

# Pemilihan Model Terbaik pada *Generalized Poisson Regression* Menggunakan *Akaike Information Criterion*

RUT ESRA<sup>1</sup>, DARNAH ANDI NOHE<sup>2</sup>, M. FATHURAHMAN<sup>3</sup>

<sup>1,2,3</sup> Program Studi Statistika Fakultas MIPA Universitas Mulawarman, Indonesia  
e-mail: darnah.98@gmail.com

## ABSTRAK

*Poisson regression* merupakan salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel respon yang berupa data *count* dengan variabel prediktor berupa data *count*, kontinu, kategorik atau campuran dengan syarat terjadi *equidispersion* yaitu nilai variansi dari variabel respon harus sama dengan nilai rata-ratanya. Namun yang sering terjadi adalah pelanggaran terhadap *equidispersion*. *Generalized Poisson Regression* (GPR) adalah suatu model regresi yang digunakan untuk menganalisis hubungan antara variabel respon yang berupa data *count* dengan satu atau lebih variabel prediktor dan mengalami *underdispersion*, *equidispersion*, atau *overdispersion*. Data tuberkulosis paru (TB paru) di Indonesia tahun 2020 mengalami *overdispersion*, sehingga GPR adalah metode yang cocok untuk memodelkan data tersebut. Tujuan penelitian ini adalah untuk mendapatkan model GPR terbaik pada data jumlah kasus TB paru di Indonesia tahun 2020 dan mengetahui faktor-faktor yang memengaruhinya. Hasil analisis menunjukkan bahwa terdapat lima belas model GPR yang terbentuk dari empat variabel prediktor yang berpengaruh terhadap jumlah kasus TB paru di Indonesia tahun 2020. Model GPR terbaik berdasarkan nilai *Akaike Information Criterion* (AIC) terkecil adalah model dengan empat variabel yang berpengaruh yaitu kepadatan penduduk, jumlah penduduk miskin, persentase lantai rumah tidak kedap air, dan persentase tempat pengelolaan pangan yang memenuhi syarat.

*Kata Kunci: AIC, Data Count, Gpr, Overdispersion, Tuberkulosis Paru*

## ABSTRACT

*Poisson regression* is a regression model that can be used to analyze the relationship between response variables in the form of count data and predictor variables in the form of count, continuous, categorical or mixed data with the condition that *equidispersion* occurs, namely the variance value of the response variable must be equal to the average value. However, what often happens is that the variance value is greater than the average value or is called *overdispersion*. *Generalized Poisson Regression* (GPR) is a regression model used to analyze the relationship between response variables in the form of count data and one or more predictor variables and occur *underdispersion*, *equidispersion* or *overdispersion*. Data for pulmonary tuberculosis in Indonesia in 2020 occurred *overdispersion*, so GPR is a suitable method to model the data. The purpose of this study was to obtain the best GPR model and to obtain the factors that significantly influence the number of pulmonary tuberculosis cases in Indonesia in 2020. The results of the analysis show that there are fifteen GPR models formed from four predictor variables that affect the number of pulmonary tuberculosis cases in Indonesia in 2020. The best GPR model based on the smallest *Akaike Information Criterion* (AIC) value is a model with four influential variables, namely population density, number of poor people, percentage of house floors that are not waterproof, and percentage of food management places that meet the requirements.

*Kata Kunci: AIC, Count Data, Gpr, Overdispersion, Pulmonary Tuberculosis*

## 1. PENDAHULUAN

Analisis regresi merupakan salah satu metode statistika yang sering digunakan untuk menyatakan hubungan antar variabel. Hubungan tersebut dapat diekspresikan dalam bentuk

persamaan yang menghubungkan variabel respon dengan variabel prediktor (Wasilaine, dkk, 2014). Persamaan regresi merupakan persamaan garis yang menyatakan adanya hubungan antara dua variabel tersebut. Analisis regresi sering digunakan sebagai metode untuk menganalisis data variabel respon berupa data kontinu. Data variabel respon yang dapat dianalisis bukan hanya kontinu, tetapi juga dapat berupa data count. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel respon Y berupa data count dengan variabel X berupa data count, kontinu, kategorik atau campuran adalah model Poisson Regression.

Poisson Regression (PR) umumnya diterapkan pada data penelitian kesehatan masyarakat, biologi, dan teknik, dimana datanya berupa data count (Simarmata & Ispriyanti, 2011; Darnah, dkk., 2019). Dalam model PR, salah satu asumsi utama adalah terjadinya equidispersion yaitu nilai variansi dari variabel respon harus sama dengan nilai rata-ratanya (Cahyandari, 2014). Namun hal yang kerap kali terjadi adalah pelanggaran akan asumsi tersebut. Salah satu cara untuk mengatasi hal tersebut adalah melakukan pemodelan dengan *Generalized Poisson Regression* (GPR). GPR adalah salah satu model regresi yang digunakan untuk menganalisis hubungan antara variabel respon yang berupa data count dengan satu atau lebih variabel prediktor dan mengalami underdispersion atau overdispersion (Ismail & Jemain, 2007; Darnah, 2016).

Dalam menentukan hubungan antara variabel respon dan prediktor juga dibutuhkan model regresi terbaik. Model regresi terbaik adalah model yang dapat menjelaskan perilaku variabel respon dengan sebaik-baiknya dengan memilih variabel-variabel prediktor dari sekian banyak variabel prediktor yang tersedia dalam data. Salah satu metode untuk memilih model regresi terbaik adalah menggunakan nilai *Akaike Information Criterion* (AIC) yang ditemukan oleh Akaike dan Schwarz. Model regresi terbaik adalah model dengan nilai AIC terkecil (Fadmi, 2015).

Model GPR pada penelitian ini diaplikasikan pada data Tuberculosis atau tuberkulosis paru (TB paru) di Indonesia. TB paru adalah suatu penyakit menular yang disebabkan oleh kuman *Mycobacterium tuberculosis*. Melalui laman resmi World Health Organization (WHO), TB paru menjadi salah satu penyakit menular dan paling membahayakan di dunia. Terdapat 4.000 orang meninggal dan 28.000 orang jatuh sakit setiap harinya dikarenakan penyakit ini. World Health Organization (2018) menyatakan Indonesia merupakan salah satu negara yang mempunyai jumlah kasus TB terbesar di dunia diantara 8 negara yaitu India (27%), Cina (9%), Indonesia (8%), Filipina (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and Afrika Selatan (3%). Pada tahun 2020, Indonesia menjadi negara penyumbang kasus TB paru terbanyak ketiga di dunia (World Health Organization, 2021). Melalui laman resmi Kementerian Kesehatan Republik Indonesia, TB paru masih menjadi perhatian khusus. Pemerintah Indonesia memiliki komitmen Sustainable Development Goals (SDGs) yaitu mengeliminasi TB paru di Indonesia pada tahun 2030.

Penelitian mengenai TB paru dan faktor-faktor yang memengaruhinya telah banyak dilakukan oleh peneliti. Faktor-faktor yang berpengaruh terhadap kasus TB paru adalah persentase kondisi lantai rumah tidak kedap air (Naben, dkk., 2013), persentase tempat umum dan pengelolaan makanan sehat berpengaruh signifikan terhadap jumlah kasus penyakit TB paru (Lestari, dkk., 2014), kepadatan hunian, ventilasi, pencahayaan, kelembapan (Perdana & Putra, 2018), persentase rumah tangga yang memiliki akses terhadap sumber air minum yang layak (Yotenka & Banapon, 2019), persentase penduduk yang memiliki akses sanitasi layak, persentase penduduk miskin, persentase tempat umum sehat, persentase pengelolaan makanan sehat, dan persentase masyarakat berperilaku Hidup Bersih dan Sehat (ber-PHBS) (Asriyanti, dkk., 2022). Tujuan penelitian ini adalah mendapatkan model GPR terbaik pada data jumlah kasus TB paru di Indonesia tahun 2020 dan mengetahui faktor-faktor yang memengaruhinya

## 2. METODE PENELITIAN

### 2.1 Generalized Linear Models

Generalized Linear Models (GLM) adalah perluasan dari model regresi umum untuk variabel respon dengan sebaran keluarga eksponensial (Astuti, dkk, 2013). Terdapat komponen penting dalam GLM yaitu

1. *Random Component*, yaitu nilai-nilai pengamatan respon Y yang saling bebas dari berdistribusi tertentu.

2. Systematic Component, yaitu kombinasi linear dari variabel X dengan parameter  $\beta$  yang dilambangkan dengan  $\mathbf{x}_i^T \beta$
3. *Link Function* atau fungsi penghubung yaitu suatu fungsi yang menjelaskan nilai ekspektasi dari variabel respon Y yang menghubungkan dengan variabel-variabel prediktor melalui persamaan linier. Dituliskan dengan  $\eta_i = g(\cdot)$ , fungsi  $g(\cdot)$  disebut dengan fungsi penghubung.

(Agresti, 2002).

### 2.2 Distribusi Poisson

Menurut Walpole (1995), distribusi Poisson adalah distribusi suatu kejadian dengan peluang kecil pada selang waktu tertentu dan hasil pengamatan berupa variabel *count*. Distribusi peluang untuk variabel acak Poisson Y adalah

$$f(y) = \frac{e^{-\mu} \mu^y}{y!} \quad \dots(1)$$

dengan  $\mu$  merupakan rata-rata banyaknya kejadian dalam selang tertentu,  $y = 1, 2, \dots, n$  dan variansi dari Y mempunyai nilai lebih dari nol. Untuk mengetahui apakah data berdistribusi Poisson atau tidak adalah dengan menggunakan uji *Kolmogorov-Smirnov*.

Hipotesis:

$H_0 : F(x) = f_0(x)$  (Data mengikuti suatu distribusi)

$H_0 : F(x) \neq f_0(x)$  (Data tidak mengikuti suatu distribusi)

Statistik uji:

$$D = \max |F(x) - f_0(x)| \quad \dots(2)$$

dengan  $F(x)$  adalah distribusi populasi dan  $f_0(x)$  adalah distribusi sampel.  $H_0$  ditolak jika  $D > D_\alpha$  atau nilai hasil *output* kurang dari nilai  $\alpha$  (Lungan, 2006).

### 2.3 Multikolinearitas

Menurut Priyatno (2013), adanya hubungan antar variabel bebas atau disebut multikolinieritas dalam model regresi dapat dideteksi menggunakan *Variance Inflation Factor* (VIF). VIF kurang dari atau sama dengan 10 maka tidak terjadi multikolinieritas pada model regresi. Nilai VIF dinyatakan sebagai berikut

$$VIF = \frac{1}{(1 - R_j^2)}; j = 1, 2, \dots, p \quad \dots(3)$$

dengan  $p$  adalah banyaknya variabel,  $R_j^2$  adalah nilai koefisien determinasi dari variabel  $X_j$  dengan variabel prediktor lainnya, yaitu

$$R_j^2 = \frac{\text{Jumlah Kuadrat Regresi}}{\text{Jumlah Kuadrat Total}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (\hat{Y}_{ij} - \bar{Y}_j)^2}{\sum_{i=1}^n \sum_{j=1}^p (Y_{ij} - \bar{Y}_j)^2} \quad \dots(4)$$

(Hocking, 1996).

### 2.4 Poisson Regression

Myers (1990) mengatakan bahwa *Poisson Regression* (PR) adalah suatu model regresi non-linear yang digunakan untuk menganalisis data berdistribusi Poisson dimana variabel respon berbentuk data *count* dan nilainya adalah *integer* tidak negatif (Cahyandari, 2014). Beberapa asumsi yang harus dipenuhi dalam penggunaan PR, yaitu variabel respon mengikuti distribusi Poisson, *equidispersion*, dan tidak terjadi multikolinieritas. PR termasuk salah satu GLM. Fungsi peluang untuk PR adalah

$$f(y) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad \dots(5)$$

untuk  $\mu_i > 0$  dengan  $\mu_i$  merupakan rataan dari variabel respon Y dan  $Var(Y_i) = E(Y_i)$  (Safrida, dkk, 2013). Estimasi parameter model PR menggunakan *Maximum Likelihood Estimation* (MLE). Fungsi *likelihood* dari persamaan 5 adalah

$$L(\mu_i) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad \dots(6)$$

dengan  $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$  maka diperoleh fungsi pada persamaan 7 sebagai berikut.

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \frac{\exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) (\exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{y_i}}{y_i!} \quad \dots(7)$$

Jika  $\ell(\boldsymbol{\beta}) = \ln L(\boldsymbol{\beta})$  maka fungsi *log likelihood* dari persamaan 7 adalah:

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \ln \left( \prod_{i=1}^n \frac{\exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) (\exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{y_i}}{y_i!} \right) \quad \dots(8) \\ &= \sum_{i=1}^n \frac{\exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) (\exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{y_i}}{y_i!} \\ &= \sum_{i=1}^n \ln \left( \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) + \ln(\exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{y_i} - \ln(y_i!) \right) \\ &= \sum_{i=1}^n [-\exp(\mathbf{x}_i^T \boldsymbol{\beta}) + y_i(\mathbf{x}_i^T \boldsymbol{\beta}) - \ln(y_i!)] \end{aligned}$$

Selanjutnya persamaan 8 diturunkan terhadap  $\boldsymbol{\beta}^T$  untuk memperoleh turunan pertama yang ditunjukkan pada persamaan 9

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^n (y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})) \mathbf{x}_i = 0 \quad \dots(9)$$

Persamaan 9 merupakan persamaan yang eksplisit sehingga memerlukan metode iterasi dalam penyelesaiannya. Metode yang digunakan adalah metode iterasi *Fisher-Scoring* (Aulele, 2012). Algoritma iterasi *Fisher-Scoring* adalah

$$\hat{\boldsymbol{\beta}}_{(m+1)} = \hat{\boldsymbol{\beta}}_{(m)} - \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_{(m)}) \mathbf{g}^T(\hat{\boldsymbol{\beta}}_{(m)}) \quad \dots(10)$$

dimana

$\hat{\boldsymbol{\beta}}_{(m)}$  adalah vektor dari estimator parameter pada iterasi ke- $m$  dengan  $m = 1, 2, \dots, k$ .

$$\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_{(m)}) = -E \begin{bmatrix} \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0^2} & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0 \partial \beta_1} & \dots & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0 \partial \beta_k} \\ \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0 \partial \beta_1} & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_1^2} & \dots & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_1 \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0 \partial \beta_k} & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_1 \partial \beta_k} & \dots & \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_k^2} \end{bmatrix}; \mathbf{g}^T(\hat{\boldsymbol{\beta}}_{(m)})_{(k+1)(k+1)} = \left( \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_0} \quad \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_1} \quad \dots \quad \frac{\partial \ell(\hat{\boldsymbol{\beta}})}{\partial \beta_k} \right)$$

Iterasi diakhiri ketika mencapai keadaan konvergen, yaitu:  $\|\hat{\boldsymbol{\beta}}_{(m+1)} - \hat{\boldsymbol{\beta}}_{(m)}\| \leq \varepsilon$ , dimana  $\varepsilon$  merupakan bilangan riil positif yang sangat kecil (Myers, 1990).

### 2.5 Pengujian Parameter Model PR

Untuk mengetahui apakah model PR yang didapatkan telah layak digunakan atau tidak, maka dilakukan uji simultan dan uji parsial. Uji simultan pada model PR untuk mengetahui apakah terdapat pengaruh dari semua variabel prediktor terhadap variabel respon (Fitrial & Fatikhurriq, 2020). Pengujian ini dilakukan menggunakan metode uji *G*. Hipotesis uji simultan pada model PR:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

(variabel prediktor secara simultan tidak berpengaruh terhadap variabel respon)

$$H_1 : \text{paling sedikit terdapat satu } \beta_j \neq 0 ; j = 1, 2, 3, \dots, p$$

(minimal terdapat satu variabel prediktor berpengaruh secara simultan terhadap variabel respon)

Statistik uji yang digunakan ditunjukkan pada persamaan 11.

$$\begin{aligned} G &= -2 \ln \Lambda \\ &= -2 \ln \frac{L(\hat{\omega})}{L(\hat{\Omega})} \quad \dots(11) \end{aligned}$$

$$= 2 \left( L(\hat{\Omega}) - L(\hat{\omega}) \right)$$

$$= 2 \sum_{i=1}^n \left( (y_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - (y_i \hat{\beta}_0 - \exp(\hat{\beta}_0)) \right)$$

Kriteria pengujiannya adalah  $H_0$  ditolak jika nilai  $G > \chi^2_{(\alpha;p)}$  (Masfian, dkk, 2016).

Selanjutnya dilakukan pengujian parameter model PR secara parsial. Uji parsial digunakan untuk menguji setiap  $\beta_j$  secara individu untuk mengetahui variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon (Hocking, 1996):

$$H_0 : \beta_j = 0$$

(variabel prediktor ke- $j$  tidak berpengaruh terhadap variabel respon)

$$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p$$

(variabel prediktor ke- $j$  berpengaruh terhadap variabel respon)

Statistik yang digunakan adalah statistik uji  $Z$  ditunjukkan oleh persamaan 12.

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{12}$$

dengan

$\hat{\beta}_j$  : taksiran parameter model PR

$SE(\hat{\beta}_j)$  : taksiran standar error dari  $\hat{\beta}_j$

Kriteria dari pengujian tersebut adalah  $H_0$  ditolak jika nilai  $|Z| > Z_{\alpha/2}$ ,  $\alpha$  merupakan taraf signifikansi yang digunakan pada model.

### 2.6 Underdispersion dan Overdispersion Pada Model PR

Menurut Irwan & Sari, D (2013), salah satu asumsi dalam analisis PR yaitu terjadinya *equidispersion* pada variabel respon. *Equidispersion* adalah rata-rata dari variabel respon sama dengan variansinya atau  $E(y_i) = Var(y_i)$ . Namun yang sering terjadi adalah kasus *overdispersion* dan *underdispersion*. *Overdispersion* terjadi jika nilai variansi lebih besar dari rata-rata ( $Var(y_i) > E(y_i)$ ), sedangkan jika nilai variansi lebih kecil dari rata-rata disebut *underdispersion* ( $Var(y_i) < E(y_i)$ ). Jika model PR diaplikasikan pada data *overdispersion* atau *underdispersion* akan menyebabkan standar error *underestimate* yang mengakibatkan beberapa variabel prediktor menjadi tidak signifikan. Menurut Darnah (2011), parameter dispersi ( $\delta$ ) dapat dihitung menggunakan persamaan berikut:

$$\delta = \frac{Dev}{db} \tag{13}$$

dengan  $db = n - p$ . Menurut Rashwan & Kamel (2011), *Dev* didefinisikan sebagai

$$Dev = 2 \sum_{i=1}^n y_i \ln \left( \frac{y_i}{\hat{\mu}} \right) \tag{14}$$

Jika  $\delta$  bernilai lebih dari satu, maka pada data terjadi *overdispersion*. Sebaliknya, jika  $\delta$  bernilai kurang dari satu, maka pada data terjadi *underdispersion*.

### 2.6 Generalized Poisson Regression

Jika diketahui hubungan antara variabel respon  $Y$  dan  $p$  buah variabel prediktor  $X_1, X_2, \dots, X_p$  dengan banyaknya sampel  $n$  yaitu  $\{(x_{1i}, x_{2i}, \dots, x_{pi}, y_i); i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p\}$ . Pengamatan ke- $i$  dari variabel  $X_1, X_2, \dots, X_p$  adalah  $x_{1i}, x_{2i}, \dots, x_{pi}$  dan pengamatan ke- $i$  dari variabel  $Y$  adalah  $y_i$ . Fungsi peluang *Generalized Poisson Regression* (GPR) adalah

$$f(y_i; \mu_i; \delta) = \left( \frac{\mu_i}{1 + \delta \mu_i} \right)^{y_i} \frac{(1 + \delta y_i)^{y_i - 1}}{y_i!} \exp \left[ \frac{-\mu_i(1 + \delta y_i)}{1 + \delta \mu_i} \right] \tag{15}$$

dengan  $E(y_i) = \mu_i$  dan  $Var(y_i) = \mu_i + (1 + \delta \mu_i)^2$ .  $\delta$  merupakan parameter dispersi (Hilbe, 2011). Sehingga model GPR sebagai berikut.

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*) + \varepsilon_i \tag{16}$$

$$\mu_i = \exp(\beta_0^* + \beta_1^* x_{1i} + \beta_2^* x_{2i} + \dots + \beta_p^* x_{pi}) + \varepsilon_i$$

Estimasi parameter model GPR menggunakan metode MLE. Fungsi *likelihood* dari fungsi peluang GPR pada persamaan 15 adalah

$$L(\boldsymbol{\beta}^*, \delta) = \prod_{i=1}^n \left( \frac{\mu_i}{1 + \delta \mu_i} \right) \prod_{i=1}^n \frac{(1 + \delta y_i)^{y_i - 1}}{y_i!} \exp \left( \sum_{i=1}^n \left[ \frac{-\mu_i(1 + \delta y_i)}{1 + \delta \mu_i} \right] \right) \quad \dots(17)$$

Selanjutnya nilai  $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)$  disubsitusikan ke dalam persamaan 17 sehingga diperoleh fungsi *likelihood*

$$L(\boldsymbol{\beta}^*, \delta) = \prod_{i=1}^n \left( \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)}{1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)} \right) \prod_{i=1}^n \frac{(1 + \delta y_i)^{y_i - 1}}{y_i!} \exp \left( \sum_{i=1}^n \left[ \frac{-\exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)(1 + \delta y_i)}{1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)} \right] \right) \quad \dots(18)$$

Jika  $\ell(\boldsymbol{\beta}^*, \delta) = \ln L(\boldsymbol{\beta}^*, \delta)$  maka fungsi *log likelihood* dari persamaan 18 adalah:

$$\begin{aligned} \ell(\boldsymbol{\beta}^*, \delta) &= \sum_{i=1}^n \left\{ y_i (\mathbf{x}_i^T \boldsymbol{\beta}^*) - y_i \ln(1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)) + (y_i - 1) \ln(1 + \delta y_i) - \ln(y_i!) + \right. \\ &\quad \left. - \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)(1 + \delta y_i) (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))^{-1} \right\} \quad \dots(19) \\ &= \sum_{i=1}^n \left\{ y_i \ln(\exp(\exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))) - y_i (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)) + (y_i - 1) \ln(1 + \delta y_i) + \right. \\ &\quad \left. - \ln(y_i!) - \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)(1 + \delta y_i)}{1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*)} \right\} \end{aligned}$$

Selanjutnya persamaan 19 diturunkan secara parsial terhadap  $\boldsymbol{\beta}^*$  yang ditunjukkan pada persamaan 20.

$$\frac{\partial \ell(\boldsymbol{\beta}^*, \delta)}{\partial \boldsymbol{\beta}^*} = \sum_{i=1}^n \left\{ y_i \mathbf{x}_i - \delta y_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*) (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))^{-1} - (1 + \delta y_i) \right\} \quad \dots(20)$$

Untuk mendapatkan penaksir parameter  $\delta$  maka persamaan 19 diturunkan terhadap  $\delta$  dan disamakan dengan nol. Sehingga bentuk turunan yang diperoleh adalah

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}^*, \delta)}{\partial \delta} &= \sum_{i=1}^n \left\{ y_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*) (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))^{-1} + y_i (y_i - 1) (1 + \delta y_i)^{-1} - \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*) \right. \\ &\quad \left. \left[ y_i (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))^{-1} - (1 + \delta y_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*) (1 + \delta \exp(\mathbf{x}_i^T \boldsymbol{\beta}^*))^{-2} \right] \right\} \quad \dots(21) \end{aligned}$$

Persamaan 19 dan persamaan 20 merupakan persamaan yang eksplisit sehingga memerlukan metode iterasi dalam penyelesaiannya seperti metode iterasi *Fisher-Scoring* pada persamaan 10 (Saputro, dkk, 2021). Pengujian hipotesis parameter model GPR meliputi uji simultan dan uji parsial. Uji simultan bertujuan untuk mengetahui adanya pengaruh variabel prediktor terhadap variabel respon secara simultan. Sedangkan, uji parsial bertujuan untuk mengetahui adanya pengaruh dari masing-masing variabel prediktor terhadap variabel respon secara individu. Uji simultan dilakukan dengan menggunakan uji  $G^*$  dengan hipotesis:

$$H_0 : \beta_1^* = \beta_2^* = \dots = \beta_p^* = 0$$

(variabel prediktor secara simultan tidak berpengaruh terhadap variabel respon)

$$H_1 : \text{paling sedikit terdapat satu } \beta_j^* \neq 0; j = 1, 2, 3, \dots, p$$

(minimal terdapat satu variabel prediktor berpengaruh secara simultan terhadap variabel respon)

Statistik uji yang digunakan ditunjukkan pada persamaan 22.

$$\begin{aligned} G^* &= -2 \ln \Lambda \quad \dots(22) \\ &= -2 \ln \frac{L(\hat{\omega}^*)}{L(\hat{\Omega}^*)} \\ &= 2 (L(\hat{\Omega}^*) - L(\hat{\omega}^*)) \\ &= 2 \left[ \sum_{i=1}^n \left\{ y_i \ln(\exp(\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^*))) - y_i (1 + \delta \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^*)) + (y_i - 1) \ln(1 + \delta y_i) + \right. \right. \\ &\quad \left. \left. - \ln(y_i!) - \frac{\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^*)(1 + \delta y_i)}{1 + \delta \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^*)} \right\} \right] - \left[ \sum_{i=1}^n \left\{ y_i \ln(\exp(\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_0))) + \right. \right. \end{aligned}$$

$$-y_i \left( 1 + \hat{\delta} \exp(\mathbf{x}_i^T \hat{\beta}_0) \right) + (y_i - 1) \ln(1 + \hat{\delta} y_i) - \ln(y_i!) - \frac{\exp(\mathbf{x}_i^T \hat{\beta}_0)(1 + \hat{\delta} y_i)}{1 + \hat{\delta} \exp(\mathbf{x}_i^T \hat{\beta}_0)} \Bigg\}$$

Kriteria pengujiannya adalah  $H_0$  ditolak jika  $G^* > \chi^2_{(\alpha;p)}$  (Masfian, dkk, 2016).

Selanjutnya dilakukan pengujian parameter model secara parsial. Statistik uji yang digunakan untuk uji parsial, yaitu uji  $Z^*$  dengan hipotesis (Hocking, 1996):

$$H_0 : \beta_j^* = 0$$

(variabel prediktor ke- $j$  tidak berpengaruh terhadap variabel respon)

$$H_1 : \beta_j^* \neq 0, j = 1, 2, \dots, p$$

(variabel prediktor ke- $j$  berpengaruh terhadap variabel respon)

Statistik yang digunakan adalah statistik uji  $Z$  ditunjukkan oleh persamaan 23

$$Z^* = \frac{\hat{\beta}_j^*}{SE(\hat{\beta}_j^*)} \quad \dots(23)$$

dengan

$\hat{\beta}_j^*$  : taksiran parameter model GPR

$SE \hat{\beta}_j^*$  : taksiran standar error dari  $\hat{\beta}_j^*$

Kriteria pengujian adalah  $H_0$  ditolak jika  $|Z^*| > Z_{\alpha/2}$ ,  $\alpha$  merupakan taraf signifikansi yang digunakan.

### 2.7 Akaike Information Criterion

Jika model regresi telah didapatkan selanjutnya adalah membandingkan model tersebut untuk mencari model terbaik yang dapat digunakan. Terdapat beberapa metode pengukuran yang bisa digunakan untuk mencari model terbaik salah satunya adalah *Akaike Information Criterion* (AIC) (Hilbe, 2011). AIC adalah salah satu metode untuk memilih model regresi terbaik yang ditemukan oleh Akaike dan Schwarz dimana metode AIC didasarkan pada metode MLE (Fathurahman, 2009). AIC dapat digunakan untuk memperoleh model terbaik dari beberapa metode dengan asumsi Poisson. AIC dinyatakan sebagai berikut:

$$AIC = -2\ell(\hat{\beta}) + 2K \quad \dots(24)$$

$\ell(\hat{\beta})$  adalah nilai *ln-likelihood* dari parameter model regresi dan  $K$  adalah banyaknya parameter dalam membangun model. Model terbaik adalah model yang memiliki nilai AIC terkecil (Fadmi, 2015).

### 2.8 Tuberkulosis Paru

*Tuberculosis* atau tuberkulosis paru (TB paru) merupakan penyakit infeksi menular yang disebabkan oleh *Mycobacterium tuberculosis* dan paling sering bermanifestasi di paru. Mikobakterium ini ditularkan melalui droplet di udara, sehingga seorang penderita TB paru merupakan sumber penyebab penularan TB paru pada populasi di sekitarnya (Kementerian Kesehatan RI, 2014).

Terdapat 2 faktor penyebab penyakit TB paru, yaitu:

1. Faktor karakteristik individu  
Faktor karakteristik individu antara lain faktor umur, jenis kelamin, tingkat pendidikan, pekerjaan, kebiasaan merokok, status gizi, kondisi sosial ekonomi, dan perilaku
2. Faktor risiko lingkungan  
Faktor risiko lingkungan antara lain kepadatan hunian, pencahayaan, ventilasi, kondisi rumah, kelembapan udara, suhu, dan ketinggian wilayah (Mariana & Chairani, 2017).

### 2.9 Faktor-Faktor Penyebab Tuberkulosis Paru (TB paru)

Kepadatan penduduk memiliki definisi yaitu membagi jumlah penduduk suatu wilayah dengan luas wilayahnya. Jika luas wilayah tidak sebanding dengan jumlah penduduk maka lingkungan

tersebut menjadi *over crowded* dan tidak sehat dikarenakan minimnya konsumsi O<sub>2</sub> bagi para penduduk serta penyebaran penyakit menular khususnya melalui udara akan semakin mudah (Mawardi, dkk, 2019).

Sesuai dengan Kepmenkes RI Nomor 829/MENKES/SK/VII/1999 tentang persyaratan kesehatan perumahan, syarat rumah sehat adalah rumah yang memiliki kriteria sanitasi lingkungan rumah minimal akses air minum, akses jamban sehat, lantai, pencahayaan dan ventilasi. Sanitasi lingkungan rumah berpengaruh terhadap kesehatan keluarga tersebut. Jika kondisi sanitasi lingkungan rumah baik maka kemungkinan kuman *Mycrobacterium tuberculosis* untuk berkembang biak semakin kecil (Suyani, dkk, 2020).

Salah satu syarat rumah sehat adalah memiliki lantai yang bersih dan kedap air seperti terbuat dari bahan kayu, semen, keramik atau ubin (Mawardi, dkk, 2019). Lantai rumah yang hanya terbuat dari tanah dan tidak dilapisi oleh bahan apapun akan menjadi tempat yang lembab dan menyebabkan masa daya tahan kuman *Mycrobacterium tuberculosis* menjadi lebih lama (Perdana & Putra, 2018).

Menurut Crofton (2002), kompleks kemiskinan adalah wilayah yang sangat mudah dalam penyebaran penyakit TB paru. Hal tersebut terjadi dikarenakan penduduk miskin sulit dalam memenuhi kebutuhan hidupnya, seperti makanan sehat, fasilitas sarana kesehatan dikarenakan tidak adanya biaya, terbatasnya perilaku hidup bersih dan sehat dan lain-lain (Mawardi, dkk, 2019).

## 2.10 Metode Analisis

Model analisis GPR diterapkan pada data TB paru di Indonesia tahun 2020. Variabel prediktor yang diduga berpengaruh terhadap jumlah kasus TB paru di Indonesia adalah kepadatan penduduk ( $x_1$ ), persentase rumah tangga yang memiliki akses sanitasi layak ( $x_2$ ), jumlah penduduk miskin ( $x_3$ ), persentase lantai rumah tidak kedap air ( $x_4$ ), persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ), dan persentase rumah tangga yang memiliki akses terhadap air minum layak ( $x_6$ ). Tahapan analisis data pada penelitian ini adalah sebagai berikut:

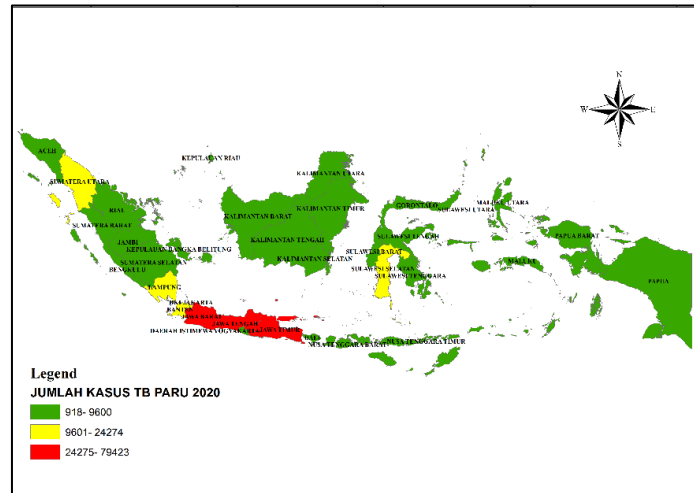
1. Mendeskripsikan gambaran umum dari data jumlah kasus TB paru serta variabel-variabel yang diduga memengaruhi.
2. Melakukan pengujian distribusi Poisson.
3. Mendeteksi adanya multikolinieritas.
4. Melakukan pemodelan PR.
5. Melakukan uji signifikansi parameter model PR yaitu uji simultan dan uji parsial.
6. Melakukan pengujian *overdispersion*.
7. Melakukan pemodelan dengan GPR.
8. Melakukan uji signifikansi parameter model GPR yaitu uji simultan dan uji parsial.
9. Memilih model GPR terbaik berdasarkan nilai AIC terkecil.
10. Menarik kesimpulan.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Statistika Deskriptif

Statistika deskriptif adalah metode statistika yang berkenaan dengan pengumpulan dan penyajian suatu data dengan hanya memberikan gambaran pada data yang ada dan tidak dapat memberikan generalisasi dan kesimpulan apapun tentang data yang lebih besar (populasi) (Nohe, 2013). Statistika deskriptif pada penelitian ini ditunjukkan pada Gambar 1.





Gambar 1. Jumlah kasus TB paru di Indonesia tahun 2020

Gambar 1 menunjukkan bahwa peneliti membagi kedalam tiga kategori yaitu zona hijau, zona kuning dan zona merah. Provinsi dengan jumlah kasus TB paru pada zona merah dimana jumlah kasus TB paru pada provinsi tersebut sangat tinggi dibandingkan provinsi lainnya yaitu sebanyak 24.275-79.423 kasus adalah Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur. Dan dari ketiga provinsi tersebut, Provinsi Jawa Barat merupakan provinsi dengan jumlah kasus TB paru tertinggi pada tahun 2022 yaitu sebanyak 79.423 kasus. Lalu, provinsi dalam kategori zona kuning dengan jumlah kasus 9.601-24.274 kasus adalah Provinsi Sumatera Utara, Provinsi Lampung, Provinsi DKI Jakarta, dan Provinsi Sulawesi Selatan. Provinsi lainnya berada pada zona hijau dimana jumlah kasus TB paru sebanyak 918-9.600 kasus

### 3.2 Uji Distribusi Poisson

Pengujian distribusi Poisson untuk mengetahui apakah data yang digunakan mengikuti distribusi Poisson atau tidak menggunakan uji *Kolmogorov-Smirnov*. Hasil uji distribusi Poisson dapat dilihat pada Tabel 1.

Tabel 1. Hasil Uji Komogorov-Smirnov

<i>D</i>	<i>D</i> <sub><math>\alpha=0,05</math></sub>
0,764705882	0,227

Hasil pengujian menunjukkan bahwa nilai  $D = 0,764705882 > D_{\alpha=0,05} = 0,227$  maka diputuskan  $H_0$  ditolak. Sehingga dapat disimpulkan bahwa data yang digunakan tidak mengikuti distribusi Poisson. Hal ini juga dapat mengindikasikan data yang digunakan mengalami *overdispersion* atau *underdispersion* dikarenakan asumsi pada distribusi Poisson yaitu nilai rata-rata dan variansi pada data respon sama tidak terpenuhi. Sebelum data dimodelkan menggunakan analisis PR, harus dilakukan pemeriksaan data terhadap multikolinearitas terlebih dahulu untuk mengetahui apakah data memenuhi asumsi yaitu data tidak mengalami multikolinieritas.

### 3.3 Pemeriksaan Multikolinieritas

Pemeriksaan multikolinieritas dilakukan untuk mengetahui apakah terdapat hubungan antar variabel prediktor. Untuk mendeteksi multikolinieritas dalam model regresi menggunakan nilai *Variance Inflation Factor* (VIF). Nilai VIF kurang dari 10 maka tidak terjadi multikolinieritas pada model regresi. Nilai VIF dari masing-masing variabel prediktor dapat dilihat pada Tabel 2.

Tabel 2. Nilai VIF Variabel Prediktor

Variabel	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
VIF	1,24	1,38	1,43	1,41	1,19	1,53

Tabel 2 menjelaskan bahwa tidak terjadi multikolinieritas antar variabel prediktor karena nilai VIF dari masing-masing variabel tidak lebih dari 10. Selanjutnya dilakukan analisis menggunakan PR dan pemeriksaan dispersi pada model PR yang didapatkan.

### 3.3 Pemodelan Kasus TB Paru di Indonesia Menggunakan Analisis PR

Analisis PR dilakukan untuk mendapatkan model PR dan lalu model tersebut diuji lebih lanjut dan mengetahui apakah data yang digunakan mengalami *overdispersion* atau *underdispersion*. Estimasi parameter model PR menggunakan metode MLE. Hasil estimasi parameter model PR menggunakan *software R* disajikan pada Tabel 3.

Tabel 3. Hasil Estimasi Parameter PR

Parameter	Estimator	Estimasi
$\beta_0$	$\hat{\beta}_0$	9,84
$\beta_1$	$\hat{\beta}_1$	$1,38 \times 10^{-4}$
$\beta_2$	$\hat{\beta}_2$	$2,71 \times 10^{-3}$
$\beta_3$	$\hat{\beta}_3$	$6,11 \times 10^{-7}$
$\beta_4$	$\hat{\beta}_4$	$-2,64 \times 10^{-2}$
$\beta_5$	$\hat{\beta}_5$	$-2,55 \times 10^{-2}$
$\beta_6$	$\hat{\beta}_6$	$8,15 \times 10^{-3}$

### 3.3 Pengujian Signifikansi Parameter Model PR

Pengujian secara simultan pada model PR menggunakan uji *G*. Hasil uji simultan model PR dengan bantuan *software R* dapat diketahui bahwa nilai  $G = 436.095,3 > \chi^2_{(0,05;6)} = 12,59$  maka  $H_0$  ditolak. Sehingga dapat disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh terhadap variabel respon.

Pengujian parsial model PR menggunakan uji *Z*. Hasil uji parsial model PR menggunakan *software R* sebagai berikut.

Tabel 4. Hasil Uji Parsial Model PR

Variabel	Nilai  Z	$Z_{\alpha/2}$	Keputusan
$x_1$	226,65	1,96	Signifikan
$x_2$	14,01	1,96	Signifikan
$x_3$	490,63	1,96	Signifikan
$x_4$	62,48	1,96	Signifikan
$x_5$	153,32	1,96	Signifikan
$x_6$	25,13	1,96	Signifikan

Berdasarkan Tabel 4, dapat disimpulkan bahwa variabel yang berpengaruh adalah kepadatan penduduk ( $x_1$ ), persentase rumah tangga yang memiliki akses sanitasi layak ( $x_2$ ), jumlah penduduk miskin ( $x_3$ ), persentase lantai rumah tidak kedap air ( $x_4$ ), persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ), dan persentase rumah tangga yang memiliki akses terhadap air minum layak ( $x_6$ ).

Berdasarkan hasil estimasi parameter pada Tabel 1, maka model PR jumlah kasus TB paru di Indonesia tahun 2020 dapat ditulis sebagai berikut.

$$\hat{\mu} = \exp(9,84 + (1,38 \times 10^{-4})x_{1i} + (2,71 \times 10^{-3})x_{2i} + (6,11 \times 10^{-7})x_{3i} + (-2,64 \times 10^{-2})x_{4i} + (-2,55 \times 10^{-2})x_{5i} + (8,15 \times 10^{-3})x_{6i})$$

Berdasarkan model PR dapat diketahui bahwa setiap pertambahan kepadatan penduduk ( $x_1$ ) sebanyak 1 jiwa per kilometer persegi dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika persentase rumah tangga yang memiliki akses sanitasi layak ( $x_2$ ) bertambah 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika rata-rata jumlah penduduk miskin ( $x_3$ ) bertambah 1 jiwa dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika persentase lantai rumah tidak kedap air ( $x_4$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus. Ketika persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus. Dan ketika persentase rumah tangga yang memiliki akses terhadap air minum layak ( $x_6$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus.

### 3.4 Pemeriksaan Overdispersion

Pemeriksaan *overdispersion* pada model PR jumlah kasus TB paru di setiap provinsi di Indonesia tahun 2020 pada Tabel 5.

Tabel 5. Hasil Uji Overdispersion

<i>Dev</i>	<i>db</i>	$\delta$
52.381	27	1.940,04

Berdasarkan Tabel 5, dapat diketahui bahwa terjadi *overdispersion* pada model PR data TB paru di Indonesia tahun 2020 dikarenakan hasil bagi nilai *deviance* terhadap derajat bebas lebih dari 1. Jika model PR data TB paru yang mengalami *overdispersion* digunakan untuk menggambarkan data TB paru Indonesia tahun 2020, maka taksiran parameter yang dihasilkan dari model PR tersebut akan tetap konsisten namun tidak akan efisien. Sehingga hasil taksiran parameter dari model PR tersebut menjadi bias dan kemungkinan untuk terjadinya kesalahan dalam menarik kesimpulan mengenai data yang diamati semakin besar. Oleh karena itu, digunakan model GPR untuk mengatasi *overdispersion* pada model PR.

### 3.5 Pemodelan kasus TB Paru di Indonesia Menggunakan Analisis GPR

GPR adalah salah satu analisis regresi untuk mengatasi kasus *overdispersion* atau *underdispersion* yaitu nilai rata-rata dan nilai variansi tidak sama. Estimasi parameter model GPR menggunakan metode MLE dengan bantuan *software R*. Hasil estimasi parameter dapat dilihat pada Tabel 6.

Tabel 6. Hasil Estimasi Parameter Poisson Regression

Parameter	Estimator	Estimasi
$\beta_0^*$	$\hat{\beta}_0^*$	8,77
$\beta_1^*$	$\hat{\beta}_1^*$	$1,25 \times 10^{-4}$
$\beta_2^*$	$\hat{\beta}_2^*$	$6,75 \times 10^{-3}$
$\beta_3^*$	$\hat{\beta}_3^*$	$5,95 \times 10^{-7}$
$\beta_4^*$	$\hat{\beta}_4^*$	$-3,56 \times 10^{-2}$
$\beta_5^*$	$\hat{\beta}_5^*$	$-2,12 \times 10^{-2}$
$\beta_6^*$	$\hat{\beta}_6^*$	$-1,06 \times 10^{-3}$

#### 1. Uji Signifikansi Parameter Secara Simultan Pada Model GPR

Pengujian secara simultan pada model GPR menggunakan uji  $G^*$ . Hasil uji simultan model PR dengan bantuan *software R* dapat diketahui bahwa nilai  $G^* = 51,33307 > \chi_{(0,05;6)}^2 = 12,59$  maka  $H_0$  ditolak. Sehingga dapat disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh terhadap variabel respon.

#### 2. Uji Signifikansi Parameter Secara Parsial Pada Model GPR

Pengujian parsial model GPR menggunakan uji  $Z^*$ . Hasil uji parsial model GPR menggunakan *software R* sebagai berikut.

Tabel 7. Hasil Uji Parsial Model GPR

Variabel	Nilai $ Z^* $	$Z_{\alpha/2}$	Keputusan
$x_1$	5,871	1,96	Signifikan
$x_2$	1,069	1,96	Tidak signifikan
$x_3$	12,883	1,96	Signifikan
$x_4$	2,259	1,96	Signifikan
$x_5$	3,692	1,96	Signifikan
$x_6$	0,102	1,96	Tidak signifikan

Berdasarkan Tabel 7, dapat disimpulkan bahwa variabel yang berpengaruh adalah kepadatan penduduk ( $x_1$ ), jumlah penduduk miskin ( $x_3$ ), persentase lantai rumah tidak kedap air ( $x_4$ ), dan persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ).

### 3.6 Pemilihan dan Interpretasi Model GPR Terbaik

Dari empat variabel prediktor yang berpengaruh dapat membentuk 15 kemungkinan model GPR dan akan dicari model GPR terbaik dengan nilai AIC terkecil. Berikut nilai AIC dari 15 model yang terbentuk dan ditunjukkan pada Tabel 8.

Tabel 8. Nilai AIC Model GPR

No	Kemungkinan Model (Y dengan X <sub>j</sub> )	Nilai AIC	Parameter yang Signifikan
1	$x_1$	683,71	$\hat{\beta}_0^*, \hat{\beta}_1^*$
2	$x_3$	655,89	$\hat{\beta}_0^*, \hat{\beta}_3^*$
3	$x_4$	686,11	$\hat{\beta}_0^*, \hat{\beta}_4^*$
4	$x_5$	684,55	$\hat{\beta}_0^*, \hat{\beta}_5^*$
5	$x_1, x_3$	650,06	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_3^*$
6	$x_1, x_4$	683,89	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_4^*$
7	$x_1, x_5$	681,30	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_5^*$
8	$x_3, x_4$	656,96	$\hat{\beta}_0^*, \hat{\beta}_3^*, \hat{\beta}_4^*$
9	$x_3, x_5$	654,30	$\hat{\beta}_0^*, \hat{\beta}_3^*, \hat{\beta}_5^*$
10	$x_4, x_5$	686,14	$\hat{\beta}_0^*, \hat{\beta}_4^*, \hat{\beta}_5^*$
11	$x_1, x_3, x_4$	651,10	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_3^*, \hat{\beta}_4^*$
12	$x_1, x_3, x_5$	644,62	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_3^*, \hat{\beta}_5^*$
13	$x_1, x_4, x_5$	682,86	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_4^*, \hat{\beta}_5^*$
14	$x_3, x_4, x_5$	653,80	$\hat{\beta}_0^*, \hat{\beta}_3^*, \hat{\beta}_4^*, \hat{\beta}_5^*$
15	$x_1, x_3, x_4, x_5$	*643,39	$\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\beta}_3^*, \hat{\beta}_4^*, \hat{\beta}_5^*$

\* Model terbaik

Dari kelima belas kemungkinan model GPR pada Tabel 8, dapat diketahui bahwa model terbaik dengan nilai AIC terkecil adalah model yang melibatkan empat variabel prediktor yaitu kepadatan penduduk ( $x_1$ ), jumlah penduduk miskin ( $x_3$ ), persentase lantai rumah tidak kedap air ( $x_4$ ), dan persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ) dengan nilai AIC sebesar 643,39. Lalu dilakukan estimasi parameter model GPR menggunakan metode MLE dengan bantuan *software R*. Hasil estimasi parameter dapat dilihat pada Tabel 9.

Tabel 9. Hasil Estimasi Parameter Model GPR Terbaik

Parameter	Estimator	Estimasi
$\beta_0^*$	$\hat{\beta}_0^*$	9,14
$\beta_1^*$	$\hat{\beta}_1^*$	$1,26 \times 10^{-4}$
$\beta_3^*$	$\hat{\beta}_3^*$	$6,08 \times 10^{-7}$
$\beta_4^*$	$\hat{\beta}_4^*$	$-3,05 \times 10^{-2}$
$\beta_5^*$	$\hat{\beta}_5^*$	$-1,96 \times 10^{-2}$

Berdasarkan hasil estimasi parameter dan pemilihan model GPR terbaik untuk jumlah kasus TB paru di Indonesia tahun 2020 dapat ditulis sebagai berikut.

$$\hat{\mu}_i = \exp(9,14 + (1,26 \times 10^{-4})x_{1i} + (6,08 \times 10^{-7})x_{3i} - (3,05 \times 10^{-2})x_{4i} - (1,96 \times 10^{-2})x_{5i})$$

Berdasarkan model GPR terbaik, dapat diketahui dapat diketahui bahwa setiap penambahan kepadatan penduduk ( $x_1$ ) sebanyak 1 jiwa per kilometer persegi dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika rata-rata jumlah penduduk miskin ( $x_3$ ) bertambah 1 jiwa dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika persentase lantai rumah tidak kedap air ( $x_4$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus, hal ini tidak sejalan dengan penelitian sebelumnya karena kondisi suatu waktu menyebabkan perbedaan data. Ketika persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus.

#### 4. SIMPULAN DAN SARAN

Model GPR terbaik pada data jumlah kasus TB paru di Indonesia tahun 2020 adalah sebagai berikut.

$$\hat{\mu}_i = \exp(9,14 + (1,26 \times 10^{-4})x_{1i} + (6,08 \times 10^{-7})x_{3i} - (3,05 \times 10^{-2})x_{4i} - (1,96 \times 10^{-2})x_{5i})$$

Berdasarkan model terbaik GPR maka dapat diketahui faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus TB paru di Indonesia pada tahun 2020 yaitu kepadatan penduduk, jumlah penduduk miskin, persentase lantai rumah tidak kedap air dan persentase tempat pengelolaan pangan yang memenuhi syarat. Setiap penambahan kepadatan penduduk ( $x_1$ ) sebanyak 1 jiwa per kilometer persegi dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika rata-rata jumlah penduduk miskin ( $x_3$ ) bertambah 1 jiwa dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami peningkatan sebesar 1 kasus. Ketika persentase lantai rumah tidak kedap air ( $x_4$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus. Dan ketika persentase tempat pengelolaan pangan yang memenuhi syarat ( $x_5$ ) bertambah sebanyak 1 persen dan nilai variabel lain tetap maka rata-rata kasus TB paru di Indonesia akan mengalami penurunan sebesar 1 kasus.

#### DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*. Second Edition. New York: John Wiley & Sons.
- Asriyanti, R., Yahya, I., Abapihi, B., Wibawa, Gusti Ngurah A., Rasdiyanah, dan Laome, L. (2022). *Penerapan Regresi Nonparametrik Spline Dalam Memodelkan Faktor-Faktor Yang Mempengaruhi Jumlah Kasus Tuberkulosis DI Sulawesi Tenggara*. Makalah dipresentasikan dalam Seminar Nasional Sains dan Terapan VI, Program Studi Statistika FMIPA Universitas Halu Oleo, Manado, 25 April.
- Astuti, C. C., Sumarminingsih, E. & Soehono, L. A. (2013). Perbandingan *Generalized Poisson Regression* dan *Negative Binomial Regression* Untuk Data Overdispersi Dan Underdispersi Pada Regresi Poisson. *Jurnal Mahasiswa Statistik*, 1(2), 105-108.
- Aulele, S. N. (2012). Pemodelan Jumlah Kematian Bayi di Provinsi Maluku Tahun 2010 Dengan Menggunakan Regresi Poisson. *Jurnal Barekeng*, 6(2), 23-27.
- Cahyandari, R. (2014). Pengujian Overdispersi Pada Model Regresi Poisson. *Statistika, Issue Statistika*, 14(2), 69-76.
- Darnah. (2011). Mengatasi Overdispersi pada Model Regresi Poisson dengan *Generalized Poisson Regression* I. *Jurnal Eksponensial*, 2,5-10.
- Darnah. (2016). Modelling of Filariasis in East Java with Poisson Regression and *Generalized Poisson Regression* Models. *AIP Conference Proceedings*, 1723, 50-62.
- Darnah, Utoyo, M. I. & Chamidah, N. (2019). Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator. *IOP Conference Series: Earth and Environmental Science*, 243, 1-7.
- Fadmi, F. R. (2015). Prediksi Jumlah Kasus Baru Kusta dengan Metode Geographically Weighted Poisson Regression (GWPR). *Jurnal Biometrika dan Kependudukan*, 4, 14-24.
- Fathurahman, M. (2009). Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. *Jurnal Informatika Mulawarman*, 4, 37-41.
- Fitrial, N. H. & Fatikhurizqi, A. (2020). *Pemodelan Jumlah Kasus Covid-19 di Indonesia dengan Pendekatan Regresi Poisson dan Regresi Binomial Negatif*. Makalah dipresentasikan dalam Seminar Nasional Official Statistics, Politeknik Statistika STIS, Jakarta, 23-24 September. 6, 65-72.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. Second Edition. Cambridge: Cambridge University Press.
- Hocking, R. (1996). *Methods and Applications of Linear Models*. New York: John Wiley and Sons, Inc.

- Irwan & Sari, D. (2013). *Pemodelan Regresi Poisson dan Binomial Negatif pada Kasus Kecelakaan Kendaraan Bermotor di Lalu Lintas Sumatera Barat*. Makalah dipresentasikan dalam Seminar Nasional Matematika dan Pendidikan Matematika 2013, Jurusan Pendidikan Matematika FMIPA Universitas Negeri Yogyakarta, Yogyakarta, 9 November.
- Ismail, N. & Jemain, A. A. (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Model. Malaysia: Causalty Actuarial Society Forum.
- Kementrian Kesehatan RI. (2014). *Pedoman Nasional Penanggulangan Tuberkulosis*. Jakarta: Kementrian Kesehatan RI.
- Lestari, R. D., Wulandari, S. P. & Purnadi. (2014). Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Penyakit Tuberkulosis di Jawa Timur dengan Pendekatan *Generalized Poisson Regression* dan Geographically Weighted Poisson Regression. *Jurnal Seni dan Seni Pomits*, 3, 188-193.
- Lungan, R., (2006). *Aplikasi Statistika dan Hitung Peluang*. Yogyakarta: Graha Ilmu.
- Mariana, D. & Chairani, M. (2017). Kepadatan Hunian, Ventilasi dan Pencahayaan Terhadap Kejadian TB Paru di Wilayah Kerja Puskesmas Binanga Kabupaten Mamuju Sulawesi Barat. *Jurnal Kesehatan Manarang*, 3, 75-80.
- Masfian, I., Yuniarti, D. & Hayati, M. N. (2016). Penerapan *Generalized Poisson Regression I* Untuk Mengatasi Overdispersi Pada Regresi Poisson (Studi Kasus: Pemodelan Jumlah Kasus Kanker Serviks di Provinsi Kalimantan Timur). *Jurnal Eksponensial*, 7(1), 59-65.
- Mawardi, Sambera, R. & Hamisah, I. (2019). Studi Hubungan Antara Faktor Lingkungan Fisik Rumah Dengan Penderita TB Paru BTA di Aceh Selatan. *Jurnal Serambi Engineering*, 4(1), 406-415.
- Myers, R. (1990). *Classical and Model Regression with Application*, Second Edition. Boston: PWSKENT Publishing Company.
- Naben, A. X., Suhartono & Nurjazuli. (2013). Kebiasaan Tinggal di Rumah Etnis Timor Sebagai Faktor Risiko Tuberkulosis Paru. *Jurnal Kesehatan Lingkungan Indonesia*, 12(1), 10-21.
- Nohe, D. A. (2013). *Biostatistika 1*. Jakarta Barat: Halaman Moeka Publishing.
- Perdana, A. & Putra, S. (2018). Hubungan Faktor Lingkungan Fisik Rumah terhadap Kejadian TB Paru di Wilayah Kerja Puskesmas Panjang. *Jurnal Kesehatan*, 9(1), 46-50.
- Priyatno, D. (2013). *Analisis Korelasi, Regresi dan Multivariate dengan SPSS*. Yogyakarta: Gava Media.
- Rashwan, N. A. & Kamel, M. M. (2011). Using Generalized Poisson Log Linear Regression Models in Analyzing Two-Way Contingency Tables. *Applied Mathematical Sciences*, 5(5), 213-222.
- Safrida, N., Ispriyanti, D. & Widiharih, T. (2013). Aplikasi Model Regresi Poisson Tergeneralisasi Pada Kasus Angka Kematian Bayi Di Jawa Tengah Tahun 2007. *Jurnal Gaussian*, 2(4), 361-368.
- Saputro, D. R. S., Susanti, A. & Pratiwi, N. B. I. (2021). The Handling of Overdispersion on Poisson Regression Model with The *Generalized Poisson Regression* Model. Makalah dipublikasikan dalam The Third International Conference On Mathematics: Education, Theory And Application, Surakarta, 20 Oktober, 2326, 1-8.
- Simarmata, R. T. & Ispriyanti, D. (2011). Penanganan Overdispersi Pada Model Regresi Poisson Menggunakan Model Regresi Binomial Negatif. *Media Statistika*, 4(2), 95-104.
- Suyani, D., Putri, S. M., Nahdiyani, F., Ramadini, E. S., Rizky, N. T., Fadila, E. N. & Nisa, D. M. (2020). Peningkatan Kepemilikan Rumah Sehat di Rt 01 dan 03 Dusun Modalan Kecamatan Banguntapan Bantul. *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian Kepada Masyarakat*, 4(2), 191-196.
- Walpole, R. (1995). *Pengantar Statistika*. Jakarta: PT Gramedia Pustaka Umum.
- Wasilaine, T. L., Talakua, M. W. & Lesnussa, Y. A. (2014). *Model Regresi Ridge Untuk Mengatasi Model Regresi Linier Berganda Yang Mengandung Multikolinieritas (Studi Kasus: Data Pertumbuhan Bayi di Kelurahan Namaelo RT 001, Kota Masohi)*. *Jurnal Berekeng*, 8(1), 31-37.
- World Health Organization. (2018). *Global Tuberculosis Report 2017*. Perancis: WHO Press.
- World Health Organization. (2021). *Global Tuberculosis Report 2020*. Perancis: WHO Press.

Yotenka, R. & Banapon, A. (2019). *Modelling the Number of Tuberculosis (TB) Cases in Indonesia using Poisson Regression and Negative Binomial Regression*. Makalah dipresentasikan dalam International Seminar on Science and Technology (ISSTEC 2019), Jakarta.