Zero inflated poisson regression analysis on imb ownership in Sidoarjo regency 2019

H Permadi, S D Maulidah
Departement of Mathematics, Faculty of Mathematics and Natural Science, Universitas Negeri Malang, Indonesia

Corresponding author: hendro.permadi.fmipa@um.ac.id,
Email: sucidindamaulidah@gmail.com

Abstract. The ownership of IMB is caused by several factors that can be seen by statistical analysis. However, in this case, the data is Poisson distribution so the modeling uses Poisson regression. The alternative to solve the overdispersion case is Zero Inflated Poisson Regression. This study uses the number of data in each village in Sidoarjo Regency with 8 factors that are suspected to influence it, namely the number of standing buildings larger 40th (X1), Land Certificate (X2), Knowledge of IMB (X3), Knowledge of Sanctions (X4), Costs (X5), Lazy or Convoluted Bureaucracy (X6), and the Feeling of Calm in owning a building (X7), and Participation in the IMB Socialization Program (X8). The results show that the IMB ZIP Ownership Regression Model is as follows:  

\[
\log(\mu_i) = 5.470 + 2.831 \times 10^{-4}X_1 - 8.976 \times 10^{-6}X_2 + 1.356 \times 10^{-4}X_3 + 2.530 \times 10^{-4}X_4 \]

\[
\logit(w_i) = 0.652 - 0.000244X_2 + 0.00715X_3 - 0.00096X_4 - 0.00155X_6 . \]

The models that most influence the ownership of the IMB is the Land Certificate (X2), Knowledge of the IMB (X3), Knowledge of Sanctions (X4), Lazy or Convoluted Bureaucracy (X6). 

1. Introduction
The three pillars that play a role in development include government, society, and the private sector. There are several development activities that are the responsibility of the government, but there are also development activities that are responsible / handed over to the community, in this case, a development whose authority is the responsibility of the community itself including housing development. Therefore, for housing development to remain orderly and controlled, and directed in accordance with the development policy of the Sidoarjo regency government, it is deemed necessary to issue a policy which is realized in the form of Sidoarjo regency number 4 of 2012 concerning licensing in building construction (IMB).

The ownership of IMB is caused by several factors that can be seen by statistical analysis. In applied statistics, there is a method to find out the relationship between the response variable and the independent variable, namely regression analysis. The relationship between these variables is expressed in a regression equation and in a linear or non-linear form, this linear relationship is called a linear regression equation.
Linear regression equations are used to analyze the dependent variable in the form of continuous random variables with normal distribution whether there is a relationship or correlation with other variables, but many cases of response or dependent variables that use random variables but follow the Poisson distribution. If the response variable follows the Poisson distribution, then the relationship of the response variable and the independent variable informing the model uses the Poisson regression model (Myers, 2010). But in reality, also the Poisson regression assumptions are not fully met. Many data are found that the value of the variance is smaller (underdispersion) or greater than the mean value (Overdispersion). Data containing Overdispersion produces a smaller standard error value than the original value which can cause invalidation [7].

The problem of underdispersion or overdispersion in Poisson Regression is usually caused by a large number of nulls in the dependent variable (excess zeros) and if overdispersion data then the Poisson regression method is not suitable for completing the data because GLM models such as Poisson regression can produce underestimate estimators so that it will be difficult to identify significant parameters. Research on underdispersion or overdispersion in the Poisson Regression in its completion has different assumptions. To overcome this problem, Zero-Inflated Poisson (ZIP) Regression is used [4].

Adeliana [2] conducted an overdispersion test on Poisson regression by applying Zero-Inflated Poisson Regression (ZIP) about factors that affect the number of cases of tetanus neonatorum in East Java in 2014. Several Zero-Inflated Poisson Regulations (ZIP) to address overdispersion data such as in Elizabeth's research (2017) about the factors that influence maternal mortality rates in East Java Province in 2015.

Based on these problems and previous studies, the authors are interested in analyzing ownership data of IMB in Sidoarjo Regency in 2019 using the Zero-Inflated Poisson (ZIP) method. The data is thought to have overdispersed due to a large number of values 0. So if using the Poisson regression model it is less appropriate. Zero-Inflated Poisson (ZIP) can also find out the factors that influence IMB ownership in Sidoarjo Regency in 2019.

Lambert Diane's experiment [5] regarding soldering defects in printed wiring boards, two sets of conditions give nearly the same number of defects, but the perfect state is more likely under one set of conditions and the average number of defects in an imperfect state is more, small under another set of conditions. That is, the ZIP regression can show not only which conditions give the lower mean number of defects but also why the means are lower.

According to Ruliana [8], the ZIP distribution is two combinations of the Poisson-Bernoulli distribution. If \( Y_i \) is the dependent variable with ZIP distribution, then the zero value in the observational data is thought to arise because of two ways. The first way is zero states with probability \( w_i \), the second way is the Poisson state with probability \( (1 - w_i) \) and the Poisson distribution with the average value is \( \lambda_i \). These two distribution process will produce the probability density function as follows:

\[
Pr = (Y_i = y_i) = f(y) = \begin{cases} 
  w_i (1 - w_i) e^{-\mu_i}, & \text{for } y_i = 0 \\
  (1 - w_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, & \text{for } y_i > 0, 0 \leq w_i \leq 1 
\end{cases}
\]

The ZIP distribution is usually written as \( Y \sim ZIP (\mu, w) \) where \( \mu \) is the Poisson distribution and \( w \) is the probability of occurrence of zero values.

The method for estimating parameters of the Zero Inflated Poisson (ZIP) regression model is MLE (Maximum Likelihood Estimation). If the probability density function of regression is known, this method is suitable for estimating the parameter values of a model.

\[
Log L(\beta, \gamma | y_i) = log \prod_{i=1}^{n} p(y_i; \beta, \gamma)
\]
\[ Log L(\beta \mid y_i) = \log \left\{ \prod_{i=1}^{n} \left( w_i + (1 - w_i) \exp (-\mu_i) \right) \right\} \]

\[ Log L(\beta \mid y_i) = \sum_{i=0}^{n} \left\{ \log \left[ \left( w_i + (1 - w_i) \exp (-\mu_i) \right) \right] y_i = 0 \right\} \]

2. Method

This research is a secondary data that works together with 9 CV, namely CV. Karya Mandiri, CV. Dhiratama Cipta persada, CV. Adhi Loka Consultant, PT. Parliament Tiga-Tiga, CV. Riptaloka Konsultan, CV. Adhirajasa Ciptana Engineering, CV. Tegar Teknik Utama, CV. Karya Karya, PT. Global Prisma Konsulindo.

Data collection techniques using incidental sampling means anyone who accidentally met in the residential area of Sidoarjo Regency. This study takes a minimum of 5 respondents from each village then analyzes the data obtained. The observations and results of the questionnaire were analyzed quantitatively both descriptive and inferential.

The observed variables are the ownership of IMB, the number of buildings standing bigger than 40 years, the Land Certificate, Knowledge of IMB, Knowledge of Sanctions, Costs, Lazy or Convoluted Bureaucracy, and Feeling of Ownership of the building, and Participation of IMB Socialization Program. The software used in this analysis is SPSS, Minitab, and Rstudio.

The stage of analyzing data is as follows:
1. Determine the Statistics Description of Each Variable
2. Determine Partial Correlations between Free and Bound Variables
3. Distribution Test
4. Multicollinearity Test
5. Equidispersion Testing
6. Estimating Zero Inflated Poisson (ZIP) Parameters
7. Test the regression coefficient simultaneously
8. Test the Regression coefficient Individually
9. Determine the suitability of the model
10. Interpret the model

3. Research Result

3.1 Determine the Value of Partial Correlation between the Independent and Bound Variables

Poisson regression is useful for knowing the correlation of the dependent variable in the form of discrete and discrete, continuous, or mixed independent variables. According to Aggression, the Poisson regression model is also a GLM (Generalized Linear Model) where the dependent variable is assumed to have a Poisson distribution. The Generalized Linear Model (GLM) is an extension of the general regression model for the response variable in the form of an exponential distribution. There is a function \( g \) in linear GLM that connects the average dependent variable with the independent variable. GLM Models as a follow

\[ g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} \]

where \( i \) is the observation and \( j \) is the variable with the following information
$x_{ij} = $ observation of the data i on variable j

$k = $ number of variables

$\beta_j = $ parameter whose value is unknown but must be estimated from the data

### Table 1. Correlation Results.

| No | Item Question                                      | Value | Sig                  |
|----|----------------------------------------------------|-------|----------------------|
| 1  | Standing Buildings > 40 years (X1)                 | 0.302 | Not Significance     |
| 2  | Land Certificate (X2)                              | 0.000 | Significance         |
| 3  | IMB knowledge and changes (X3)                     | 0.002 | Significance         |
| 4  | Knowledge of Sanctions (X4)                        | 0.000 | Significance         |
| 5  | Cost (X5)                                          | 0.945 | Not Significant      |
| 6  | Lazy, Convoluted Bureaucracy (X6)                 | 0.000 | Significance         |
| 7  | Feeling Calm (X7)                                 | 0.003 | Significance         |
| 8  | IMB Participation in Socialization (X8)            | 0.011 | Significance         |

Partial Correlation Test Results for each question with the ownership of IMB in the residential area of Sidoarjo Regency in 2019 are Land Certificate (X2), IMB Knowledge and changes (X3), Knowledge of IMB Sanctions (X4), Lazy, Convoluted Bureaucracy (X6) and Participation in the Socialization of IMB (X7) is said to be significant because of the value of sig (0.000) < 0.05. The conclusion of this significance test is that there is a relationship between each of these variables.

### 3.2 Distribution Test

### Table 2. Distribution tes of IMB Ownership Variables.

| No | Distribution   | Parameter       | Anderson Darling | Statistic | Rank |
|----|----------------|-----------------|------------------|-----------|------|
| 1  | Poisson        | $\lambda = 184.96$ | 48.755           | 1         |
| 2  | D.Uniform      | $a = -262 b = 631$ | 98.737           | 2         |
| 3  | Geometric      | $\rho = 0.00538$  | 340.01           | 3         |

The Poisson distribution is determined by a single parameter $\mu$ wherein the incidence rate is the average rare event per one exposure. Exposure can be time, space, distance, area, volume, or population size. So for a period of time, the symbol $t$ represents exposure and when no exposure value is given, it is considered a parameter so that the event probability $y$ is as follows:

$$
Pr = (Y = y|\mu, t)) = \frac{e^{-\mu t}(\mu t)^y}{y!} \quad y = 0,1,2, ...$$

Based on Table 2. Test Results Distribution of Variable Ownership of IMB Sidoarjo Regency is poisson distribution because the rank or sequence of data using easyfit software which occupies the highest is the Poisson distribution.

### 3.3 Multicollinearity Test

### Table 3. Multicollinearity Test Results.

| No | Item Question                                      | VIF  |
|----|----------------------------------------------------|------|
| 1  | Standing Buildings > 40 years (X1)                 | 1.448|
| 2  | Land Certificate (X2)                              | 1.734|
| 3  | IMB Knowledge and changes (X3)                     | 1.831|
| 4  | Knowledge of Sanctions (X4)                        | 1.333|
| 5  | Cost (X5)                                          | 1.429|
| 6  | Lazy, Convoluted Bureaucracy (X6)                 | 1.282|
| 7  | Feeling Calm (X7)                                 | 10.408|
| 8  | IMB Participation in Socialization (X8)            | 9.263|
Based on the calculation results, it is known that the multicollinearity calculation between variables meets the specified criteria, namely VIF value <10, the decision taken is to accept and reject. The conclusion is that between independent variables there is a problem of multicollinearity in the variable feeling of calm (X7)

3.4 Equidispersion Test

The problem of underdispersion / Overdispersion in poisson regression can be detected by looking at the variance and average values of the response variables and with the Poisson regression model deviance value divided by the degree of freedom, if the division is greater than 1, it can be said to occur overdispersion of the data and if the division is smaller than 1, it can be said to occur underdispersion of the data. Of the 6 independent variables used, the possible model is 11. The model is presented in the following table.

| No | Model     | Estimated Parameter Value | Devians |
|----|-----------|--------------------------|---------|
| 1. | \( X_1 \) | \( \beta_0 + \beta_1 \times 10^{-4} \) | 119784  |
| 2. | \( X_2 \) | \( \beta_0 + \beta_2 \times 10^{-4} \) | 114600  |
| 3. | \( X_3 \) | \( \beta_0 + \beta_3 \times 10^{-4} \) | 117244  |
| 4. | \( X_4 \) | \( \beta_0 + \beta_4 \times 10^{-4} \) | 115304  |
| 5. | \( X_5 \) | \( \beta_0 + \beta_5 \times 10^{-4} \) | 120163  |
| 6. | \( X_6 \) | \( \beta_0 + \beta_6 \times 10^{-4} \) | 116242  |
| 7. | \( X_1 + X_2 \) | \( \beta_0 + \beta_1 + \beta_2 \times 10^{-4} \) | 113451  |
| 8. | \( X_1 + X_2 + X_3 \) | \( \beta_0 + \beta_1 + \beta_2 + \beta_3 \times 10^{-4} \) | 112598  |
| 9. | \( X_1 + X_2 + X_3 \) | \( \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 \times 10^{-4} \) | 110608  |
| 10. | \( X_1 + X_2 + X_3 + X_4 \) | \( \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 \times 10^{-4} \) | 109095  |
| 11. | \( X_1 + X_2 + X_3 \) | \( \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 \times 10^{-4} \) | 107652  |

From Table 4, there are 11 models that are formed that are suitable to be used with each deviance where the model group consists of groups of models with 1 independent variable, 2 independent variables, 3 independent variables, 4 independent variables, and 5 independent variables.

| No | Model                      | Devians | dB  | Devians/Db |
|----|---------------------------|---------|-----|------------|
| 1. | \( \exp(\beta_0 + \beta_1) \) | 119764  | 351 | 341.21     |
| 2. | \( \exp(\beta_0 + \beta_2) \) | 114600  | 351 | 326.50     |
| 3. | \( \exp(\beta_0 + \beta_3) \) | 117244  | 351 | 334.03     |
| 4. | \( \exp(\beta_0 + \beta_4) \) | 115304  | 351 | 328.50     |
| 5. | \( \exp(\beta_0 + \beta_5) \) | 120163  | 351 | 342.34     |
| 6. | \( \exp(\beta_0 + \beta_6) \) | 116242  | 351 | 331.17     |
| 7. | \( \exp(\beta_0 + \beta_1 + \beta_2) \) | 113451  | 350 | 324.14     |
| 8. | \( \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3) \) | 112598  | 349 | 322.63     |
| 9. | \( \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4) \) | 110608  | 348 | 317.84     |
| 10. | \( \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5) \) | 109095  | 347 | 314.39     |
| 11. | \( \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6) \) | 107652  | 346 | 311.13     |

From Table 5 the value of the deviance is divided by the degree of freedom greater than 1. So that overdispersion results in a poisson regression model that is less appropriate if used to model the IMB Ownership data in Sidoarjo district. By using Zero-Inflated Poisson Regression (ZIP), it is expected to be able to overcome and improve the case of overdispersion with lots of zero data.
3.5 Modeling Zero-Inflated Poisson Regression (ZIP)

**Table 6.** Estimated Zero-Inflated Poisson Regression Parameters (ZIP).

| No  | Model | $\beta_0$  | $\beta_1$  | $\beta_2$  | $\beta_3$  | $\beta_4$  | $\beta_5$  | $\beta_6$  |
|-----|-------|------------|------------|------------|------------|------------|------------|------------|
| 1.  | $X_1$ | 5.968      | 4.310^{-4} | 0          | 0          | 0          | 0          | 0          |
| 2.  | $X_2$ | 5.566      | 3.410^{-4} | 0          | 0          | 0          | 0          | 0          |
| 3.  | $X_3$ | 5.858      | 2.810^{-4} | 0          | 0          | 0          | 0          | 0          |
| 4.  | $X_4$ | 5.906      | 0          | 0          | 2.310^{-4} | 0          | 0          | 0          |
| 5.  | $X_5$ | 5.976      | 0          | 0          | 0          | 7.410^{-4} | 0          | 0          |
| 6.  | $X_6$ | 5.816      | 0          | 0          | 0          | 0          | 3.10^{-4}  | 0          |
| 7.  | $X_1 + X_2$ | 5.571  | -1.110^{-4} | 3.410^{-4} | 0          | 0          | 0          | 0          |
| 8.  | $X_1 + X_2 + X_3$ | 5.568  | 2.010^{-4}  | 2.910^{-4} | 1.410^{-4} | 0          | 0          | 0          |
| 9.  | $X_1 + X_2 + X_3 + X_4$ | 5.532  | 2.510^{-5}  | 2.910^{-4} | 9.310^{-4} | 1.310^{-4} | 0          | 0          |
| 10. | $X_1 + X_2 + X_3 + X_4 + X_5$ | 5.515  | -1.710^{-4} | 3.610^{-4} | 1.710^{-4} | 3.810^{-4} | -3.910^{-4}| 0          |
| 11. | $X_1 + X_2 + X_3 + X_4 + X_5 + X_6$ | 5.458  | 3.310^{-4}  | 4.910^{-4} | 7.610^{-4} | -2.810^{-4} | 2.210^{-4} | 0          |
| 12. | $X_2 + X_3$ | 5.575  | 2.910^{-4}  | 1.310^{-4} | 0          | 0          | 0          | 0          |
| 13. | $X_2 + X_3 + X_4$ | 5.541  | 3.10^{-4}   | 9.10^{-5}  | 1.210^{-4} | 0          | 0          | 0          |
| 14. | $X_2 + X_3 + X_4 + X_5$ | 5.470  | 2.810^{-4}  | -8.910^{-4} | 1.410^{-4} | 0          | 2.510^{-4} | 0          |

| No  | Model | $Y_0$  | $Y_1$  | $Y_2$  | $Y_3$  | $Y_4$  | $Y_5$  | $Y_6$  |
|-----|-------|--------|--------|--------|--------|--------|--------|--------|
| 1.  | $X_1$ | -0.025 | 3.610^{-4} | 0      | 0      | 0      | 0      | 0      |
| 2.  | $X_2$ | 0.515  | -3.10^{-4} | 0      | 0      | 0      | 0      | 0      |
| 3.  | $X_3$ | 0.246  | 0      | -2.710^{-4} | 0      | 0      | 0      | 0      |
| 4.  | $X_4$ | 0.391  | 0      | 0      | -1.10^{-4} | 0      | 0      | 0      |
| 5.  | $X_5$ | 0.115  | 0      | 0      | 0      | 1.910^{-4} | 0      | 0      |
| 6.  | $X_6$ | 0.236  | 0      | 0      | 0      | 0      | -2.10^{-4} | 0      |
| 7.  | $X_1 + X_2$ | 0.365  | 4.810^{-4} | -4.10^{-4} | 0      | 0      | 0      | 0      |
| 8.  | $X_1 + X_2 + X_3$ | 0.396  | 4.710^{-4} | -4.10^{-4} | -1.310^{-4} | 0      | 0      | 0      |
| 9.  | $X_1 + X_2 + X_3 + X_4$ | 0.465  | 4.810^{-4} | -0.00033 | 1.710^{-4} | -1.10^{-3} | 0      | 0      |
| 10. | $X_1 + X_2 + X_3 + X_4 + X_5$ | 4.810^{-1} | 4.710^{-4} | -4.110^{-4} | 3.910^{-4} | -8.910^{-4} | 6.310^{-4} | 0      |
| 11. | $X_1 + X_2 + X_3 + X_4 + X_5 + X_6$ | 0.509  | 4.810^{-4} | -4.10^{-4} | 9.110^{-4} | -9.10^{-4} | 6.10^{-4} | -1.10^{-4} |
| 12. | $X_2 + X_3$ | 0.550  | 0      | -3.10^{-4} | -2.10^{-4} | 0      | 0      | 0      |
| 13. | $X_2 + X_3 + X_4$ | 0.616  | 0      | -3.10^{-4} | 1.210^{-4} | -1.10^{-3} | 0      | 0      |
| 14. | $X_2 + X_3 + X_4 + X_5$ | 0.652  | 0      | -2.10^{-4} | 1.710^{-4} | 1.10^{-3}  | 0      | -4.10^{-4} |

**Table 7.** $G_{hitung}$ and AIC ZIP Model.

| No  | Model | Log-Likelihood | $G_{hitung}$ | AIC  |
|-----|-------|----------------|---------------|------|
| 1.  | $X_1$ | -10890         | 21780         | 121028 |
| 2.  | $X_2$ | -9465          | 18930         | 115863 |
| 3.  | $X_3$ | -10150         | 20300         | 116507 |
| 4.  | $X_4$ | -10590         | 21180         | 116567 |
| 5.  | $X_5$ | -10890         | 21780         | 121426 |
| 6.  | $X_6$ | -9689          | 19378         | 117505 |
| 7.  | $X_1 + X_2$ | -9462  | 18924         | 114716 |
| 8.  | $X_1 + X_2 + X_3$ | -9311  | 18622         | 113865 |
| 9.  | $X_1 + X_2 + X_3 + X_4$ | -9233  | 18466         | 111877 |
| 10. | $X_1 + X_2 + X_3 + X_4 + X_5$ | -8956  | 17912         | 110366 |
| 11. | $X_1 + X_2 + X_3 + X_4 + X_5 + X_6$ | -8473  | 16946         | 108925 |
| 12. | $X_2 + X_3$ | -9315  | 18630         | 114630 |
| 13. | $X_2 + X_3 + X_4$ | -9238  | 18476         | 112581 |
| 14. | $X_2 + X_3 + X_4 + X_6$ | -8616  | 17232         | 110654 |
Based on Table 3.5.2 we get a simultaneous test of the model where \( G_{Hitung} > X^2_{(a,2p)} \) = 21,026 so that overall ownership of the Sidoarjo Regency IMB in 2019 is influenced by the variable length of the building with a 40th \((X1)\) land certificate \((X2)\), Knowledge of IMB and its Changes \((X3)\), knowledge of IMB sanctions \((X4)\), Costs \((X5)\), Lazy and Bureaucratic Convolution \((X6)\). Whereas in the partial test described in Table 4.5.2 it can be seen that the p-value of all six variables are all significant \((> 0.05)\)

3.6 Selection of the Best Model

The selection of the best model for the IMB ownership case in Sidoarjo Regency in 2019 can be seen in table 5.2. The model is a model that is influenced by the variable that has the smallest AIC value, namely the 11th model with variables consisting of the length of the building with a 40th \((X1)\) large land certificate \((X2)\), IMB Knowledge and Changes \((X3)\), knowledge of IMB sanctions \((X4)\), Costs \((X5)\), Lazy and Convoluted Bureaucracy \((X6)\) with AIC of 108925. But keep in mind that the best model is not only seen in the AIC value only the smallest, but the significance of the parameters are also seen, so based on table 1 variables The length of the building has a greater 40th \((X1)\) and Cost \((X5)\) the parameters are not significant so that the models contained variables \(X1\) and \(X5\) are not the best models, so the model the best model is the variable land certificate \((X2)\), IMB Knowledge and Changes \((X3)\), knowledge of IMB sanctions \((X4)\), Lazy and Bureaucracy convoluted \((X6)\) with a log-likelihood of -8616. The ZIP Regression Model is shown as follows:

\[
\log(\mu_i) = 5,470 + 2,831 \times 10^{-4}X_2 - 8,876 \times 10^{-6}X_3 + 1,356 \times 10^{-4}X_4 + 2,530 \times 10^{-4}X_6 \\
\logit(w_i) = 0.652 - 0.000244X_2 + 0.000175X_3 - 0.000996X_4 - 0.00155X_6
\]

The ZIP Regression Log Model explains that a 1% increase in ownership of a land certificate will increase the ownership of the IMB by \(2,831 \times 10^{-4}\) this is appropriate because the requirement to have an IMB is a must-have land certificate. While a 1% increase in knowledge about IMB will increase ownership of IMB by \(-8,876 \times 10^{-6}\) this is also appropriate because everyone who wants to do something must understand the purpose of the benefits or losses, how can someone take care of the licensing if knowledge of IMB does not exist. Likewise with the IMB Sanction can increase someone has an IMB letter of \(1,356 \times 10^{-4}\). Another factor that also influences ownership of IMB is that the procedure or convolution is bureaucratic, making people lazy to administer the IMB, so the easier the procedure or requirements to take care of the IMB will increase the level of ownership of the IMB by \(2,530 \times 10^{-4}\).

As for the logit model, the ZIP Regression, in this case, explains the opportunity of not having a BMI, meaning an increase in a significant variable will reduce the absence of BMI. The land certificate variable will reduce by \(100(e^{-0.000244} - 1) = 36.78\%\) IMB knowledge by 36.79\%, IMB sanctions by 36.75\%, and bureaucracy or procedures by 36.73%.

4. Conclusion

Based on the results of the research contained in the analysis and discussion, the following conclusions can be obtained

ZIP regression models for IMB ownership in Sidoarjo Regency in 2019 are:

\[
\log(\mu_i) = 5,470 + 2,831 \times 10^{-4}X_2 - 8,876 \times 10^{-6}X_3 + 1,356 \times 10^{-4}X_4 + 2,530 \times 10^{-4}X_6 \\
\logit(w_i) = 0.652 - 0.000244X_2 + 0.000175X_3 - 0.000996X_4 - 0.00155X_6
\]

The ZIP Regression Log Model explains that for an increase of 1% \(X2\) will increase ownership of the IMB by \(2,831 \times 10^{-4}\), \(X3\) by \(-8,876 \times 10^{-6}\), \(X4\) by \(1,356 \times 10^{-4}\) dan \(X6\) by \(2,530 \times 10^{-4}\). Where the results of ZIP Regression modeling obtained factors that have a significant effect on IMB Ownership in Sidoarjo Regency in 2019 are Land Certificates \((X2)\), Knowledge of IMB \((X3)\), Knowledge of Sanctions \((X4)\), Lazy or Convoluted Bureaucracy \((X6)\)
Based on the results of the ZIP Regression modeling, it was found that the factors that had a significant effect on IMB ownership in Sidoarjo Regency in 2019 were Land Certificate (X2), Knowledge of IMB (X3), Knowledge of Sanctions (X4), Lazy or Convoluted Bureaucracy (X6).

According to Adarabioyo, M. I., & Ipinyomi, R. A. (2019) ZIP regression models are easy to approximate. One method for estimating parameters in the known distribution of regression is the Maximum Likelihood Estimation (MLE) method, where the maximum likelihood estimate (MLE) approaches the normal distribution with a large number of samples, and the confidence interval can be calculated by the ratio test or the MLE normality estimate. The analysis showed that the confidence interval based on the likelihood ratio test was better. Finally, the ZIP regression model is not only easy to interpret, but it can also lead to more refined data analysis.

References
[1] Adarabioyo M I & Ipinyomi R A 2019 Comparing Zero-inflated Poisson, Zero-inflated Negative Binomial and Zero-inflated Geometric in Count Data with Excess Zero Asian Journal of Probability and Statistics 4(2) 1-10.https://doi.org/10.9734/ajpas/2019/v4i230113 More Citation Formats
[2] Adeliana 2018 Estimasi Parameter Geographically weighted Zero Inflated Poisson Regression (GWZIPR) dengan Pembobot Fixed Bisquare Kernel. Malan :Electronic theses UINMA
[3] Badan Pusat Statistika Kabupaten Sidoarjo Tahun 2019
[4] Kusuma Rahmainiar Dwinta & Purwono Yogo 2018 Zero-Inflated Poisson Regression Analysis On Frequency Of Health Insurance Claim PT.XYZ. (DOI : https://dx.doi.org/10.2991/icbmr-18.2019.52)
[5] Lambert Diane 1992 Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing, Technometrics, 34:1, 1-14
[6] Myers R H 1990 Classical and Modern Regression with Application (2nd ed.) Boston: PWS- KENT Publishing Company
[7] McCullagh P & Nelder JA 1992 Generalized Linear Models, 2nd Ed. London: Chapman % Hall.
[8] Ruliana 2015 Generalized Poisson Regression (GPR) Modeling to Overcome Equidispersion Violation on Poisson Regression in Measles Cases in Semarang City Semarang Thesis Faculty of Mathematics and Natural Sciences Semarang State University Semarang