

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Regresi Parametrik

Regresi parametrik merupakan salah satu model regresi yang digunakan dalam menganalisis hubungan variabel respon (Y) dengan variabel prediktor (X). Dalam regresi terdapat kurva yang terbentuk berdasarkan data yang digunakan. Regresi parametrik memiliki bentuk kurva yang berpola, misal kurva membentuk pola linier, kuadratik, atau eksponensial, dan lain-lain. Dalam regresi parametrik dapat digunakan untuk mengetahui hubungan yang linier terhadap model yang terbentuk. Berikut ini merupakan model regresi berganda jika diberikan  $n$  observasi dengan  $p$  variabel prediktor (Hardle, 1994) :

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i ; i = 1, 2, \dots, n ; k = 1, 2, \dots, p \quad (1)$$

Keterangan :

- $y_i$  = Variabel respon pengamatan ke-i
- $x_{ik}$  = Variabel prediktor ke-k pengamatan ke-i
- $\beta_k$  = Koefisien regresi pada  $x_k$
- $\beta_0$  = Intercept
- $\varepsilon_i$  = Residual ke-i

Apabila ditulis dalam bentuk matriks yaitu sebagai berikut :

$$Y = X\beta + \varepsilon \quad (2)$$

## 2.2. Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu model regresi selain regresi parametrik. Regresi nonparametrik sangat tepat digunakan untuk mendeteksi adanya sifat nonlinieritas dan hubungan nonlinier dalam model (Eubank, 1999). Nonlinier tersebut dapat terlihat pada kurva yang terbentuk. Kurva regresi nonparametrik yang terbentuk tidak mengandung pola tertentu atau bisa dikatakan bahwa bentuk kurvanya tidak diketahui. Berikut merupakan model umum dari regresi nonparametrik yaitu sebagai berikut :

$$y_i = m(t_i) + \varepsilon_i \quad (3)$$

Keterangan :

$y_i$  = Variabel respon pengamatan ke-i

$t_i$  = Variabel prediktor pengamatan komponen nonparametrik ke-i

$m$  = Fungsi regresi yang tidak diketahui

$\varepsilon_i$  = Residual ke-i

## 2.3. Regresi Semiparametrik

Regresi semiparametrik merupakan gabungan dari regresi parametrik dan nonparametrik (Ruppert, 2003). Dalam regresi semiparametrik mengandung komponen regresi parametrik dan nonparametrik. Komponen dari regresi parametrik yaitu diketahui bahwa terdapat variabel yang memiliki kurva berpola. Sedangkan, komponen regresi nonparametrik yaitu terdapat variabel yang tidak diketahui bentuk kurvanya. Berikut merupakan model dari regresi semiparametrik sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_1 + f(t_i) + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (4)$$

$$Y = X\beta + f + \varepsilon \quad (5)$$

Keterangan :

$y_i$  = Variabel respon dari data ke-i

$\beta_0, \beta_1$  = Parameter

$x_i$  = Variabel prediktor dari data ke-i untuk komponen parametrik

$f(t_i)$  = Fungsi regresi yang tidak diketahui bentuknya untuk komponen nonparametrik

$t_i$  = Variabel prediktor dari data ke-I untuk komponen nonparametric

$\varepsilon_i$  = error ke-i yang diasumsikan menyebar  $\sim N(0, \sigma^2)$

Dari model diatas terdapat komponen parametrik yaitu  $\beta_0 + \beta_1 x_1$  dan komponen nonparametriknya yaitu  $f(t_i)$  (Ruppert, 2003). Apabila ditulis dalam bentuk matriks yaitu sebagai berikut :

$$Y = X\beta + TM + \varepsilon \quad (6)$$

#### 2.4. Regresi Kernel

Regresi kernel merupakan teknik statistika nonparametrik yang digunakan untuk mengestimasi fungsi  $m(x)$  yang terdapat dalam model regresi nonparametrik. Nadaraya dan Watson pada tahun 1994 mendefinisikan estimator regresi kernel, sehingga disebut estimator Nadaraya-Watson (Hardle, 1994). Berikut merupakan estimator Nadaraya-Watson :

$$m(x) = \frac{\sum_{i=1}^n Kh\left(\frac{x-x_i}{h}\right)y_i}{\sum_{i=1}^n Kh\left(\frac{x-x_i}{h}\right)} \quad (7)$$

Ada tujuh teknik *smoothing* yang dapat digunakan, yaitu histogram, kernel, deret orthogonal, spline, k-NN, deret fourier, dan wavelet (Eubank, 1998). Estimator kernel adalah salah satu pendekatan nonparametrik dalam mengestimasi kurva yang *smooth*. Estimator kernel merupakan pengembangan dari estimator histogram yang diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962), sehingga disebut densitas kernel Rosenblatt-Parzen (Hardle, 1994).

Secara umum kernel  $K$  dengan *bandwidth* ( $h$ ) didefinisikan (Wand dan Jones, 1995) :

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \text{ untuk } -\infty < x < \infty, h > 0 \quad (8)$$

Serta memenuhi :

(i)  $K(x) \geq 0$ , untuk semua  $x$

(ii)  $\int_{-\infty}^{\infty} K(x) dx = 1$

(iii)  $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$

(iv)  $\int_{-\infty}^{\infty} x K(x) dx = 0$

Maka, estimator densitas kernel untuk fungsi densitas  $f(x)$  adalah :

$$\hat{f}_h = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n \left( \frac{x - x_i}{h} \right) \quad (9)$$

$\hat{f}_h$  tergantung pada fungsi kernel  $K$  dan parameter  $h$ . Bentuk bobot kernel ditentukan oleh fungsi kernel  $K$ , sedangkan ukuran bobotnya ditentukan oleh

parameter pemulus  $h$  yang disebut *bandwidth*. Peran bandwidth seperti lebar interval pada histogram. Ada tujuh fungsi kernel (Hardle, 1994) :

1. Kernel Uniform :  $K(x) = \frac{1}{2}I(|x| \leq 1)$
2. Kernel Segitiga :  $K(x) = (1 - |x|)I(|x| \leq 1)$
3. Kernel Epanechnikov :  $K(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$
4. Kernel Kuartik :  $K(x) = \frac{15}{16}(1 - x^2)^2I(|x| \leq 1)$
5. Kernel Triweight :  $K(x) = \frac{35}{32}(1 - x^2)^3I(|x| \leq 1)$
6. Kernel Cosinus :  $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right)I(|x| \leq 1)$
7. Kernel Gaussian :  $K(x) = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{2}(-x^2)\right), -\infty < x < \infty$

dengan  $I$  adalah indikator.

## 2.5. Estimasi Regresi Semiparametrik Kernel

Pada persamaan model semiparametrik diatas, dapat dilihat bahwa  $m$  sebagai fungsi dan  $\beta$  sebagai parameter. Fungsi dan parameter tersebut akan diestimasi dari data yang digunakan dalam penelitian. Fungsi  $m$  akan diestimasi menggunakan kernel, sedangkan parameter  $\beta$  akan diestimasi dengan *Weighted Least Square* (WLS). Metode *Weighted Least Square* (WLS) memiliki nama lain yaitu metode kuadrat terkecil tertimbang. Metode *Weighted Least Square* (WLS) merupakan salah satu metode estimasi parameter yang digunakan untuk menentukan nilai taksiran parameter yang

meminimumkan jumlah kuadrat residu dengan dilakukan pembobotan pada suatu faktor yang tepat. Nilai MSE pada metode *Weighted Least Square* (WLS) diperoleh paling terkecil jika dibandingkan dengan metode *least square* yang lainnya. Berikut merupakan perumusan dari metode *Weighted Least Square* (WLS) :

$$WLS = \sum_{i=1}^n (y_i^* - \sum_{j=1}^2 m_j(t_j))^2 K_h(t_i - t) \quad (10)$$

Misalkan, terdapat  $n$  observasi,  $p$  variabel prediktor parametrik, dan  $t$  variabel prediktor nonparametrik sehingga persamaannya dapat ditulis :

$$y_i - \sum_{k=1}^p x_k \beta_{ik} = m_j(t_j) + \varepsilon_i \quad (11)$$

Jika  $y_i^* = y_i - \sum_{k=1}^p \beta_k x_{ik}$ , dengan  $m(t)$  ekuivalen dengan nilai harapan dari variabel respon apabila nilai  $T=t$  diketahui. Dengan asumsi variabel respon dan variabel prediktor merupakan variabel random.

## 2.5. Pemilihan *Bandwidth* (h)

*Bandwidth* (h) adalah parameter pemulus (*smoothing*) yang memiliki fungsi untuk mengontrol kurva yang sedang diestimasi. *Bandwidth* (h) yang terlalu kecil akan menyebabkan fungsi yang sedang diestimasi menjadi sangat kasar dan sangat fluktuatif (*under-smoothing*). Sebaliknya, jika nilai *bandwidth* (h) terlalu lebar akan menyebabkan kurva sangat halus (*over-smoothing*). Oleh karena itu, diperlukan *bandwidth* (h) optimal dengan

metode *Generalized Cross Validation* (GCV). *Generalized Cross Validation* (GCV) merupakan salah satu metode dalam analisis regresi untuk menentukan *bandwidth* yang optimal (Wu, dan Zang. 2006). Jika  $Y$  diestimasi maka :

$$\hat{Y} = H(h) Y \quad (12)$$

Dimana  $H(h) = W(t) + P_{\hat{X}}(I - W(t))$  dengan  $P_{\hat{X}} = \hat{X}(\hat{X}^T \hat{X})^{-1} \hat{X}^T$

$$GCV(h) = \frac{n^{-1} \sum_{i=1}^n [y_i - \hat{y}_i]^2}{\{1 - \text{tr}(H(h)) / n\}^2}, \quad i=1, 2, \dots, n \quad (13)$$

Keterangan :

GCV(h) : nilai GCV pada bandwidth h

n : banyak subjek

$y_i$  : data aktual subjek ke i

$\hat{y}_i$  : hasil estimasi subjek ke i

Tr(H) : jumlah elemen dari diagonal utama matriks penghalus ukuran nxn

Nilai *bandwidth* (h) yang optimal adalah nilai GCV (h) yang minimum.

Untuk melihat tingkat ukuran kesalahan estimator dengan metode *Mean Square Error* (MSE) (Aydin, 2007) :

$$MSE(h) = N^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

Semakin kecil nilai *Mean Square Error* (MSE) maka hasil estimasi akan semakin mendekati fungsi aslinya.

## 2.5. Produksi Padi

Tanaman pangan yang utama di Indonesia yaitu tanaman padi. Orang Indonesia lebih memilih mengonsumsi hasil olahan padi sebagai makanan pokok, tetapi juga mengonsumsi bahan makanan pokok lainnya sebagai makanan sampingan. Indonesia sebagai negara agraris seharusnya memiliki hasil produksi padi yang melimpah, sehingga tidak perlu mengadakan impor beras dari negara tetangga.

Padi memiliki nama latin yaitu *Oryza Sativa*. Padi merupakan tanaman pokok yang ada di Indonesia, sebagaimana negara agraris. Di Indonesia beras merupakan mata dagangan yang sangat penting sebab beras merupakan bahan makanan pokok dan merupakan sumber kalori bagi sebagian besar penduduk dan situasi beras secara tidak langsung dapat mempengaruhi bahan konsumsi yang lain (Djiwandi, 1980).

BPS dan Deptan (1999) mendefinisikan luas panen merupakan luas lahan sawah yang biasa diambil hasilnya. Luas tanam merupakan luas lahan sawah yang ditanami. Sedangkan produktivitas merupakan hasil yang diperoleh tiap satuan luas, dan produksi merupakan suatu besaran berat yang mengukur hasil total padi yang diperoleh, juga merupakan hasil kali antara produktivitas dan luas panen.

Menurut Darwanto (2010) bahwa variabel luas lahan memiliki koefisien positif. Tanda positif tersebut mengandung arti bahwa semakin luas lahan yang digunakan untuk usaha tani padi, maka semakin meningkat pula produksi padinya.



Faktor-faktor yang mempengaruhi produksi padi diantaranya yaitu :

1. Luas panen padi

Definisi luas panen menurut Badan Pusat Statistika (BPS) yaitu luasan tanaman yang dipungut hasilnya setelah cukup umur. Satuan yang digunakan yaitu Hektar (Ha). Dalam panen berhasil ini termasuk juga tanaman yang hasilnya sebagian saja dapat dipungut (paling sedikit sampai dengan 11%) yang mungkin disebabkan karena mendapat serangan organisme pengganggu tumbuhan atau bencana alam.

2. Produktivitas

Dalam ilmu ekonomi pertanian produktivitas merupakan perbandingan antara hasil yang diharapkan akan diterima pada waktu panen (penerimaan) dengan biaya (pengorbanan) yang harus dikeluarkan. Pengertian produktivitas ini merupakan penggabungan antara konsepsi efisiensi usaha (fisik) dengan kapasitas tanah (Mubyarto, 1989). Satuan yang digunakan yaitu kwintal/hektar (Kwa/Ha).

3. Luas puso/kerusakan

Luas rusak adalah jika tanaman mengalami serangan organisme pengganggu tumbuhan, bencana alam, sedemikian rupa sehingga hasilnya kurang dari 11% keadaan normal. Satuan yang digunakan yaitu hektar (Ha).