BAB II

TINJAUAN PUSTAKA

Bab ini akan dibahas konsep-konsep yang mendasari penelitian ini antara lain: analisis regresi parametrik, analisis regresi nonparametrik, *spline*, pemilihan titik *knot* optimal, kriteria pemilihan model.

2.1 Analisis Regresi Parametrik

Analisis regresi merupakan analisis statistika yang digunakan untuk melihat hubungan antara variabel respons dengan satu atau lebih variabel bebas (prediktor). Dalam analisis regresi variabel terikat (respons) biasanya dinotasikan dengan Y dan variabel bebas (prediktor) dinotasikan dengan X. Misalkan terdapat data berpasangan (X_i,Y_i) yang saling berkorelasi untuk n pengamatan maka hubungan antara variabel X_i dan variabel Y_i dapat dituliskan sebagai berikut (Kutner et al., 2004):

$$Y_i = f(X_i) + \varepsilon_i \quad ; \ i = 1,2,3, \dots, n \tag{2.1}$$

dengan Y_i menyatakan variabel respons pada amatan ke- i, X_i menyatakan variabel prediktor pada amatan ke-i, $f(X_i)$ menyatakan fungsi regresi yang sudah diketahui bentuknya, ε_i suku galat yang diasumsikan bebas dengan nilai tengah nol dan variansnya σ^2 pada amatan ke-i, dan n menyatakan banyaknya amatan.

Regresi parametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respons dengan variabel prediktor apabila bentuk kurva regresinya diketahui. Secara umum model regresi dengan lebih dari satu variabel prediktor $(X_1, X_2, \dots, X_j)$ dapat dituliskan sebagai berikut (Neter et al., 1990):

$$Y_i = \theta_0 + \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \theta_m X_{ij} + \varepsilon_i \; ; i = 1,2,3, \dots, n \qquad (2.2)$$

dengan $\theta_0, \theta_1, \theta_2, \dots, \theta_m$ menyatakan parameter koefisien regresi, $X_{i1}, X_{i2}, \dots, X_{ij}$ menyatakan variabel prediktor yang nilainya diketahui, ε_i suku galat yang diasumsikan saling bebas dengan nilai tengah nol dan variansnya σ^2.

Persamaan (2.2) dapat pula disajikan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1j} \\ 1 & X_{21} & X_{22} & \dots & X_{2j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nj} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \text{atau}$$

$$Y = X\theta + \varepsilon \qquad (2.3)$$

dengan:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1j} \\ 1 & X_{21} & X_{22} & \dots & X_{2j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nj} \end{bmatrix}, \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (2.4)$$

dengan Y menyatakan vektor kolom untuk variabel respons berukuran $n \times 1$, X menyatakan matriks konstanta berukuran $n \times (j+1)$, θ menyatakan vektor kolom untuk parameter berukuran $(j+1) \times 1$, dan ε menyatakan vektor kolom untuk variabel acak normal bebas dengan nilai harapan $E\{\varepsilon\} = 0$ dan matriks ragam $\sigma^2\{\varepsilon\} = \sigma^2$ berukuran $(n+1) \times 1$.

2.2 Analisis Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respons dengan variabel prediktor di mana tidak terdapat informasi sebelumnya mengenai pola data atau kurva regresinya. Secara umum model regresi nonparametrik adalah sebagai berikut (Eubank, 1999):

$$Y_i = f(t_i) + \varepsilon_i; \quad i = 1,2,\dots,n \tag{2.5}$$

dengan Y_i menyatakan variabel respons ke-i, t_i menyatakan variabel prediktor ke-i, $f(t_i)$ menyatakan fungsi regresi yang tidak diketahui bentuk kurvanya atau polanya tidak diketahui, dan ε_i menyatakan suku galat yang diasumsikan bebas dengan nilai tengah nol dan varians σ^2.

Dalam regresi nonparametrik data diharapkan mencari sendiri bentuk kurva regresinya. Sehingga bentuk kurva regresi nonparametrik tidak dipengaruhi oleh subjektivitas dari peneliti. Fungsi regresi nonparametrik f diasumsikan mulus (*smooth*) sehingga dalam mengestimasi fungsi regresi memiliki fleksibilitas yang tinggi. Terdapat beberapa teknik estimasi dalam regresi nonparametrik antara lain deret Fourier, *kernel*, *wavalets*, dan *spline*.

2.3 Analisis Regresi Nonparametrik Spline

Spline merupakan model polinomial yang tersegmen atau terbagi-bagi yang dapat menghasilkan fungsi regresi sesuai dengan pola data. *Spline* juga mempunyai kemampuan yang baik dalam menangani data pada sub-sub interval tertentu. Penduga *spline* tergantung pada titik *knot*. Titik *knot* merupakan titik fokus dalam fungsi *spline* atau suatu titik perpaduan yang disebabkan oleh perubahan pola perilaku dari suatu fungsi dalam selang yang berbeda. Pada persamaan (2.5) fungsi $f(t_i)$ merupakan fungsi yang tidak diketahui bentuk dan sifat pemulusnya. Oleh karena itu, untuk mengestimasi fungsi $f(t_i)$ dapat digunakan model regresi *spline*. Secara umum fungsi *spline* f berorde *h* dengan titik *knot* $(k_1, k_2, \dots, k_r)$ dapat dinyatakan sebagai berikut (Eubank, 1999):

$$f(t_i) = \beta_0 + \beta_1 t_i^1 + \beta_2 t_i^2 + \cdots + \beta_{h-1} t_i^{h-1} + \sum_{l=1}^{r} \beta_{(h-1+l)} (t_i - k_l)_+^{h-1}$$

$$= \sum_{j=0}^{h-1} \beta_j t_i^j + \sum_{l=1}^{r} \beta_{(h-1+l)} (t_i - k_l)_+^{h-1} \quad (2.6)$$

dengan $\beta_0, \beta_1, \beta_2, \dots, \beta_{h-1}, \beta_{(h-1+l)}$ menyatakan koefisien regresi, $t_i^1, t_i^2, \dots, t_i^{h-1}$ menyatakan variabel prediktor yang nilainya diketahui, $(t_i - k_l)_+^{h-1}$ merupakan fungsi potongan (*truncated*) yang dijabarkan sebagai berikut:

$$(t_i - k_l)_+^{h-1} = \begin{cases} (t_i - k_l)^{h-1}, & t_i \geq k_l \\ 0, & t_i < k_l \end{cases}. \quad (2.7)$$

Apabila persamaan (2.6) disubtitusikan kepersamaan (2.5) maka akan diperoleh regresi nonparametrik *spline* sebagai berikut :

$$Y_i = \sum_{j=0}^{h-1} \beta_j t_i^j + \sum_{l=1}^{r} \beta_{(h-1+l)} (t_i - k_l)_+^{h-1} + \varepsilon_i; \quad i = 1,2, \dots, n \quad (2.8)$$

Misalkan terdapat n pengamatan, model regresi nonparametrik *spline* pada persamaan (2.8) dapat pula disajikan dalam bentuk matriks yaitu sebagai berikut:

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & t_1^1 & t_1^2 & \dots & t_1^{h-1} & (t_1-k_1)_+^{h-1} & \dots & (t_1-k_r)_+^{h-1} \\ 1 & t_2^1 & t_2^2 & \dots & t_2^{h-1} & (t_2-k_1)_+^{h-1} & \dots & (t_2-k_r)_+^{h-1} \\ 1 & t_3^1 & t_3^2 & \cdots & t_3^{h-1} & (t_3-k_1)_+^{h-1} & \cdots & (t_3-k_r)_+^{h-1} \\ & & & & & \vdots & & \\ 1 & t_n^1 & t_n^2 & \cdots & t_n^{h-1} & (t_n-k_1)_+^{h-1} & \cdots & (t_n-k_r)_+^{h-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_h \\ \beta_{(h-1+1)} \\ \vdots \\ \beta_{(h-1+r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

atau

$$
Y = t\beta + \varepsilon \tag{2.9}
$$

dengan

$$
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix},
$$

$$
t = \begin{bmatrix} 1 & t_1^1 & t_1^2 & \dots & t_1^{h-1} & (t_1-k_1)_+^{h-1} & \dots & (t_1-k_r)_+^{h-1} \\ 1 & t_2^1 & t_2^2 & \dots & t_2^{h-1} & (t_2-k_1)_+^{h-1} & \dots & (t_2-k_r)_+^{h-1} \\ 1 & t_3^1 & t_3^2 & \cdots & t_3^{h-1} & (t_3-k_1)_+^{h-1} & \cdots & (t_3-k_r)_+^{h-1} \\ & & & & & \vdots & & \\ 1 & t_n^1 & t_n^2 & \cdots & t_n^{h-1} & (t_n-k_1)_+^{h-1} & \cdots & (t_n-k_r)_+^{h-1} \end{bmatrix}, \tag{2.10}
$$

dan

$$
\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_h \\ \beta_{(h-1+1)} \\ \vdots \\ \beta_{(h-1+r)} \end{bmatrix}.
$$

Estimasi parameter β pada regresi nonparametrik *spline* dapat dilakukan dengan metode *maximum likelihood estimator* (MLE). Jika galat pada persamaan (2.5) diasumsikan berdistribusi normal, maka Y_i juga berdistribusi normal dengan nilai tengah $f(t_i)$ dan variansnya σ^2. Sehingga fungsi densitas peluang Y_i menjadi:

$$f(Y\,;f(t),\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(Y-f(t)\right)^2}{2\sigma^2}\right], f(t) > 0, \sigma^2 > 0 \qquad (2.11)$$

Sehingga, fungsi *likelihood* dapat dituliskan sebagai berikut:

$$L(Y,f) = \prod_{i=1}^{n} f(Y_i; f(t_i), \sigma^2)$$

$$= f(Y\,; f(t_1), \sigma^2) \ldots f(Y\,; f(t_n), \sigma^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(Y_1 - f(t_1)\right)^2}{2\sigma^2}\right] \cdots \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(Y_n - f(t_n)\right)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\left(\sqrt{2\sigma^2}\right)^n} exp\left\{-\left[\frac{\left(Y_1-f(t_1)\right)^2}{2\sigma^2} + \cdots + \frac{\left(Y_n-f(t_n)\right)^2}{2\sigma^2}\right]\right\}$$

$$= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} exp\left[-\sum_{i=1}^{n} \frac{\left(Y_i-f(t_i)\right)^2}{2\sigma^2}\right]$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - f(t_i)\right)^2\right]. \qquad (2.12)$$

Selanjutnya maksimumkan fungsi *likelihood* $L(Y,f)$ agar mendapatkan estimasi titik f yang dapat diuraikan sebagai berikut:

$$\max_f\{L(Y,f)\} = \max_{\beta \epsilon R^{h+r}} \left\{(2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \sum_{j=0}^{h-1} \beta_j t_i^j + \sum_{l=1}^{r} \beta_{h-1+l}\,(t_i - k_l)_+^{h-1}\right)^2\right)\right\}. \qquad (2.13)$$

Lanjutkan dengan menerapkan transformasi logaritma pada persamaan (2.13). sehingga diperoleh:

$$\log L(Y,\beta) = \log\left\{(2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \sum_{j=0}^{h-1}\beta_j t_i^j \right.\right.\right.$$

$$\left.\left.\left. + \sum_{l=1}^{r}\beta_{h-1+l}\,(t_i - k_l)_+^{h-1}\right)^2\right]\right\}$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{h-1}\beta_j t_i^j + \sum_{l=1}^{r}\beta_{h+l}\,(t_i - k_l)_+^{h-1}\right)^2$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - t\beta)'(Y - t\beta). \tag{2.14}$$

Kemudian persamaan (2.14) diturunkan secara parsial terhadap β dan disamakan dengan nol di sisi kanannya maka akan didapatkan hasil penjabaran sebagai berikut:

$$\frac{\partial(\log(Y,\beta))}{\partial\beta} = \frac{\partial\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - t\beta)'(Y - t\beta)\right)}{\partial\beta}$$

$$= \frac{\partial}{\partial\beta}\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - t\beta)'(Y - t\beta)\right)$$

$$= -\frac{1}{2\sigma^2}\left(\frac{\partial}{\partial\beta}(Y'Y - Y't\beta - \beta't'Y + \beta't't\beta)\right).$$

Karena $Y't\beta$ menghasilkan matriks berukuran 1×1, $Y't\beta$ dapat diubah menjadi $(Y't\beta)' = \beta't'Y$ sehingga:

$$\frac{\partial(\log(Y,\beta))}{\partial\beta} = -\frac{1}{2\sigma^2}\left(\frac{\partial}{\partial\beta}(Y'Y - 2\beta't'Y + \beta't't\beta)\right)$$

$$= -\frac{1}{2\sigma^2}(-2t'Y + 2t't\beta)$$

$$= -2t'Y + 2t't\beta = 0$$

dengan demikian diperoleh:

$$2t'Y = 2t'tβ.$$

Sehingga

$$\hat{\beta} = (t't)^{-1}t'Y \quad (2.15)$$

Estimasi dari $\hat{Y}$ dapat dinyatakan sebagai berikut:

$$\hat{Y} = t\hat{\beta} + \varepsilon$$

$$\hat{Y} = t(t't)^{-1}t'Y + \varepsilon$$

$$= A(k)Y + \varepsilon \quad (2.16)$$

dengan $A(k)$ menyatakan matriks yang digunakan untuk menghitung nilai *generalized cross validation* (GCV) untuk memilih titik *knot* optimal.

2.4 Pemilihan Titik *Knot* Optimal

Pemilihan titik *knot* yang optimal akan memberikan penduga *spline* yang baik. Titik *knot* merupakan titik fokus dalam fungsi *spline* sehingga kurva yang terbentuk tersegmen pada titik tersebut atau titik perpaduan bersama yang terjadi karena terdapat perubahan pola perilaku fungsi atau kurva pada suatu interval. Untuk mendapatkan titik *knot* yang optimal dapat menggunakan metode *generalized cross validation* (GCV) (Wahba, 1990). Pemilihan titik *knot* optimal menggunakan fungsi GCV dapat didefinisikan sebagai berikut:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}tr[I-A(k)])^2} \quad (2.17)$$

dengan $MSE(k) = n^{-1}\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2$, $\hat{Y} = t(t't)^{-1}t'Y, I$ menyatakan matriks identitas, *n* menyatakan banyaknya data, *k* menyatakan titik *knot* $(k_1, k_2, \dots, k_r)$, dan $A(k) = t(t't)^{-1}t'$.

2.5 Kriteria Pemilihan Model Terbaik

Salah satu tujuan dalam analisis regresi adalah untuk memperoleh hubungan yang signifikan antara variabel prediktor dengan variabel respons. Ada beberapa kriteria dalam pemilihan model regresi terbaik antara lain: koefisien determinasi (R^2), *cross validation* (CV), *mean square error* (MSE), dan *generalized coss validation* (GCV). Dalam penelitian ini, kriteria pemilihan model terbaik menggunakan nilai koefisien determinasi (R^2) yang tertinggi dan nilai GCV yang terkecil.

2.6 Angka Kematian Bayi

Angka kematian bayi adalah banyaknya kematian bayi usia di bawah satu tahun, per 1.000 kelahiran hidup pada satu tahun tertentu (BPS, 2015). Angka kematian bayi sangat berguna dalam menggambarkan keadaan kesehatan dan keadaan ekonomi masyarakat. Banyak faktor yang memengaruhi kematian bayi antara lain berat badan bayi lahir rendah, persalinan dengan bantuan non medis, ibu hamil mendapatkan tablet Fe3, rumah tangga berprilaku hidup bersih dan sehat (PHBS), dan ibu hamil melaksanakan program pelayanan *antenatal* sesuai dengan standar paling sedikit empat kali (K4). Kematian bayi merupakan kematian yang terjadi dari bayi baru lahir sampai dengan bayi berumur satu tahun. Secara umum dari segi penyebabnya, kematian bayi dibedakan menjadi dua macam yaitu kematian bayi endogen atau *neonatal* dan kematian bayi eksogen atau post neo-natal.

Kematian bayi endogen atau *neonatal* merupakan kematian bayi yang disebabkan oleh faktor-faktor bawaan anak yang didapat oleh orang tuanya sejak lahir yang diperoleh selama bulan kehamilan. Kematian bayi endogen atau neonatal terjadi pada bulan pertama setelah bayi dilahirkan. Kematian bayi eksogen atau post neo-natal merupakan kematian bayi yang disebabkan oleh faktor-faktor lingkungan luar seperti keadaan ekonomi, penolong pertama saat kelahiran, pelayanan kesehatan dan sebagainya. Kematian bayi eksogen atau post neo-natal terjadi setelah usia bayi satu bulan sampai bayi berusia satu tahun.