BAB II

TINJAUAN PUSTAKA

2.1 Malaria

Malaria merupakan penyakit yang ditularkan oleh nyamuk nyamuk *Anopheles* betina yang membawa parasit *plasmodium* yang mengalir dalam aliran darah dan akhirnya hinggap di hati setelah menggigit korban. Secara umum gejala orang yang terkena malaria yaitu demam, menggigil, bekeringat, sakit kepala, mual atau muntah (infodatin malaria, 2016). Penyakit infeksi ini banyak dijumpai di daerah tropis. *World Malaria Report 2015* menyebutkan bahwa malaria telah menyerang 106 negara di dunia. Malaria ini adalah penyebab sebagian besar kematian.

2.2 Faktor- Faktor yang Mempengaruhinya

Malaria ini adalah penyebab sebagian besar kematian. Oleh karena itu diperlukan upaya untuk mengurangi jumlah kasus malaria dengan menganalisis faktor-faktor yang mempengaruhi jumlah kasus malaria, diantaranya yaitu:

2.2.1 Perilaku Hidup Bersih

Faktor lingkungan dengan memberdayakan masyarakat untuk berperilaku hidup bersih dan sehat sebagai usaha pencegahan dan pemberantasan malaria yang efektif. Perilaku hidup bersih dan sehat merupakan faktor terbesar kedua setelah faktor lingkungan yang mempengaruhi kesehatan individu, kelompok, atau masyarakat dalam memberantas malaria (Infodatin malaria, 2016).

2.2.2 Pengobatan *Arthemisinin based Combination Therapy* (ACT)

Pengobatan *Arthemisinin based Combination Therapy* (ACT) kepada penderita malaria. Dimana melalui *Global Malaria Programme* (GMP), ACT dijadikan target untuk menghentikan penyebaran dan mengurangi kejadian insiden malaria pada tahun 2015 sebagai indikator dari target pembangunan *Millenium Development Goals* (MDGs), dimana pemerintah Indonesia memiliki program nasional untuk bebas dari malaria pada tahun 2030.

2.2.3 Tempat Tinggal Kumuh

Faktor lainnya untuk mengatasi tingginya jumlah kasus malaria yaitu memperhatikan rumah tangga kumuh. Rumah tangga kumuh merupakan rumah yang tidak memenuhi persyaratan keselamatan, bangunan, dan kecukupan minimum luas bangunan serta memenuhi syarat bagi kesehatan penghuninya. Lingkungan tempat tinggal yang kumuh dapat meningkatkan jumlah kasus penyakit, diantaranya kasus malaria (Kemenkes RI, 2018).

2.2.4 Kepadatan Penduduk

Kepadatan penduduk merupakan indikator dari tekanan penduduk di suatu daerah (BPS, 2016). Pertumbuhan penduduk dalam satu wilayah dengan kecenderungan peningkatan penggunaan energi dan kegiatan dapat memperburuk kondisi kesehatan lingkungan (Achmadi, 2005). Keadaan hunian yang padat dan hunian yang ditempati oleh berbagai macam orang akan besar pengaruhnya terhadap timbulnya risiko penularan. Pertumbuhan penduduk yang tidak memiliki pola tertentu dengan munculnya area yg sifatnya kumuh dan urbanisasi yang tidak terkontrol menjadi faktor yang berperan dalam munculnya penyebaran kasus malaria.

2.3 Statistika Deskriptif

Statistika deskriptif merupakan analisis yang berkaitan dengan pengumpulan dan penyajian data sehingga dapat memberikan informasi yang berguna. Analisis ini bertujuan menguraikan tentang sifat-sifat atau karakteristik dari suatu keadaan dan untuk membuat deskripsi atau gambaran yang sistematis dan akurat mengenai fakta-fakta, sifat-sifat dari fenomena yang diselidiki (Nurina, 2015).

2.4 Multikolineritas

Salah satu syarat yang harus dipenuhi dalam pembentukan model regresi dengan beberapa variabel prediktor adalah tidak ada kasus multikolinearitas. Pendeteksian kasus multikolinieritas dilakukan menggunakan kriteria nilai VIF (Hocking, 1996). Jika nilai VIF (Variance Inflation Factor) lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel prediktor. Nilai VIF dinyatakan sebagai berikut :

$$VIF_k = \frac{1}{1-R_k^2} \tag{2.1}$$

Keterangan:

R_k^2 : koefisien determinasi antara satu variabel prediktor dengan variabel prediktor lainnya. (Hocking dalam Hidayanti, 2015).

2.5 Regresi Poisson

Regresi Poisson merupakan model regresi nonlinear yang sering digunakan untuk memodelkan data *count.* Jika variabel random diskrit (y) merupakan distribusi Poisson dengan parameter μ maka fungsi probilitas dari distribusi Poisson dapat dinyatakan sebagai berikut.

$$P(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!}; \quad y = 0,1,2, \ldots dan\ \mu > 0 \tag{2.2}$$

Keterangan:

μ : Rata-rata variabel respon yang berdistribusi Poisson .

y : Variabel dependen

e : Bilangan natural = 2.718

(Pratama, 2015)

Sedangkan untuk nilai harapan dan ragam pada Poisson adalah dimana nilai rata-rata dan varian dari y mempunyai nilai lebih dari 0 sebagai berikut:

$$E[y] = Var[y] = \mu \tag{2.3}$$

(Casella, 2012)

Persamaan model regresi Poisson dapat ditulis sebagai berikut.

$$\hat{\mu}_i = \exp\left(\hat{\beta}_0 + X_{1i} + \hat{\beta}_2 X_{2i} + \cdots \hat{\beta}_p X_{pi}\right) \tag{2.4}$$

Keterangan:

μ_i : Merupakan rata-rata jumlah kejadian yang terjadi dalam interval waktu tertentu.

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$: Parameter

$X_{1i}, X_{2i}, \ldots, X_{pi}$: Variabel Independen

2.5.1 Uji Estimasi dan Signifikasi Parameter

Metode yang digunakan untuk mengestimasi parameter regresi Poisson adalah metode *Maximum Likelihood Estimation* (MLE). Dengan fungsi dirumuskan sebagai berikut :

$$lnL(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{e}^{\boldsymbol{X_i^T \beta}} + \sum_{i=1}^{n} y_i X_i^T \beta - \sum_{i=1}^{n} ln(y_i!) \qquad (2.5)$$

Untuk pengujian signifikansi parameter regresi Poisson terdiri atas uji serentak dan uji parsial menggunakan Maximum Likelihood Ratio Test (MLRT).

1. Uji signifikansi secara serentak

   Uji signifikansi secara serentak menggunakan hipotesis sebagai berikut :

   $H_0 = \ \beta_1 = \beta_2 = \cdots = \beta_p = 0$

   $H_1 =$ minimal ada satu $\beta_k \neq 0; k = 1,2, \ldots p$

   Statistik Uji :

   $$D(\boldsymbol{\beta}) = -2ln\lambda = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \qquad (2.6)$$

   Keterangan:

   $L(\widehat{\omega})$: Fungsi *likelihood* tanpa melibatkan variabel predictor

   $L(\widehat{\Omega})$: Fungsi *likelihood* dengan melibatkan variabel predictor.

   Kriteria Pengujian:

   Tolak H_0 jika $D(\beta) > x^2_{(a,p)}$ yang berarti minimal ada satu parameter yang berpengaruh secara signifikan.

2. Uji signifikansi secara parsial

Kemudian dilakukan pengujian parameter secara parsial dengan hipotesis sebagai berikut :

$H_0 = \beta_k = 0$

$H_1 = \beta_k \neq 0; \; k = 1,2, \ldots .p$

Statistik Uji :

$$z_{hitung} = \frac{\bar{\beta}_k}{SE(\bar{\beta}_k)} \qquad (2.7)$$

Tolak H_0 jika $|Z_{hitung}|>Z_{\alpha/2}$ dengan α merupakan tingkat signifikansi yang ditentukan. Tolak H0 artinya bahwa parameter ke-k signifikan terhadap model regresi Binomial Negatif.

2.5.2 Overdispersi Regresi Poisson

Regresi Poisson dikatakan overdispersi apabila nilai variansnya lebih besar dari nilai rata-ratanya. Yang artinya sifat *equaidispersion* dari asumsi regresi Poisson tidak terpenuhi. *Overdispersion* menyebabkan taksiran parameter model bias dan tidak efisien. Selain itu *overdispersion* menybabkan tingkat kesalahan model semakin besar dan regressi Poisson menjadi tidak sesuai (Camero, 1998).

Keberadaan *overdispersion* pada model regresi Poisson dapat di uji dengan nilai disperse *pearson Chi-square* atau *deviance* yang dibagi dengan derajat bebasnya, diperoleh nilai lebih besar dari 1. (Hardin, 2015).

2.6 Regresi Binomial Negatif

Regresi binomial negative termasuk dalam kelompok *Generalized Linear Model* (GLM). GLM merupakan perluasan dari proses pemodelan linier untuk pemodelan data yang mengikuti distribusi poisson dan lain-lain (McCullagh dan

Nelder, 1989). Distribusi binomial negative sering digunakan untuk memodelkan data *count* yang merupakan variabel dependen yang mengalami overdispersi. Analisis regresi binomial negative digunakan untuk mengetahui hubungan antar dua variabel atau lebih yang didasarkan pada distribusi binomial negatif. Adapun Model regresi binomial negtif dinyatakan sebagai berikut :

$$\hat{y}_i = exp\left[\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} \ldots . \hat{\beta}_p X_{pi}\right] + \varepsilon_i \qquad (2.8)$$

Keterangan:

$\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_p$: Parameter

ε_i : Galat untuk pengamatan ke-i

(Bouk, 2015)

2.6.1 Pengujian Parameter Regresi Binomial Negatif

Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk pendugaan parameter dalam regresi Binomial Negatif. Untuk pengujian signifikansi parameter pada regresi binomial negatif terdiri atas uji serentak dan uji parsial.

1. Uji signifikansi secara serentak

   Uji signifikansi secara serentak untuk estimasi parameter model regresi Binomial Negatif menggunakan uji devians dengan hipotesis sebagai berikut:

   $H_0 = \beta_1 = \beta_2 = \cdots = \beta_p = 0$

   $H_1 =$ minimal ada satu $\beta_k \neq 0; k = 1,2, \ldots p$

   Statistik Uji :

$$D(\boldsymbol{\beta}) = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) = 2\ln(L(\widehat{\Omega}) - L(\widehat{\omega})) \qquad (2.9)$$

Keterangan:

$L(\widehat{\omega})$: Fungsi *likelihood* tanpa melibatkan variabel predictor

$L(\widehat{\Omega})$: Fungsi *likelihood* dengan melibatkan variabel predictor.

Kriteria Pengujian:

Tolak H_0 jika $D(\beta) > x^2_{(a,p)}$ yang berarti minimal ada satu parameter yang berpengaruh secara signifikan.

2. Uji signifikansi parsial

   Pengujian parameter secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh yang memberikan pengaruh signifikan terhadap model dengan hipotesis sebagai berikut :

   $H_0 = \beta_k = 0$

   $H_1 = \beta_k \neq 0; \; k = 1,2, \ldots . p$

   Statistik Uji :

$$W_k = \frac{\overline{\beta}_k}{SE(\overline{\beta}_k)} \tag{2.10}$$

   Tolak H_0 jika $|Z_{hitung}| > Z_{\alpha/2}$ dengan α merupakan tingkat signifikansi yang ditentukan.

2.7 Aspek Spatial

Pengujian Aspek Data Spasial Regresi spasial merupakan salah satu metode yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor dengan memperhatikan aspek lokasi atau spasial. Aspek spasial yang dimaksud adalah data yang digunakan memiliki error saling berkorelasi dan memiliki heterogenitas spasial (Anselin, 1988).

2.7.1 Uji Heterogenitas Spasial

Wahendra & Sri Pingit (2015) dalam penelitiannya menyebutkan bahwa pengujian heterogenitas spasial memiliki fungsi untuk menentukan apakah ada sesuatu yang khusus pada setiap lokasi pengamatan, sehingga menghasilkan parameter regresi yang berbeda-beda secara spasial. Pengujian ini menggunakan uji *Breusch-Pagan* (BP) dengan hipotesis :

$H_0 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_\infty^2 = \sigma^2$ (variansi antarlokasi sama)

$H_1 = minimal\ ada\ satu\ \sigma_1^2\ \neq\ \sigma^2$ (variansi antarlokasi berbeda)

Menggunakan statistic uji *Breusch-Pagan* (BP) sebagai berikut :

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f}^t\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f} \tag{2.11}$$

Dengan elemen vektor *f* adalah :

$$f_t = \left(\frac{e_i^2}{\sigma^2} - 1\right)$$

$$e_i = y_i - \hat{y}_i$$

Z **:** Matriks berukuran *nx*(*k*+1) berisi vektpr yang sudah di normal standardkan untuk setiap observasi.

e_i : Error untuk pengamatan ke i

σ^2 : Ragam error (e_i)

Kriteria keputusan tolak H_0 jika nilai BP > $X^2_{(a,k)}$ atau *p-value* < α yang artinya terjadi heteroskedastisitas dalam model (variansi berbeda antarlokasi).

2.8 Penentuan Bandwidth

Penentuan Bandwidth dan Pembobot Optimum Secara teoritis bandwidth merupakan luasan dengan radius b dari titik pusat lokasi yang digunakan sebagai dasar menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut. Pengamatan-pengamatan yang terletak di dalam radius b masih dianggap berpengaruh terhadap model pada lokasi tersebut sehingga akan diberi bobot tergantung pada fungsi yang digunakan. Selain itu, bandwidth menjadi pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan data. Nilai bandwidth yang sangat kecil menyebabkan varians semakin besar. Hal ini dikarenakan jika nilai bandwidth sangat kecil maka akan semakin sedikit pengamatan yang berada dalam radius b, sehingga model yang diperoleh akan sangat kasar (under smoothing). Sebaliknya nilai bandwidth yang besar akan menimbulkan bias yang semakin besar karena semakin banyak pengamatan yang berada dalam radius b, sehingga model yang diperoleh akan terlampau halus (over smoothing). Pemilihan bandwidth optimum menjadi sangat penting karena akan mempengaruhi ketepatan model terhadap data, yaitu mengatur varians dan bias dari model. Penentuan bandwidth optimum dilakukan menggunakan metode Cross Validation (CV) sebagai berikut :

$$CV(b) = \sum_{i=1}^{n}(y_i - \hat{y}_{\neq i}(b))^2 \qquad (2.12)$$

Keterangan:

$\hat{y}_{\neq i}(b)$: Nilai penaksir y_i dengan pengamatan lokasi (u_i, v_i).

y_i : Variabel dependen

2.9 Matriks Pembobot

2.9.1 *Adaptive Kernel*

Menurut Fotheringham, dkk (2002), penggunaan metode Adaptive Kernel sangat cocok apabila suatu pengamatan tersebar dengan pola tidak beraturan dan berkelompok. Metode adaptive kernel memungkinkan untuk mendapatkan nilai bandwith yang berbeda untuk setiap titik pengamatan. Hal ini dikarenakan metode adaptive kernel dapat menyesuaikan dengan kondisi titik pengamatan. Bila titik pengamatan tersebar di sekitar pengamatan lokasi ke i yang diperoleh rlatif lebih kecil . Adapun jenis fungsi pembobot adaptive kernel yang digunakan yaitu :

1. Adaptive Gaussian Kernel

$$W_j\ (u_i, v_i) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] \tag{2.13}$$

2. Adaptive Tricube Kernel

$$W_j\ (u_i, v_i) = \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3 \tag{2.14}$$

Dengan:

$d_{ij} = \sqrt{\left(u_i, u_j\right)^2 + \left(v_i, v_j\right)^2}$ adalah jarak Euclidean antar lokasi (u_i, v_i) dan lokasi $\left(u_j, v_j\right)$

h : Nilai bandwith optimum

2.10 *Geographically Weighted Negative Binomial Regression* (GWNBR)

GWNBR merupakan metode dari Regresi Binomial Negatif yang dikembangkan dalam menduga data yang memiliki spasial heterogenitas untuk data cacah yang memiliki overdispersi. Hasil yang didapat dari model ini merupakan estimasi parameter lokal dengan masing-masing lokasi akan memiliki parameter yang berbeda (Silva & Rodrigues dalam Rini, 2018).

$$y_i \sim BN(\mu_i, \theta_i)$$

Dengan :

$$\mu_i = \exp(\textstyle\sum_k \beta_k(u_i, v_i) x_{ik}, \theta_i(u_i, v_i)) \qquad (2.15)$$

Sehingga model GWNBR dapat dirumuskan sebagai berikut :

$$y_i \sim BN[\exp(\textstyle\sum_k \beta_k(u_i, v_i) x_{ik}, \theta_i(u_i, v_i))] \qquad (2.16)$$

Keterangan :

y_i : Nilai pengamatan respon ke-*i*

x_{ik} : Nilai pengamatan variabel prediktor ke-*k* pada pengamatan lokasi (u_i, v_i)

$\beta_k(u_i, v_i)$: Parameter regresi variabel prediktot ke-*k* untuk setiaplokasi

$\theta_i(u_i, v_i)$: Parameter disperse setiap lokasi (u_i, v_i)

(u_i, v_i) : Lokasi(koordinat lintang dan bujur) dari titik lokasi ke-*i*

Dimana y_i = 0,1,2,…

i = *j*= 1,2,…,*n*

k = 1,2,…*p*

Fungsi sebaran Binomial Negatif pada setiap lokasi dirumuskan sebagai berikut :

$$f(y_i|\boldsymbol{x_i},\boldsymbol{\beta}(u_i,v_i),\theta_i) = \frac{\Gamma(y_i+\theta_i^{-1})}{\Gamma(\theta_i^{-1})y_i!}\left(\frac{1}{1+\theta_i\mu_i}\right)^{\theta_I^{-1}}\left(\frac{\theta_i\mu_i}{1+\theta_i\mu_i}\right)^{y_i} \sim BN(\mu_i,\theta_i)$$

(2.17)

Dimana :

y_i = 0,1,2,…

μ_i = exp($\boldsymbol{x_i^T}\boldsymbol{\beta}(u_i,v_i)$)

θ_i = $\theta_i(u_i,v_i)$

2.10.1 Pengujian Kesamaan Model GWNBR dengan Regresi Binomial Negatif

Pengujian kesamaan model GWNBR dengan regresi Binomial Negatif dilakukan untuk melihat terdapat perbedaan yang signifikan atau tidak antara model GWNBR dengan Regresi Binomial Negatif dengan hipotesis sebagai berikut :

$H_0 = \beta_\mathrm{k}\left(u_{\mathrm{i}\,,}v_\mathrm{i}\right) = \beta_\mathrm{k}; i = 1,2,..,\mathrm{n}; \mathrm{k} = 1,2,\dots,\mathrm{p}$

$H_1 = \beta_k\left(u_{\mathrm{i}\,,}v_\mathrm{i}\right) \neq \beta_\mathrm{k}$

Statistik Uji :

$$F_{hit} = \frac{\frac{Devians\ model_A}{df_A}}{\frac{Devians\ model_B}{df_B}} \qquad (2.18)$$

Tolak H_0 jika $F_{hit} > F_{tabel}$ yang artinya bahwa ada perbedaan yang signifikan antara model Binomial Negatif dengan GWNBR.

(Ricardo dan Carvalho, 2013)

2.10.2 Pengujian Signifikan Parameter Model GWNBR

Pengujian signifikan parameter model GWNBR terdiri atas uji serentak dan uji parsial.

1. Uji signifikansi secara serentak

   Uji signifikansi secara serentak mengunakan metode *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis sebagai berikut :

   $H_0 = \beta_{\mathrm{i}}(u_{\mathrm{i}}, v_{\mathrm{i}}) = \beta_2 (u_{\mathrm{i}}, v_{\mathrm{i}}) = \ldots = \beta_{\mathrm{p}}(u_{\mathrm{i}}, v_{\mathrm{i}}) = 0$

   $H_1 =$ minimal ada satu $\beta_{\mathrm{k}}(u_{\mathrm{i}}, v_{\mathrm{i}}) \neq 0; k = 1,2, \ldots p$

   Statistik Uji :

   $$D(\boldsymbol{\beta}) = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) = 2\ln(L(\widehat{\Omega}) - L(\widehat{\omega})) \qquad (2.19)$$

   Keterangan:

   $L(\widehat{\omega})$: Fungsi *likelihood* tanpa melibatkan variabel predictor

   $L(\widehat{\Omega})$: Fungsi *likelihood* dengan melibatkan variabel predictor.

   Kriteria Pengujian:

   Tolak H_0 jika $D(\beta) > x^2_{(a,p)}$ yang berarti minimal ada satu parameter yang berpengaruh secara signifikan.

2. Uji signifikansi parsial

   Pengujian ini untuk mengetahui parameter saja yang signifikan terhadap variabel respon pada tiap-tiap lokasi dengan hipotesis sesbagai berikut :

   $H_0 = \beta_{\mathrm{k}}(u_{\mathrm{i}}, v_{\mathrm{i}}) = 0$

   $H_1 = \beta_{\mathrm{k}}(u_{\mathrm{i}}, v_{\mathrm{i}}) \neq 0; \;\; k = 1,2, \ldots . p$

Statistik Uji :

$$Z_{hit} = \frac{\bar{\beta}_k(u_{\mathrm{i}}, v_{\mathrm{i}})}{SE(\bar{\beta}_k(u_{\mathrm{i}}, v_{\mathrm{i}}))} \tag{2.20}$$

Tolak H_0 jika $|Z_{hitung}|>Z_{\alpha/2}$ artinya bahwa parameter tersebut berpengaruh siginifkansi terhadap variabel respon di lokasi pada tiap lokasi.

2.11 Pemilihan Model Terbaik

Pemilihan Model Terbaik Salah satu kriteria informasi dalam pemilihan model terbaik yang paling sering digunakan adalah Akaike's Information Criteria (AIC).

$$AIC = -2\ln \widehat{L(\hat{\theta})} + 2k\ ) \tag{2.21}$$

Keterangan :

$\widehat{L(\hat{\theta})}$ = nilai likelihood

k = Jumlah parameter pada tiap model yang terbentuk