

BAB II

TINJAUAN PUSTAKA

2.1 Statistik Deskriptif

Statistika deskriptif merupakan alat, teknik, atau prosedur yang digunakan untuk menggambarkan atau mendeskripsikan kumpulan data atau hasil pengamatan menjadi sebuah informasi (Walpole, 1995). Hasil informasi berdasarkan statistika deskriptif kurang lengkap dan akurat. Akan tetapi, penampilan yang menarik dan mudah dipahami oleh orang awam. Statistika deskriptif dibagi menjadi dua yaitu ukuran pemusatan dan penyebaran data sedangkan penyajian statistika deskriptif seperti tabel, histogram, boxplot, diagram, grafik dan lain-lainnya.

2.2 Peta Tematik

Peta tematik adalah peta yang menyajikan tema tertentu dan untuk kepentingan tertentu dengan menggunakan peta rupa bumi yang telah disederhanakan sebagai dasar untuk meletakkan informasi tematiknya (Badan Informasi Geospasial, 2017). Pewarnaan pada peta ditujukan untuk membedakan wilayah satu dengan lainnya.

Metode pembagian warna pada peta menggunakan *Natural Breaks* yang merupakan pengelompokan pola data, dengan nilai dalam kelas memiliki batas-batas yang ditentukan berdasarkan nilai jangkauan terbesar. Proses pada metode ini berulang-ulang dan menggunakan *break* yang berbeda dalam dataset yang memiliki varians terkecil. Metode ini baik untuk pemetaan nilai yang tidak merata (Crisana, 2014).

2.3 Multikolinearitas

Salah satu syarat yang harus dipenuhi dalam pembentukan model regresi dengan beberapa variabel prediktor adalah tidak ada kasus multikolinieritas. Pendeteksian kasus multikolinieritas yaitu dengan koefisien korelasi dan kriteria nilai VIF (*Variance Inflation Factor*). Terdapat kasus multikolinieritas jika nilai VIF lebih besar dari 10. Nilai VIF dinyatakan sebagai berikut.

$$VIF_k = \frac{1}{1 - R_k^2} \quad (2.1)$$

dimana $R_k^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (y_i - \bar{y})_k^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

dengan R_k^2 adalah koefisien determinasi antara satu variabel prediktor dengan variabel prediktor lainnya (Hocking, 1996).

2.4 Distribusi Poisson

Distribusi Poisson sering digunakan untuk memodelkan peristiwa yang memiliki peluang kejadian kecil pada interval waktu tertentu (Osgood, 2000). Regresi Poisson merupakan model regresi nonlinier yang tepat digunakan untuk memodelkan data cacahan (Agresti, 2002).

2.5 Overdispersi

Regresi Poisson dikatakan mengandung overdispersi apabila nilai variansnya lebih besar dari nilai meannya. Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi yang akan menyebabkan anak dugaan dari parameter regresinya tetap konsisten namun tidak efisien. Hal ini berdampak pada nilai standar error yang menjadi *under estimate*, sehingga kesimpulan yang

diperoleh menjadi tidak valid (Cameron & Trivedi, 1990). Pemeriksaan overdispersi menggunakan statistik uji *Pearson's Chi-Square* sebagai berikut.

$$\sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma_i^2} \sim \chi_{(n-p)}^2 ; \text{ dengan } p = k + 1 \quad (2.2)$$

Data mengalami overdispersi jika statistik uji *Pearson's ChiSquare* dibagi dengan derajat bebas akan menghasilkan nilai lebih besar dari 1.

2.6 Aspek Spasial

Pemodelan pada data spasial dapat dikelompokkan berdasarkan tipe data spasial yang digunakan yaitu spasial titik dan spasial area. Masing-masing tipe data spasial tersebut dapat dikelompokkan lagi berdasarkan jenis data yang digunakan yaitu *cross sectional* dan *time series*. Pemodelan data spasial selalu melibatkan matriks pembobot spasial. Sedangkan efek spasial pada data dapat berupa error yang saling berkorelasi (dependensi spasial) maupun keragaman (heterogenitas) spasial antar lokasi.

Pengujian dependensi spasial menggunakan uji Moran's I dengan hipotesis sebagai berikut.

$H_0 : I = 0$ (tidak terdapat dependensi spasial)

$H_0 : I \neq 0$ (terdapat dependensi spasial)

Statistik uji yang digunakan sebagai berikut.

$$Z = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{Var}(\hat{I})}} \quad (2.3)$$

dimana $\hat{I} = \frac{e^T W e}{e^T e}$ dengan e merupakan vektor residual dan W merupakan matriks pembobot spasial antar lokasi. Rumus mean dan varians dari Moran's I sebagai berikut.

$$E(\hat{I}) = \frac{tr(MW)}{(n-k)} \quad (2.4)$$

$$Var(\hat{I}) = \frac{tr(MWMM^T) + tr(MW)^2 + (tr(MW))^2}{d - E(\hat{I})^2} \quad (2.5)$$

dengan $d = (n - k)(n - k - 2)$; $M = 1 - X(X^T X)^{-1}X^T$. Tolak H_0 jika nilai $|Z| > Z_{(\alpha/2)}$ yang artinya terdapat dependensi spasial dalam model.

Pengujian heterogenitas spasial dilakukan untuk melihat ciri khas pada setiap lokasi pengamatan yang akan mengakibatkan parameter regresi yang dihasilkan berbeda secara spasial. Pengujian heterogenitas spasial dilakukan dengan statistik uji *Breusch-Pagan* (BP) dengan hipotesis sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (variansi antar lokasi sama)}$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 \text{ (variansi antar lokasi berbeda)}$$

Statistik uji yang digunakan adalah sebagai berikut.

$$BP = \left(\frac{1}{2}\right) f^T Z(Z^T Z)^{-1} Z^T f \quad (2.6)$$

dengan elemen vektor f adalah $f_i = \frac{e_i^2}{\sigma^2} - 1$ dan $e_i = y_i - \hat{y}_i$. Z merupakan matriks berukuran $n \times (k+1)$ yang berisi vektor yang sudah dinormal standarkan untuk setiap observasi. Tolak H_0 jika nilai dari $BP < \chi_{(\alpha, k)}^2$ yang artinya terjadi heteroskedastisitas dalam model atau variansi antar lokasi berbeda.

2.7 Matriks Pembobot Spasial

Matriks pembobot spasial menunjukkan keragaman spasial antara lokasi yang satu dengan yang lain. Elemen dari matriks pembobot spasial (\mathbf{W}) merupakan fungsi dari jarak *Euclidian* antar lokasi. Pembentukan fungsi pembobot dari jarak *Euclidian* salah satunya dapat menggunakan fungsi *Adaptive Bisquare Kernel*. Fungsi *Adaptive Bisquare Kernel* merupakan fungsi kernel dengan bandwidth yang berbeda pada setiap lokasi pengamatan. Fungsi *Adaptive Bisquare Kernel* dinyatakan sebagai berikut.

$$w_{ii^*} = \begin{cases} \left(1 - \left(\frac{d_{ii^*}}{h_i}\right)^2\right)^2 & ; \text{ untuk } d_{ii^*} \leq h_i \\ 0 & \end{cases} \quad (2.7)$$

dengan $d_{ii^*} = \sqrt{(u_i - u_{i^*})^2 + (v_i - v_{i^*})^2}$

d_{ii^*} adalah jarak *Euclidian* antara lokasi ke- i dan lokasi ke- i^* .

Sedangkan h_i adalah parameter penghalus atau yang disebut sebagai *bandwidth* dari lokasi ke- i . Pemilihan *bandwidth* optimum diperlukan untuk mengatur variansi dan bias dari penaksir yang dihasilkan. Pemilihan bandwidth optimum dapat dilakukan menggunakan metode *Cross Validation* (CV). Metode CV ini didefinisikan oleh persamaan :

$$CV(h) = \sum_{i=1}^n (Y_i - Y_{\neq i}(h))^2 \quad (2.8)$$

dengan

$Y_{\neq i}(h)$: nilai penaksir y_i ketika pengamatan di lokasi (u_i, v_i) tidak diikutsertakan pada penaksiran

V_I : transpose dari matriks S, matriks yang mewakili nilai estimasi model

2.8 Regresi Binomial Negatif

Regresi binomial negatif merupakan salah satu solusi untuk mengatasi adanya kasus overdispersi. Model regresi binomial negatif memiliki fungsi massa peluang sebagai berikut (Greene, 2008).

$$f(y, \mu, \theta) = \frac{\Gamma(y + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y+1)} \left(\frac{1}{1 + \theta\mu}\right)^{\theta^{-1}} \left(\frac{\theta\mu}{1 + \theta\mu}\right)^y, y = 0, 1, 2, \dots \quad (2.9)$$

Model binomial negatif merupakan paduan antara distribusi Poisson dan Gamma.

Berikut ini merupakan model regresi binomial negatif.

$$\mu_i = \exp(\beta_0(u_i, v_i)) + \beta_1(u_i, v_i)x_{i1} + \dots + \beta_p(u_i, v_i)x_{ip} = 0 \quad (2.10)$$

Fungsi peluang dari distribusi binomial negatif adalah.

$$f(y, \mu, \theta) = \exp \left\{ \ln \frac{\Gamma(y_j + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_j + 1)} + \frac{1}{\theta} \ln \left(\frac{1}{1 + \theta\mu_j} \right) + y \ln \left(\frac{\theta\mu_j}{1 + \theta\mu_j} \right) \right\} \quad (2.11)$$

Pendugaan parameter model regresi binomial negatif menggunakan metode *Maximum Likelihood Estimation* (MLE). Kemudian untuk mendapatkan nilai dari parameternya digunakan iterasi Newton Raphson. Fungsi likelihood dari regresi binomial negatif adalah sebagai berikut.

$$L(\beta, \theta) = \prod_{j=1}^n \left\{ \frac{\Gamma(y_j + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_j + 1)} \left(\frac{1}{1 + \theta\mu_j}\right)^{\theta^{-1}} \left(\frac{\theta\mu_j}{1 + \theta\mu_j}\right)^{y_j} \right\} \quad (2.12)$$

2.9 Geographically Weighted Negative Binomial Regression (GWNBR)

Geographically Weighted Negative Binomial Regression (GWNBR) merupakan salah satu metode yang cukup efektif menduga data yang memiliki heterogenitas spasial untuk data count yang over dispersi. Model GWNBR

merupakan pengembangan dari model regresi binomial negatif. Model GWNBR menghasilkan parameter lokal dengan masing-masing lokasi akan memiliki parameter yang berbeda-beda. Model GWNBR dapat dirumuskan sebagai berikut (Ricardo & Carvalho, 2013)

$$y_i \sim NB \left[\exp(\beta_j(u_i, v_i) x_{ik}), \theta(u_i, v_i) \right], i = 1, 2, \dots, n \quad (2.13)$$

dimana,

y_i : nilai observasi respon ke- i

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

$\beta_j(u_i, v_i)$: koefisien regresi variabel prediktor ke- j untuk setiap lokasi (u_i, v_i)

$\theta(u_i, v_i)$: parameter disperse untuk setiap lokasi (u_i, v_i)

Fungsi distribusi Binomial Negatif untuk setiap lokasi dapat ditulis dalam bentuk persamaan sebagai berikut.

$$f(y_i | \beta_j(u_i, v_i), \theta_i) = \frac{\Gamma\left(y_i + \frac{1}{\theta_i}\right)}{\Gamma\left(\frac{1}{\theta_i}\right) \Gamma(y_i + 1)} \left(\frac{1}{1 + \theta_i \mu_i}\right)^{\frac{1}{\theta_i}} \left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i}\right)^{y_i}, y_i = 0, 1, 2, \dots \quad (2.14)$$

dimana, $\mu_i = \exp(x_i^T \beta_j(u_i, v_i))$

$$\theta_i = \theta(u_i, v_i)$$

Estimasi model GWNBR menggunakan *Maximum Likelihood Estimation* (MLE). Langkah awal dari metode Likelihood adalah membentuk fungsi *Likelihood* sebagai berikut.

$$L(\beta(u_i, v_i), \theta_i, i = 1, 2, \dots, n) = \prod_{i=1}^n \left(\prod_{r=0}^{y_i-1} \left(r + \frac{1}{\theta_i}\right) \right) \frac{1}{y_i!} \left(\frac{1}{1 + \theta_i \mu_i}\right)^{\frac{1}{\theta_i}} \left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i}\right)^{y_i} \quad (2.15)$$

Pengujian signifikansi parameter model GWNBR terdiri dari uji serentak dan parsial. Uji signifikansi serentak menggunakan *Maximum Likelihood Ratio Test* (MLRT).

2.12 Kriteria Kebaikan Model

Salah satu tujuan dalam analisis regresi adalah untuk mendapatkan model terbaik. Model terbaik adalah model yang mampu menjelaskan hubungan antara variabel prediktor dengan variabel respon berdasarkan kriteria tertentu. Kriteria yang sering digunakan dalam pemilihan model terbaik adalah nilai AIC.

Akaike Information Criterion (AIC) adalah kriteria kesesuaian model dalam menduga model secara statistik. Kriteria AIC digunakan apabila pemodelan regresi bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap model. Besarnya nilai AIC sejalan dengan nilai devians dari model. Semakin kecil nilai devians maka akan semakin kecil pula tingkat kesalahan yang dihasilkan model sehingga model yang diperoleh menjadi semakin tepat. Nilai devians akan semakin kecil apabila rasio antara fungsi *likelihood* di bawah H_0 dengan fungsi *likelihood* di bawah populasi semakin besar. Hal ini mengindikasikan bahwa parameter yang diuji semakin mendekati nilai parameter populasi yang sebenarnya yang berarti dugaan model semakin baik. Oleh karena itu, model terbaik adalah model dengan AIC terkecil dan dengan devians terkecil pula. Nilai AIC dirumuskan sebagai berikut.

$$AIC = -2 \ln L(\hat{\beta}) + 2k \quad (2.16)$$

dimana $L(\hat{\beta})$ adalah *likelihood* dari model GWNBR, sedangkan k adalah jumlah parameter dalam model.

2.10 *Human Immunodeficiency Virus (HIV)*

HIV/AIDS merupakan penyakit menular yang disebabkan oleh infeksi Human Immunodeficiency Virus yang menyerang sistem kekebalan tubuh. Infeksi tersebut menyebabkan penderita mengalami penurunan ketahanan tubuh sehingga sangat mudah untuk terinfeksi berbagai macam penyakit lain. Sebelum memasuki fase AIDS, penderita terlebih dulu dinyatakan sebagai HIV positif. Jumlah HIV positif yang ada di masyarakat dapat diketahui melalui 3 metode, yaitu pada layanan Voluntary, Counseling, and Testing (VCT), sero survey, dan Survei Terpadu Biologis dan Perilaku (STBP) (Bappenas, 2014).

HIV dan virus-virus sejenisnya umumnya ditularkan melalui kontak langsung antara lapisan kulit dalam (membran mukosa) atau aliran darah, dengan cairan tubuh yang mengandung HIV, seperti darah, air mani, cairan vagina, cairan preseminal, dan air susu ibu. Penularan dapat terjadi melalui hubungan intim (vaginal, anal, ataupun oral), transfusi darah, jarum suntik yang terkontaminasi, serta bentuk kontak lainnya dengan cairan-cairan tubuh tersebut.

Permasalahan HIV dan AIDS menjadi tantangan kesehatan hampir di seluruh dunia, termasuk di Indonesia. Sejak pertama kali ditemukan sampai dengan Juni 2018, HIV/ AIDS telah dilaporkan keberadaannya oleh 433 (84,2%) dari 514 kabupaten/kota di 34 provinsi di Indonesia.

Jumlah kumulatif infeksi HIV yang dilaporkan sampai dengan Juni 2018 sebanyak 301.959 jiwa (47% dari estimasi ODHA jumlah orang dengan HIV AIDS tahun 2018 sebanyak 640.443 jiwa) dan paling banyak ditemukan di kelompok umur 25-49 tahun dan 20-24 tahun. Adapun provinsi dengan jumlah infeksi HIV

tertinggi adalah DKI Jakarta (55.099), diikuti Jawa Timur (43.399), Jawa Barat (31.293), Papua (30.699), dan Jawa Tengah (24.757). Jumlah kasus HIV yang dilaporkan terus meningkat setiap tahun, sementara jumlah AIDS relatif stabil. Hal ini menunjukkan keberhasilan bahwa semakin banyak orang dengan HIV /AIDS (ODHA) yang diketahui statusnya saat masih dalam fase terinfeksi (HIV positif) dan belum masuk dalam stadium AIDS (Depkes RI, 2018)

Diagnosis infeksi HIV/AIDS dapat dilakukan berdasarkan klasifikasi klinis WHO dan CDC. Gejala mayor dan gejala minor infeksi HIV/AIDS menurut WHO dijelaskan pada Tabel 2.1 sebagai berikut.

Tabel 2.1 Gejala HIV/AIDS

Gejala Mayor	Gejala Minor
Berat badan menurun >10% dalam 1 bulan	Batuk menetap >1 bulan
Diare kronis berlangsung >1 bulan	Dermatitis generalisata
Demam berkepanjangan >1 bulan	Herpes simpleks kronis progresif
Demensia/HIV ensefalopati	Infeksi jamur berulang pada alat kelamin wanita
	<i>Retinitis Cytomegalovirus</i>
	<i>Kandidiasis Orofaringeal</i>

Seseorang yang terinfeksi virus HIV berpotensi menularkan meski tidak memiliki ciri yang dapat dilihat secara kasat mata (fisik). Status HIV seseorang hanya dapat diketahui dengan melakukan cek/pemeriksaan darah di laboratorium. Karena itu, Jika merasa pernah melakukan perilaku berisiko atau merasa berisiko tertular segera lakukan tes HIV.

Upaya pencegahan dan pengendalian HIV -AIDS bertujuan untuk mewujudkan target Three Zero pada 2030, yaitu: 1) Tidak ada lagi penularan infeksi

baru HIV, 2) Tidak ada lagi kematian akibat AIDS, dan 3) Tidak ada lagi stigma dan diskriminasi pada orang dengan HIV AIDS (ODHA).

