BAB III

REGRESI SPASIAL DENGAN PENDEKATAN

GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR)

3.1 Regresi Poisson

Regresi Poisson merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data tentang peristiwa yang jarang terjadi. Regresi Poisson mengasumsikan bahwa variabel acak Y berdistribusi Poisson dengan fungsi peluangnya sebagai berikut:

$$f(y_i;\mu_i) = \frac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!}, \qquad y_i = 0,1,2,\ldots$$

dimana parameter $\mu_i > 0$. Model regresi Poisson dapat ditulis sebagai berikut:

$$y_i = \mu_i + \varepsilon_i, \quad i = 1,2,\ldots,n$$

Diasumsikan terdapat fungsi g yang menghubungkan mean dari variabel respon dengan prediktor linear, yaitu

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k = \boldsymbol{x_i^T\beta}$$

Dimana

X_i : Variabel prediktor ke-i

$\boldsymbol{\beta}$: matriks parameter regresi ukuran $(k+1)\times 1$

$\boldsymbol{x_i^T}$: matriks transpose dari X_i ukuran $(k+1)\times n$

Fungsi g merupakan fungsi penghubung (*link function*). Hubungan antara mean dan prediktor linear adalah

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x_i^T\beta})$$

Fungsi penghubung yang paling cocok digunakan untuk regresi Poisson adalah fungsi penghubung log, karena fungsi log menjamin bahwa nilai variabel yang diharapkan dari variabel responnya akan bernilai non negatif. Fungsi penghubung log berbentuk

$$\eta_i = \ln \mu_i$$

Hubungan antara mean variabel respon dengan prediktor linear adalah

$$\ln \mu_i = \boldsymbol{x_i^T\beta}$$

$$e^{\ln \mu_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$$

$$\mu_i = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$$

Sehingga fungsi penghubung untuk model regresi Poisson mempunyai logaritma seperti berikut:

$$\ln \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

$$\mu_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)$$

Persamaan distribusi Poisson dapat dinyatakan sebagai

$$f(y_i; \mu_i) = \frac{e^{-\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta} y_i)}{y_i!}, \qquad y_i = 0,1,2, \ldots$$

Selanjutnya model regresi Poisson dapat ditulis sebagai:

$$y_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \varepsilon_i$$

Penaksiran parameter dari model regresi Poisson tersebut dengan menggunakan metode kemungkinan maksimum (*Maksimum Likelihood Estimator*). Fungsi kemungkinannya adalah sebagai berikut:

$$L(y; \beta) = \prod_{i=1}^{m} \frac{e^{-\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta} y_i)}{y_i!}$$

Persamaan di atas akan dimaksimumkan dengan menggunakan teknik iteratif yang menghasilkan penaksir kemungkinan maksimum untuk koefisien regresi dalam $\hat{\beta}$

$$\ln L(y; \beta) = \ln \prod_{i=1}^{m} \frac{e^{-\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta} y_i)}{y_i!}$$

$$= \sum_{i=1}^{m} \ln \left[\frac{e^{-\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta} y_i)}{y_i!} \right]$$

$$= \sum_{i=1}^{m} \left[\ln \left(e^{-\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})} \right) + \ln(\exp(\boldsymbol{x}_i^T \boldsymbol{\beta} y_i)) - \ln y_i! \right]$$

$$= \sum_{i=1}^{m} [\boldsymbol{x}_i^T \boldsymbol{\beta} y_i - \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) - \ln y_i!]$$

Untuk mendapatkan nilai maksimum likelihood-nya, maka digunakan metode perhitungan iterasi numerik Newton Raphson karena persamaan tersebut implisit.

3.2 *Akaike Information Criterion* (AIC)

Akaike Information Criterion (AIC) merupakan pengukuran untuk kualitas relatif dari model statistik dari data yang diberikan untuk pemilihan model terbaik dari beberapa model yang ada. Perhitungan AIC dapat dilakukan dengan rumus:

$$AIC = 2k - 2\ln(likelihood)$$

dimana :

k : banyaknya parameter yang akan ditaksir.

$\ln(likelihood)$: nilai maksimum likelihood model.

Untuk ukuran sampel yang terbatas digunakan AICc, yaitu nilai AIC yang telah dikoreksi

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

dengan n merupakan ukuran sampel. Selain itu, jika nilai k yang semakin besar atau variabel yang akan ditaksirnya semakin banyak, maka penggunaan nilai AICc ini jauh lebih baik dibandingkan dengan menggunakan nilai AIC, misalnya ketika n tidak lebih besar dari k^2. Namun, ketika n semakin besar, nilai AICc akan konvergen ke nilai AIC, sehingga AICc dapat dihiraukan.

Model yang terbaik yaitu model yang memiliki nilai AIC atau AICc terkecil.

3.3 Pembobotan

Matriks pembobot untuk titik regresi ke-i adalah

$$W(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}$$

Misalkan w_{ij} merupakan pembobotan geografis dari data ke-j pada regresi ke-i dan d_{ij} merupakan jarak antara observasi ke-j dan regresi ke-i. Pada model regresi global tanpa pembobotan geografis, masing-masing observasi memiliki nilai pembobotan sebesar 1.

$$w_{ij} = 1, \quad i,j = 1,2,\ldots,n$$

dengan j menunjukkan titik pada daerah dimana data diobservasi dan i merupakan sebuah titik pada ruang untuk penakasiran parameter. Pada model GWR, pembobotan bervariasi sesuai lokasi pada titik regresi ke-i, dimana $0 \leq w_{ij} \leq 1$ dan w_{ij} semakin kecil ketika jarak d_{ij} bertambah. Artinya, observasi yang dekat dengan titik regresi akan memberikan bobot yang besar dibandingkan yang jauh dari titik regresi.

Tabel 3.1 Fungsi Pembobot Spasial pada *Geographically Weighted Regression*

Fungsi pembobot	**Bentuk Fungsi**
Fixed Gaussian	$w_{ij} = \exp\left(-d_{ij}^2/h^2\right)$
Fixed Bisquare	$w_{ij} = \left(1 - \frac{d_{ij}^2}{h^2}\right)^2 ; d_{ij} < h$ $= 0; d_{ij} > h$
Adaptive Bisquare	$w_{ij} = \left(1 - \frac{d_{ij}^2}{h_i^2}\right)^2 ; d_{ij} < h_i$ $= 0; d_{ij} > h_i$
Adaptive Gaussian	$w_{ij} = \exp\left(-d_{ij}^2/h_i^2\right)$

w_{ij} : nilai bobot dari observasi pada lokasi ke-j untuk penaksiran koefisien pada lokasi ke-i

d_{ij} : jarak *euclidean* antara i dan j

h : ukuran *bandwidth* yang ditetapkan didefinisikan dengan ukuran matriks jarak

h_i : ukuran *bandwidth* yang diadaptasi sebagai jarak tetangga terdekat ke-i

Jarak *euclidean* merupakan jarak antara titik regresi ke-i dengan lokasi ke-j $(i \neq j)$ yang dirumuskan dengan

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

dimana:

u_i : *Longitude* pada lokasi i

u_j : *Longitude* pada lokasi j

v_i : *Latitude* pada lokasi i

v_j : *Latitude* pada lokasi j

Bandwidth merupakan radius suatu lingkaran dengan pusat titik lokasi i, sehingga jika terdapat titik-titik lokasi yang berada dalam lingkaran tersebut, maka titik-titik lokasi tersebut dianggap memiliki pengaruh terhadap penaksiran koefisien regresi pada titik lokasi i. *Bandwidth* yang digunakan yaitu *bandwidth* dengan nilai optimum yang dapat dicari dengan menggunakan metode Cross Validation (CV) dengan rumus:

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

dimana $\hat{y}_{i \neq i}(b)$ merupakan nilai penaksir y_i denan radius b, tetapi pengamatan di titik i dihilangkan dari proses penaksiran. *Bandwidth* yang optimum didapat ketika nilai CV minimum.

Pada fungsi *Fixed*, *bandwidth* yang optimum ditentukan sama pada setiap wilayah yang dianalisis, baik itu penyebaran data yang padat atau tidak, pembobotan akan diberikan nilai yang sama. Sedangkan pada fungsi *adaptive* akan memiliki *bandwidth* yang berbeda-beda sesuai dengan kepadatan data pada wilayah analisis. Ketika data padat, *bandwidth* akan kecil, sedangkan ketika data jarang, *bandwidth* akan semakin besar. Fungsi ini mampu menyesuaikan ukuran varians data. Pembobotan yang terbaik yaitu pembobotan dengan hasil perhitungan AICc yang terkecil.

3.4 Penaksiran Parameter

Model GWPR merupakan pengembangan dari model regresi Poisson yang menghasilkan penaksiran parameter model yang bersifat lokal untuk setiap lokasi tempat data tersebut didapat. Variabel respon dan variabel prediktornya bergantung pada lokasi dimana data tersebut didapat.

Penaksiran parameter dilakukan dengan menggunakan metode kemungkinan maksimum. Letak geografis merupakan pembobot pada model ini, sehingga pembobot diberikan pada bentuk log-likelihoodnya untuk model GWPR untuk pengamatan ke-j pada lokasi ke-i, yaitu:

$$\ln L(\beta(u_i,v_i)) = \sum_{j=1}^{n} \left[y_j x_j \beta(u_j,v_j) - \exp\left(x_j\beta(u_j,v_j)\right) - \ln y!\right] W_{ij}(u_j,v_j)$$

Penaksiran parameter didapat dengan menurunkan fungsi log-likelihood, kemudian menurunkannya terhadap $\beta(u_i,v_i)$

$$\frac{\partial \ln L(\beta(u_i,v_i))}{\partial\beta(u_j,v_j)} = \frac{\partial \sum_{j=1}^{n} \left[y_j x_j \beta(u_j,v_j) - \exp\left(x_j\beta(u_j,v_j)\right) - \ln y!\right] W_{ij}(u_j,v_j)}{\partial\beta(u_i,v_i)}$$

$$= \sum_{j=1}^{n} \left[y_j x_j - x_j \exp\left(x_j\beta(u_j,v_j)\right)\right] W_{ij}(u_j,v_j)$$

Kemudian untuk memaksimumkan fungsi log likelihood tersebut persamaan di atas dibuat sama dengan nol, sehingga diperoleh:

$$\frac{\partial \ln L(\beta(u_i,v_i))}{\partial\beta(u_j,v_j)} = \sum_{j=1}^{n} \left[y_j x_j - x_j \exp\left(x_j\beta(u_j,v_j)\right)\right] W_{ij}(u_j,v_j) = 0$$

Persamaan di atas merupakan persamaan implisit, sehingga untuk menyelesaikannya digunakan prosedur iterasi numerik dengan menggunakan pendekatan metode iterasi Newton Raphson. Rumus iterasinya sebagai berikut:

$$\beta^{(m+1)}(u_<,v_i) = \beta_m(u_i,v_i) - H_m^{-1}\left(\beta_m(u_i,v_i)\right) g_m\left(\beta_m(u_i,v_i)\right)$$

dimana

$$g_m\left(\beta_m(u_i,v_i)\right) = \frac{\partial \ln L(\beta(u_i,v_i))}{\partial\beta(u_i,v_i)}$$

$$= \sum_{i=1}^{n} x_i y_i W_{ij}(u_j,v_j) - \sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) \exp(x_i\beta(u_i,v_i))$$

$$= \sum_{i=1}^{n} [x_i y_i W_{ij}(u_j,v_j) - x_i W_{ij}(u_j,v_j) \exp(x_i\beta(u_i,v_i))]$$

$$= -\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) \exp\left(x_i\beta(u_i,v\_)\right) + \sum_{i=1}^{n} x_i y_i W_{ij}(u_j,v_j)$$

$$H_m^{-1}\left(\beta_m(u_i,v_i)\right) = \frac{\partial^2 \ln L(\beta(u_i,v_i))}{\partial\beta(u_i,v_i)\,\partial\beta^T(u_i,v_i)}$$

$$= -\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) x_i^T \exp\left(x_i^T \beta(u_i,v_i)\right)$$

Sehingga diperoleh:

$$\beta^{(m+1)}(u_i,v_i) = \beta_m(u_i,v_i) - H_m^{-1}\left(\beta_m(u_i,v_i)\right) g_m\left(\beta_m(u_i,v_i)\right)$$

$$= \beta_m(u_i,v_i) - \left[\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) x_i^T \exp\left(x_i^T\, \beta(u_i,v_i)\right)\right]^{-1}$$

$$\left[-\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) \exp(x_i\beta(u_i,v_i)) + \sum_{i=1}^{n} x_i y_i W_{ij}(u,v_j)\right]$$

$$= \beta_m(u_i,v_i) - \left[\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) x_i^T \exp\left(x_i^T \beta(u_i,v_i)\right)\right]^{-1}$$

$$\left[-\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) \exp(x_i\beta(u_i,v_i)) + \sum_{i=1}^{n} x_i y_i W_{ij}(u_j,v_j)\right]$$

misalkan, $\exp\left(x_i^T \beta(u_i,v_i)\right) = \hat{\mu}_i$, sehingga:

$$\beta^{(m+1)}(u_i,v_i) = \beta_m(u_i,v_i) - \left[\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) x_i^T \hat{\mu}_i\right]^{-1}$$

$$\left[-\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j)\hat{\mu}_i + \sum_{i=1}^{n} x_i y_i W_{ij}(u_j,v_j)\right]$$

$$= \left[\sum_{i=1}^{n} x_i W_{ij}(u_i,v_i)\hat{\mu}_i x_i^T\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_i,v_i)\hat{\mu}_i x_i^T\right] \beta_m(u_i,v_i)$$

$$- \left[\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j) x_i^T \hat{\mu}_i\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_j,v_j)\hat{\mu}_i\right]$$

$$+ \left[\sum_{i=1}^{n} x_i W_{ij}(u,v_j) x_i^T \hat{\mu}_i\right]^{-1} \left[\sum_{i=1}^{n} x_i y_i W_{ij}(u_j,v_j)\right]$$

$$= \left[\sum_{i=1}^{n} x_i W_{ij}(u_i, v_i)\hat{\mu}_i x_i^T\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_i, v_i)\hat{\mu}_i x_i^T\right] \beta_m(u_i, v_i)$$

$$- \left[\sum_{i=1}^{n} x_i W_{ij}(u_j, v_j) x_i^T \hat{\mu}_i\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_j, v_j)\hat{\mu}_i\right] \left[1 - \frac{y_i}{\hat{\mu}_i}\right]$$

$$= \left[\sum_{i=1}^{n} x_i W_{ij}(u_i, v_i)\hat{\mu}_i x_i^T\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_j, v_j)\hat{\mu}_i\right]$$

$$\left\{x_i^T \beta_m(u_i, <_i) - \left[1 - \frac{y_i}{\hat{\mu}_i}\right]\right\}$$

$$= \left[\sum_{i=1}^{n} x_i W_{ij}(u_i, v_i)\hat{\mu}_i x_i^T\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_j, v_j)\hat{\mu}_i\right]$$

$$\left\{x_i^T \beta_m(u_i, v_i) + \left[\frac{y_i}{\hat{\mu}_i} - 1\right]\right\}$$

$$\beta^{(m+1)}(u_i, v_i) = \left[\sum_{i=1}^{n} x_i W_{ij}(u_i, v_i)\hat{\mu}_i x_i^T\right]^{-1} \left[\sum_{i=1}^{n} x_i W_{ij}(u_j, v_j)\hat{\mu}_i\right]$$

$$\left\{x_i^T \beta_m(u_i, v_i) + \left[\frac{y_i - \hat{\mu}_i}{\gtrless_i}\right]\right\}$$

Dengan mengulang prosedur iterasi untuk setiap titik regresi ke-i dengan $m = 1$, maka penaksiran lokal akan didapatkan. Iterasi akan dihentikan pada saat konvergen, yaitu dengan syarat $\left|\beta^{(m+1)}(u_i, v_i) - \beta^{m}(u_i, v_i)\right| \leq \epsilon$, dimana ϵ merupakan bilangan positif yang sangat kecil.