Analysis Of Overdispersed Count Data By Poisson Model

Ismah[1], Anang Kurnia[2], Kusman Sadik[3]

[1]*Mathematics Education, Universitas Muhammadiyah Jakarta, Banten, Indonesia*
[2]*Statistics, Institut Pertanian Bogor, Bogor,Indonesia*

Corresponding author: anangk@apps.ipb.ac.id

Abstract: Lack assumption that commonly happens in Poisson model is over-dispersion. Over-dispersion is a condition in which the variance value is larger than mean of response variable. The aim of this research is to analyze Poisson models, i.e. Poisson Regression (POI), Zero-Inflated Poisson Regression (ZIP), Generalized Poisson Regression (GP) and Zero-Inflated Generalized Poisson Regression (ZIGP) of over-dispersion data. The data used in this research is Indonesian Demographic and Health Survey (SKDI) Data in 2017. Total number of 17.212 families with response variable of child mortality in these families become the objects of the study. The estimator of parameter model is Maximum likelihood estimator (MLE). The results analysis of those four models aforementioned above show that over-dispersion case causes the usage of POI model becomes less appropriate, while GP model can be used for over-dispersion case, however if the case of over-dispersion is caused by zero excess in the data, GP will be better than ZIP and ZIGP. It can be seen in the minimum of AIC value reached by each model through the data of SDKI with zero excess (having >50% of zero numbers), in which POI =13922, GP = 13578, ZIP = 13589 and ZIGP = 13588. Thus, it can be concluded that in over-dispersion data with zero excess (with big numbers of zero), ZIGP is less appropriate to be applied, because range of data is short (0-6).

Keywords: Poisson Model, Over-dispersion, *Maximum likelihood estimator*

1. INTRODUCTION

Finding the relationship between dependent variable and the independent one, or the influence of independent variable on dependent variable can be done through linier model, with the assumption that the response variable independently and identically distributed normally. However, dependent variable which is normally distributed is difficult to gain when facing real data. Thus, the alternative to accomplish the relationship or influence between independent and dependent variables if the dependent variable is not normally distributed, is through generalized linier modelling, by having larger assumption, i.e. identically and independently distributed dependent variable is resulted from exponential family.

In a case in which dependent variable is discrete data or count data, in a generalize linear modeling can be finished by using Poisson model. The assumption should be fulfilled in Poisson model is that the average and variance of dependent variable are similar. However, if the average value gained is smaller than the variance one then it is called over-dispersion. The consequence of over-dispersion data existence is including standard deviation of regression parameter which becomes smaller and resulted incorrect conclusion. Therefore, the application of Poisson model on the case of over-dispersion is less appropriate. Thus, another model needs to be applied; the one which is not sensitive when over-dispersion case exists.

The regression of Generalized Poisson (GP) is the development of Poisson regression which can be used to solve the problem of over-dispersion (Wang & Famoye, 1997), which was firstly introduced by Consul and Jain (1970). One of the causes of over-dispersion is big numbers of excess zeros, such as the data in total number of students who do not pass national examination, or total number of child mortality who suffered from tuberculosis, and many other examples. Zero Inflated Poisson (ZIP) which was proposed by Lambert (1992) can also be used as one of the methods to overcome over-dispersion with excess zeros in its dependent variable. ZIP then being developed into Zero Inflated Generalized Poisson (ZIGP) (Famoye and Singh, 2006).

There were numerous numbers of research conducted to deal with over-dispersion data with excess zeros, such as Yang et.al. (2009) and Y. N. Phang and E. F. Loh (2013) who applied model of Zero Inflated Poisson to test the data of over-dispersion, besides that, Hall et.al,. (2010) used regression of Zero Inflated Poisson as robust estimation. Another research related to another model implemented to solve the problem of over-dispersion data are Paola Zaninotto and Emanuela Falaschelli (2010) who compared 4 models, i.e. Poisson Model, Negative Binomial, Zero Inflated Poisson and Zero Inflated Negative Binomial, Hidayatul Fitriyah, et.al. (2016) who applied the method of Regression Binomial Negative and the approach of quasi-likelihood to analyze factors which influenced under-5 year child mortality rate in the province of West Java, Handayani, et.al. (2017) compare Laplace, Penalized Quasi likelihood (PQL) and Adaptive Gaussian Quadrature (AGQ) approximations, while Felix Farnoye and Karan P. Singh (2006) applied the regression of Zero Inflated Generalized Poisson on the data of domestic violence.

In this study, a research on Poisson was investigated including Poisson Regression (POI), GP, ZIP and ZIGP on the data of over-dispersion of child mortality in families in which the data were gained from the data of Indonesian Demographic and Health Survey (SKDI) in 2017. The purpose of analyzing those 4 models is to find out which model would be appropriate to be applied in analyzing excess zeros data of child mortality number in each family on SKDI data, 2017.

2. MATERIALS

Poisson Regression was used to modeling the counting data of dependent variable. While the function of Poisson distribution parameter μ is as follows:

$$P(Y = y|\mu) = \frac{e^{-\mu}\mu^{y}}{y!}$$

in which y = 0, 1, 2, ... and $\mu > 0$, and expectation $E(y_i) = \mu_i = \exp(x_i\beta)$, so that it is a model of Poisson regression (Long 1997 in Aji Setiawan, 2012).

$$\ln(\mu_i) = x_i\beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

β_j is the coefficient regression with $j = 1, 2, \ldots, k$, x_i is independent variable with $i = 1, 2, \ldots, n$, n refers to total number of observation and k refers to total number of independent variable.

For $Y \sim GP(\mu, \varphi)$ then there will be 3 probabilities, namely $Var(Y) > (=, <) E(Y) \Leftrightarrow \varphi > (=, <) 1$, so that it can be said that model can experience over-dispersion or under-dispersion. For the case of under-dispersion, a distribution contains μ and φ is needed, in which it is difficult when estimating μ and φ, and then in the context of regression, a functional connected is needed for φ which contains μ. The distribution of ZIGP is analogous with distribution of ZIP (Mullahy,1986) namely by adding parameter zero-inflation ω. Therefore, there are three parameters in ZIGP distribution which is notated as $Y \sim GP(\mu, \varphi, \omega)$. While the differences among those four models, i.e. Poisson Regression, GP, ZIP and ZIGP can be seen in the following model description table 1:

Table 1. Probability Density Function, Mean and Variance of Four Poisson Models of POI, GP, ZIP and ZIGP

Model	Criterion	
	Probability Density Function	Mean and Varians
POI	$P(Y_i = y_i) = \frac{e^{-\mu}\mu^y}{y!}$	$E(y) = Var(y) = \mu$
GP	$P(Y_i = y_i) = \left(\frac{\mu_i}{1+\omega\mu_i}\right)^{y_i} \frac{(1+\omega y_i)^{y_i-1}}{y_i!} e^{\left(-\frac{\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right)}$	$E(y) = \mu$; $Var(y) = \mu_i(1+\omega\mu_i)^2$
ZIP	$P(Y_i = y_i) = \begin{cases} \varphi_i + (1-\varphi_i)e^{-\mu_i}, y_i = 0 \\ (1-\varphi_i)\frac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!}, \quad y_i = 1, 2, \ldots, ; 0 \le \varphi_i \le 1 \end{cases}$	$E(y) = (1-\varphi_i)\mu$; $Var(y) = (1-\varphi_i)({\mu_i}^2+\mu_i) - (1-\varphi_i)^2{\mu_i}^2$
ZIGP	$P(Y_i = y_i) = \begin{cases} \varphi_i + (1-\varphi_i)e^{\frac{-\mu_i}{(1+\omega\mu_i)}}, y_i = 0 \\ (1-\varphi_i)\left(\frac{-\mu_i}{(1+\omega\mu_i)}\right)^{y_i} e^{\left(\frac{-\mu_i(1+\omega y_i)}{1+\omega y_i}\right)} \frac{(1+\omega y_i)^{y_i-1}}{y_i!}, \quad y_i = 1, 2, \ldots \end{cases}$	$E(y) = (1-\varphi_i)\mu$; $Var(y) =$

		$(1-\varphi_i)(\mu_i^2+\mu_i(1+\omega\mu_i)^2)-(1-\varphi_i)^2\mu_i^2$

Lamberg (1992) in Dewanti, Made Susilawati and I Gusti Ayu Made Srinadi (2016) said that if $\varphi_i = 0$ on probability density function ZIGP model then it will become GP model, and if $\omega = 0$ then it will become ZIP model. If φ_i has positive value then it will be presented as ZIGP model, while if φ_i has negative value then it will be presented as b zero deflation generalized Poisson, in which it rarely happens.

The estimation paramater of the four models of POI, GP, ZIP and ZIGP applied a method of maximum likelihood estimation (MLE). This MLE used a technique of maximizing the likelihood function. The likelihood function of those four models can be seen in the following table 2:

Table 2. Parameter Estimated (*Maximum Likelihood*) and Equation Modeling of POI, GP, ZIP and ZIGP Models

Model	Criterion	
	Parameter Estimated (*Maximum Likelihood*)	Equation Modeling
POI	$\frac{\prod_{i=1}^{n}\mu_i^{y_i}e^{\left(-\sum_{i=1}^{n}\mu_i\right)}}{\prod_{i=1}^{n}y_i!}$	$\ln(\mu_i) = x_i\beta$
GP	$\prod_{i=1}^{n}\left\{\left(\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}}}{1+\omega e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}}}\right)^{y_i}\frac{(1+\omega y_i)^{y_i-1}}{y_i!}e^{\left(-\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}}(1+\omega y_i)}{1+\omega e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}}}\right)}\right.$	$\ln(\mu_i) = x_i\beta$
ZIP	for $y_i = 0$ $\prod_{y_i=0}\frac{e^{(X_i^T\gamma)}+e^{\left(-e^{(X_i^T\beta)}\right)}}{\left(1+e^{(X_i^T\gamma)}\right)}$ for $y_i > 0$ $\prod_{y_i>0}\frac{e^{\left(-e^{(X_i^T\beta)}\right)}\left(e^{(X_i^T\beta)}\right)^{y_i}}{\left(1+e^{(X_i^T\gamma)}\right)y_i!}$	$\ln\mu_i = \beta_0+\sum_{j=1}^{k}x_{ij}\beta_j$ $; i = 1, 2, \dots, n$ $\text{logit}\,\varphi_i = \gamma_0+\sum_{j=1}^{k}x_{ij}\gamma_j$ $; i = 1, 2, \dots, n$
ZIGP	for $y_i = 0$ $\prod_{y_i=0}\frac{1}{1+\mu_i^{-\tau}}+\left(\mu_i^{-\tau}+e^{\frac{-\mu_i}{1+\omega\mu_i}}\right)$ for $y_i > 0$	$\ln\mu_i = \beta_0+\sum_{j=1}^{k}x_{ij}\beta_j$ $; i = 1, 2, \dots, n$

	$\prod_{y_i>0} \frac{1}{1+\mu_i{}^{-\tau}} \left(\frac{-\mu_i}{1+\omega\mu_i}\right)^{y_i} \frac{(1+\omega y_i)^{y_i{}^{-1}}}{y_i} e^{\frac{-\mu_i(1+\omega y_i)}{1+\omega\mu_i}}$	$\text{logit}\,\varphi_i = \log\left(\frac{\varphi_i}{1+\varphi_i}\right) = -\tau \sum_{j=1}^{k} x_{ij}\beta_j \; ; i = 1, 2, \dots, n$

Significance test towards parameter of each model used Wald test with the following equation:

$$W = \left(\frac{\hat{\beta}_k}{SE\ (\beta_k)}\right)^2$$

$\hat{\beta}_k$ refers to MLE of parameter β_k and $SE\ (\beta_k)$ is standard error for β_k.

The selection of goodness of fit model was based on the value of Akaike's Information Criterion (AIC) with the following formula:

$$AIC = -2 \log L(\hat{\beta}) + 2k$$

In which $L(\hat{\beta})$ refers to likelihood value and k refers to total number of parameters. The criteria to choose the best model based on the lowest value of AIC or close to zero.

3. METHODS

The data used in this research was total number of child mortality in families as dependent variable (Y), while the independent variable consisted of mother's education level (X_1), father's education level (X_2), types of contraception used (X_3), place of delivering the baby (X_4), wealth index combined (X_5) and marital status (X_6). The data were taken from Indonesian Demographic and Health Survey (SKDI), year 2017. The decision of having both dependent and independent variables were based on the research done by Lin Dai et. al. (2018) who modeled the excess zeros and the heterogeneity of counting data with complex survey design applied for the survey of demographic health in sub-Sahara Afrika.

The following table 3 described the data used in this research:

Table 3. The Description of Research Variables

No.	Variable	Description
1	Child mortality (Y)	0 = No child died
2	Mother's education level (X_1)	0 = no education 1 = primary 2 = secondary 3 = higher
3	Husband/partner's education level (X_2)	0 = no education 1 = primary 2 = secondary 3 = higher 4 = don't know
4	Contraceptive used and intention (X_3)	1 = using modern method 2 = using traditional methods

		3 = Non-user- intend to use later 4 = does not intend to use 5 = Never had sex
5	Place of delivery (X_4)	10 = home 11 = respondent's home 12 = other home 20 = public sector 21 = goverment hospital 22 = goverment clinic 23 = goverment health centre 26 = other public sector 30 = private sector 31 = private hospital/clinic 37 = other private sector 96 = other
6	Wealth index combined (X_5)	1 = poorest 2 = poorer 3 = middle 4 = richer 5 = richest
7	Current marital status (X_6)	1 = married 2 = single

The steps of conducting the research consisted of:

1. Screening the data of SDKI data, 2017
2. Doing over-dispersion test, if the value of deviance or chi square value divided by degree of freedom is more than one, then the data experinces over-dispersion, if the data do not experience any over-dispersion then that data will only be analyzed through POI
3. Estimating the parameter on the four models
4. Having goodness of fit on the four models, based on the value of AIC or using vuong.

4. RESULTS AND DISCUSSION

The data were analyzed by using software of R with packages ZIGP. Child mortality in families (y) had the minimum value of 0 and the maximum of 6, with total number of 17212 observation after screening process by erasing the missing data, then the data were grouped based on the families. Excess zeros of the dependent variable was 89% of the observation number or equal with 15364 observations. Data description can be seen in the following table 4:

Table 4. Data Description

Variable	MIN	Max	Percentage		Variance
			MIN	Max	
Y	0	6	89.16%	0.01%	0.180155
X_1	0	3	1.52%	17.73%	0.478266
X_2	0	9	1.60%	15.11%	0.60766
X_3	1	4	63.47%	9.77%	1.161213
X_4	11	96	26.52%	0.17%	90.74011
X_5	1	5	30.13%	15.61%	2.093489
X_6	1	2	98.04%	1.52%	0.019253

Table 4 above presented total number of mothers who did not have any education ($X_1 = 0$) in Indonesia as noted in the data of SDKI 2017 were 1.52% or 262 moms, meanwhile, mothers who had the highest education ($X_1 = 3$) were 14.95% or 3051 persons. And the number of fathers who did not have any education ($X_2 = 0$) were more than mother, namely 276 and those who had the highest education were 15.11% or 2600 persons.

Mothers who used contraception with modern technique ($X_3 = 1$) were high enough, i.e. 63.47%, while those who did not use any contraception ($X_3 = 4$) were 9.77%. For delivery place, mothers who delivered the baby at their own houses ($X_4 = 11$) were quite many, i.e. 26.52%, while the option for other places of delivering the baby not listed in the answer ($X_4 = 96$) was only 0.17%, or in other words, Mothers' places of delivery were almost all in the list of the answer.

Wealth index combined ($X_5 = 1$) was high enough, namely 30.13%, while the number of the richest category ($X_5 = 5$) was half of the poorest category, i.e. 15.61%. The last variable was marital status of the respondents when being interviewed; those who admitted that they were married ($X_6 = 1$) were 98.04%, while those who were single ($X_6 = 2$) because they did not marry yet or they were divorced were 1.52%.

The results of data analysis through regression Poisson (POI) which were analyzed by using software R can be seen below:

Table 5. POI Model

MU REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)	
b0	Intercept	-2.65073	0.21791	-12.1641	<2e-16	***
b1	X_1	-0.61859	0.08699	-7.1115	1.15E-12	***
b2	X_2	-0.28668	0.11979	-2.3932	0.0167	*
b3	X_3	-0.49046	0.08497	-5.7724	7.82E-09	***
b4	X_4	-0.02215	0.00264	-8.3779	<2e-16	***
b5	X_5	5.15602	0.37491	13.7527	<2e-16	***
b6	X_6	0.59286	0.1671	3.5481	0.0004	***

Based on the results gained through POI, it was known that all independent variables were significant with $\alpha = 5\%$.

ISSN 2515-8260 Volume 07, Issue 10, 2020

The analysis results with ZIP model divided the observation into two models. The first model used to determine the probability of dependent variable was having zero value which was then called as logit model, while the second model was the one used to determine the probability of dependent variable of an observation (Long 1997 in Aji Setiawan 2012). Data analysis by using ZIP resulted the following data:

Table 6. ZIP Model

MU REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)	
b0	Intercept	-1.41467	0.54492	-2.5961	0.0094	**
b1	X_1	-0.2618	0.19019	-1.3765	0.1687	
b2	X_2	-0.83016	0.26085	-3.1826	0.0015	**
b3	X_3	-0.02032	0.1849	-0.1099	0.9125	
b4	X_4	-0.02275	0.00625	-3.6392	0.0003	***
b5	X_5	2.88076	0.84224	3.4203	0.0006	***
b6	X_6	0.7468	0.44848	1.6652	0.0959	.
OMEGA REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)	
g0	Intercept	0.94123	0.93788	1.00357	0.3156	
g1	X_1	0.53838	0.28476	1.89063	0.0587	.
g2	X_2	-0.90238	0.39338	-2.29393	0.0218	*
g3	X_3	0.74936	0.28489	2.6304	0.0085	**
g4	X_4	-0.00154	0.00931	-0.16528	0.8687	
g5	X_5	-3.71411	1.2512	-2.96844	0.003	**
g6	X_6	0.26125	0.82545	0.31649	0.7516	

Independent variables of X_2, X_4, and X_5 influenced dependent variable (Y) significantly with $\alpha = 5\%$, and insignificantly for variables X_1, X_3 dan X_6. On the second model, a slightly different data analysis than the first model was found in which the significant independent variables were X_2, X_3, and X_5, and the insignificant ones were variables X_1, X_4 and X_6.

GP analysis achieved the following results:

Table 7. GP Model

MU REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)	
b0	Intercept	-2.68656	0.23904	-11.2389	<2e-16	***
b1	X_1	-0.62006	0.09492	-6.5326	6.46E-11	***
b2	X_2	-0.18705	0.13095	-1.4284	0.1532	
b3	X_3	-0.52512	0.0927	-5.6646	1.47E-08	***
b4	X_4	-0.01978	0.00285	-6.9295	4.22E-12	***
b5	X_5	5.06549	0.40873	12.3932	<2e-16	***
b6	X_6	0.57991	0.18443	3.1443	0.0017	**
PHI REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)	
a0	Intercept	-2.27813	0.72756	-3.13117	0.0017	**
a1	X_6	0.31065	0.74727	0.41571	0.6776	

It was found that all variables were significant except X_2 at $\alpha = 5\%$. While data analysis with ZIGP showed that all variables were not significant, as can be seen in table 8 below:

Table 8. ZIGP Model

MU REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)
b0	Intercept	-2.57398	2.35176	-1.0945	0.2737
b1	X_1	-0.59557	0.46785	-1.273	0.203
b2	X_2	-0.24314	0.58662	-0.4145	0.6785
b3	X_3	-0.49016	0.42497	-1.1534	0.2487
b4	X_4	-0.02015	0.01657	-1.2159	0.224
b5	X_5	5.17759	2.66837	1.9404	0.0523 .
b6	X_6	0.6534	2.23636	0.2922	0.7702
PHI REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)
a0	Intercept	-2.14239	1.55015	-1.38205	0.167
a1	X_6	0.00002	1.58984	0.00001	1
OMEGA REGRESSION		Estimate	Std. Error	z value	Pr(>|z|)
g0	Intercept	-1.46249	12.36414	-0.11828	0.9058
g1	X_1	0.00001	2.33836	0.00001	1
g2	X_2	-0.00001	2.90144	0	1
g3	X_3	0.00002	2.11097	0.00001	1
g4	X_4	0.00006	0.08368	0.00071	0.9994
g5	X_5	0.00001	13.61928	0	1
g6	X_6	0.00003	11.83979	0	1

Those four models were tested further to determine the criteria of goodness of fit model based on AIC value and the analysis results with vuong at R.

Table 9. Goodness of fit model of POI, GP, ZIP and ZIGP

	POI	GP	ZIP	ZIGP
POI	AIC = 13922		Favour model 2. P-value: 2.83e-12	
GP	Favour model 1. P-value: 0.0004	AIC = 13578	Favour model 1. P-value: <2e-16	
ZIP			AIC = 13589	
ZIGP				AIC = 13588

Among those four models, it was found that GP model was better than the other three models with the value of AIC 13578. In the research, an alternative model was previously conducted to change the variations from the function of i μ, φ and ω, but the model applied was the best one based on achieved AIC value compared to the other alternative models.

The conclusion that can be explained was that excess zeros on the data did not mean that the model of zero inflated would become the appropriate solution to accomplish it, the range of counting data was also needed to be observed in the dependent variable; child mortality data in families (Y) had the range of 0 until 6, which gave the interpretation that most number of child mortality being observed was 6. That range was considered having very small differences so that the used of zero inflated model was less appropriate. As the real example in McCulagh (1989) which found that range numbers of broken ships due to sea-wave were between the data of 0 to 40.000, and this range was big enough to apply zero inflated as the appropriate solution compared to POI or GP model.

5. REFERENCES

[1] Consul, P. C. and G. C. Jain (1970). On the generalization of Poisson distribution. *Ann. Math. Statist*. 41, 1387.

[2] Dewanti N. P. P., Made. S and I Gusti A. M. S. (2016). Perbandingan Regresi Zero Inflated Poisson (ZIP) Dan Regresi Zero Inflated Negative Binomial (ZINB) Pada Data Overdispersion (Studi Kasus: Angka Kematian Ibu Di Provinsi Bali). *E-Jurnal Matematika*. Vol. 5 (4), November 2016, pp. 133-138

[3] Felix Famoye and Karan P. Singh. (2006). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science* **4**(2006), 117-130

[4] Hidayatul Fitriyah, Anang Kurnia, Mochamad Farit Afendi (2016). Metode Regresi Binomial Negatif Dan Pendekatan *Quasi-Likelihood* Untuk Menganalisis Faktor-Faktor Yang Mempengaruhi Jumlah Kematian Balita Di Jawa Barat. Tesis. Institut Pertanian Bogor.

[5] Hall, D. B. and Shen, J. (2010). Robust Estimation for Zero-Inflated Poisson Regression, *Scandinavian Journal of Statistics*, Volume 37, Issue 2, 237–252.

[6] Ismail, N., dan Jemain, A. A, (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. *Casualty Actuarial Society Forum*: 103-158.

[7] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34, 1-14.

[8] Lin Dai1, Michael D Sweat2, and Mulugeta Gebregziabher, 2018. Modeling excess zeros and heterogeneity in count data from a complex survey design with application to the demographic health survey in sub-Saharan Africa. *Stat Methods Med Res*. 2018 January; 27(1): 208–220. doi:10.1177/0962280215626608

[9] Mullahy, J. (1986) Specification and Testing of Some Modified Count Data Models. *Journal of Econometrics*, 33, 341-365. https://doi.org/10.1016/0304-4076(86)90002-3

[10] Mccullagh p, nelder ja. (1989). *Generalized linear models. 2nd edition*. Chapman & Hall, London.

[11] Paola Zaninotto, Emanuela Falaschetti. Comparison of methods for modelling a count outcome with excess zeros: an application to Activities of Daily Living (ADL-s). *Journal of Epidemiology and Community Health, BMJ Publishing Group*, 2010, 65 (3), pp.205. ff10.1136/jech.2008.079640ff. ffhal00563516f

[12] Settyawan A. (2012). Perbandingan Antara Regresi Poisson, Binomial Negatif, dan Zero Inflated Poisson pada Data Overdispersi. *Skripsi*. Institut Pertanian Bogor. Bogor

[13] Yang, Z., Hardin, J W., and Addy, C. L. (2009). Testing Overdispersion In The Zero-Inflated Poisson Model. *Journal of Statistical Planning and Inference*, Volume 139, Issue 9, 3340–3353.

[14] Wang, W. & Famoye, F. 1997. Modeling household fertility decision with generalized Poisson regression. *Journal of Population Economics*, 10, 3, 273- 283.