II. TINJAUAN PUSTAKA

2.1 Regresi Linear

Istilah regresi diperkenalkan pertama kali pada tahun 1886 oleh Sir Francis Galton, secara umum analisis regresi bertujuan untuk mengestimasi serta memprediksi nilai rata-rata variabel dependen berdasarkan nilai variabel independen. Menurut Supangat (2007), model regresi adalah suatu persamaan yang menyatakan hubungan antara variabel dependen dengan satu atau lebih variabel independen, jika parameter pada model tersebut berhubungan secara linear dengan variabel dependen maka disebut regresi linear. Berdasarkan kepada hubungan antar varibel independen, regresi linear dapat dibagi menjadi 2, yaitu regresi linear sederhana dan regresi linear berganda.

2.1.1 Regresi Linear Berganda

Jika regresi linear sederhana adalah persamaan yang difokuskan untuk mengetahui hubungan antar dua variabel, maka regresi linear berganda adalah persamaan yang digunakan untuk mengetahui hubungan beberapa variabel independen terhadap variabel dependen (Suyono, 2015).

Adapun bentuk persamaan dari regresi linear berganda yaitu (Draper & Smith, 1998):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

dengan:

Y : variabel dependen

β_0 : konstanta

$\beta_1, \beta_2, \cdots, \beta_k$: koefisian regresi

$X_1, X_2, \cdots, X_k$: variabel independen

ε : galat/eror

Misalkan y_i adalah observasi dari variabel dependen y untuk pengamatan ke$-i$, x_{it} adalah nilai observasi independen ke$-t$ untuk pengamatan ke$-i$ dan ε_i merupakan galat dari pengamatan ke$-i$ untuk $i = 1,2,3, \cdots, n$. Misalkan terdapat sebanyak k variabel independen dan n pengamatan, maka persamaan regresi dapat dituliskan sebagai berikut:

$$\begin{gathered} y_1 = \beta_1 + x_{12}\beta_2 + \cdots + x_{1k}\beta_k + \varepsilon_1 \\ y_2 = \beta_1 + x_{22}\beta_2 + \cdots + x_{2k}\beta_k + \varepsilon_2 \\ \vdots \\ y_n = \beta_n + x_{n2}\beta_2 + \cdots + x_{nk}\beta_k + \varepsilon_n \end{gathered}$$

atau dapat ditampilkan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

dengan:

Y : vektor observasi untuk variabel dependen berukuran $n \times 1$

X : matriks k variabel independen atau variabel regressor berukuran $n \times k$

β : vektor parameter berukuran $k \times 1$

ε : vektor eror berukuran $n \times 1$

Ordinary Least Square (Motode Kuadrat Terkecil) adalah metode untuk mendapatkan garis regresi yang baik, yaitu sedekat mungkin dengan datanya sehingga menghasilkan prediksi yang baik. Pada dasarnya metode ini meminimumkan jumlah kuadrat *error* (Draper & Smith, 1998). Adapun proses estimasi dengan menggunakan metode ini adalah sebagai berikut:

1. Bentuk umum : $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$
2. Nyatakan sebagai $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\widehat{\beta}}$, dengan $\boldsymbol{\widehat{\beta}}$ merupakan suatu vektor kolom $k + 1$ dan ε merupakan suatu vektor kolom $n \times 1$ dari n residual. Dimana $\sum_{i=1}^{n} \varepsilon_i^2$ merupakan jumlah kuadrat residual. Dalam notasi matriks, ini sama dengan meminimumkan $\varepsilon\varepsilon^T$ karena

$$\boldsymbol{\varepsilon\varepsilon^T} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

$$\boldsymbol{\varepsilon\varepsilon^T} = \left(\boldsymbol{Y} - \boldsymbol{X\widehat{\beta}}\right)\left(\boldsymbol{Y} - \boldsymbol{X\widehat{\beta}}\right)^{\boldsymbol{T}}$$

$$\boldsymbol{\varepsilon\varepsilon^T} = \boldsymbol{YY^T} - \boldsymbol{2Y\widehat{\beta}^T X^T} + \boldsymbol{X\widehat{\beta}\widehat{\beta}^T X^T} \tag{3}$$

3. Menggunakan sifat-sifat transpose suatu matriks, yaitu $\left(\boldsymbol{X\widehat{\beta}}\right)^{\boldsymbol{T}} = \boldsymbol{\widehat{\beta}^T X^T}$ dan $\boldsymbol{Y\widehat{\beta}^T X^T}$ adalah suatu saklar atau angka *real*, sehingga bentuk itu sama dengan transposenya $\boldsymbol{X\widehat{\beta}Y^T}$. Sehingga dapat dinyatakan dengan:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3k} \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

$$\boldsymbol{\varepsilon_i} = \boldsymbol{Y_i} - \boldsymbol{\widehat{\beta}_0} - \boldsymbol{\widehat{\beta}_1 X_{i1}} - \boldsymbol{\widehat{\beta}_2 X_{i2}} - \cdots - \boldsymbol{\widehat{\beta}_k X_{ik}}$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)^2$$

4. Jika diturunkan atau dideferensialkan JKG (jumlah kuadrat galat) secara berurutan terhadap $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ maka diperoleh:

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)(X_{i1})$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_2} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)(X_{i2})$$

$$\vdots$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_k} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)(X_{ik})$$

5. Kemudian di samakan dengan 0 maka diperoleh persamaan sebagai berikut:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 X_{i1} - \sum_{i=1}^{n} \hat{\beta}_2 X_{i2} - \cdots - \sum_{i=1}^{n} \hat{\beta}_0 X_{ik}$$

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ik}$$

$$\sum_{i=1}^{n} X_{i1} Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_{i1} X_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{i1} X_{ik}$$

$$\sum_{i=1}^{n} X_{i2} Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} X_{i1} X_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}^2 + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{i2} X_{ik}$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ik} Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} X_{ik} X_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} X_{ik} X_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ik}^2$$

6. Dinyatakan ke dalam bentuk matriks, persamaan di atas akan menjadi:

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^{2} & \sum_{i=1}^{n} X_{i1} X_{i2} & \cdots & \sum_{i=1}^{n} X_{i1} X_{ik} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1} X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \cdots & \sum_{i=1}^{n} X_{i2} X_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{ik} X_{i1} & \sum_{i=1}^{n} X_{ik} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \widehat{\beta_k} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

$$(\boldsymbol{X^T X})\boldsymbol{\widehat{\beta}} = \boldsymbol{X^T Y} \tag{4}$$

7. Persamaan (4) diperoleh dari menurunkan persamaan matriks (4) terhadap $\boldsymbol{\widehat{\beta}}$, sehingga diperoleh:

$$\frac{\partial(\boldsymbol{\varepsilon\varepsilon^T})}{\partial \boldsymbol{\widehat{\beta}}} = \boldsymbol{-2X^T Y + 2X^T X\widehat{\beta}}$$

8. Kemudian samakan hasil persamaan di atas dengan nol, sehingga diperoleh:

$$-2\boldsymbol{X^TY} + 2\boldsymbol{X^TX\widehat{\beta}} = 0$$

$$2\boldsymbol{X^TX\widehat{\beta}} = 2\boldsymbol{X^TY}$$

$$\boldsymbol{X^TX\widehat{\beta}} = \boldsymbol{X^TY}$$

9. Kedua sisi pada persamaan diatas dikalikan dengan $(X^TX)^{-1}$, maka diperoleh:

$$(\boldsymbol{X^TX})^{-\mathbf{1}}(\boldsymbol{X^TX})\boldsymbol{\widehat{\beta}} = (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TY}$$

$$\frac{\mathbf{1}}{(\boldsymbol{X^TX})}(\boldsymbol{X^TX})\boldsymbol{\widehat{\beta}} = (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TY}$$

$$\boldsymbol{\widehat{\beta}} = (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TY} \tag{5}$$

2.2 Uji Asumsi Klasik

Menurut Ghozali (2018), uji asumsi klasik adalah suatu syarat bagi analisis regresi berganda yang berbasis *ordinary least square*. Uji asumsi klasik ini bertujuan untuk menghasilkan model regresi yang memenuhi kriteria *BLUE (Best Linear Unbiased Estimator)*, model regresi yang memenuhi kriteria *BLUE* dapat digunakan sebagai estimator yang terpecaya dikarenakan estimator tersebut dinyatakan tidak bias, berdistribusi normal, efisien dan juga konsisten. Sebuah model regresi dikatakan telah memenuhi kriteria BLUE apabila telah melewati serangkaian uji, yaitu uji normalitas, uji multikolinearitas, uji heteroskedastisitas serta uji autokorelasi.

2.2.1 Uji Normalitas

Uji normalitas bertujuan untuk memastikan bahwa residual (ε_i) yang telah distandarisasi telah berdistribusi normal atau dapat dilambangkan dengan $\varepsilon_i \sim N(0, \sigma^2)$. Nilai residual dikatakan berdistribusi normal jika nilainya sebagian besar mendekati nilai dari rata-ratanya, sehinggga bila residual tersebut berdistribusi normal dan digambarkan dalam kurva maka kurva tersebut akan berbentuk lonceng dimana kedua sisinya melebar hingga tidak terhingga (Suliyanto, 2008).

Menurut Ghozali (2018) dengan asumsi normalitas terpenuhi, estimator *ordinary least square* mempunyai sifat yang tak bias, konsisten serta efisien. Dan salah satu cara untuk melakukan uji normalitas adalah dengan uji *Kolmogorov Smirnov* dengan hipotesis sebagai berikut:

1. Hipotesis:

   H_0 : residual berdistribusi normal

   H_1 : residual tidak berdistribusi normal
2. Taraf signifikansi : $\alpha = 0{,}05$
3. Statistik uji :

$$KS_{hitung} = \max|F_s(x) - F_t(x)| \tag{6}$$

dengan:

$F_s(x)$: distribusi frekuensi kumulatif sampel

$F_t(x)$: distribusi frekuensi kumulatif teoritis

4. Kriteria pengambilan keputusan : H_0 akan ditolak jika nilai $KS_{hitung} > KS_{tabel}(n, \alpha)$ atau $sig. < 0{,}05$.

Jika asumsi normalitas tidak terpenuhi maka konsekuensinya ialah nilai prediksi yang diperoleh akan tidak konsisten dan bias. Cara untuk mengatasi asumsi normalitas yang tidak terpenuhi dapat dilakukan beberapa cara, yaitu: menambah jumlah data, melakukan tranformasi data atau menghilangkan data yang dianggap sebagai penyebab data tidak normal (Suliyanto, 2008).

2.2.2 Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menunjukkan bahwa dalam model regresi ditemukan adanya korelasi antara variabel independen atau tidak (Ghozali, 2018). Model yang baik seharusnya tidak terjadi korelasi yang tinggi diantara variabel independen. Cara untuk mendeteksi ada atau tidaknya multikolinearitas di dalam model regresi dapat diketahui dari nilai *tolerance* dan nilai *Variance Inflation Factor (VIF)*. Tolerance mengukur variabilitas variabel independen yang terpilih yang tidak dapat dijelaskan oleh variabel bebas lainnya. Hipotesis untuk melakukan uji multikolinearitas sebagai berikut:

1. Hipotesis:

   H_0 : terjadi multikolinearitas pada variabel independen

   H_1: tidak terjadi multikoliearitas pada variabel independen
2. Taraf signifikansi: $\alpha = 0{,}05$
3. Statistik uji yang digunakan:

$$VIF_j = \frac{1}{1-R_j^2} \tag{7}$$

   dengan:

   R_j^2 : koefisien determinasi antara X_j dengan variabel independen lainnya pada persamaan.
4. Kriteria pengambilan keputusan: H_0 akan ditolak jika nilai $VIF \leq 10$ atau $tol \geq 0{,}1$.

Menurut Suliyanto (2008), jika model regresi terdapat masalah multikolinearitas maka akan mengakibatkan penaksir *ordinary least square* tidak dapat ditentukan, nilai *t-statistik* koefisien dari satu atau beberapa variabel bebas secara statistik tidak signifikan sehingga dapat menyebabkan dikeluarkannya variabel bebas dari model dan mungkin R^2 bisa tinggi namun sangat sedikit taksiran koefisien regresi yang signifikan. Cara untuk mengatasi masalah multikolinearitas yang terjadi adalah dengan menghilangkan satu atau lebih variabel bebas, melakukan transformasi data atau menggabungkan data *time series* dan data *cross section*.

Multikolinearitas dapat diatasi dengan berbagai metode, termasuk *Partial Least Square (PLS)* dan *Principal Component Regression (PCR)*. PLS adalah suatu metode yang menggabungkan karakteristik dari komponen utama dan regresi linear berganda. Tujuan dari PLS adalah untuk mengestimasi dan menganalisis variabel terikat berdasarkan variabel bebas. Dalam konteks ini, PLS mengurangi dimensi variabel bebas dengan menciptakan variabel baru yang merupakan kombinasi linear dari variabel bebas dengan dimensi yang lebih kecil. Sementara itu, PCR adalah salah satu metode analisis regresi yang menggunakan komponen utama untuk menangani masalah multikolinearitas dalam regresi berganda (Supriyadi et al., 2017).

2.2.3 Uji heteroskedastisitas

Menurut Sunyoto (2013), jika residualnya mempunyai varian yang sama maka disebut terjadi homoskedastisitas dan jika berbeda disebut terjadi heteroskedastisitas, serta persamaan regresi yang baik adalah jika tidak terjadi heteroskedastisitas. Ciri-ciri tidak terjadi Heteroskedastisitas adalah titik-titik residual menyebar disekitar angka 0, titik-titik tidak mengumpul di satu tempat, serta penyebaran titik-titik residual tidak membentuk pola. Sehingga tidak terjadi heteroskedastisitas apabila ciri-ciri tidak terpenuhi. Salah satu cara dalam pengujian heteroskedastisitas dapat dilakukan dengan membuat alur sebaran (scatterplot) antara residual dan nilai prediksi dari variabel terikat yang telah distandarisasi. Adapun hipotesis sebagai berikut:

1. Hipotesis:
   H_0 : residual homoskedastisitas
   H_1 : residual heteroskedastis
2. Taraf signifikansi: $\alpha = 0{,}05$
3. Statistik uji yang digunakan:
$$t_{hitung} = \frac{\beta_j}{S(\beta_j)} \tag{8}$$
4. Kriteria pengambilan keputusan: H_0 ditolak jika nilai $|t_{hitung}| > t_{\left(\frac{\alpha}{2};n-k-1\right)}$ atau $sig. < 0{,}05$

Menurut Suliyanto (2008), konsekuansi dari terdapatnya maslaah heteroskedastisitas adalah formulasi untuk menafsir varian dari estimasi *ordinary least square* secara umum akan bias dan prediksi tidak efisien. Cara untuk mengatasi masalah heteroskedastisitas yang muncul adalah dengan melakukan transformasi data.

2.3 Pencilan (*Outlier*)

Pencilan atau *outlier* merupakan suatu pengamatan yang ekstrim, dimana residual nilai mutlaknya jauh lebih besar daripada yang lain dan bisa jadi terletak tiga atau empat simpangan baku dari rata-ratanya sehingga menyebabkan data sebagai pencilan (Montgomery et al., 2012). Pencilan adalah titik-titik data yang tidak setipe dengan titik data yang lainnya. Didalam himpunan data biasanya terdapat 10% *outlier* dan jumlah maksimum dari *outlier* yang diperbolhkan terdapat dalam suatu data adalah sebesar 50% (Hampel et al., 2005).

Secara umum data *outlier* merupakan suatu data yang berbeda dari data-data lainnya pada suatu hasil yang kemungkinan nilainya lebih besar dari nilai pengamatan pada umumnya ataupun lebih kecil. *Outlier* adalah suatu keganjilan yang menandakan suatu titik yang khas jika dibandingkan dengan pengamatan lainnya. Suatu *outlier* baru bisa untuk ditolak jika setelah ditelusuri ternyata *outlier* tersebut merupakan akibat dari beberapa kesalahan seperti kerusakan yang terjadi pada alat hitung, ketidaktepatan data dan analisis yang salah. Jikalau *outlier* bukan terjadi karena masalah tersebut maka penyelidikan lebih lanjut harus dilakukan, karena jika diabaikan dan dihapus data yang tersebut untuk memperbaiki persamaan akan berbahaya. Tindakan tersebut dapat menyebabkan kesalahan dalam memprediksi atau mengestimasi data (Draper & Smith, 1998).

Menurut Draper & Smith (1998), *outlier* harus dilihat terhadap posisi dan sebaran data lainnya sehingga akan diambil keputusan apakah *outlier* tersebut perlu dihilangkan atau tetap dibiarkan. Terdapat berbagai macam metode dalam mendeteksi *outlier* diantaranya *boxplot, levarage value, standardized residual, breakdown point* dan *cook's distance*. Dalam penelitian ini akan digunakan metode *boxplot*.

Metode *boxplot* adalah metode yang menggunakan nilai kurtil dan jangkauan untuk mendeteksi keberadaan dari *outlier*. Kuartil $1, 2$ dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan atau *IQR (Interquartile Range)* didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3. *Outlier* (pencilan) dapat diketahui berdasarkan nilainya, yaitu nilai yang kurang dari $1{,}5 \times IQR$ terhadap kuartil 1 dan nilai yang lebih besar dari $1{,}5 \times IQR$ terhadap kuartil 3. Adapun persamaan untuk mencari nilai IQR sebagai berikut (Paludi, 2009):

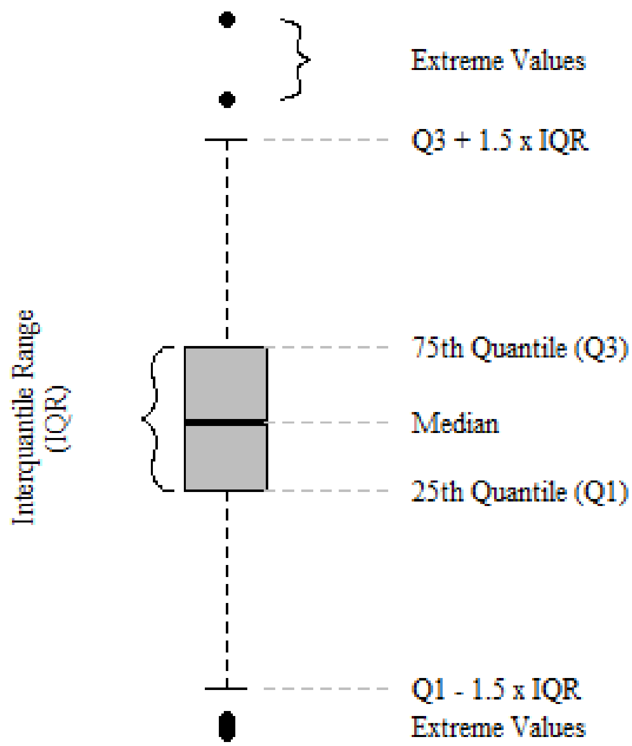
$$IQR = Q_3 - Q_1 \tag{9}$$

dengan:

IQR : jarak antar kuartil

Q_3 : kuartil ke-3

Q_1 : kuartil ke-1



Gambar 1. Pendeteksian *Outlier* Dengan *Boxplot*

Sumber : https://swampthingecology.org/blog/knitr_files/2020-01-24-Boxplot_files/figure-html/unnamed-chunk-5-1.png

2.4 Uji Keseluruhan Variabel

Dalam uji kelesulurahan variabel akan dugunakan uji F. Uji F digunakan untuk menguji apakah variabel-variabel independen secara bersamasama signifikan berpengaruh terhadap variabel dependen (Ghozali, 2018). Adapun hipotesis uji F sebagai berikut:

1. Hipotesis:

   $H_0: \beta_j = 0; j = 1,2,\cdots,k$ (seluruh variabel independen tidak berpengaruh secara simultan terhadap variabel dependen)

$H_1 = \beta_j \neq 0; j = 1,2,\cdots,k$ (seluruh variabel independen berpengaruh secara simultan terhadap veriabel dependen)

2. Taraf signifikansi: $\alpha = 0{,}05$
3. Satistik uji yang digunakan:

$$F_{hitung} = \frac{MSR}{MSE} \quad (10)$$

dengan:

$MSR = \frac{SSR}{k}$

$MSE = \frac{SSE}{n-k-1}$

SSR : jumlah kuadrat regresi = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

SSE : jumlah kuadrat *error* = $\sum_{i=1}^{n}(y_i - \bar{y})^2$

4. Kriteria keputusan: H_0 ditolak jika nilai $F_{hitunh} > F_{(\alpha;k,n-k-1)}$ atau $Sig. < 0{,}05$

2.5 Uji Masing-Masing Variabel

Uji masing-masing variabel dilakukan dengan menggunakan uji statistik t. Uji statistik t pada dasarnya digunakan untuk melihat seberapa jauh pengaruh satu variabel independen secara individual terhadap variabel dependen (Ghozali, 2018). Adapun hipotesis sebagai berikut:

1. Hipotesis:

$H_0: \beta_j = 0; j = 1,2,\cdots,k$ (variabel independen j tidak berpengaruh secara individual terhadap veriabel dependen)

$H_0: \beta_j \neq 0; j = 1,2,\cdots,k$ (variabel independen j berpengeruh secara individual terhadap variabel dependen)

2. Taraf signifikansi: $\alpha = 0{,}05$
3. Statistik uji yang digunakan:

$$t_{hitung} = \frac{\beta_j}{s_e(\beta_j)} \quad (11)$$

dengan:

β_j : koefisien regresi variabel ke-j ; $j = 1,2,\cdots,k$

$s_e(\beta_j)$: standar *error* koefisien regresi variabel k

4. Kriteria keputusan: H_0 ditolak jika nilai $|t_{hitung}| > t_{\left(\frac{\alpha}{2};n-k-1\right)}$ atau $sig. < 0{,}05$

2.6 Regresi Spasial

Menurut Anselin (1988) regresi spasial adalah pengembangan dari metode regresi linear klasik yang terjadi karena adanya pengaruh spasial terhadap data yang dianalisis. Peggunaan regresi spasial dalam menganalisis data yang dengan efek spasial dikarenakan jika menggunakan regresi linear sederhana ataupun regresi linear berganda maka model yang akan dihasilkan nantinya akan kurang akurat dan menyebabkan kesimpulan yang kurang tepat karena asumsi *error* saling bebas tidak terpenuhi.

Regresi spasial merupakan analisis yang digunakan untuk memodelkan data yang memiliki informasi spasial. Model umum regresi spasial dinyatakan dalan persamaan Anselin (1988):

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{12}$$

$$\boldsymbol{u} = \lambda\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon} \tag{13}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

dengan:

$\boldsymbol{y}$: vektor variabel dependen yang berukuran $n \times 1$

$\boldsymbol{X}$: matriks variabel independen berukuran $n \times (k+1)$

$\boldsymbol{\beta}$: vektor koefisien parameter regresi yang berukuran $(k+1) \times 1$

ρ : koefisien autoregresi lag spasial

λ : koefisien autoregresi lag pada *error* yang bernilai $|\lambda| < 1$

$\boldsymbol{u}$: vektor *error* yang diasumsikan ada autokorelasi berukuran $n \times 1$

$\boldsymbol{\varepsilon}$: vektor *error* yang berukuran n x 1, yang berdistribusi normal dengan mean 0 dan varians $\sigma^2 I$

$\boldsymbol{W}$:matriks pembobot spasial yang berukuran $n \times 1$

n : banyaknya amatan atau lokasi

k : jumlah variabel independen

I : matriks identitas dengan ukuran $n \times n$

Terdapat empat model yang dapat dibentuk dari model umum regresi yaitu:

1. *Geographically Weighted Regression* (GWR)

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

2. *Spatial Lag Model* (SLM) atau *Spatial Autogressive model* (SAR)
   jika $\rho \neq 0, \lambda = 0$ maka model umum regresi pertama menjadi:

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

3. *Spatial Error Model* (SEM)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \lambda\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

4. *General Spatial model* atau *Spatial Autoregressive Moving Avarage* (SARMA)

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \boldsymbol{u} = \lambda\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

2.6.1 *Spatial Autoregressive Model (SAR)*

Menurut Anselin (1988), *Mixed-Autoregressive* atau *Spatial Lag Model (SLM)* atau disebut juga *Spatial Autoregressive Model (SAR)* merupakan sebuah model yang mengkombinasikan model regresi sederhana dengan lag spasial pada variabel dependen dengan menggunakan data *cross section*. *Spatial Autoregressive Model (SAR)* terbentuk jika pada persamaan (2.12) nilai $\rho \neq 0$ dan $\lambda = 0$, secara matematis bentuk umum dari *Spatial Autoregressive Model (SAR)* dapat dituliskan sebagai berikut:

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{14}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I) \tag{15}$$

dengan:

$\boldsymbol{y}$: vektor variabel dependen berukuran $n \times 1$

ρ : koefisien parameter spasial lag dari variabel dependen

$\boldsymbol{W}$: matriks pembobot spasial berukuran $n \times 1$

$\boldsymbol{\varepsilon}$: vektor *error* berukuran $n \times 1$

$\boldsymbol{X}$: matriks variabel independen berukuran $n \times 1$

$\boldsymbol{\beta}$: vektor koefisien parameter variabel independen

Model persamaan diatas mengasumsikan bahwa pada varibael depenten terjadi proses *autoregressive* dan model *Spatial Autoregressive Model* merupakan model yang tepat untuk digunakan pada pola spasial dengan pendekatan area. Bentuk pendugaan dari metode SAR yaitu:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(1 - \rho\boldsymbol{W})\boldsymbol{y} \tag{16}$$

dan penduga untuk ρ yaitu

$$\hat{\rho} = (\boldsymbol{y}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{y})^{-1}\boldsymbol{y}^T\boldsymbol{W}^T\boldsymbol{y} \tag{17}$$

Persamaan (16) diperoleh dari meminimumkan jumlah kuadrat *error* yaitu:

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \tag{18}$$

menjadi

$$\sum \varepsilon_i^2 = \sum (Y_i - \rho W y - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_k X_k)^2$$

dalam notasi matriks dituliskan menjadi

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]\begin{bmatrix}\varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n\end{bmatrix} = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

dari persamaan di atas diperoleh

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = (\boldsymbol{I} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = \left[\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T} - \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}^{\boldsymbol{T}}\right]\left[\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) - \boldsymbol{X}\boldsymbol{\beta}\right]$$

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = \left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) - \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) - \left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta}$$

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon} = \left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) - 2\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) + \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta}$$

Dengan menggunakan sifat matriks dimana

$$\left[\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)\right] = [\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}]\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T} = \left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T}\boldsymbol{X}\boldsymbol{\beta}$$

Langkah selanjutnya menurunkan persamaan $\varepsilon^{T}\varepsilon$ terhadap β

$$\frac{\partial(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = \frac{\partial\left(\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)^{T}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) - 2\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) + \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}}$$

$$\frac{\partial(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = -2\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right)\boldsymbol{X}^{\boldsymbol{T}} + 2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta}$$

Setelah hasil dari penurunan diperoleh, langkah selanjutnya adalah hasil tersebut di sama dengankan dengan nol sehingga:

$$-2\boldsymbol{X}^{\boldsymbol{T}}\left((\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}\right) + 2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta} = 0$$

$$2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta} = 2\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}$$

$$\boldsymbol{\beta} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{I}-\rho\boldsymbol{W})\boldsymbol{y}$$

Pada persamaan (15) ε diasumsikan menyebar secara normal, bebas stokastik serta identik dengan nilai tengah 0 dan ragam σ^2, ε sendiri merupakan *error* pada lokasi. Fungsi kepekatan peluang dari ε adalah

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right]$$

Fungsi kepekatan peluang bersama dari n variabel acak $\varepsilon_1, \varepsilon_2, \cdots \varepsilon_n$.

$$\begin{aligned} f(\varepsilon) &= f(\varepsilon_1).f(\varepsilon_2).\cdots.f(\varepsilon_n) \\ &= \left[\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\varepsilon_1^2}{2\sigma^2}\right]\right).\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\varepsilon_2^2}{2\sigma^2}\right]\right).\cdots.\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\varepsilon_n^2}{2\sigma^2}\right]\right)\right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp\left[-\frac{\sum_{i=1}^{n}\varepsilon_i^2}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp\left[-\frac{\varepsilon^T\varepsilon}{2\sigma^2}\right] \end{aligned}$$

Fungsi kepekatan bersama variabel dependen y diperoleh dengan mentransformasi ruang ε berdimensi n ke sebuah ruang y berdimensi n. Dari persamaan (13) diperoleh persamaan:

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$$

Fungsi kepekatan peluang bersama dari n variabel independen y

$$\begin{aligned} f(y) &= f(\epsilon)[J] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp\left[-\frac{\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\varepsilon}}{2\sigma^2}\right]\left[\frac{d\varepsilon}{dy}\right] \end{aligned}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right][\boldsymbol{I} - \rho\boldsymbol{W}]$$

Fungsi kemungkinan variabel dependen y:

$$L(\beta, \rho, \sigma^2; y) = f(y; \beta, \rho, \sigma^2)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right]$$

Pendugaan parameter model diperoleh dengan memaksimalkan fungsi kemungkinan yabg ekuivalen dengan memaksimalkan logaritma dari fungsi kemungkinan pada persamaan:

$$J = L(\beta, \rho, \sigma^2; y)$$

$$= \ln\left(\frac{[\boldsymbol{I} - \rho\boldsymbol{W}]}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right]\right)$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln[\boldsymbol{I} - \rho\boldsymbol{W}] - \frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}$$

Selanjutnya untuk mendapatkan estimasi dari σ^2 persamaan di atas diturunkan terhadap σ^2 dan samakan dengan 0, sehingga diperoleh

$$\frac{\partial \ln(L)}{\partial \sigma^2} = 0$$

$$\frac{\partial\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln[\boldsymbol{I} - \rho\boldsymbol{W}] - \frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right)}{\partial \sigma^2} = 0$$

$$0 = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left((\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$$

$$0 = \frac{-n\sigma^2 + \left((\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)}{2(\sigma^2)^2}$$

$$0 = \frac{-n\sigma^2 + \left((\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)}{2\sigma^4}$$

$$\frac{n\sigma^2}{2\sigma^4} = \frac{(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^4}$$

$$\frac{2\sigma^4 . n\sigma^2}{2\sigma^4} = \frac{2\sigma^4 . (\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^4}$$

$$n\sigma^2 = (\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\sigma^2 = \frac{1}{n}(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

2.6.2 *Spatial Error Model*

Menurut Anselin (1988), *Spatial Error Model (SEM)* adalah model spasial *error* dimana pada *error* terdapat korelasi spasial serta memiliki kelebihan yaitu memberikan model yang lebih baik untuk pengamatan yang saling berhubungan.

Spatial Error Model (SEM) terbentik jika $\rho = 0$ dan $\lambda \neq 0$, sehingga model ini mengasumsikan bahwa proses autoregressive hanya pada *error* model. Secara matematis bentuk umum dari *Spatial Error Model (SEM)* adalah

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \tag{19}$$
$$\boldsymbol{u} = \lambda \boldsymbol{Wu} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

dengan:

$\boldsymbol{y}$: vektor variabel dependen berukuran $n \times 1$

$\boldsymbol{W}$: matriks pembobot spasial berukuran $n \times n$

$\boldsymbol{X}$: matriks variabel independen berukuran $n \times k$

$\boldsymbol{\beta}$: vektor koefisien parameter variabel independen

$\boldsymbol{\varepsilon}$: vektor *error* berukuran $n \times 1$

Penduga bagi parameter pada model regresi dengan metode *Spatial Error Model (SEM)* sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = \left[\left(\boldsymbol{X} - \hat{\lambda}\boldsymbol{WX}\right)^T\left(\boldsymbol{X} - \hat{\lambda}\boldsymbol{WX}\right)\right]^{-1}\left(\boldsymbol{X} - \hat{\lambda}\boldsymbol{WX}\right)^T\left(\boldsymbol{Y} - \hat{\lambda}\boldsymbol{WY}\right) \tag{20}$$

persamaan (20) diperoleh dengan meminimumkan jumlah kuadrat *error* yaitu

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta} - ((\boldsymbol{I} - \boldsymbol{\lambda W})^{-1})$$

dari persamaan diatas diperoleh

$$\begin{aligned}
\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} &= \left((\boldsymbol{Y} - \boldsymbol{X\beta})(\boldsymbol{I} - \lambda\boldsymbol{W})\right)^T((\boldsymbol{Y} - \boldsymbol{X\beta})(\boldsymbol{I} - \lambda\boldsymbol{W}))\\
&= [\boldsymbol{Y}(\boldsymbol{I} - \lambda\boldsymbol{W}) - \boldsymbol{X}(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{\beta}]^T[\boldsymbol{Y}(\boldsymbol{I} - \lambda\boldsymbol{W}) - \boldsymbol{X}(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{\beta}]\\
&= \left[((\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{Y})^T - \left((\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}\right)^T\right][(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{Y} - (\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}]\\
&\qquad [(\boldsymbol{Y} - \lambda\boldsymbol{WY})^T - (\boldsymbol{X} - \lambda\boldsymbol{WX})^T\boldsymbol{\beta}^T][(\boldsymbol{Y} - \lambda\boldsymbol{WY}) - (\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}]\\
&= (\boldsymbol{Y} - \lambda\boldsymbol{WY})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) - (\boldsymbol{Y} - \lambda\boldsymbol{WY})^T(\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta} - (\boldsymbol{X} - \lambda\boldsymbol{WX})^T\boldsymbol{\beta}^T(\boldsymbol{Y} - \lambda\boldsymbol{WY})\\
&\qquad + (\boldsymbol{X} - \lambda\boldsymbol{WX})^T\boldsymbol{\beta}^T(\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}\\
&= (\boldsymbol{Y} - \lambda\boldsymbol{WY})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) - 2\boldsymbol{\beta}^T(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) + (\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}^T\boldsymbol{\beta}
\end{aligned}$$

dengan menggunakan sifat-sifat matriks dimana,

$$\begin{aligned}
[\boldsymbol{\beta}^T(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY})] &= [\boldsymbol{\beta}^T(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY})]^T\\
&= (\boldsymbol{Y} - \lambda\boldsymbol{WY})^T(\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}
\end{aligned}$$

langkah selanjutnya menurunkan persamaan $\varepsilon^T\varepsilon$ terhadap β dan disama dengankan dengan 0, sehingga

$$\frac{\partial(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = 0$$

$$\frac{\partial((\boldsymbol{Y} - \lambda\boldsymbol{WY})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) - 2\boldsymbol{\beta}^T(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) + (\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{X} - \lambda\boldsymbol{WX})\boldsymbol{\beta}^T\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 0$$

$$-2(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY}) + 2\widehat{\boldsymbol{\beta}}(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{X} - \lambda\boldsymbol{WX}) = 0$$

$$2\widehat{\boldsymbol{\beta}}(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{X} - \lambda\boldsymbol{WX}) = 2(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY})$$

$$\widehat{\boldsymbol{\beta}} = [(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{X} - \lambda\boldsymbol{WX})]^{-1}(\boldsymbol{X} - \lambda\boldsymbol{WX})^T(\boldsymbol{Y} - \lambda\boldsymbol{WY})$$

2.7 Matriks Pembobot Spasial

Matriks pembobot spasial merupakan matriks berukuran $n \times n$ yang menyatakan hubungan antara observasi depedensi spasial dengan notasi W (Lee & Wong, 2000). Secara umum bentuk matriks W sebagai berikut:

$$\boldsymbol{W} = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nn} \end{bmatrix}$$

dengan:

w_{ij} : mendefinisikan bobot antara daerah ke-i dan ke-j

$i = 1,2,\cdots,n$

$j = 1,2,\cdots,n$

Anselin (1988) menyebutkan bahwa terdapat tiga pendekatan untuk mendefenisikan matriks pembobot spasial, yaitu *contiguity*, *distance* dan general. Matriks yang berdasarkan persentuhan batas wilayah atau *contiguity* menyatakan bahwa interaksi spasial terjadi antar wilayah yang bertetangga. Matriks kontiguitas spasial **C** adalah matriks biner berukutan $n\,x\,n$ yang berelemen 0 dan 1. Bentuk umum dari matriks pembobot spasial berdasarkan kontiguitas, yaitu:

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

dengan

$$W_{ij} = \frac{C_{ij}}{\sum_{j=1}^{k} C_{ij}} \tag{21}$$

sehingga,

$$\boldsymbol{W} = \begin{bmatrix} \frac{C_{11}}{\sum_{j=1}^{k} C_{ij}} & \frac{C_{12}}{\sum_{j=1}^{k} C_{ij}} & \cdots & \frac{C_{1n}}{\sum_{j=1}^{k} C_{ij}} \\ \frac{C_{21}}{\sum_{j=1}^{k} C_{ij}} & \frac{C_{22}}{\sum_{j=1}^{k} C_{ij}} & \cdots & \frac{C_{2n}}{\sum_{j=1}^{k} C_{ij}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{C_{n1}}{\sum_{j=1}^{k} C_{ij}} & \frac{C_{n2}}{\sum_{j=1}^{k} C_{ij}} & \cdots & \frac{C_{nn}}{\sum_{j=1}^{k} C_{ij}} \end{bmatrix}$$

Menurut Lesage (1999), informasi lokasi dapat diketahui dari dua sumber yaitu hubungan ketetanggaan dan jarak. Matriks pembobot spasial digunakan untuk menentukan bobot antar lokasi yang diamati berdasarkan hubungan ketetanggaan antar lokasi. Jenis-jenis matriks yang digunakan untuk matriks pembobot spasial yaitu:

1. Persinggungan sisi *(rook contiguity)*, menentukan daerah pengamatan berdasarkan kepada sisi yang saling bersinggungan dan sudut tidak diperhitungkan.
2. Persinggungan sudut *(bishop contiguity)*, menentukan daerah pengamatan berdasarkan sudut yang saling bersinggungan dan sisi tidak diperhitungkan.
3. Persinggungan sisi-sudut *(queen contiguity)*, menentukan daerah pengamatan berdasarkan sisi yang bersinggungan dan sudut juga diperhitungkan.

2.8 Efek Dependensi Spasial

Dependensi spasial terjadi akibat adanya dependensi dalam data wilayah. Dependensi spasial muncul berdasarkan hukum Tobler yaitu segala sesuatu yang saling berhubungan dengan hal yang lain tetapi sesuatu yang lebih dekat mempunyai pengaruh yang besar serta penyelesaian yang dapat dilakukan jika terdapat efek dependensi spasial adalah dengan melakukan pendekatan area (Anselin, 1988).

Anselin (1988) menyatakan bahwa untuk mengetahui dependensi spasial di dalam *error* suatu model dapat dilakukan dengan melakukan uji, uji tersebut adalah dengan menggunakan uji Moran’s I dan uji Langrange Multiplier.

1. Uji Moran’s I

Morans’s I merupakan sebuah uji statistik untuk melihat nilai autokorelasi spasial, yang digunakan untuk mengidentifikasi suatu lokasi dari pengelompokan spasial atau autokorelasi spasial. Uji statistik Moran’s I dibatasi oleh 1 dan -1. Menurut Lee & Wong (2000), Moran’s I dapat diukur dengen menggunakan persamaan:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (\sum_{i=1}^{n} (x_i - \bar{x})^2)}$$

dengan:

n : banyaknya pengamatan

$\bar{x}$: nilai rata-rata dari $\{x_i\}$ dari n lokasi

x_j : nilai pada lokasi ke-j

x_i :nilai pada lokasi ke-i

w_{ij} : elemen matriks pembobot spasial

Jika $I > I_0$, maka nilai autokorelasi akan bernilai positif, hal ini berarti bahwa pola data membentuk kelompok, $I = I_0$ artinya tidak terdapat autokorelasi spasial dan $I < I_0$ artinya nilai autokorelasi akan bernilai negatif, hal ini berarti pola data menyebar (Lee & Wong, 2000).

Hipotesis yang akan digunakan adalah,

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

$H_1: I \neq 0$ (ada dependensi spasial)

Statistik uji yang digunakan adalah sebagai berikut:

$$z_{hitung} = \frac{I - I_0}{\sqrt{var(I)}} \sim N(0,1) \tag{22}$$

dengan,

$$var(I) = \frac{n^2\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2\right) - n\left(\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2\right) + 3\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2}{(n^2 - 1)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2} - [E(I)]^2$$

$$E(I) = I_0 = -\frac{1}{n-1}$$

Pengembilan keputusan : tolak H_0 jika $\left|Z_{hitung}\right| > Z_{\frac{a}{z}}$

2. Uji Langrange Multiplier (LM)

Uji langrange multiplier merupakan sebuah uji untuk mengetahui apakah sebuah model memiliki efek spasial atau tidak. Langrange multiplier yang mana pada tes ini, nilai sisa diperoleh dari kuadrat terkecil dan hitungan matriks bobot spasial yang digunakan adalah W (Anselin, 1988).

Uji langrange multiplier ini digunakan untuk memilih model regresi spasial yang paling sesuai. Uji ini terdiri atas LM_{lag} dan LM_{error}. Apabila LM_{lag} signifikan maka model yang sesuai adalah SAR, jika LM_{error} signifikan maka model yang sesuai adalah SEM.

Bentuk tes langrange multiplier, yaitu (Rati et al., 2013):

1. Pada *Spatial Autoregressive Model* (SAR)

   Hipotesis yang digunakan adalah:

   $H_0: \rho = 0$ (tidak ada efek spasial pada lag)

   $H_1: \rho \neq 0$ (ada efek spasial pada lag)

   Uji statistik yang digunakan:

$$LM_{lag} = \frac{(\boldsymbol{\varepsilon}^T \boldsymbol{W}\boldsymbol{y})^2}{\boldsymbol{s}^2((\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta})^T M(\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{N}\boldsymbol{s}^2)} \tag{23}$$

dengan,

$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$

$\boldsymbol{N} = tr[(\boldsymbol{W}^T + \boldsymbol{W})\boldsymbol{W}]$

$\boldsymbol{s}^2 = \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}}{n}$

$\boldsymbol{I}$: matriks identitas

tr : *trace*

$\boldsymbol{\varepsilon}$: vektor *error*

$\boldsymbol{W}$: matriks pembobot spasial

$\boldsymbol{y}$: vektor variabel dependen

$\boldsymbol{X}$: matriks variabel independen

$\boldsymbol{\beta}$: vektor koefisien parameter variabel independen

Pengambilan keputusan: jika $LM_{lag} > \chi^2_{(1,\propto)}$ maka tolak H_0

2. Pada *Spatial Error Model* (SEM)

Hipotesis yang digunakan adalah:

$H_0: \lambda = 0$ (tidak ada efek spasial pada *error*)

$H_1: \lambda \neq 0$ (ada efek spasial pada *error*)

Uji statistik yang digunakan:

$$LM_{error} = \frac{\left(\frac{\varepsilon^t W \varepsilon}{s^2}\right)^2}{N} \quad (24)$$

Pengambilan keputusan: jika $LM_{error} > \chi^2_{(1,\propto)}$ maka tolak H_0

2.9 Uji Signifikansi Parameter Regresi Spasial

Pengujian isignifikansi parameter pemodelan spasial pada penelitian ini akan menggunakan uji Wald seperti yang ditunjukkan oleh (Anselin, 1988). Menurut Ramadani et al. (2013), untuk menguji parameter regresi spasial digunakan hipotesis sebagai berikut:

$H_0: \rho = 0$ (tidak terdadat korelasi spasial)

$H_1: \rho \neq 0$ (terdapat korelasi spasial)

Statistik uji : $Wald = \frac{\hat{\theta}^2}{var(\hat{\theta})}$ (25)

dengan:

θ : parameter taksiran ρ dan β

Pengambilan keputusan : jika $\chi^2_{a,1}$ maka tolak H_o

2.10 Kriteria Pemilihan Model

Ada beberapa kriteria untuk memilih model dalam penelitian. Kriteria pemilihan model yang digunakan dalam penelitian ini adalah koefisien determinasi (R^2) dan *Akaike information Criterion (AIC)* (Lesage, 1999).

1. Koefisien determinasi (R^2), dinotasikan dengan:

$$R^2 = \frac{SSR}{SST} \tag{26}$$

   dengan:

   SSR : *Sum Square Regression* (jumlah kuadrat regresi)

   SST : *Sum Square Total* (jumlah kuadrat total)

   Nilai R^2 yang semakin besar menunjukkan kepercayaan terhadap model semakin besar pula.

2. *Akaike Information Criterion (AIC)*, dinotasikan dengan:

$$AIC = -2Lm + 2m \tag{27}$$

   dengan:

   Lm = maksimum log – likelihood

   m = jumlah parameter dalam model

   Penentuan model terbaik dilihat dari nilai AIC yang dilakukan dengan melihat nilai AIC yang paling kecil dari model yang terbentuk.