. The model with the smallest AIC value is the best model. In Table 3, it is shown the possibility of a negative binomial regression model with the smallest AIC value for each combination of variables ranging from one to six combinations of predictor variables using a significance level of 5%.

Table 3. Parameter Estimation of Negative Binomial Regression Model

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.5652	0.4100	-2.3688	-0.7616	14.57	0.0001
PBGB	1	0.0412	0.1254	-0.2046	0.2871	0.11	0.7423
TDK_IMUNISASI	1	-0.0031	0.0148	-0.0322	0.0260	0.04	0.8344
RPUS	1	0.0569	0.0139	0.0297	0.0840	16.83	<.0001
RTKS	1	0.0038	0.0018	0.0004	0.0073	4.67	0.0307
RKFM	1	-0.0045	0.0040	-0.0124	0.0033	1.30	0.2546
PRTLH	1	0.0010	0.0054	-0.0097	0.0117	0.03	0.8519
PTKBNM	1	-0.0042	0.0058	-0.0156	0.0073	0.51	0.4753
PRTSTL	1	-0.0066	0.0028	-0.0121	-0.0012	5.73	0.0167
RLS	1	-0.2101	0.0336	-0.2760	-0.1443	39.07	<.0001
Dispersion	1	0.0317	0.0088	0.0144	0.0490		

Table 1 shows that the 5% significance level can be seen that the p-value of all the parameters is smaller than 0.05. so that the parameters intersept, RPUS, RTKS, PRTSTL and RLS have significant effect on the model. Then Negative Binomia regression models generated by:

$$\hat{\mu} = \exp(-1.5652 + 0.0569X3 + 0.0038X4 - 0.0066X8 - 0.2101X9$$

Predictor variables that influence the number of child cases in West Java is, present of puskesmas (RPUS), healt officer (RTKS), present sanitation not worthy (PRTSTL) and average of school duration (RLS).

Table 4. Value of Deviance and Pearson Poisson Regression

Criterion	DF	Value	Value/DF
Deviance	16	26.1555	1.6347
Scaled Deviance	16	26.1555	1.6347
Pearson Chi-Square	16	23.3100	1.4569
Scaled Pearson X2	16	23.3100	1.4569

Based on table 4 shows that the value of deviance / df and Pearson chi-square/df ≈ 1 the Negative Binomial regression can overcoming overdispersion and methods can accommodate the dispersion parameter.

CONCLUSION

Poisson regression model and Negative Binomial regression comparisonswas conducted to determine a better model used in modelling the number of cases of cervical cancer each district / city in West Java. The criteria for selection of the best model used is AIC. Best model is the model that has the smallest AIC value.

Table 5. Criteria For Assessing Goodness Of Fit

Criterion	Value
AIC (Poisson)	4638.7315
AIC (Negative Binomial)	449.8338

Based on AIC values in Table 5, the smallest AIC value is a negative binomial regression model. Then the best model for the number child mortality obtained from the Negative Binomial regression model. This suggests that the negative binomial regression model is more appropriate in the case overdispersion poisson regression. With the resulting model is as follows:

$$\mu = \exp\ (-0{,}0726+0{,}074X2+0{,}1604X6+0{,}0001X8)$$

Altough based on table 4 shows that the value of deviance / df and Pearson chi-square/df ≈ 1 the Negative Binomial regression can overcoming overdispersion. Need to develop other method such as Generalized Poisson Regression (GPR) and Quasi-Likelihood method.

REFERENCES

Bozdogan, H. (2000). *Akaike's Information Criterion and Recent Developments in Information Complexity*. Mathematical Psychology. 44. 62-91.

Cameron, A.C. dan Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Greene, W., *Functional Forms For The Negative Binomial Model For Count Data. Foundations and Trends in Econometrics*. Working Paper, Department of Economics, Stern School of Business, New York University, 2008: 585-590.

Hardin JW dan Hilbe JM. (2007). *Generalized Linier Models and Extensions*. Texas: A Strata Perss Publication.

McCullagh P, Nedler JA. 1983. *Generalized Linear Models*. London: Chapman and Hall.

McCulloch, Charles E, Shayle RS. 2001. *Genelralized, Linear and Mixed Models.* New York : John Wiley and Sons Inc.