Desimal: Jurnal Matematika Vol 7 No 2 (2024) 311-320

Contents lists available at DJM

DESIMAL: JURNAL MATEMATIKA

p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm

http://ejournal.radenintan.ac.id/index.php/desimal/index

Modeling tuberculosis in children under five using poisson and negative binomial regression

Ahmad Fajri*, Nur Rahmi, Putri Ayu Maharani, Muhammad Ikhlashul Amal

Institut Teknologi Bacharuddin Jusuf Habibie, Indonesia

ARTICLE INFO

Article History
Received : 17-04-2024
Revised : 17-05-2024
Accepted : 20-06-2024
Published : 30-08-2024

Keywords:
Negative Binomial Regression; Overdispersion; Poisson Regression; Tuberculosis.

*Correspondence: E-mail: ahmadfajri@ith.ac.id

Doi: 10.24042/djm.v7i2.23464

ABSTRACT

Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis. Indonesia is the country with the second highest number of tuberculosis cases after India. The Ministry of Health stated that there has been a significant increase in cases of tuberculosis among children in Indonesia where the increase in cases of tuberculosis among children has reached more than 200 percent. The number of tuberculosis cases can be reduced if the factors that affect the number of tuberculosis patients are known. Therefore, efforts should be made to model the number of cases of Tuberculosis among children under five years of age to provide useful information to prevent and control Tuberculosis. The relationship between these factors and the number of people with Tuberculosis can be determined using Poisson regression analysis because the number of cases of Tuberculosis is calculated data. Tuberculosis data contain overdispersion, so another approach is used to overcome it, which is by using a negative binomial regression model. The best model obtained based on the AIC value is the Negative Binomial regression model with an AIC value of 184.095. For further research, it is suggested to test the spatial effect and modeling using the Negative Binomial geographic weighted regression method to find out whether the characteristics of one region and the other influence the geographic location on the model.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis (Vidyastari et al., 2019). The cause of tuberculosis transmission is through the air (Smeltzer, 2016). Mycobacterium tuberculosis enters the human body mainly through the lungs but can also through the skin, urinary tract, and food tract (Puspasari & Fina, 2019). In addition, Mycobacterium tuberculosis bacteria can be transmitted to other humans through sputum splashes when a person with active pulmonary tuberculosis coughs or sneezes (Price & Wilson, 2006). When a person with tuberculosis speaks, coughs, or sneezes, the person indirectly releases approximately 3,000 droplets of sputum into the air containing germs (Aditama et al., 2011), and droplets of saliva fly into the air, are inhaled by healthy people and enter into the lungs which then causes pulmonary tuberculosis (Widayanti & Bintari, 2013). The symptoms caused by dry cough are coughing up phlegm for two weeks or more, which can be accompanied by blood-tinged phlegm, coughing up blood, shortness of breath, weakness, loss of appetite, weight loss, lethargy, night sweats without physical activity, and a higher fever for more than one month (Pralambang & Setiawan, 2021). Pulmonary tuberculosis can cause death if you do not take the medicine regularly for up to 6 months. In addition to affecting the individual, it also affects the sufferer's family, namely psychological effects in the form of anxiety, decreased support, and low self-confidence (Astuti, 2013).

Tuberculosis cases are still a global health problem. Indonesia has the second highest number of tuberculosis cases after India (World Health Organization, 2023). In 2022, the Ministry of Health together with all health workers managed to detect more than 700 thousand cases of tuberculosis. This figure is the highest achievement since tuberculosis was declared a national priority program. Currently, the number of tuberculosis cases in Indonesia is 969 thousand, and deaths are 93 thousand every year, or equivalent to 11 deaths per hour (Ministry of Health, 2023). The Ministry of Health stated that there has been a significant increase in cases of tuberculosis among children in Indonesia. The increase in cases of tuberculosis in children reached more than 200 percent. In 2021, tuberculosis cases in children reached 42,187, then in 2022 reached 100,726. In March 2023, a total of 18,144 children were infected with tuberculosis. Meanwhile, cumulatively, the Ministry of Health detected 443,235 cases of tuberculosis in 2021, and this increased to 717,941 cases in 2022 (Prabowo, 2023).

The number of tuberculosis cases can be reduced if the factors that affect the number of tuberculosis sufferers are known significantly. The relationship between these factors and the number of people with tuberculosis can be determined using Poisson regression analysis because the number of people with tuberculosis is calculated data. One of the characteristics of the Poisson distribution is the presence of an equidispersion condition, but in its use it is often characterized by an overdispersion condition (McCullagh, 2019). Another approach to overcome overdispersion is to use the Negative Binomial regression model because it allows the parameters to describe the variance of the data (Hardin & Hilbe, 2014). Some studies that develop this model include Cameron and Trivedi (1998) comparing the Poisson regression model with other regression models (Cameron & Trivedi, 1998), Dean (1992) regarding the overdispersion test in Poisson and Negative Binomial regression models (Draper et al., 1992), (Ismail & Jemain, 2005) in the insurance sector to determine customer opportunities in submitting motor vehicle insurance claims in Malaysia, (Ismail & Jemain, 2007) in the insurance sector to overcome overdispersion using negative binomial and Poisson's Quadrant. Therefore, we propose research to model tuberculosis cases using Poisson and Negative Binomial regression models.

METHOD

The type of data used in this research is secondary data obtained from the South Sulawesi Provincial Health Service. The data consisted of 6 variables, with one dependent variable and six independent variables. The dependent variable in this study is the number of tuberculosis cases of children under five in South Sulawesi in 2023. The independent variable used in this study is the percentage of malnourished children under five (X_1), the percentage of low birth weight babies. (X_2), the percentage of children under five who received complete basic immunization (X_3), the percentage of healthy homes (X_4), and the percentage of active smokers (X_5).

Before analyzing using the Poisson regression model, a multicollinearity check between the independent variables by looking at the Variance Inflation Factor (VIF) value is necessary. If the VIF value >10, then there is multicollinearity (Kutner et al., 2005) with the formula

$$VIF_i = \frac{1}{1-R_2^i}$$

where R_2^i is the coefficient of determination between X_i and other independent variables. Independent variables that have a value of $VIF_i > 10$ are not included in the model construction.

The Poisson distribution is often used to model rare events such as the number of people with liver cancer in an area in a certain time period, the number of traffic accidents in a location per year, etc. (Kleinbaum et al., 1988). A discrete random variable Y follows a Poisson distribution with parameter $\mu > 0$, where μ represents the average number of events per unit time if the probability mass function is

$$P(y|\mu) = \frac{e^{-\mu}\mu^y}{y!}, y = 0,1, \ldots$$

where $E(Y) = Var(Y) = \mu$ (Cameron & Trivedi, 1998).

A random variable with a Poisson distribution is written as $Y \sim$Poisson (μ). For example, $Y_1, Y_2, \ldots, Y_n$ are random samples of Poisson random variables with mean μ_i, then the probability mass function can be expressed as

$$P(Y_i = y_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, y_i = 0,1,2, \ldots.$$

The Poisson regression model is a nonlinear regression model that is often used to analyze discrete data that describes the relationship between the dependent variable and the independent variable. The natural logarithm function (ln) is a link function commonly used in Poisson regression models. The statement of the Poisson regression model using the link function is as follows (Cahyandari, 2014):

$$\mu_i = e^{(x_i{}^T\beta)} = e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_k x_{ki}}.$$

Informatian:
$i = 1,2, \ldots, n$
$j = 1,2, \ldots, k$
$n =$ number of observations
$k =$ number of independent variables
$\mu_i =$ the average number of events in a certain time period
$x_{ik} =$ the kth independent variable and the ith observation
$\beta_0 =$ constant
$\beta_k =$ regression coefficient of the kth independent variable
$x_i =$ vector of independent variables
$\beta =$ vector of Poisson regression parameters

Poisson regression parameter estimation was conducted using the Maximum Likelihood Estimation (MLE) method (Myers et al., 2012). The Poisson regression probability function is

$$L(y_i|\beta) = \prod_{i=1}^n P(Y_i = y_i|\beta)$$
$$= \frac{e^{-\sum_{i=1}^n e^{(\mathrm{x_i}^T\beta)}} \prod_{i=1}^n \left(e^{\mathrm{x_i}^T\beta}\right)^{y_i}}{\prod_{i=1}^n y_i!}$$

and the natural logarithm of the Poisson regression likelihood function is

$$\ln L(y_i|\beta) = -\sum_{i=1}^n e^{(\mathrm{x_i}^T\beta)} + \sum_{i=1}^n y_i \ln e^{\mathrm{x_i}^T\beta} - \sum_{i=1}^n \ln y_i!.$$

The estimation of β from the above equation can be obtained using Newton Raphson iteration or the literatively weighted least squares (ILWS) algorithm.

(Scoot Long, 1997) in (Jackman, 2007) states that overdispersion occurs due to unobserved sources of variation in the data or the influence of other variables, which result in the probability of an event occurring depending on previous events. (McCullagh, 2019) states that data calculated for Poisson regression is said to contain overdispersion if the variance is greater than the mean. The standard deviation of the estimated parameter is biased downward (underestimate), and the importance of the influence of the independent variable is biased upward (overestimate). An estimate of dispersion can be measured from the ratio between the deviations or Pearson Chi-Square divided by the degrees of freedom. This ratio is called the dispersion ratio. The dispersion estimate is said to be overdispersed if the dispersion ratio is $>$ 1 and underdispersed if the dispersion ratio is < 1.

Negative Binomial Regression is one solution to overcome overdispersion. The Negative Binomial regression model has the following probability mass function (Greene, 2008):

$$f(y,\mu,\theta) = \frac{\Gamma(y+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y+1)}\left(\frac{1}{1+\theta\mu}\right)^{\theta-1}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y}, y = 0,1,2 \ldots$$

where $\Gamma(\theta) = \int_0^\infty t^{\theta-1}\, e^{-t} dt = (\theta-1)!$.

Information:

$y =$ value of the dependent variable
$\mu =$ expected value of y
$\theta =$ dispersion parameter

The relationship between the expected value and the variance of the dependent variable in the Negative Binomial regression model can be written as follows (Hilbe, 2011):

$$E(Y) = \mu \text{ and } Var(Y) = \mu + \theta\mu^2.$$

The Negative Binomial regression model with the link function is

$$\mu_i = e^{(\mathbf{x_i}^{\mathrm{T}}\boldsymbol{\beta})} = e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_k x_{ki}}.$$

Informatian:

$i = 1,2,\ldots,n$
$j = 1,2,\ldots,k$
$n =$ number of observations
$k =$ number of independent variables
$\mu_i =$ the average number of events in a certain time period
$x_{ik} =$ the kth independent variable and the ith observation
$\beta_0 =$ constant
$\beta_k =$ regression coefficient of the kth independent variable
$x_i =$ vector of independent variables
$\beta =$ vector of Negative Binomial regression parameters

Negative Binomial regression parameters were estimated using the Maximum Likelihood Estimation (MLE) method. The Negative Binomial regression probability function is

$$L(y_i|\beta,\theta) = \prod_{i=1}^{n}\left\{\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i+1)}\left(\frac{1}{1+\theta\mu_i}\right)^{\theta-1}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}\right\}$$

and the natural logarithm of the Negative Binomial regression likelihood function is

$$ln\, L(y_i|\beta,\theta) = \sum_{i=1}^{n}\left\{ln\left(\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i+1)}\right) + y_i\, ln(\theta\mu_i) - \left(\frac{1}{\theta}+y_i\right) ln(1+\theta\mu_i)\right\}.$$

The solution to the above equation can be solved using the literal weighted least squares (ILWS) algorithm.

One test that can be used to test the partial significance of each independent variable is the Wald test. The Wald test hypothesis is as follows:

$H_0: \beta_j = 0$ (the jth independent variables has no significant effect)
$H_1: \beta_j \neq 0$ (the jth independent variables has a significant effect)

with Wald test statistics

$$W = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right]^2, j = 1,2,\ldots,k$$

where $\hat{\beta}_j$ is the estimated parameter and $SE(\hat{\beta}_j)$ is the standard error for the estimated parameter β_j. Reject H_0 at the siginificance level α if $W > \chi^2_{(\alpha,1)}$ or if the the p value is smaller than the siginificance level α (Myers et al., 2012).

The most frequently used measure of goodness of fit of a maximum likelihood model is the Akaike Information Criterion (AIC). Akaike defines the AIC calculation as follows:

$$\mathrm{AIC} = -2\ln L(y|\hat{\mu}) + 2\mathrm{p}$$

where $\ln L(y|\hat{\mu})$ is the natural logarithm of the model involving all independent variables and p is the number of parameters. Figure 1 shows the flow of the study that has been carried out.

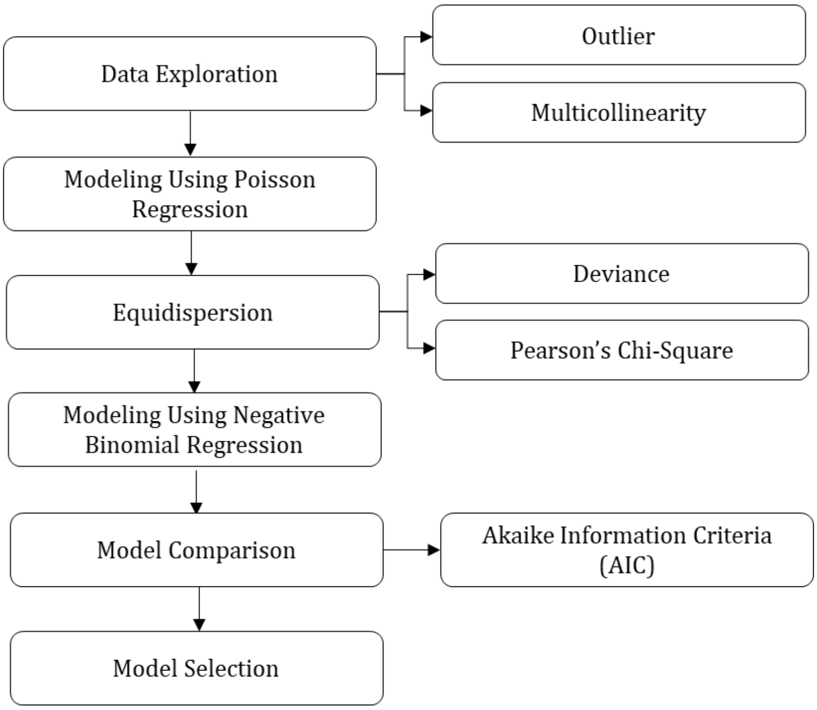

Figure 1. Research Flowchart

RESULTS AND DISCUSSION

South Sulawesi Province has a total of 553 cases of tuberculosis in children under the age of five that occurred in 2023. Figure 2 depicts the distribution map of tuberculosis cases in children under the age of five in South Sulawesi in 2023.

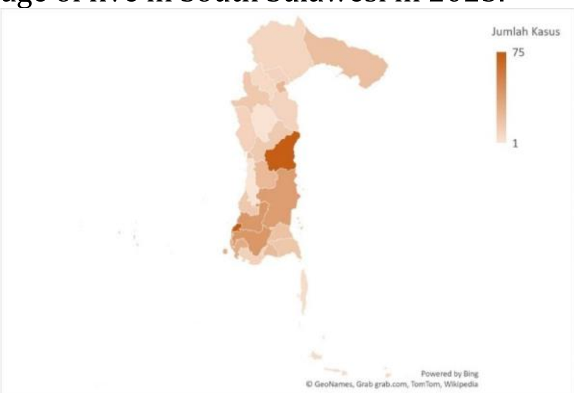

Figure 2. Map of the distribution of the number of tuberculosis cases among children under five in South Sulawesi

Based on Figure 2, the dark orange color indicates districts/cities with very high cases of tuberculosis among children under five years of age. The darker the orange color in the district/city, the higher the number of tuberculosis cases in children under five. On the other hand, the orange color fades in the district/city, the lower the number of cases of tuberculosis in children under five years occurs.

The highest number of tuberculosis cases in children under the age of five occurred in Makassar City and Wajo District, which is shown in dark orange. Meanwhile, other districts/cities appear to have somewhat different numbers of tuberculosis cases among children under five, but they are not significantly different and are still below 75 cases, as indicated by the fainter orange color.

Table 1 shows the descriptive statistics of tuberculosis cases in children under five years of age in South Sulawesi. Tuberculosis cases among children under five in South Sulawesi have an average of 23.04 cases and a variance of 375.78 cases. The highest number of cases of tuberculosis in children under the age of five occurred in Makassar city with 75 cases, while the lowest number of cases of tuberculosis in children under five occurred in Barru Regency with 1 case. The average number of tuberculosis cases, which is greater than the median, shows that the distribution of tuberculosis among children under five in South Sulawesi is trending to the right.

Table 1. Descriptive statistics of the number of tuberculosis cases among children under five in South Sulawesi

Descriptive Statistics	Value
Mean	23.04
Median	18.00
Variance	375.78
Minimum Value	1
Maximum Value	75

Figure 3 and Figure 4 contain a histogram and boxplot of the number of tuberculosis among children under the age of five in South Sulawesi.

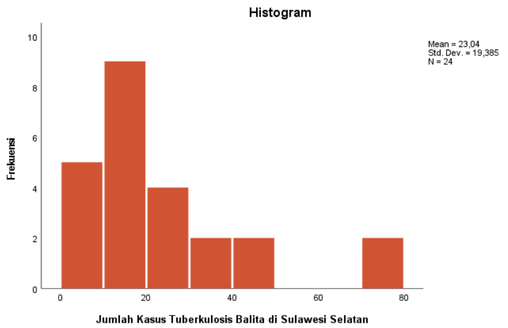

Figure 3. Histogram of the number of tuberculosis cases among children under five in South Sulawesi

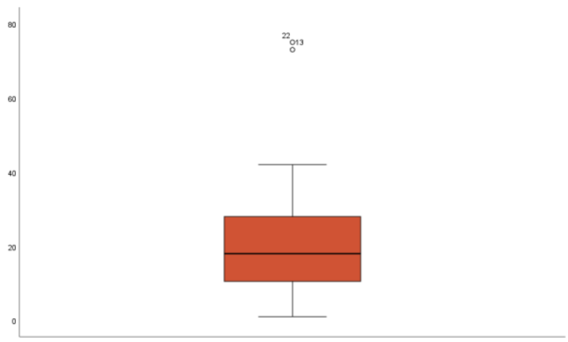

Figure 4. Boxplot of the number of tuberculosis cases among children under five in South Sulawesi

Based on Figure 3 and Figure 4, the histogram of the number of tuberculosis cases among children under five in South Sulawesi seems to point to the right, with the outliers in the boxplot occurring in Makassar City and Wajo Regency with 75 and 73 cases respectively. . The presence of outliers in the data will affect the overall distribution of the data. Therefore, the cases that occurred in Makassar City and Wajo Regency will be considered as exceptional cases in South Sulawesi, and they will not be included in further data processing.

The distribution map, descriptive statistics, histogram, and boxplot of the number of cases of Tuberculosis in children under five years old in South Sulawesi without Makassar City and Wajo Regency can be seen in Figure 5, Table 2, Figure 6, and Figure 7.

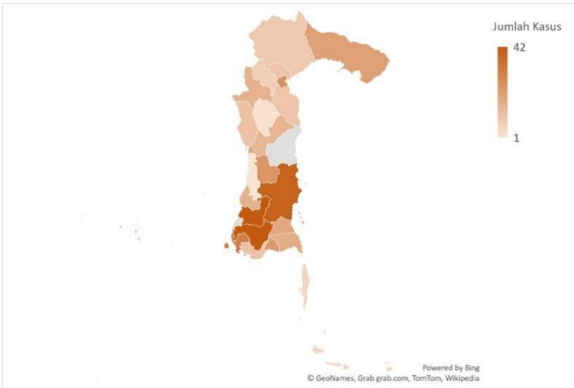

Figure 5. Map of the distribution of the number of tuberculosis cases among children under five in South Sulawesi without Makassar City and Wajo Regency

Table 2. Descriptive statistics of the number of tuberculosis cases among children under five in South Sulawesi without Makassar City and Wajo Regency

Descriptive Statistics	Value
Mean	18.41
Median	16.50
Variance	141.68
Minimum Value	1
Maximum Value	42

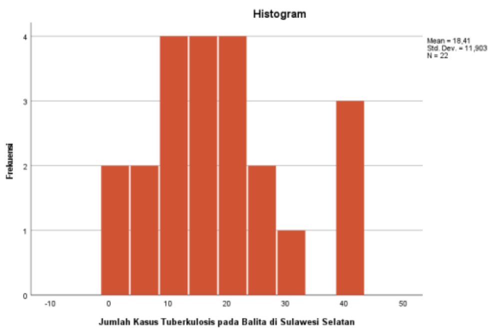

Figure 6. Histogram of the number of tuberculosis cases among children under five in South Sulawesi without Makassar City and Wajo Regency

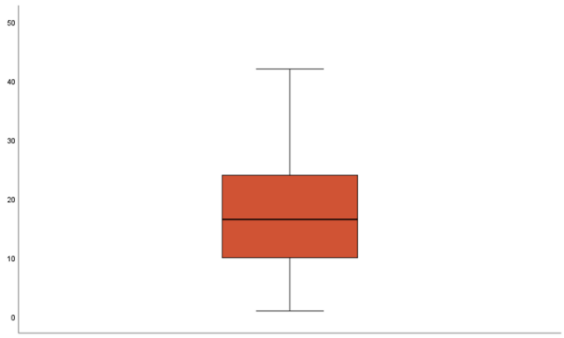

Figure 7. Boxplot of the number of tuberculosis cases among children under five in South Sulawesi without Makassar City and Wajo Regency

Based on Figure 5, Figure 6, and Figure 7, by not including Makassar City and Wajo Regency, the spread of tuberculosis cases among children under five in South Sulawesi can be seen more clearly, and there are no outliers in the data. The highest cases of tuberculosis occur in Gowa and Maros districts. Meanwhile, the lowest cases of tuberculosis occurred in Barru and Enrekang districts. In addition to affecting the visualization of data distribution, the exclusion of Makassar City and Wajo Regency also affects data variation. whose value fell to 141.68 cases of tuberculosis in children under five years of age. Other descriptive statistics can be seen again in Table 2.

The relationship between the number of cases of tuberculosis in children under five years of age and the factors affecting it can be determined using regression analysis. The regression model formed is a model using five independent variables, namely the percentage of children under five years of age who are malnourished (X_1), the percentage of low birth weight babies (X_2), the percentage of children under five years who receive complete. important immunizations (X_3), percentage of healthy households (X_4), and percentage of active smokers (X_5). The regression analysis used is Poisson regression because the number of tuberculosis cases can be assumed to have a Poisson distribution. A multicollinearity check was conducted to meet the modeling assumption in the regression, i.e. there is no relationship between the independent variables. The criterion used to detect multicollinearity is using the VIF of the independent variable. If the VIF value is $>$ 10, it is said that there is multicollinearity in the data, and this independent variable is not included in the model preparation. VIF values for each independent variable can be shown in Table 3.

Table 3. Multicollinearity Check

Independent Variables	VIF
X_1	1.096
X_2	1.159
X_3	1.037
X_4	1.115
X_5	1.091

Based on Table 3, each independent variable has a VIF value < 10, which indicates that there is no multicollinearity in the data and all independent variables are included in the model preparation.

Modeling the number of tuberculosis cases among children under five in South Sulawesi in 2023 was conducted using Poisson regression analysis. The obtained Poisson regression model is as follows:

$$\widehat{\mu_i} = e^{(1.602+0.015X_{1i}-0.014X_{2i}+0.008X_{3i}+0.003X_{4i}+0.043X_{5i})}.$$

The results of partial parameter testing of the Poisson regression model at a significance level of $\alpha = 10\%$ can be seen in Table 4. This test uses the Wald W test statistic, compared with the value $\chi^2_{(\alpha;1)}$ with a significance level of $\alpha = 10\%$, namely $\chi^2_{(0.1;1)} = 2.706$.

Table 4. Results of testing the Poisson regression model parameters

Parameter	Estimated Value	Standard Error	W
β_0	1.602	0.3483	21.1552
β_1	0.015	0.0266	0.3179
β_2	-0.014	0.0078	3.2215
β_3	0.008	0.0025	10.2400
β_4	0.003	0.0027	1.2345
β_5	0.043	0.0137	9.8513

Table 4 shows that the independent variables X_2, X_3, and X_5 have a Wald test statistical value $W > \chi^2_{(0.1;1)} = 2.706$, which means that the independent variables X_2, X_3, and X_5 have a significant effect on the model. One of the assumptions that must be met when carrying out Poisson regression is the condition of equidispersion, where the mean and variance have the same value. One violation of equidispersion is overdispersion, where the variance has a value greater than the average. The test statistic used to determine the occurrence of overdispersion is the ratio between the deviation or Pearson's Chi-square and the degrees of freedom, called the dispersion ratio. If the dispersion ratio is > 1, then the data is overdispersed. The results of the overdispersion test can be seen in Table 5.

Table 5. Overdispersion Testing

Criteria	Value	Degree of Freedom	Test Statistics
Deviance	142.994	16	8.937
Person Chi-Square	132.271	16	8.267

Table 5 shows that the dispersion ratio is > 1 based on the deviation or Pearson's Chi-square criterion for data to suffer from overdispersion. Poisson regression modeling that contains overdispersion causes the conclusions obtained to be inaccurate because the standard error value is smaller than the actual value (underestimate), and the importance of the influence of the independent variable is biased upwards (overestimate).

Based on the results of the overdispersion test in Poisson regression modeling, it is known that there is overdispersion in the data. Another approach that can be used if the data is overdispersed is the Negative Binomial regression model. The Negative Binomial regression model obtained is $\widehat{\mu_i} = e^{(1.725+0.008X_{1i}-0.015X_{2i}+0.008X_{3i}+0.003X_{4i}+0.042X_{5i})}$. The results of partial parameter testing of the Negative Binomial regression model at a significance level of α=10% can be seen in Table 6. This test uses the Wald W test statistic, compared with the value $\chi^2_{(\alpha;1)}$ with a significance level of $\alpha = 10\%$ that is $\chi^2_{(0.1;1)} = 2.706$.

Table 6. Results of testing the parameters of the Negative Binomial regression model

Parameter	Estimated Value	Standart Error	W
β_0	1.725	1.5405	1.2538
β_1	0.008	0.1432	0.3121
β_2	-0.015	0.0292	0.2638
β_3	0.008	0.0109	0.5386
β_4	0.003	0.1115	0.0007
β_5	0.042	0.0696	0.3641

Table 6 shows that all independent variables have Wald test statistical values $W < \chi^2_{(0.1;1)} = 2.706$, which means that all independent variables have no significant effect on the model.

The selection of the best model is based on the AIC value. The best model selected is the model with the smallest AIC value. A comparison of AIC values for each model is presented in Table 7.

Table 7. Comparison of AIC Values

Regression Model	AIC
Poisson	253.570
Negative Binomial	184.095

Table 7 shows that the Negative Binomial regression model has a smaller AIC value. Therefore, the following Negative Binomial regression model $\widehat{\mu_i} = e^{(1.725+0.008X_{1i}-0.015X_{2i}+0.008X_{3i}+0.003X_{4i}+0.042X_{5i})}$ was chosen as a better model to model the number of tuberculosis cases among children under five in South Sulawesi in 2023 with an AIC value of 184.095. This proves that Negative Binomial regression can overcome cases of overdispersion in Poisson regression.

CONCLUSIONS AND SUGGESTIONS

The best model to model the number of Tuberculosis cases among children under five in South Sulawesi in 2023 based on the AIC value is the Negative Binomial regression model with an AIC value of 184.095. The resulting Negative Binomial regression model is $\widehat{\mu_i} = e^{(1.725+0.008X_{1i}-0.015X_{2i}+0.008X_{3i}+0.003X_{4i}+0.042X_{5i})}$.

For further research, it is suggested to test the spatial effects and modeling using the Negative Binomial geographically weighted regression method to find uot whether the characteristics of one region and another influence geographical location on the model.

REFERENCES

Aditama, T. Y., Subuh, M., & Mustikawati, D. E. (2011). Pedoman nasional pengendalian tuberkulosis. *Direktorat Jendral Pengendalian Penyakit dan Penyehatan Lingkungan Kementrian Kesehatan RI. Jakarta: Kemenkes RI.*

Astuti, S. (2013). Hubungan tingkat pengetahuan dan sikap masyarakat terhadap upaya pencegahan penyakit Tuberkulosis di rw 04 Kelurahan Lagoa Jakarta Utara Tahun 2013.

Cahyandari, R. (2014). Pengujian overdispersi pada model regresi poisson (Studi kasus: laka lantas mobil penumpang di Provinsi Jawa Barat). *Statistika*, *14*(2), 69-76. https://doi.org/10.29313/jstat.v14i2.1204

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (No. 53). Cambridge university press.

Draper, N., Smith, H., & Sumantri, B. (1992). *Analisis Regresi Terapan.* Jakarta: Gramedia Pustaka Utama.

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters*, *99*(3), 585-590. https://doi.org/10.1016/j.econlet.2007.10.015

Hardin, J. W., & Hilbe, J. M. (2014). Regression models for count data based on the negative binomial (p) distribution. *The Stata Journal*, *14*(2), 280-291. https://doi.org/10.1177/1536867X1401400203

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Ismail, N., & Jemain, A. A. (2005). Generalized Poisson regression: An alternative for risk classification. *Jurnal Teknologi*, 39â-54. http://dx.doi.org/10.11113/jt.v43.770

Ismail, N., & Jemain, A. A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. In *Casualty actuarial society forum* (Vol. 2007, pp. 103-58). Citeseer.

Jackman, S. (2007). Models for Unordered Outcomes Political Science

150C/350C. *Political Science*, *150*, 350C.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1988). *Applied regression analysis and other multivariable methods* (Vol. 601). Belmont, CA: Duxbury press.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. McGraw-hill.

Kemenkes. (2023, Apr 11). *Indonesia Raih Rekor Capaian Deteksi TBC Tertinggi di Tahun 2022*. Retrieved from Kementerian Kesehatan RI: https://ayosehat.kemkes.go.id/indonesia-raih-rekor-capaian-deteksi-tbc-tertinggi-di-tahun-2022#:~:text=Saat%20ini%20diketahui%20bahwa%20Indonesia,dengan%2011%20kematian%20per%20jam. Akses pada 6 Februari 2024.

McCullagh, P. (2019). *Generalized linear models*. Routledge.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). *Generalized linear models: with applications in engineering and the sciences*. John Wiley & Sons.

Prabowo, A. (2023, Mar 18). *Kasus TBC pada Anak Indonesia Naik 200 Persen*. Retrieved from Radio Republik Indonesia: https://www.rri.co.id/kesehatan/191033/kasus-tbc-pada-anak-indonesia-naik-200-persen#:~:text=Lebih%20lanjut%2C%20Imran%20menyebutkan%20Maret,118.438%20kasus%20TBC%20di%20Indonesia. Akses pada 6 Februari 2024.

Pralambang, S. D., & Setiawan, S. (2021). Faktor risiko kejadian tuberkulosis di Indonesia. *Jurnal Biostatistik, Kependudukan, Dan Informatika Kesehatan*, *2*(1), 60-71. https://doi.org/10.51181/bikfokes.v2i1.4660

Price, S. A., & Wilson, L. M. (2006). Patofisiologi konsep klinis proses-proses penyakit. *Jakarta: Egc*, *4*(2), 1127-1128.

Puspasari, S. F. A., & Fina, S. (2019). Asuhan keperawatan pada pasien dengan gangguan sistem pernapasan. *Yogyakarta: Pustaka Baru*.

Scott Long, J. (1997). Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences*, *7*.

Smeltzer, S. C. (2016). Smeltzer and Bare's Textbook of Medical-Surgical Nursing: Vol. 1 And 2 (M. Farrell, Ed.).

Vidyastari, Y. S., Riyanti, E., & Cahyo, K. (2019). Faktor–Faktor Yang Mempengaruhi Pencapaian Target Cdr (Case Detection Rate) Oleh Koordinator P2tb Dalam Penemuan Kasus di Puskesmas Kota Semarang. *Jurnal Kesehatan Masyarakat*, *7*(1), 535-544. https://doi.org/10.14710/jkm.v7i1.23077

Widayanti, E., & Bintari, S. H. (2013). Uji Resistensi Mycobacterium Tuberculosis terhadap Obat Anti Tuberkulosis (Oat) dengan Metode Penipisan. *Life Science*, *2*(1).

World Health Organization. (2023, November 7). *Global Tuberculosis Report 2023*. Retrived from World Helath Organization: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023. Akses pada 6 Februari 2024.