

**ESTIMASI MODEL REGRESI SEMIPARAMETRIK
MENGUNAKAN PENDUGA NADARAYA-WATSON
DENGAN KERNEL EPANECHNIKOV**

(Skripsi)

Oleh

Resti Novalia



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019**

ABSTRACT

SEMIPARAMETRIC REGRESSION MODEL ESTIMATION USING NADARAYA-WATSON ESTIMATOR WITH EPANECHNIKOV KERNEL

By

RESTI NOVALIA

Semiparametric regression model is an estimation that combines both parametric and nonparametric regression model. Independent variables that satisfy parametric assumptions can be predicted by least square method, whereas that does not meet the parametric assumptions is estimated by the method nonparametric. This study aims to analyze semiparametric regression model using epanechnikov kernel estimator with dengue fever case and simulated data. Selection of the optimal bandwidth will produce a smooth regression curve estimation in accordance with the pattern data. Selection of the optimum bandwidth is determined based on the criteria that the minimum value of GCV. The result shows that semiparametric regression model with epanechnikov kernel is able to estimate the real and simulated data very well.

Keywords: *Semiparametric Regression, Kernel, Bandwidth, GCV*

ABSTRAK

ESTIMASI MODEL REGRESI SEMIPARAMETRIK MENGGUNAKAN PENDUGA NADARAYA-WATSON DENGAN KERNEL EPANECHNIKOV

Oleh

RESTI NOVALIA

Estimasi model regresi semiparametrik adalah estimasi yang menggabungkan model regresi parametrik dan nonparametrik. Variabel independen yang memenuhi asumsi parametrik dapat diprediksi dengan metode kuadrat terkecil, sedangkan yang tidak memenuhi asumsi parametrik dapat diduga oleh metode nonparametrik. Tujuan dari penelitian ini adalah untuk menganalisis model regresi semiparametrik dengan menggunakan estimator kernel epanechnikov dari kasus demam berdarah dan data simulasi. Pemilihan *bandwidth* optimal akan menghasilkan kurva regresi yang mulus estimasi sesuai dengan pola data. Pemilihan bandwidth optimal ditentukan berdasarkan kriteria nilai GCV minimum. Hasil dari penelitian ini menunjukkan bahwa model regresi semiparametrik dengan kernel epanechnikov dapat mengestimasi data asli dan data simulasi dengan sangat baik.

Kata Kunci: Regresi Semiparametrik, Kernel, *Bandwidth*, GCV

**ESTIMASI MODEL REGRESI SEMIPARAMETRIK
MENGUNAKAN PENDUGA NADARAYA-WATSON
DENGAN KERNEL EPANECHNIKOV**

Oleh

Resti Novalia

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019**

Judul Skripsi

: **ESTIMASI MODEL REGRESI
SEMIPARAMETRIK MENGGUNAKAN
PENDUGA NADARAYA-WATSON DENGAN
KERNEL EPANECHNIKOV**

Nama Mahasiswa

: **Resti Novafia**

Nomor Pokok Mahasiswa

: 1517031090

Jurusan

: Matematika

Fakultas

: Matematika dan Ilmu Pengetahuan Alam



Ir. Netti Herawati, M.Sc., Ph.D.
NIP. 196501251990032001

Drs. Suharsono S., M.S., M.Sc., Ph.D.
NIP. 196205131986031003

2. Ketua Jurusan Matematika

A handwritten signature in blue ink, consisting of stylized, overlapping loops and lines.

Prof. Dra. Wamliana, M.A., Ph.D.
NIP. 196311081989022001

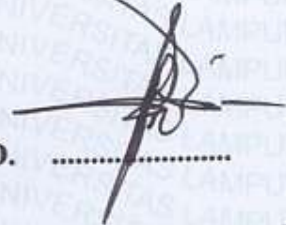
MENGESAHKAN

1. Tim. Penguji

Ketua : Ir. Netti Herawati, M.Sc., Ph.D.

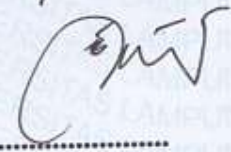


Sekretaris : Drs. Suharsono S., M.S., M.Sc., Ph.D.



Penguji

Bukan Pembimbing : Drs. Eri Setiawan, M.Si.



Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Drs. Suratman, M.Sc.
NIP. 196406041990031002

Tanggal Lulus Ujian Skripsi : 17 Oktober 2019

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini :

Nama : **Resti Novalia**

Nomor Pokok Mahasiswa : **1517031090**

Jurusan : **Matematika**

Judul : **ESTIMASI MODEL REGRESI
SEMIPARAMETRIK MENGGUNAKAN
PENDUGA NADARAYA-WATSON DENGAN
KERNEL EPANECHNIKOV**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 17 Oktober 2019

Yang Menyatakan,



Resti Novalia

NPM. 1517031090

RIWAYAT HIDUP

Penulis dilahirkan di Kota Bandar Lampung, Provinsi Lampung pada tanggal 26 Mei 1997, sebagai anak pertama dari tiga bersaudara, dari pasangan bapak Nurdin dan ibu Laili Yustina. Penulis mulai menempuh pendidikan pertamanya di Taman Kanak-Kanak Dharma Wanita Unila pada tahun 2002. Pada tahun 2003 penulis melanjutkan pendidikannya di Sekolah Dasar Negeri 3 Rajabasa, Bandar Lampung. Kemudian, penulis melanjutkan pendidikan di Sekolah Menengah Pertama Negeri 8 Bandar Lampung pada tahun 2009. Pada tahun 2012 penulis melanjutkan pendidikannya di Sekolah Menengah Atas Negeri 14 Bandar Lampung. Pada tahun 2015 penulis tercatat sebagai salah satu mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam di Universitas Lampung melalui jalur masuk Penerimaan Mahasiswa Perluasan Akses Pendidikan (PMPAP) yang merupakan salah satu beasiswa yang disediakan oleh Universitas Lampung.

Selama menjadi mahasiswa penulis aktif di Organisasi tingkat jurusan yaitu Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) menjabat sebagai anggota pada Biro Dana dan Usaha HIMATIKA pada dua periode kepengurusan yaitu pada tahun 2016 dan 2017.

Pada tahun 2018 penulis melakukan Kerja Praktik (KP) di Badan Pusat Statistik Provinsi Lampung. Sebagai bentuk pengabdian mahasiswa, penulis telah melaksanakan Kuliah Kerja Nyata (KKN) di Desa Wawasan Kecamatan Tanjung Sari, Kabupaten Lampung Selatan.

Kata Inspirasi

Jika berani naik pohon harus berani turun juga, jika berani daftar kuliah harus
berani lulus juga

- Bambang Widodo

Anda mungkin bisa menunda, tapi waktu tidak akan menunggu

- Benjamin Franklin

Ilmu tanpa agama adalah lumpuh, agama tanpa ilmu adalah buta

- Albert Einstein

PERSEMBAHAN

Dengan mengucapkan Alhamdulillah,
puji dan syukur kepada Allah SWT atas segala rahmat dan karunia-Nya, serta
suri tauladan Nabi Muhammad SAW.

Kupersembahkan sebuah karya sederhana ini untuk:
Ayahanda Nurdin & Ibunda Laili Yustina

Terimakasih Bak, Emak untuk semua limpahan kasih sayang, pengorbanan, doa,
dan dukungan selama ini. Karena atas ridho kalianlah Allah memudahkan setiap
langkah-langkah yang aku tapaki.

Mungkin karya ini tak sebanding dengan pengorbanan yang telah kalian lakukan.
Tapi percayalah ini sebuah titik awal perjuangan baktiku untuk kalian, karena
kalian adalah motivasi terbesar dalam hidupku.

Kedua adikku *Ariki Ardiansyah* dan *Sahrul Ramadhan* dan keluarga dekat yang
senantiasa berdoa untuk keberhasilanku.

Dosen Pembimbing dan Penguji yang sangat berjasa dalam mengarahkan, membimbing,
menasihati, dan memberi motivasi kepada penulis.

Serta,
Almamaterku tercinta

SANWACANA

Syukur Alhamdulillah penulis panjatkan kehadiran Allah SWT karena atas berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi ini. Skripsi yang berjudul “Estimasi Model Regresi Semiparametrik Menggunakan Penduga Nadaraya-Watson dengan Kernel Epanechnikov” disusun sebagai salah satu syarat memperoleh gelar Sarjana Sains (S.Si.) di Universitas Lampung. Dalam kesempatan ini rasa terima kasih yang setulus-tulusnya penulis ucapkan kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D. selaku dosen pembimbing I yang selalu mengarahkan, membimbing dan memotivasi penulis dalam menyelesaikan skripsi ini.
2. Bapak Drs. Suharsono, M.S., M.Sc., Ph.D. selaku dosen pembimbing II, terima kasih untuk bimbingan dan masukannya selama penyusunan skripsi ini.
3. Bapak Drs. Eri Setiawan, M.Si. selaku dosen penguji yang telah memberikan saran dan nasehatnya dalam menyelesaikan skripsi ini.
4. Bapak Dr. Lazakaria, S.Si., M.Sc. selaku Pembimbing Akademik atas bimbingan dan pembelajarannya selama ini.
5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Drs. Suratman, M.Sc. selaku Dekan FMIPA Universitas Lampung.

7. Seluruh Dosen dan Staf Jurusan Matematika yang telah memberikan Ilmu dan bantuan yang berguna bagi penulis.
8. Bak dan Emak tercinta yang senantiasa mendoakan, menyayangi, memberi semangat dan nasehat, serta kedua adik dan keluarga dekat yang senantiasa berdoa untuk keberhasilan penulis.
9. Sekar, Dinda, Rahma, Pipin, Rima, Loves terimakasih atas kebersamaan, keceriaan dan dukungannya selama perkuliahan ini.
10. Sahabat-sahabatku tersayang Ayu Meta Sari, Firda Aulia Larasati, Nanda Rizqi Caesarani, M. Rizky Saputra dan Dimas Ponco Wirianto terimakasih atas kebersamaan, keceriaannya dan waktu luang yang telah diberikan untuk penulis.
11. Anggun, Agung, dan Tirania terimakasih sudah selalu ada saat penulis membutuhkan pertolongan.
12. Teman – Teman HIMATIKA dan angkatan 2015 yang tidak dapat disebutkan satu persatu terimakasih atas keakraban dan kebersamaan selama ini.
13. Seluruh pihak yang telah membantu yang tidak dapat disebutkan satu persatu, atas dukungan dan doanya selama penyelesaian skripsi ini.

Bandar Lampung, November 2019

Penulis,

Resti Novalia

DAFTAR ISI

	Halaman
DAFTAR TABEL	i
DAFTAR GAMBAR	ii
I. PENDAHULUAN	1
1.2 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	2
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	4
2.1 Analisis Regresi	4
2.2 Regresi Parametrik	5
2.3 Uji Normalitas	7
2.4 Uji Multikolinearitas.....	7
2.5 Uji Heteroskedastisitas	7
2.6 Regresi Nonparametrik.....	8
2.7 Regresi Semiparametrik	9
2.8 Estimasi Model Semiparametrik Kernel.....	9
2.8.1 <i>Partially Linear Model</i>	9
2.8.2 Estimasi Parameter Variabel Parametrik.....	10
2.8.3 Estimasi Kurva Regresi Nonparametrik	10
2.9 Fungsi Kernel	11
2.9.1 Definisi Fungsi Kernel.....	11
2.9.2 Macam-macam Fungsi Kernel	11
2.10 Metode Kernel dengan Fungsi Epanechnikov	13
2.11 Pemilihan <i>Bandwidth</i> Optimal.....	13
2.12 Ukuran Kebaikan Model	15

III. METODOLOGI PENELITIAN	16
3.1 Waktu dan Tempat Penelitian.....	16
3.2 Data Penelitian.....	16
3.3 Metodologi Penelitian.....	17
IV.HASIL DAN PEMBAHASAN.....	19
4.1 Penentuan Komponen Parametrik dan Nonparametrik Pada Data Asli	19
4.2 <i>Partially Linear Model</i> pada Data Asli	21
4.2.1 Estimasi Parameter Regresi Variabel Parametrik	21
a. Uji Normalitas Galat	21
b. Uji Multikolinearitas	22
c. Uji Heteroskedastisitas Galat	23
4.4.2 Estimasi Kurva Regresi Nonparametrik	24
4.3 Hasil Estimasi Regresi Semiparametrik pada Data Asli	26
4.4 Penentuan Komponen Parametrik dan Nonparametrik Pada Data Bangkitan	27
4.5 <i>Partially Linear Model</i> pada Data Bangkitan.....	29
4.5.1 Estimasi Parameter Regresi Variabel Parametrik	29
a. Uji Normalitas Galat	29
b. Uji Multikolinearitas	30
c. Uji Heteroskedastisitas Galat	31
4.5.2 Estimasi Kurva Regresi Nonparametrik	32
4.6 Hasil Estimasi Regresi Semiparametrik pada Data Bangkitan .	34
V. KESIMPULAN.....	35
DAFTAR PUSTAKA	36
LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Pengujian Normalitas Galat	21
2. Tingkat Korelasi Antarvariabel Bebas	22
3. Uji <i>Breusch Pagan</i> pada Data Asli	23
4. Nilai Estimasi Parameter Beta	23
5. Nilai <i>Bandwidth</i> dan GCV pada Data Asli	24
6. Pengujian Normalitas Galat	29
7. Tingkat Korelasi Antarvariabel Bebas	30
8. Uji <i>Breusch Pagan</i> pada Data Bangkitan.....	31
9. Nilai Estimasi Parameter Beta	31
10. Nilai <i>Bandwidth</i> dan GCV pada Data Bangkitan.....	32

DAFTAR GAMBAR

Gambar	Halaman
1. Diagram Pencar Antara Y dengan X_1	19
2. Diagram Pencar Antara Y dengan X_2	20
3. Diagram Pencar Antara Y dengan X_3	20
4. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> terlalu kecil ($h = 1$).....	25
5. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> optimal ($h = 1.6$).....	25
6. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> terlalu besar ($h = 2$).....	26
7. Diagram Pencar Antara Y dengan X_1	27
8. Diagram Pencar Antara Y dengan X_2	28
9. Diagram Pencar Antara Y dengan X_3	28
10. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> terlalu kecil ($h = 1$).....	33
11. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> optimal ($h = 1.6$).....	33
12. Kurva dugaan regresi nonparametrik dengan kernel Epanechnikov pada data Y dengan variabel X_3 untuk nilai <i>bandwidth</i> terlalu besar ($h = 2$).....	34

I. PENDAHULUAN

1.1. Latar Belakang dan Masalah

Analisis regresi merupakan salah satu teknik statistika yang digunakan untuk menggambarkan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor. Pada regresi harus ada variabel yang ditentukan dan variabel yang menentukan atau dengan kata lain adanya ketergantungan antara variabel yang satu dengan yang lainnya. Kedua variabel biasanya bersifat kausal atau mempunyai hubungan sebab akibat yaitu saling berpengaruh. Sehingga dengan demikian, regresi merupakan bentuk fungsi tertentu antara variabel tak bebas (Y) dengan variabel bebas (X). Jika bentuk kurva regresi diketahui, maka untuk menduga fungsi regresi dapat menggunakan pendekatan parametrik. Namun, jika kurva regresi tidak diketahui, maka untuk menduga fungsi regresi dapat menggunakan pendekatan nonparametrik (Green & Silverman, 2000).

Pendekatan model regresi semiparametrik merupakan pendekatan model yang mengombinasikan model regresi parametrik dan regresi nonparametrik. Pada regresi semiparametrik, sebagian variabel penjelasnya bersifat parametrik dan sebagian lain bersifat nonparametrik. Regresi semiparametrik digunakan jika pola

hubungan antara sekumpulan variabel bebas dan variabel terikat diketahui dan ada pula yang tidak diketahui. Diberikan model regresi semiparametrik:

$$y_i = \sum_{k=1}^n \beta_k x_{ik} + m(x_i) + \varepsilon_i \quad ; i = 1, 2, \dots, n \quad (1.1)$$

y_i adalah variabel respon ke- i , $\sum_{k=1}^n \beta_k x_{ik}$ adalah komponen parametrik, $m(x_i)$ adalah fungsi regresi yang tidak diketahui, dan ε_i adalah galat acak. Terdapat beberapa teknik *smoothing* dalam model regresi semiparametrik salah satunya adalah kernel. Penduga kernel fleksibel dan secara matematik mudah diselesaikan. Pada penduga kernel yang terpenting adalah pemilihan parameter pemulus (*bandwidth*) yang optimal untuk mendapatkan kurva regresi yang optimal. Pada teknik pemulusan menggunakan estimator kernel, terdapat tiga macam estimasi kernel yaitu Nadaraya-Watson, Priestly-Chao dan Gasser-Müller. Dalam metode ini dibutuhkan fungsi kernel, fungsi kernel antara lain kernel Uniform, Triangle, Epanechnikov, Gaussian, Kuadratik, Cosinus, dan Tricube. Kernel Epanechnikov merupakan fungsi kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding fungsi kernel yang lainnya. Berdasarkan uraian tersebut, peneliti tertarik untuk menduga model regresi semiparametrik menggunakan penduga Nadaraya-Watson dengan kernel Epanechnikov (Hastie & Tibshirani, 1990).

1.2. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengestimasi model regresi semiparametrik menggunakan estimator kernel dengan fungsi Epanechnikov.

1.3. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Dapat menjelaskan model analisis regresi kernel dengan fungsi Epanechnikov.
2. Dapat menjadi referensi bagi pembaca apabila ingin melakukan penelitian mengenai model regresi menggunakan estimator kernel dengan fungsi Epanechnikov.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan salah satu teknik statistika yang sering digunakan untuk mengetahui hubungan kausal antara variabel respon dan variabel prediktor. Untuk data yang satu variabel prediktor dan satu variabel respon, analisis regresi disebut juga dengan analisis regresi linear sederhana. Sedangkan untuk data yang memiliki lebih dari variabel prediktor dan satu variabel respon disebut dengan analisis regresi linear berganda. Jadi secara umum analisis regresi adalah metode analisis data yang menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor. Selain digunakan untuk mengetahui hubungan antara data berpasangan x dan y , fungsi lain dari regresi adalah menggunakan model umum regresi untuk melakukan prediksi secara matematis (Green & Silverman, 2000).

Ada dua pendekatan yang dapat digunakan untuk mengestimasi fungsi regresi atau kurva regresi, yaitu secara parametrik dan nonparametrik. Dalam pendekatan parametrik, bentuk hubungan antara variabel respon dan variabel prediktor diketahui atau diperkirakan dari bentuk kurva regresi. Sedangkan pada regresi

nonparametrik, bentuk kurva tidak dapat langsung diketahui atau diperkirakan (Wand & Jones, 1995).

Sedang garis regresi didefinisikan sebagai garis lurus yang ditarik dari titik – titik diagram pencar (*scattered diagram*) dari nilai variabel terikat dan variabel bebas sehingga garis tersebut menggambarkan hubungan linier antara variabel-variabel tersebut. Garis ini dapat digambarkan dari nilai-nilai persamaan regresi dalam bentuk yang paling sederhana yaitu:

$$Y_i = \beta_0 + \beta_1 X_i \quad (2.1)$$

Sumbu vertikal dari diagram pencar digunakan untuk menggambarkan nilai-nilai variabel respon sedang sumbu horizontal menggambarkan nilai prediktor.

Intercept merupakan titik sumbu vertikal yang merupakan nilai variabel respon yang diprediksi saat nilai prediktor atau variabel bebas sebesar nol. Nilai yang diprediksi akan sebesar 0 jika koefisien regresi baku digunakan (Green & Silverman, 2000).

2.2 Regresi Parametrik

Regresi parametrik merupakan metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon.

Secara umum model regresi parametrik dengan variabel prediktor ke-k adalah:

$$y_i = \sum_{k=1}^n \beta_k x_{ik} + \varepsilon_i \quad \text{dengan} \quad i = 1, 2, \dots, n \quad (2.2)$$

Bentuk hubungan antara variabel respon dan variabel prediktor diperkirakan dari bentuk kurva regresi, misalkan berbentuk pola linear, kuadrat, eksponensial, dan

polinomial. Dalam pendekatan regresi parametrik, data yang digunakan harus memenuhi asumsi kenormalan galat, yaitu galat berdistribusi normal dengan rata-rata nol dan ragam konstan. Dalam kasus parametrik, metode yang paling populer untuk menduga fungsi regresi adalah metode kuadrat terkecil (Hardle, 1994).

Dalam regresi parametrik terdapat asumsi yang sangat kaku dan kuat yaitu bentuk kurva regresi diketahui, misalnya linear, kuadratik, kubik, polinomial derajat-p, eksponen, dan lain-lain. Untuk memodelkan data menggunakan regresi parametrik linear, kuadrat, kubik atau yang lain, umumnya dimulai dengan membuat diagram pencar. Pendekatan regresi parametrik memiliki sifat yang sangat baik dari pandangan statistika inferensi, seperti sederhana, mudah interpretasinya, parsimoni, estimatornya tidak bias, tergolong estimator linear, efisien, konsisten, BLUE (*Best Linear Unbiased Estimator*), yang sangat jarang dimiliki oleh pendekatan regresi lain seperti regresi nonparametrik dan regresi semiparametrik. Dalam pandangan regresi nonparametrik, data mencari sendiri bentuk estimasi dari kurva regresinya tanpa harus dipengaruhi oleh faktor subyektifitas si peneliti. Ini berarti pendekatan model regresi nonparametrik sangatlah fleksibel dan sangat obyektif. Pendekatan model regresi parametrik yang sering digunakan oleh para peneliti adalah pendekatan model regresi linier sederhana dan model regresi linier berganda, sedangkan pada model regresi nonparametrik, pendekatan model regresi yang sering digunakan adalah pendekatan regresi spline dan regresi kernel (Wand & Jones, 1995).

2.3 Uji Normalitas

Uji Normalitas bertujuan untuk menguji apakah dalam model regresi variabel terikat dan variabel bebas keduanya mempunyai distribusi normal atau tidak. Cara yang bisa ditempuh untuk menguji kenormalan data adalah dengan menggunakan Grafik Normal P-P Plot dengan cara melihat penyebaran datanya. Jika pada grafik tersebut penyebaran datanya mengikuti pola garis lurus, maka datanya normal. Jika pada tabel test of normality dengan menggunakan Kolmogorov Smirnov nilai $\text{sig} > 0.05$, maka data berdistribusi normal (Sugiyono, 2013).

2.4 Uji Multikolinearitas

Uji Multikolinieritas bertujuan untuk menguji apakah model regresi ditemukan adanya korelasi antar variabel bebas (variabel independen). Model regresi yang baik seharusnya tidak terjadi korelasi antar variabel independen. Multikolinieritas dapat dilihat dari nilai Tolerance dan nilai Variance Inflation Factor (VIF). Jika nilai Tolerance $> 0,1$ atau sama dengan nilai VIF < 10 , maka dapat disimpulkan bahwa tidak ada multikolinieritas antar variabel independen dalam model (Ghozali, 2013).

2.5 Uji Heteroskedastisitas

Uji heteroskedastisitas bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan variance dari residual satu pengamatan ke pengamatan lain.

Salah satu statistik uji yang dapat digunakan untuk menguji apakah ragam dari error bersifat homoskedastik atau tidak adalah *Breusch-Pagan Test*. Uji ini mengasumsikan bahwa komponen error adalah independen dan tersebar normal (Ghozali, 2013).

2.6 Regresi Nonparametrik

Metode regresi nonparametrik adalah metode analisis data yang sangat populer semenjak akhir tahun 1990-an. Kebutuhan terhadap analisis yang dapat menganalisis jenis data yang beragam dan belum tentu memenuhi asumsi-asumsi parametrik membuat teknik ini makin banyak digunakan. Karena sifatnya yang fleksibel, pendekatan model regresi nonparametrik berperan penting dalam memeriksa data longitudinal (Wu & Zhang, 2006).

Regresi nonparametrik disebut fleksibel dikarenakan data tidak harus memenuhi asumsi-asumsi tertentu seperti pada regresi parametrik. Pada metode regresi nonparametrik, bentuk fungsi tidak diketahui dan tidak tergantung pada asumsi bentuk kurva tertentu (Hardle, 1994).

Misalkan X adalah variabel prediktor dan Y adalah variabel respon untuk n pengamatan berpasangan $\{(x_i, y_i)\}$, maka hubungan linear antara kedua variabel dapat diketahui dengan model umum regresi yaitu:

$$y_i = m(x_i) + \varepsilon_i \quad \text{dengan} \quad i = 1, 2, \dots, n \quad (2.3)$$

Dengan y_i adalah variabel tidak bebas pada pengamatan ke- i , ε_i merupakan

galat/sisaan/kesalahan ke- i yang diasumsikan menyebar normal dengan mean sama dengan 0 dan ragam tetap σ^2 , serta $m(x_i)$ merupakan fungsi regresi atau kurva regresi (Eubank, 1998).

2.7 Regresi Semiparametrik

Regresi semiparametrik merupakan salah satu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor dimana sebagian diketahui bentuk polanya, dan sebagian tidak diketahui bentuknya.

Analisis regresi semiparametrik merupakan gabungan dari regresi parametrik dan regresi nonparametrik, sehingga estimasi model semiparametrik ekuivalen dengan estimasi parameter-parameter pada komponen parametrik dan estimasi kurva pada komponen parametrik. Model regresi semiparametrik dapat ditulis sebagai berikut:

$$y_i = \sum_{k=1}^n \beta_k x_{ik} + m(x_i) + \varepsilon_i ; i = 1, 2, \dots, n \quad (2.4)$$

dimana y_i adalah variabel respon ke- i , $\sum_{k=1}^n \beta_k x_{ik}$ adalah komponen parametrik, $m(x_i)$ adalah fungsi regresi yang tidak diketahui, dan ε_i adalah galat acak (*random error*) dimana $\varepsilon_i \sim N(0, \sigma^2)$ (Hastie & Tibshirani, 1990).

2.8 Estimasi Model Semiparametrik Kernel

2.8.1 Partially Linear Model

Pada model persamaan (2.4) dilakukan estimasi parameter untuk variabel parametrik pada model $\hat{y}_i = \sum_{k=1}^n \beta_k x_{ik}$ dengan x adalah variabel parametrik,

kemudian mengestimasi kurva regresi pada model $\hat{y}_i = m(x_i)$ dengan x adalah variabel nonparametrik (Hastie & Tibshirani, 1990).

2.8.2 Estimasi Parameter Variabel Parametrik

Menurut Searle (1971), estimasi parameter variabel parametrik dilakukan dengan menggunakan metode kuadrat terkecil dengan meminimumkan jumlah kuadrat galat sehingga diperoleh rumus:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.7)$$

Atau dapat pula dijabarkan sebagai berikut:

$$\begin{bmatrix} \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 X_1 & \sum X_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum X_1 Y \\ \sum X_2 Y \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad (2.8)$$

2.8.3 Estimasi Kurva Regresi Nonparametrik

Menurut Hastie & Tibshirani (1990), estimasi m didapatkan menggunakan estimator kernel. Adapun penduga fungsi regresi semiparametrik adalah sebagai berikut:

$$\hat{m}(x) = \sum_{i=1}^n w_{hi}(x) y_i \quad (2.5)$$

dengan

$$w_{hi}(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)} \quad (2.6)$$

2.9 Fungsi Kernel

2.9.1 Definisi Fungsi Kernel

Suatu fungsi kernel harus merupakan fungsi kontinu, berharga riil, simetris dan terbatas. Menurut Hardle (1994), secara umum fungsi kernel didefinisikan sebagai berikut:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \quad \text{untuk } -\infty < x < \infty \text{ dan } h > 0 \quad (2.9)$$

dengan : K = fungsi kernel

h = *bandwith* atau *smoothing parameter*

Fungsi kernel diatas harus memenuhi beberapa syarat yaitu:

- (i) $K(x) \geq 0$, untuk semua x
- (ii) $\int_{-\infty}^{\infty} K(x)dx = 1$
- (iii) $\int_{-\infty}^{\infty} K(x)dx = \sigma^2 > 0$
- (iv) $\int_{-\infty}^{\infty} K(x)dx = 0$

2.9.2 Macam-macam Fungsi Kernel

Menurut Hadijati & Budiantara (2012), pada estimator kernel terdapat beberapa fungsi kernel yang umum digunakan untuk estimasi data, yaitu:

1. Kernel Uniform : $K(x) = \frac{1}{2}$; $x \in [-1, 1]$, 0 selainnya
2. Kernel Triangle : $K(x) = (1 - |x|)$; $x \in [-1, 1]$, 0 selainnya
3. Kernel Epanechnikov : $K(x) = \frac{3}{4}(1 - x^2)$; $x \in [-1, 1]$, 0 selainnya
4. Kernel Quartik : $K(x) = \frac{15}{16}(1 - x^2)^2$; $x \in [-1, 1]$, 0 selainnya

5. Kernel Triweight : $K(x) = \frac{35}{32}(1-x^2)^3$; $x \in [-1, 1]$, 0 selainnya
6. Kernel Cosinus : $K(x) = \frac{\pi}{4} \cos(\frac{\pi}{2}x)$; $x \in [-1, 1]$, 0 selainnya
7. Kernel Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$; $-\infty < x < \infty$
8. Kernel Tricube : $K(x) = \frac{70}{81}(1-x^3)^3$; $x \in [-1, 1]$, 0 selainnya

Estimator densitas kernel untuk fungsi densitas $m(x_i)$ didefinisikan:

$$\hat{m}_h(x_i) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (2.10)$$

Dari persamaan di atas, dapat dilihat bahwa \hat{m}_h dipengaruhi oleh fungsi kernel K dan parameter pemulus h . parameter pemulus (*bandwidth*) dalam fungsi densitas kernel berfungsi untuk mengatur kehalusan kurva yang akan diestimasi. Peran *bandwidth* ini diasumsikan seperti lebar interval pada histogram (Hadijati & Budiantara, 2012).

Perkalian taksiran kepadatan kernel jika *bandwidth*-nya dinyatakan dengan h_i atau $h = h_1, \dots, h_d$, maka taksiran kepadatannya adalah:

$$\hat{m}_h(x_i) = \frac{1}{n} \sum_{i=1}^n \left(\prod_{j=1}^d h_j^{-1} \right) K_h\left(-\frac{x_1-x_{i1}}{h_1}, \dots, \frac{x_d-x_{id}}{h_d}\right) \quad (2.11)$$

Sehingga perkalian taksiran kepadatan kernel adalah perkalian dari kernel marginal yaitu:

$$\hat{m}_h(x_i) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{j,h_j}(x_j - X_{ij}) \quad (2.12)$$

dimana $K_{j,h_j}(\cdot)$ adalah kepadatan kernel komponen ke- j dalam dimensi ke- d dengan *bandwidth* h_j (Hadijati & Budiantara, 2012).

2.10 Metode Kernel dengan Fungsi Epanechnikov

Kernel Epanechnikov merupakan fungsi kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding fungsi kernel yang lainnya.

Estimator kernel Epanechnikov didapatkan dengan cara mensubstitusikan rumus fungsi kernel Epanechnikov pada estimator Nadaraya-Watson sehingga menjadi:

$$\hat{m}(x_i) = \frac{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x-X_i}{h} \right)^2 \right) IY_i \right)}{\sum_{i=1}^n \frac{3}{4} \left(1 - \left(\frac{x-X_i}{h} \right)^2 \right) I} \quad (2.13)$$

dan persamaannya akan menjadi:

$$\hat{y} = \frac{\sum_{i=1}^n \left(\frac{3}{4} \left(1 - \left(\frac{x-X_i}{h} \right)^2 \right) IY_i \right)}{\sum_{i=1}^n \frac{3}{4} \left(1 - \left(\frac{x-X_i}{h} \right)^2 \right) I} + \varepsilon_i \quad (2.14)$$

dengan: $i = 1, 2, 3, \dots, n$.

Persamaan (2.14) yang telah diperoleh kemudian digunakan untuk mengestimasi data. Hal terpenting yang perlu diperhatikan adalah pemilihan *bandwidth* (Hardle, 1994).

2.11 Pemilihan *Bandwidth* Optimal

Bandwidth (h) adalah parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Menurut Silverman (1986), rumus *bandwidth* optimum adalah sebagai berikut:

$$h = 0,79(\text{IQR})n^{-1/5} \quad (2.15)$$

dengan:

$$IQR = \text{Inter Quartil Range } (Q_3 - Q_1)$$

n = banyaknya data

Proses pemilihan *bandwidth* yang sesuai (parameter *smoothing*) adalah bagian yang penting dari regresi nonparametrik. Telah diketahui secara umum, bahwa permasalahan utama pada kernel *smoothing* bukan terletak pada pemilihan kernel tetapi pada pemilihan *bandwidth*. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi yang kurang *smooth* (*under smooth*), sebaliknya jika *bandwidth* yang dipilih terlalu besar maka akan menghasilkan estimasi yang sangat *smooth* (*over smooth*) yang tidak sesuai dengan pola sebaran data.

Sehingga harus dipilih nilai *bandwidth* yang optimal agar dihasilkan estimasi terbaik. Tujuan estimasi kurva tidak hanya untuk memperoleh kurva yang mulus tetapi juga memiliki tingkat kesalahan yang tidak terlalu besar. Berdasarkan hal itu perlu dipilih nilai *bandwidth* optimal sehingga didapatkan kurva yang mulus dengan kesalahan yang minimum (Gujarati, 2003).

Menurut Craven & Wahba (1979), salah satu metode untuk mendapatkan h optimal diperoleh dengan menggunakan kriteria *Generalized Cross Validation*(GCV), yang didefinisikan sebagai berikut:

$$GCV = \frac{MSE}{\left(\frac{1}{n} \text{tr}(I-H(h))\right)^2} \quad (2.16)$$

dengan:

n = banyaknya data

I = matriks identitas

$$H(h) = X(X'X + nhI)^{-1} X'$$

$$h = \text{bandwidth}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

2.12 Ukuran Kebaikan Model

R-Square atau koefisien determinasi merupakan salah satu ukuran yang sederhana dan sering digunakan untuk menguji kualitas suatu persamaan garis regresi (Gujarati, 2004). Nilai R-Square memberikan gambaran tentang kesesuaian variabel independen dalam memprediksi variabel dependen. Adapun perhitungan nilai R-Square adalah sebagai berikut:

$$R^2 = \frac{JKR}{JKT} \quad (2.17)$$

Sifat dari *R-Square* adalah:

- a. Merupakan besaran yang non-negatif
- b. Batasnya adalah $0 \leq R^2 \leq 1$

Untuk mengetahui metode estimasi yang memberikan hasil yang lebih baik, maka kriteria yang digunakan adalah dengan membandingkan nilai *R-Square* (R^2) yang menunjukkan seberapa besar proporsi variasi variabel dependen yang dijelaskan oleh variabel independen. Apabila nilai koefisien determinasi semakin besar, maka semakin besar kemampuan semua variabel independen dalam menjelaskan varians dari variabel dependen (Ghozali, 2013).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2018-2019 dan bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini yaitu data simulasi dan diaplikasikan pada data rekam medis pasien DBD periode bulan Januari sampai bulan Agustus 2018 yang diambil dari RSUD Abdul Moeloek Bandar Lampung. Populasi dari penelitian ini adalah pasien DBD yang pernah menjalani rawat inap di RSUD Abdul Moeloek Bandar Lampung. Peubah respons yaitu Lama Kesembuhan Pasien DBD (hari) dan peubah bebas yaitu Suhu Tubuh (X_1), Kadar Hemoglobin (X_2), serta Kadar Hematokrit (X_3).

3.3 Metode Penelitian

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan data asli, Lama Kesembuhan pasien DBD dan faktor yang diduga mempengaruhi yaitu data Suhu Tubuh (X_1), Kadar Hemoglobin (X_2) dan kadar Hematokrit (X_3). Begitupun dengan data bangkitan, mendeskripsikan data X_1 , X_2 dan X_3 .
2. Melakukan simulasi data dengan langkah-langkah sebagai berikut:
 - a. Membangkitkan data dengan variabel bebas (X_i) $p = 3$, $n = 30$, $\beta_1 = 0.8666508$, $\beta_2 = 1.1225459$ dengan $X_1 \sim \text{Normal}(37.6167, 0.4939)$, $X_2 \sim \text{Normal}(31.3, 3.75)$ dan $X_3 \sim \text{Weibull}(10.9, 41.65)$.
 - b. Mendapatkan nilai Y_i dari model $Y_i = \beta_1 X_1 + \beta_2 X_2 + m(X_i) + \epsilon_i$
3. Membuat diagram pencar pada masing-masing variabel prediktor dengan variabel respon untuk mengetahui bentuk pola data.
4. Melakukan uji asumsi klasik terhadap galat pada variabel parametrik
5. Mengestimasi parameter regresi pada variabel parametrik menggunakan metode kuadrat terkecil pada persamaan (2.7).
6. Menghitung nilai *bandwith* menggunakan rumus lebar jendela silverman pada persamaan (2.15).
7. Menentukan parameter pemulus (*bandwidth*) optimal menggunakan metode Nadaraya-Watson dengan kernel Epanechnikov dan ditentukan dengan GCV yang memenuhi kriteria minimum.
8. Mengestimasi kurva regresi nonparametrik menggunakan *bandwidth* optimal

9. Mendapatkan estimasi model regresi semiparametrik berdasarkan hasil estimasi parameter regresi parametrik dan estimasi kurva regresi nonparametrik.
10. Interpretasi model yang diperoleh.

KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan bahwa berdasarkan data asli pada data rekam medis pasien DBD di RSUD Abdul Moeloek Provinsi Lampung dan data simulasi didapatkan bahwa penduga Nadaraya-Watson dengan kernel Epanechnikov dapat diduga dengan baik dan menghasilkan model masing-masing sebagai berikut:

1. $\hat{y} = 0.8671058 X_1 + 1.1225459 X_2 + \hat{m}_h(X_3)$

2. $\hat{y} = 0.8668485 X_1 + 1.1219249 X_2 + \hat{m}_h(X_3)$

DAFTAR PUSTAKA

- Alifia, S. 2008. Penentuan Kurva Regresi Nonparametrik dengan Menggunakan Metode Nadaraya Watson. Skripsi. Jurusan Matematika FMIPA Universitas Lampung, Bandar Lampung.
- Draper, N.R. & Smith, H. 1992. *Analisis Regresi Terapan*. Ed. ke-2. Gramedia Pustaka Utama, Jakarta.
- Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- Gujarati, D.N. 2003. *Basic Econometrics*. McGraw-Hill Irwin, New York.
- Green, W. & Silverman, B.W. 2000. *Density Estimation for Statistics and Data Analysis*. Arrowsmith Ltd., Bristol.
- Ghozali, I. 2013. *Aplikasi Analisis Multivariate dengan Program SPSS*. Ed. ke-7. Badan Penerbit Universitas Diponegoro, Semarang.
- Fitriani, A., Srinadi, I.G.A.M. & Susilawati, M. 2015. Estimasi Regresi Semiparametrik Menggunakan Estimator Kernel Uniform. *E-Jurnal Matematika*. **4**(4): 176-180.
- Hadijati, M. & Budiantara, I.N. 2012. *Estimasi Kernel dalam Regresi Nonparametrik dengan Error Berkorelasi*. Institut Teknologi Surabaya, Surabaya.
- Hardle, W. 1994. *Applied Nonparametric Regression*. Cambridge University Press, New York.

Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized Additive Models*. Chapman and Hall, New York.

Purwanto. 2004. *Dasar-dasar Statistika*. Penerbit Alfabeta, Bandung.

Searle, S.R. 1971. *Linear Models*. Jhon Wiley and Sons, New York.

Sugiyono. 2013. *Statistika Matematika*. Graha Ilmu, Yogyakarta.

Wand, M. & Jones, M.M.C. 1995. *Kernel Smoothing*. Chapman & Hall, London.

Wu, D. & Zhang, A. 2006. *Regression Analysis*. Marcel Dekker, New York.