

# ANALISIS JUMLAH KASUS MALARIA DI WILAYAH SUMATERA MENGUNAKAN GEOGRAPHICALLY WEIGHTED ZERO- INFLATED POISSON REGRESSION\*

Rahmat Kevin Praditia<sup>1</sup>, Dian Agustina<sup>2‡</sup>, and Dyah Setyo Rini<sup>3</sup>

<sup>1</sup>Program Studi S1 Statistika Universitas Bengkulu (UNIB), Indonesia, rahmatkevin3@gmail.com

<sup>2</sup>Program Studi S1 Statistika Universitas Bengkulu (UNIB), Indonesia, dianagustina@unib.ac.id

<sup>3</sup>Program Studi S1 Statistika Universitas Bengkulu (UNIB), Indonesia, dyah.setyorini@unib.ac.id

‡corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**

**Vol 4 No 4 (2020), 638 - 648**

Copyright © 2020 Rahmat Kevin Praditia, Dian Agustina, and Dyah Setyo Rini. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **Abstract**

A method that can be used if there is a spatial factor and if overdispersion happens in a count data is Geographically Weighted Zero-Inflated Poisson Regression (GWZIPR). This research aimed to analyze the number of malaria cases in every regency/city of Sumatra Land using the GWZIPR method and distribution mapping of factors affecting the number of malaria cases in Sumatra Land. Data involved in this research was the number of malaria cases as the response variable and the predictor variable as a percentage of households that have access to proper sanitation, a percentage of households that have access to proper water resources, and a percentage of the number of public health centers. The results were for each area which had distinctive models based on significant variables. The distribution mapping of factors affecting the number of malaria cases in every regency/city was commonly divided into three groups based on significant variables on ln and logit models. The mapping did not shape a spreading pattern or each regency/city in that group because the geographical locations were close to each other. GWZIPR method in this research was better than the ZIP Regression method because it produced the least AIC value.

**Keywords:** AIC, GWZIPR, spatial, overdispersion, ZIP Regression.

## **1. Pendahuluan**

Malaria adalah penyakit yang diakibatkan oleh parasit *Plasmodium* ketika menginfeksi sel darah merah (eritrosit). Parasit akan menularkan dari individu ke individu lainnya

---

\* Received Oct 2020; Accepted Des 2020; Published online on Des 2020

melalui gigitan nyamuk *Anopheles* betina. Parasit yang menyebabkan malaria berasal dari genus *Plasmodium*, yang terdiri dari 4 spesies: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, dan *Plasmodium ovale*. Jenis *Plasmodium* terbaru yang ditemukan yaitu *Plasmodium knowlesi* melalui metode *Polymerase Chain Reaction* (PCR) ([DIRJEN PPPL], 2014).

Penyakit malaria masih sering ditemukan diseluruh Provinsi Indonesia khususnya di Wilayah Sumatera. Wilayah Sumatera tergolong dalam stratifikasi sedang untuk jumlah kasus malaria dengan pulau lainnya seperti Kalimantan dan Sulawesi ([KEMENKES], 2011). Jumlah kasus malaria adalah variabel respon yang banyak mengandung nilai nol (*excess zeroes*) di Wilayah Sumatera, sehingga variabel respon tersebut menyebar Poisson. Analisis regresi Poisson dapat melihat adanya hubungan variabel respon yang menyebar Poisson dan variabel prediktor (Agresti, 1990).

Banyaknya kejadian nol terhadap jumlah kasus malaria membuat asumsi equidispersi tidak terpenuhi, artinya bahwa nilai ragam tidak sama dengan nilai rata-rata sehingga bisa menyebabkan overdispersi. Jika terdapat kasus yang mengalami overdispersi, maka model regresi *Zero Inflated-Poisson* (ZIP) lebih disarankan untuk menganalisis data yang bersifat cacah (*count*) dan kejadian nol dalam proporsi besar (*zero inflation*) (Lambert, 1992).

Regresi ZIP merupakan sebuah analisis regresi yang menghasilkan penduga parameter yang bersifat global. Data yang diambil dari beberapa wilayah mewakili kondisi yang berbeda dari masing-masing wilayah. Malaria di Indonesia masih ditemukan tersebar luas di semua pulau yang ada di Indonesia dengan derajat dan keparahan infeksi yang bervariasi (Siswanto & Thamrin, 2020).

Terdapatnya faktor spasial atau geografis dalam penyebaran malaria di berbagai wilayah. Perbedaan kondisi geografis di setiap wilayah, karakteristik masyarakat dan perekonomian masing-masing wilayah dapat menyebabkan adanya heterogenitas spasial (Purhadi *et al.*, 2015). Untuk kasus yang terdapat heterogenitas spasial, maka metode tersebut dikembangkan dengan menambahkan faktor spasial yaitu *Geographically Weighted Zero-Inflated Poisson Regression* (GWZIPR). Metode tersebut dapat digunakan untuk melihat persebaran kejadian malaria di Wilayah Sumatera. Oleh karena itu penggunaan *Geographically Weighted Zero-Inflated Poisson Regression* akan menghasilkan hasil yang lebih baik.

## 2. Metodologi

### 2.1 Bahan dan Data

Data penelitian merupakan data sekunder yang diperoleh melalui *website* Badan Pusat Statistik masing-masing Provinsi. Data penelitian ini adalah data kabupaten/kota di setiap Provinsi di Wilayah Sumatera tahun 2018. Data tersebut terdiri dari jumlah kasus malaria ( $Y$ ), persentase rumah tangga yang memiliki akses terhadap sanitasi layak ( $X_1$ ), persentase rumah tangga yang memiliki akses sumber air minum layak ( $X_2$ ), dan persentase jumlah puskesmas ( $X_3$ ).

### 2.2 Metode Penelitian

Tahapan analisis penelitian ini adalah sebagai berikut:

### 1. Mengidentifikasi asumsi overdispersi dengan melihat nilai ragam dan nilai rata-rata dari variabel respon

Overdispersi adalah suatu keadaan ketika nilai ragam lebih besar dibandingkan dengan nilai rata-ratanya pada variabel respon (Cameron & Trivedi, 1990). Estimasi parameter yang dilakukan pada data diskrit yang mengalami overdispersi akan mengakibatkan koefisien regresi tetap konsisten tetapi tidak efisien karena berdampak pada nilai standar *error* yang tinggi. Overdispersi atau underdispersi akan menghasilkan model yang kurang tepat karena nilai devians model yang dihasilkan sangat besar (Hilbe, 2011).

### 2. Mengidentifikasi asumsi multikolinieritas dengan melihat kriteria VIF untuk semua variabel prediktor

Untuk mendeteksi terdapatnya asumsi multikolinieritas pada suatu data adalah dengan mengetahui nilai *Varian Inflation Factor* (VIF). Jika salah satu nilai VIF melebihi angka 5 atau 10, hal ini mengindikasikan bahwa koefisien regresi merupakan penduga yang tidak baik dikarenakan multikolinieritas (Montgomery et al., 2012). Persamaan untuk mencari nilai VIF dituliskan sebagai berikut (Kutner et al., 2005):

$$VIF_j = \frac{1}{1-R_j^2} \quad (1)$$

Keterangan:

$R_j^2$ : koefisien determinasi antara  $X_j$  dengan variabel prediktor yang lain.

$$R_j^2 = \frac{[\sum_{j=1}^p (X_j - \bar{X}_j)(X_{j'} - \bar{X}_{j'})]}{\sum_{j=1}^p (X_j - \bar{X}_j)^2 (X_{j'} - \bar{X}_{j'})^2}, j = 1, 2, \dots, p \text{ dan } j \neq j'$$

### 3. Pemodelan jumlah kasus malaria menggunakan Regresi ZIP

Regresi *Zero-Inflated Poisson* (ZIP) merupakan modifikasi dari regresi Poisson yang memungkinkan untuk pencacahan dengan kelebihan nilai nol dalam data. Kejadian dengan kelebihan nilai nol seperti ini sangat sering terjadi dalam praktek. Model regresi ZIP merupakan kombinasi dari sebaran Poisson dengan sebaran *Zero-Inflated* (Cameron & Trivedi, 1990). Menurut Lambert (1992), model regresi ZIP untuk setiap pengamatan variabel respon  $Y_1, Y_2, \dots, Y_n$  yang saling bebas adalah:

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } (\omega_i) \\ \text{Poisson}(\mu), & \text{dengan peluang } (1 - \omega_i) \end{cases} \quad (2)$$

Fungsi kepekatan peluang ZIP adalah:

$$P(Y = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu} & , y_i = 0 \\ \frac{(1 - \omega_i)e^{-\mu}\mu^{y_i}}{y_i!}, & , y_i > 0 \end{cases} \quad (3)$$

Sehingga,  $Y \sim ZIP(\mu, \omega_i)$ . Karena  $\omega_i$  adalah besaran peluang, maka nilainya pada rentang  $0 \leq \omega_i \leq 1$ . Dengan parameter  $\mu = [\mu_1 \mu_2 \dots \mu_n]^T$  dan  $\omega = [\omega_1 \omega_2 \dots \omega_n]^T$ , maka model gabungan antara  $\mu$  dan  $\omega$  adalah:

$$\ln(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta} \text{ dan } \text{logit}(\omega_i) = \ln\left(\frac{\omega_i}{1 - \omega_i}\right) = \mathbf{X}_i^T \boldsymbol{\gamma} \quad (4)$$

$$\mu_i = e^{\mathbf{X}_i^T \boldsymbol{\beta}} \quad (5)$$

$$\omega_i = \frac{e^{\mathbf{X}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{X}_i^T \boldsymbol{\gamma}}} \quad (6)$$

$$(1 - \omega_i) = \frac{1}{1 + e^{x_i^T \gamma}} \tag{7}$$

**4. Mengidentifikasi asumsi dependensi spasial dengan uji Moran's I**

Dependensi spasial mengindikasikan terdapatnya pengaruh pada pengamatan antar lokasi. Untuk melakukan uji dependensi spasial yaitu menggunakan uji Moran's I. Rentang nilai uji Moran's I adalah -1 dan 1.

Hipotesis yang digunakan adalah:

$$H_0 : \lambda = 0 \text{ (tidak terdapat dependensi spasial)}$$

$$H_1 : \lambda \neq 0 \text{ (terdapat dependensi spasial)}$$

Statistik uji Moran's I adalah:

$$Z_{hit} = \frac{I - E(I)}{\sqrt{var(I)}} \tag{8}$$

Tolak  $H_0$  jika nilai  $|Z_{hit}| > Z_{\alpha/2}$  atau nilai  $p - value < \alpha$ .

**5. Mengidentifikasi asumsi heterogenitas spasial dengan uji Koenker-Baset**

Terdapatnya perbedaan karakteristik pengamatan antar lokasi maka akan dilakukan uji heterogenitas spasial. Uji yang digunakan untuk melihat apakah adanya heterogenitas spasial bisa menggunakan uji *Koenker-Baset*. Uji *Breusch-Pagan* sensitif terhadap asumsi normalitas.

Menurut Gujarati (2004), uji *Koenker-Baset* adalah pengujian yang berdasarkan hasil residual dugaan kuadrat ( $\hat{\epsilon}_i^2$ ). Hasil dari residual dugaan kuadrat tersebut akan diregresi terhadap kuadrat hasil dugaan variabel respon dari regresi model awal. Sehingga akan menghasilkan model regresi baru.

Hipotesis yang digunakan adalah:

$$H_0 : \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_n = \sigma^2 \text{ (tidak terdapat heterogenitas spasial)}$$

$$H_1 : \sigma^2_i \neq \sigma^2 \quad ; i = 1, 2, \dots, n \text{ (terdapat heterogenitas spasial)}$$

Statistik uji *Koenker-Baset* adalah:

$$Z_{hit} = \frac{\hat{\alpha}_i}{SE(\hat{\alpha}_i)} \tag{9}$$

Tolak  $H_0$  jika nilai  $|Z_{hit}| > Z_{\alpha/2}$  atau nilai  $p - value < \alpha$ .

**6. Menghitung jarak *Euclidean* antar lokasi berdasarkan koordinat  $(u_i, v_i)$  atau *latitude* dan *longitude* setiap kabupaten/kota di Wilayah Sumatera**

Sebelum menentukan matriks pembobot, diperlukan jarak antar pengamatan lokasi  $(u_i, v_i)$  dengan lokasi  $(u_j, v_j)$ . Jarak yang digunakan dalam menentukan fungsi pembobot bisa menggunakan jarak *Euclidean*. Persamaan jarak *Euclidean* sebagai berikut:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{10}$$

Keterangan :

$d_{ij}$  : jarak *Euclidean*

$u_i$  : *latitude* (garis bujur) pada lokasi ke- $i$

$v_i$  : *longitude* (garis lintang) pada lokasi ke- $i$

**7. Menentukan *bandwidth* optimum untuk setiap lokas di Wilayah Sumatera dengan metode *Cross Validation* (CV)**

*Bandwidth* merupakan sebuah ukuran jarak dan ukuran besarnya pengaruh antar lokasi yang diartikan sebagai radius dari suatu lingkaran, sehingga sebuah titik

yang terdapat dalam lingkaran tersebut memiliki pengaruh (Fotheringham *et al.*, 2014). Metode untuk mendapatkan suatu nilai *bandwidth* yang optimum adalah metode *Cross Validation* (CV), dimana nilai *bandwidth* yang optimum diperoleh dari nilai CV yang minimum. Persamaan CV sebagai berikut:

$$CV = \sum_{i=1}^n [Y_i - \hat{Y}_{\neq i}(h)]^2 \quad (11)$$

Keterangan :

- $Y_i$  : nilai pengamatan pada lokasi ke- $i$   
 $\hat{Y}_{\neq i}(h)$  : nilai penduga  $Y_i$  dimana pengamatan di lokasi  $(u_i, v_i)$  dihilangkan dari proses pendugaan

### 8. Menentukan matriks pembobot metode GWZIPR menggunakan pendekatan fungsi pembobot *Adaptive Bisquare Kernel*.

Salah satu jenis pembobot spasial yang dapat digunakan ketika menduga parameter yaitu *Adaptive Kernel*. *Adaptive Kernel* cocok untuk suatu pengamatan yang menyebar dengan pola yang tidak teratur dan berkelompok karena dapat menyesuaikan dengan kondisi titik pengamatan (Fotheringham *et al.*, 2014). Hal tersebut menghasilkan nilai *bandwidth* optimum yang berbeda antar lokasi pengamatan. Adapun salah satu jenis *Adaptive Kernel* yaitu *Adaptive Bisquare Kernel* yang memiliki persamaan sebagai berikut:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & d_{ij} \leq h \\ 0, & d_{ij} > h \end{cases} \quad (12)$$

Keterangan:

- $w_{ij}$  : fungsi pembobot *Adaptive Bisquare Kernel*  
 $d_{ij}$  : jarak *Euclidean*  
 $h$  : parameter non negatif atau nilai *bandwidth*

### 9. Memodelkan data dengan metode GWZIPR

Model GWZIPR untuk variabel respon  $Y_i$  yang berasal dari setiap pengamatan pada variabel respon dari lokasi  $(u_i, v_i)$  yang tidak sama yaitu  $Y_1, Y_2, \dots, Y_n$  memiliki peluang yang berbeda untuk  $y_i = 0$  dan  $y_i > 0$  dengan fungsi kepekatan peluangnya adalah:

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i} & , y_i = 0 \\ \frac{(1 - \omega_i)e^{-\mu_i} \mu_i^{y_i}}{y_i!} & , y_i > 0 \end{cases} \quad (13)$$

Dengan:

$$\mu_i = \exp\left(\sum_{k=0}^p \mathbf{X}_{ik} \boldsymbol{\beta}_k(u_i, v_i)\right) \quad (14)$$

$$\omega_i = \frac{\exp\left(\sum_{k=0}^p \mathbf{X}_{ik} \boldsymbol{\gamma}_k(u_i, v_i)\right)}{\left(1 + \exp\left(\sum_{k=0}^p \mathbf{X}_{ik} \boldsymbol{\gamma}_k(u_i, v_i)\right)\right)} \quad (15)$$

$\boldsymbol{\beta}_k(u_i, v_i)$  dan  $\boldsymbol{\gamma}_k(u_i, v_i)$  adalah vektor regresi yang akan diduga terletak pada lokasi  $(u_i, v_i)$  yang berukuran  $(k + 1) \times 1$ . Sedangkan,  $\mathbf{X}_{ik}$  yaitu matriks variabel prediktor berbeda yang berhubungan dengan peluang *zero state* ( $y_i = 0$ ) dan rata-rata *Poisson state* ( $y_i > 0$ ) yang berukuran  $1 \times (k + 1)$ .

### 3. Hasil dan Pembahasan

Overdispersi adalah suatu keadaan ketika nilai ragam lebih besar dari nilai rata-rata untuk variabel respon. Untuk mengidentifikasi adanya overdispersi pada variabel respon dapat juga menggunakan nilai *deviance*. Hasil pengujian overdispersi sebagai berikut:

Tabel 1: Uji Asumsi Overdispersi.

Variabel	Rata-rata	Ragam	<i>Deviance</i>	db	<i>deviance/db</i>
Y	70,02	55315,39	18410,58	150	122,74

Berdasarkan Tabel 1, diketahui bahwa nilai ragam yaitu 55315,39 lebih besar dibandingkan nilai rata-rata yaitu 70,02. Selain itu, nilai *deviance* dibagi dengan derajat bebas adalah lebih dari satu. Menunjukkan bahwa pada jumlah kasus malaria di Wilayah Sumatera mengalami overdispersi.

Deteksi multikolinieritas digunakan untuk mengetahui hubungan antar variabel prediktor. Salah satu teknik untuk melihat adanya multikolinieritas dalam data adalah melihat nilai *Varian Inflation Factor* (VIF). Nilai VIF pada kasus ini disajikan pada Tabel 2. Berdasarkan Tabel 2, nilai VIF semua variabel prediktor diperoleh bawa kurang dari 10. Dapat disimpulkan bahwa semua variabel prediktor tidak mengalami multikolinieritas. Asumsi multikolinieritas terpenuhi maka akan dilakukan pemodelan menggunakan regresi ZIP.

Tabel 2: Nilai VIF untuk semua variabel predictor.

Variabel	Nilai VIF
$X_1$	1,032
$X_2$	1,040
$X_3$	1,048

Model regresi *Zero-Inflated Poisson* (ZIP) adalah pengembangan dari regresi Poisson untuk data cacah yang mengalami overdispersi. Hal tersebut dikarenakan adanya kelebihan nilai nol dalam data. Estimasi parameter regresi ZIP dapat dilakukan menggunakan program R 3.6.1 yang dirangkum dalam Tabel 3.

Tabel 3: Estimasi parameter Regresi ZIP

Parameter	<i>Estimate</i>	<i>Standard Error</i>	$Z_{hit}$	<i>p-value</i>
$\hat{\beta}_0$	3,0523	0,0444	68,7400	$< 2e^{-16}$ ***
$\hat{\beta}_1$	0,0049	0,0005	10,3000	$< 2e^{-16}$ ***
$\hat{\beta}_2$	0,0112	0,0004	25,0500	$< 2e^{-16}$ ***
$\hat{\beta}_3$	0,0940	0,0022	43,1800	$< 2e^{-16}$ ***
$\hat{\gamma}_0$	-0,2289	0,7381	-0,3100	0,7565
$\hat{\gamma}_1$	0,0012	0,0085	0,1460	0,8841
$\hat{\gamma}_2$	0,0052	0,0082	0,6350	0,5257
$\hat{\gamma}_3$	-0,1411	0,0518	-2,7250	0,0064 **

Berdasarkan Tabel 3 dan setelah dilakukan uji signifikansi parameter, semua variabel prediktor signifikan pada model ln dan hanya variabel  $X_3$  yang signifikan pada

model logit. Sehingga diperoleh model untuk Regresi ZIP sebagai berikut:

$$\ln(\hat{\mu}_i) = 3,0523 + 0,0049X_1 + 0,0112X_2 + 0,0940X_3$$

$$\text{logit}(\hat{\omega}_i) = -0,1411X_3$$

Interpretasi dari model ln yaitu setiap perubahan 1% pada variabel persentase rumah tangga memiliki akses sanitasi layak meningkatkan rata-rata jumlah kasus malaria sebesar  $e^{0,0049} = 1,0049$  kali dan variabel lainnya dianggap konstan. Perubahan 1% untuk variabel persentase rumah tangga memiliki akses sumber air minum layak akan meningkatkan rata-rata jumlah kasus malaria sebesar  $e^{0,0112} = 1,0113$  kali dengan syarat variabel lain dianggap konstan. Setiap perubahan 1% pada variabel persentase jumlah puskesmas meningkatkan rata-rata jumlah kasus malaria sebesar  $e^{0,0940} = 1,0986$  kali dengan syarat variabel lain dianggap konstan.

Model logit menjelaskan bahwa setiap perubahan 1% pada variabel persentase jumlah puskesmas menurunkan peluang jumlah kasus malaria sebesar 0,8684 kali. Berdasarkan model ln dan model logit yang didapatkan, variabel prediktor yang berpengaruh pada *Poisson state* adalah persentase rumah tangga memiliki akses sanitasi layak ( $X_1$ ), persentase rumah tangga memiliki akses sumber air minum layak ( $X_2$ ), dan persentase jumlah puskesmas ( $X_3$ ). Sedangkan variabel prediktor yang berpengaruh pada *Zero state* hanya persentase jumlah puskesmas ( $X_3$ ).

Berdasarkan uji dependensi spasial, diperoleh nilai uji Moran's I adalah 1,7299 yang lebih kecil dari  $Z_{\alpha/2} = 1,96$ . Selain itu nilai  $p - value = 0,0836$  yang lebih besar dari  $\alpha = 0,05$  maka  $H_0$  gagal ditolak. Artinya tidak terdapat dependensi spasial. Demikian juga berdasarkan uji heterogenitas spasial, diperoleh nilai uji *Koenker-Basset* adalah 2,3693 yang lebih besar dari  $Z_{\alpha/2} = 1,96$ . Selain itu nilai  $p - value = 0,015$  yang kurang dari  $\alpha = 0,05$  sehingga  $H_0$  ditolak. Maka pada penelitian ini terdapat heterogenitas spasial.

Matriks pembobot spasial berbeda untuk setiap kabupaten/kota. Matriks pembobot spasial tersebut digunakan untuk menduga parameter di setiap kabupaten/kota di Wilayah Sumatera. Matriks pembobot untuk Simeulue ( $u_1, v_1$ ) adalah sebagai berikut:

$$W_{ij}(u_1, v_1) = \text{diag}(1, 0,9754, 0,9848, \dots, 0,3631)$$

$$= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0,9754 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0,3631 \end{bmatrix}$$

Model *Geographically Weighted Zero-Inflated Poisson Regression* (GWZIPR) adalah model untuk menduga parameter yang terdapat pengaruh faktor lokasi pada kasus overdispersi. Pemodelan GWZIPR menggunakan matriks pembobot spasial dengan fungsi *Adaptive Bisquare Kernel*. Matriks tersebut diperoleh dari perhitungan nilai *bandwidth* optimum dan jarak *Euclidean*.

Tabel 4: Nilai *deviance* regresi ZIP dan GWZIPR.

Model	<i>Deviance</i>	df	<i>Deviance/df</i>	$F_{hitung}$
Regresi ZIP	30560	8	3820	30,29
GWZIPR	18410,58	146	126,10	

Berdasarkan Tabel 4, didapatkan nilai  $F_{hitung} = 30,29$  dan  $F_{(0,05;8;146)} = 2,002$  maka  $H_0$  ditolak karena nilai  $F_{hitung} > F_{(0,05;8;146)}$ . Artinya terdapatnya perbedaan signifikan antara model Regresi ZIP dan GWZIPR pada taraf pengujian 5%. Sehingga, dapat dilanjutkan dengan pengujian signifikansi parameter model GWZIPR. Berdasarkan pengujian didapatkan nilai uji  $G = 18410,58$  dan  $\chi^2_{(5\%;6)} = 12,59$ . Karena nilai uji  $G$  lebih besar dari nilai  $\chi^2_{(5\%;6)}$  maka  $H_0$  ditolak. Artinya paling sedikit satu variabel prediktor di setiap kabupaten/kota memberikan pengaruh yang signifikan pada model GWZIPR. Dapat juga diartikan bahwa model GWZIPR layak digunakan.

Contoh pengujian parameter secara parsial untuk Kepulauan Mentawai ( $u_{57}, v_{57}$ ) disajikan pada Tabel 5. Berdasarkan Tabel 5, diperoleh bahwa untuk variabel yang signifikan di Kepulauan Mentawai pada model In adalah rata persentase rumah tangga memiliki akses sanitasi layak ( $X_1$ ), persentase rumah tangga memiliki akses sumber air minum layak ( $X_2$ ), dan persentase jumlah puskesmas ( $X_3$ ). Sedangkan pada model logit, didapatkan bahwa untuk variabel yang signifikan di Kepulauan Mentawai adalah persentase jumlah puskesmas ( $X_3$ ). Model GWZIPR yang terbentuk untuk Kepulauan Mentawai adalah:

$$\ln(\hat{\mu}_{57}) = 3,047 + 0,006X_1 + 0,010X_2 + 0,092X_3$$

dan

$$\text{logit}(\hat{\omega}_{57}) = -0,156X_3$$

Tabel 5: Uji parsial parameter model GWZIPR

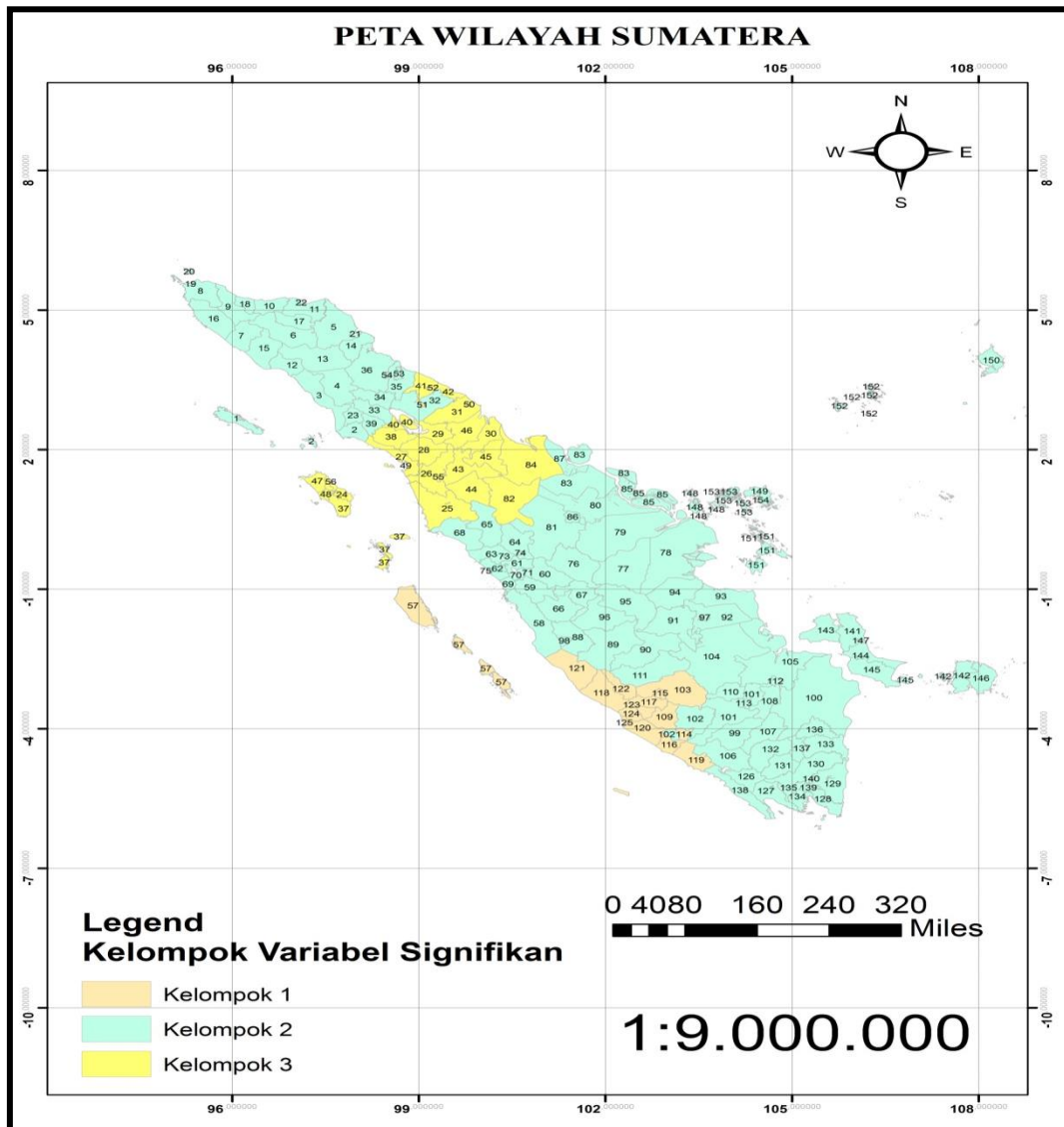
Kab/kot	Model	Parameter	Estimate	Pr(>  Z )	Kesimpulan
Kepulauan Mentawai	ln( $\mu_i$ )	$\hat{\beta}_0$	3,047	0,000	Signifikan
		$\hat{\beta}_1$	0,006	0,000	Signifikan
		$\hat{\beta}_2$	0,010	0,000	Signifikan
		$\hat{\beta}_3$	0,092	0,000	Signifikan
	logit( $\omega_i$ )	$\hat{\gamma}_0$	-0,385	0,683	Tidak Signifikan
		$\hat{\gamma}_1$	0,001	0,936	Tidak Signifikan
		$\hat{\gamma}_2$	-0,008	0,488	Signifikan
		$\hat{\gamma}_3$	-0,156	0,035	Signifikan

Interpretasi dari model In yaitu setiap perubahan 1% pada variabel persentase rumah tangga memiliki akses sanitasi layak meningkatkan rata-rata kasus malaria di Kepulauan Mentawai sebesar  $e^{0,006} = 1,006$  kali dan variabel lainnya dianggap konstan. Perubahan 1% pada variabel persentase rumah tangga memiliki akses sumber air minum layak meningkatkan rata-rata jumlah kasus malaria di Kepulauan Mentawai sebesar  $e^{0,010} = 1,010$  kali dengan syarat variabel lain dianggap konstan. Setiap perubahan 1% pada variabel persentase jumlah puskesmas meningkatkan rata-rata jumlah kasus malaria di Kepulauan Mentawai sebesar  $e^{0,092} = 1,096$  kali dengan syarat variabel lain dianggap konstan.

Model logit menjelaskan bahwa setiap perubahan 1% pada variabel persentase jumlah puskesmas menurunkan peluang jumlah kasus malaria sebesar 0,8556 kali. Variabel yang signifikan pada model logit hanya variabel persentase jumlah puskesmas ( $X_3$ ) di beberapa kabupaten/kota di Wilayah Sumatera sehingga variabel



tersebut bersifat lokal. Berdasarkan model ln dan model logit yang didapatkan, variabel prediktor yang berpengaruh pada *Poisson state* adalah persentase rumah tangga memiliki akses sanitasi layak ( $X_1$ ), persentase rumah tangga memiliki akses sumber air minum layak ( $X_2$ ), dan persentase jumlah puskesmas ( $X_3$ ). Sedangkan variabel prediktor yang berpengaruh *Zero state* hanya persentase jumlah puskesmas ( $X_3$ ).



Gambar 1: Peta persebaran variabel yang signifikan setiap Kabupaten/kota di Wilayah Sumatera.

Berdasarkan hasil uji signifikansi parameter, variabel prediktor yang signifikan pada model GWZIPR berbeda-beda pada tiap lokasi pengamatan. Sehingga, kabupaten/kota di Wilayah Sumatera dikelompokkan ke dalam tiga kelompok. Pengelompokan tersebut berdasarkan kesamaan variabel prediktor yang signifikan.

Tabel 6: Pengelompokan variabel yang signifikan model GWZIPR

Kelompok	Jumlah Kab/Kot	Variabel Signifikan
Kelompok 1	15	Ln = $X_1, X_2, X_3$ Logit = $X_3$
Kelompok 2	112	Ln = $X_1, X_2, X_3$ Logit = Tidak ada
Kelompok 3	27	Ln = $X_2, X_3$ Logit = Tidak ada

Berdasarkan Tabel 6, maka gambaran penyebaran faktor-faktor yang berpengaruh terhadap kasus malaria di setiap kabupaten/kota di Wilayah Sumatera untuk model GWZIPR pada Gambar 1.

#### 4. Simpulan dan Saran

Dari pembahasan penelitian di atas, disimpulkan bahwa faktor-faktor yang signifikan dikelompokkan menjadi tiga kelompok. Kelompok 1 adalah semua variabel prediktor signifikan pada model Ln dan hanya variabel persentase jumlah puskesmas ( $X_3$ ) yang signifikan pada model logit yang tersebar di 15 kabupaten/kota. Kelompok 2 adalah semua variabel prediktor signifikan pada model Ln dan tidak ada variabel yang signifikan pada model logit tersebar di 112 kabupaten/kota. Kelompok 3 terdiri dari variabel persentase rumah tangga memiliki akses sumber air minum layak ( $X_2$ ) dan persentase jumlah puskesmas ( $X_3$ ) signifikan pada model Ln dan dan tidak ada variabel yang signifikan pada model logit tersebar di 27 kabupaten/kota. Pemetaan penyebaran faktor-faktor yang mberpengaruh terhadap kasus malaria di setiap kabupaten/kota secara umum terdiri dari tiga kelompok berdasarkan variabel yang signifikan pada model Ln dan logit. Pemetaan tersebut memiliki letak geografis yang berdekatan satu sama lain sehingga tidak membentuk pola menyebar.

#### Daftar Pustaka

- Agresti, A. (1990). *Categorical data analysis*. New York (US): John Wiley & Sons.
- Cameron, A. C., & Trivedi, P. K. (1990). *Regression Analysis of Count Data*. New York (US): Cambridge University Press.
- [DIRJEN PPPL] Direktur Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan. (2014). *Pedoman Manajemen Malaria*. Jakarta (ID): Mitra Besari.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2014). *Geographically Weighted Regression The Analysis of Spatially Varying Relationships*. The Atrium Southern Gate England (UK): John Wiley & Sons.
- Gujarati, D. N. (2004). *Basic Econometrics (4th ed)*. New York (US): The McGraw-Hill Companies.

- Hilbe, J. M. (2011). *Negative Binomial Regression (2nd ed)*. New York (US): Cambridge University Press.
- [KEMENKES] Kementerian Kesehatan RI. (2011). *Epidemiologi Malaria di Indonesia*. Jakarta (ID): Jendela Data dan Informasi Kesehatan.
- Kutner, M. H., Nachsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models (5th ed)*. New York (US): Mc Graw Hill.
- Lambert, D. (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, 32(1): 1–14.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis (5th ed)*. New Jersey (US): John Wiley & Sons.
- Purhadi, Dewi, Y. S., & Amaliana, L. (2015). Zero Inflated Poisson and Geographically Weighted Zero-Inflated Poisson Regression Model: Application to Elephantiasis (Filariasis) Counts Data. *Journal of Mathematics and Statistics*, 11(2): 52–60.
- Siswanto, S., & Thamrin, S. A. (2020). Penentuan faktor-faktor potensial yang mempengaruhi kejadian malaria di Provinsi Papua dengan epidemiologi spasial. *Indonesian Journal of Statistics and Its Applications*, 4(3): 498–509.