ANALISIS FUNGSI *HAZARD RATE* DENGAN MENGGUNAKAN ESTIMATOR KERNEL NORMAL DAN LOGNORMAL

(Skripsi)

Oleh

MAZIATUN NISA
1817031039

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

ABSTRACT

ANALYSIS OF THE HAZARD RATE FUNCTION USING NORMAL AND LOGNORMAL KERNEL ESTIMATORS

By

MAZIATUN NISA

Hazard rate is include in the survival analysis section that can applied in life by estimating the probability of survival, death, recurrence, and other events within a certain period of time. The hazard rate function will be analyzed using normal and lognormal kernel estimators which are nonparametric regression approaches. This studi aimed to compare the normal and lognormal kernel estimators on hazard rate function and get the best estimate of the hazard rate function based on hazard function graph, density function graph and the smallest MSE value. The result showed that the lognormal kernel estimator method is the best method compared to the normal kernel estimator on the hazard rate function.

Keywords: Nonparametric Regression, Hazard Rate Function, Kernel Estimator, Normal Kernel Estimator, Lognormal Kernel Estimator.

ABSTRAK

ANALISIS FUNGSI *HAZARD RATE* DENGAN MENGGUNAKAN ESTIMATOR KERNEL NORMAL DAN LOGNORMAL

Oleh

MAZIATUN NISA

Hazard rate termasuk dalam bagian analisis kelangsungan hidup yang dapat diterapkan di kehidupan dengan mengestimasi probabilitas kelangsungan hidup, kematian, kekambuhan, dan kejadian-kejadian lainnya dalam periode waktu tertentu. Fungsi *hazard rate* akan dianalisis menggunakan estimator kernel normal dan lognormal yang merupakan metode pendekatan regresi nonparametrik. Penelitian ini bertujuan untuk membandingkan metode estimator kernel normal dan lognormal pada fungsi *hazard rate* dan mendapatkan dugaan fungsi *hazard rate* lebih baik berdasarkan grafik fungsi *hazard*, grafik fungsi densitas dan nilai MSE terkecil. Hasil penelitian ini menunjukkan bahwa metode estimator kernel lognormal merupakan metode yang lebih baik digunakan dibandingkan estimator kernel normal pada fungsi *hazard rate*.

Kata Kunci: Regresi Nonparametrik, Fungsi *Hazard Rate*, Estimator Kermel, Estimator Kernel Normal, Estimator Kernel Lognormal.

ANALISIS FUNGSI *HAZARD RATE* DENGAN MENGGUNAKAN ESTIMATOR KERNEL NORMAL DAN LOGNORMAL

Oleh

MAZIATUN NISA

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika Dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

Judul Skripsi : **ANALISIS FUNGSI *HAZARD RATE* DENGAN MENGGUNAKAN ESTIMATOR KERNEL NORMAL DAN LOGNORMAL**

Nama Mahasiswa : *Maziatun Nisa*

Nomor Pokok Mahasiswa : **1817031039**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Agus Sutrisno, S.Si., M.Si.**
NIP 19650125 199003 2 001	NIP 19700831 199903 1 002

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Ir. Netti Herawati, M.Sc., Ph.D.**	....
Sekretaris	: **Agus Sutrisno, S.Si., M.Si.**	....
Penguji Bukan Pembimbing	: **Dr. Khoirin Nisa, S.Si., M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **08 Juni 2022**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama Mahasiswa	: Maziatun Nisa
Nomor Pokok Mahasiswa	: 1817031039
Jurusan	: Matematika
Judul Skripsi	: **ANALISIS FUNGSI *HAZARD RATE* DENGAN MENGGUNAKAN ESTIMATOR KERNEL NORMAL DAN LOGNORMAL**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung,

Maziatun Nisa

RIWAYAT HIDUP

Penulis bernama Maziatun Nisa dilahirkan di Kota Gajah, Lampung Tengah pada tanggal 29 November 2000 sebagai anak keempat dari empat bersaudara dari Bapak Fauzan Adnan dan Ibu Maisaroh.

Menempuh pendidikan di Taman Kanak-Kanak (TK) Islam Syarief Hidayatulloh Kota Gajah dari tahun 2005-2006. Pendidikan Sekolah Dasar (SD) diselesaikan di SDN 2 Kota Gajah, Lampung Tengah pada tahun 2012. Kemudian, penulis menyelesaikan Sekolah Menengah Pertama (SMP) di SMPN 1 Kota Gajah, Lampung Tengah pada tahun 2015. Pada tahun 2018, penulis menyelesaikan Sekolah Menengah Atas (SMA) di SMAN 1 Kota Gajah. Pada tahun 2018, penulis terdaftar sebagai sebagai mahasiswa Program Studi S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN.

Selama menjadi mahasiswa, penulis aktif dalam mengikuti kegiatan organisasi Natural FMIPA UNILA periode 2019. Pada tahun 2020, penulis melakukan kegiatan Kuliah Kerja Praktik (KP) di Badan Penyelenggara Jaminan Sosial (BPJS) Kota Metro serta mengikuti kegiatan Kerja Kuliah Nyata (KKN) di Desa Kota Gajah Timur, Lampung Tengah.

KATA INSPIRASI

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan"

(Q.S Al-Insyirah : 5)

"Allah tidak membebani seorang hamba melainkan sesuai dengan kemampuannya"

(Q.S Al-Baqarah : 286)

"Dan barangsiapa bertakwa kepada Allah, niscaya Dia menjadikan kemudahan baginya dalam urusannya"

(Q.S At-Thalaq : 4)

"Berpikirlah positif, tidak peduli seberapa keras kehidupanmu"

(Ali bin Abi Thalib)

"Tujuanmu disini bukan untuk menjadi orang hebat, tapi untuk menemukan jati dirimu yang sebenarnya"

(Maziatun Nisa)

PERSEMBAHAN

Alhamdulillah, puji dan syukur kepada Allah SWT atas nikmat serta hidayahnya sehingga skripsi ini dapat terselesaikan dengan baik dan tepat pada waktunya.

Oleh karena itu, dengan rasa syukur dan bahagia saya persembahkan rasa terima kasih kepada :

Bapak Fauzan Adnan dan Ibu Maisaroh

Terimakasih kepada kedua orang tuaku atas segala pengorbanan, motivasi, doa dan ridho kalian serta dukungannya selama ini. Terimakasih telah memberikan pelajaran berharga kepada anakkmu ini tentang makna perjalanan hidup yang sebenarnya sehingga kelak bisa manjadi orang yang bermanfaat bagi semua orang.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terimakasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasinya, serta doa-doanya dan senantiasa memberikan dukungan dalam hal apapun

Almamater Tercinta Universitas Lampung

SANWACANA

Segala puji dan syukur penulis ucapkan kepada Allah SWT atas segala nikmat dan karunia-Nya yang tak terhingga sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Analisis Fungsi *Hazard Rate* dengan Menggunakan Estimator Kernel Normal dan Lognormal”**. Dalam penulisan skripsi ini tidak dapat terselesaikan tanpa adanya bimbingan, bantuan dan dukungan dari berbagai pihak. Sehingga, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D. selaku dosen pembimbing I yang senantiasa membimbing, memberi masukan serta saran serta mendukung penulis dalam menyelesaikan skripsi ini.
2. Bapak Agus Sutrisno, S.Si., M.Si. selaku dosen pembimbing II memberikan bimbingan, pengarahan, serta saran sehingga penulis dapat menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga skripsi ini dapat terselesaikan.
4. Bapak Dr.Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.dan dosen pembimbing akademik yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Kedua Orang Tuaku, Bapak Fauzan Adnan dan Ibu Maisaroh yang selalu memberikan motivasi serta dukunganya.
7. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

8. Keluarga sekalian yang selalu memberikan semangat kepada penulis serta doa-doanya.
9. Semua teman sejurusan matematika 2018, teman kelas A serta sahabat dari luar kampus yang telah membantu serta memberikan semangat kepada penulis yang mana tidak bisa disebutkan satu persatu.
10. Dian yang telah menemani, memberikan semangat dan mendengarkan keluh kesah kepada penulis.
11. Teman seperjuangan Maydia, Virda, Muflihah, Ratu, Intan, Slivi dan orang-orang baik yang namanya tidak bisa saya sebutkan satu persatu yang telah menjadi teman terbaik penulis yang selalu memberikan semangat dan menemani penulis dalam keadaan apapun serta telah memberikan pengalaman dan banyak cerita selama masa perkuliahan.
12. Teman-teman seperbimbingan Shofi, Ferdy, Rendi, Ridho, Refi dan teman lainnya yang namanya tidak bisa saya sebutkan satu persatu yang selalu memberikan dukungan dan motivasi serta doa-doanya.
13. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis mengharapkan masukan serta saran untuk dijadikan pelajaran kedepannya.

Bandar Lampung, 02 Juni 2022
Penulis,

Maziatun Nisa

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Suatu masalah dianalisis secara bersungguh-sungguh dengan berpikir kritis agar memperoleh suatu kebenaran dari apa yang telah diasumsikan. Analisis merupakan proses suatu masalah yang harus dipecahkan dengan cara melakukan asumsi-asumsi tertentu sampai terbukti kebenarannya melalui suatu pengamatan, percobaan, dan sebagainya (Salim & Salim, 2002).

Hazard rate termasuk dalam bagian analisis kelangsungan hidup yang dapat diterapkan di kehidupan dengan mengestimasi probabilitas kelangsungan hidup, kematian, kekambuhan, dan kejadian-kejadian lainnya dalam periode waktu tertentu. Fungsi *hazard rate* merupakan perbandingan dari fungsi kepadatan peluang terhadap kelangsungan hidup.

Kemudian, fungsi *hazard rate* akan dianalisis menggunakan estimator kernel yang merupakan metode pendekatan regresi nonparametrik. Dalam statistik khususnya dalam aplikasi data nonparametrik, estimator kernel menyediakan teknik yang berguna dalam masalah estimasi fungsi. Estimator kernel banyak digunakan di banyak bidang penelitian statistik.

Estimator kernel mempunyai adanya fleksibilitas, efektivitas, dan kemampuan untuk menghasilkan pengaruh variabel respon pada fungsi yang akan diestimasi tanpa memperhatikan asumsi distribusi data tertentu. Untuk mendapatkan hasil yang lebih baik dengan tingkat kesalahan yang paling kecil disarankan dengan menggunakan metode estimator kernel dalam pemodelan nonparametrik (Saidi *et al.*, 2021).

Menurut Salha (2021), estimator kernel lognormal memiliki hasil estimasi yang lebih baik. Dan dari sekian banyak estimator kernel, penulis tertarik menggunakan estimator kernel normal dan lognormal untuk dilakukan perbandingan pada fungsi yang akan digunakan. Karena, perbandingan antara estimator-estimator kernel tersebut jarang diteliti oleh para peneliti sebelumnya ke dalam fungsi *hazard rate*.

Dalam pengoperasian pada estimator kernel dibutuhkan pemilihan *bandwidth* optimal pada metode estimator kernel yang akan diteliti dengan menggunakan salah satu metode yaitu *Generalized Cross Validation* (GCV). Masalah pemilihan *bandwidth* atau berapa banyak yang harus dihaluskan sangat penting dalam estimasi densitas (Herawati & Nisa, 2009).

Berdasarkan uraian yang telah dipaparkan, penulis ingin membandingkan metode estimator kernel normal dengan lognormal dalam fungsi *hazard rate*. Kemudian, memilih metode yang lebih baik dari kedua estimator tersebut dengan melihat hasil akurasi yang tinggi dari berbagai aspek yang dilakukan dan menghasilkan pendugaan fungsi *hazard rate* yang lebih baik.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian adalah :

1. Untuk mengetahui hasil perbandingan metode estimator kernel normal dan lognormal pada fungsi *hazard rate*.
2. Untuk mendapatkan dugaan fungsi *hazard rate* lebih baik.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Dapat menambah wawasan dan memahami lebih dalam tentang analisis fungsi *hazard rate* menggunakan estimator kernel normal dan lognormal.
2. Dapat mengetahui perbandingan hasil estimator kernel dan hasil dari dugaan fungsi *hazard rate*.
3. Dapat menjadi referensi dalam pengembangan ilmu di bidang statistik.

II. TINJAUAN PUSTAKA

2.1 Analisis Kelangsungan Hidup

Menurut Kleinbaum & Klein (2005), analisis kelangsungan hidup merupakan metode statistika yang berhubungan dengan waktu, mulai dari waktu awal yang sudah ditentukan sampai waktu akhir dalam suatu penelitian tentang mempelajari ketahanan hidup. Peristiwa yang terjadi dapat berupa peristiwa kambuhnya suatu penyakit, kematian, respon terhadap suatu percobaan atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti.

Menurut Jakperik & Ozoje (2012), analisis kelangsungan hidup merupakan suatu metode yang berhubungan dengan waktu, mulai dari waktu awal pencatatan sampai dengan terjadinya suatu peristiwa khusus atau waktu akhir pencatatan. Di dalam analisis kelangsungan hidup, ada tiga faktor yang dibutuhkan:

a. Waktu awal pencatatan yang merupakan waktu awal dimana dilakukannya pencatatan untuk menganalisis suatu kejadian.
b. Waktu akhir pencatatan yang berguna untuk mengetahui status tersensor maupun tidak tersensor suatu data.
c. Skala waktu pengukuran yang jelas dari waktu kejadian dari awal sampai akhir kejadian. Skala diukur dalam hari, minggu atau tahun.

Menurut Jakperik & Ozoje (2012), analisis statistik yang digunakan pada data kelangsungan hidup, yaitu analisis data uji hidup atau yang biasa disebut dengan analisis data waktu hidup. Data yang digunakan dalam analisis data uji hidup biasanya disebut data waktu hidup. Analisis kelangsungan hidup memiliki beberapa tujuan, sebagai berikut:

a. Mengestimasi dan menginterpretasikan fungsi kelangsungan hidup dengan fungsi *hazard*.
b. Membandingkan fungsi kelangsungan hidup dan fungsi *hazard*.
c. Mengestimasi hubungan antara variabel bebas dengan waktu kelangsungan hidup.

Data waktu tergantung pada suatu peubah acak dalam analisis kelangsungan hidup. Setiap peubah acak membentuk sebuah distribusi dan ditandai oleh tiga fungsi:

1) Fungsi kepadatan peluang
2) Fungsi kelangsungan hidup
3) Fungsi *hazard rate*

Ketiga fungsi diatas ini ekuivalen, yang artinya jika satu dari ketiganya diberikan maka dua lainnya bisa diperoleh (Liu, 2012).

2.1.1 Fungsi Kepadatan Peluang

Menurut Lee & Wang (2003), waktu kelangsungan hidup (T) mempunyai fungsi kepadatan peluang yang merupakan probabilitas kegagalan dalam interval kecil per satuan waktu, yaitu:

$$
\begin{aligned}
f(t) &= \lim_{\Delta t \to 0} \frac{\mathrm{P}(\text{kegagalan individu dalam interval } (t, t+\Delta t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathrm{P}(t \leq T < t+\Delta t)}{\Delta t} \qquad (2.1)
\end{aligned}
$$

dimana $f(t)$ adalah fungsi non negatif, yaitu:

1. $f(t) \geq 0, \forall t \geq 0$
2. $f(t) = 0, t < 0$

2.1.2 Fungsi Kelangsungan Hidup

Menurut Amran & Faruk (2015), misalkan T dinotasikan sebagai waktu kelangsungan hidup. Dan t menyatakan setiap nilai spesifik untuk variabel acak T. Fungsi kelangsungan hidup dinotasikan dengan $S(t)$ yang merupakan probabilitas kelangsungan hidup melebihi waktu t, yaitu:

$$S(t) = \mathrm{P}(T > t) \tag{2.2}$$

Waktu kelangsungan hidup T mempunyai fungsi distribusi probabilitas dengan fungsi kepadatan peluang. Sehingga definisi fungsi distribusi kumulatif $F(t)$ bagi T ditulis sebagai:

$$F(t) = P(T \leq t) = \int_{-\infty}^{t} f(u)\, du \tag{2.3}$$

karena $t > 0$ maka,

$$F(t) = \int_{0}^{t} f(u)\, du \tag{2.4}$$

Dari definisi fungsi distribusi kumulatif $F(t)$ bagi T maka,

$$S(t) = 1 - P(T \leq t) = 1 - F(t) \tag{2.5}$$

dimana fungsi kelangsungan hidup merupakan sebuah fungsi tidak naik terhadap waktu t dengan sifat yaitu:

$$S(t) = \begin{cases} 1, \text{untuk } t = 0 \\ 0, \text{untuk } t = \infty \end{cases} \tag{2.6}$$

Artinya peluang yang masih hidup paling sedikit pada waktu $t = 0$ adalah 1 dan peluang yang masih hidup pada waktu $t = \infty$ adalah 0.

2.1.3 Fungsi *Hazard Rate*

Menurut Klein & Moeschberger (2003), dalam analisis kelangsungan hidup terdapat fungsi *hazard*. Fungsi ini juga dikenal sebagai laju kegagalan yang dinotasikan dengan $h(t)$ menyatakan probabilitas setelah beberapa saat pada selang waktu usia $(t, t + \Delta t)$. Dan dengan kondisi suatu sistem masih bertahan hingga usia t. Berikut persamaan fungsi *hazard rate* secara matematis:

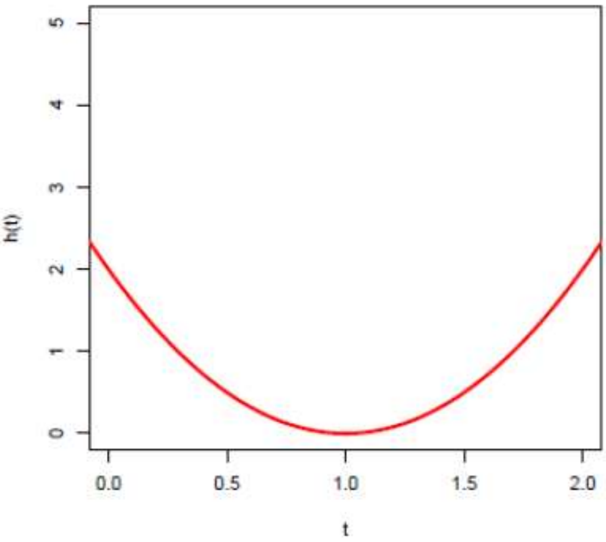
$$
\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, T \geq t)}{\Delta t . P(T \geq t)} \\
&= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t) P(T \geq t)}{\Delta t . P(T \geq t)} \\
&= \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t) - P(T < t)}{\Delta t . P(T \geq t)} \\
&= \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t) - P(T < t)}{\Delta t . [1 - P(T < t)]} \\
&= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t . [1 - F(t)]} \\
&= \frac{F'(t)}{1 - F(t)} \\
&= \frac{f(t)}{1 - F(t)}
\end{aligned}
$$

$$h(t) = \frac{f(t)}{S(t)} \tag{2.7}$$

dimana, *hazard rate* $h(t)$ untuk model distribusi laju kegagalan kontinu mempunyai sifat sebagai berikut:

1. $h(t) > 0$
2. $\int_0^\infty h(t)\, dt = \infty$

Menurut Danardono (2012), fungsi *hazard* ini juga dapat diinterpretasikan sebagai tingkat (*rate*) terjadinya suatu peristiwa. Fungsi *hazard* bukan merupakan probabilitas, sehingga dimungkinkan ketika nilainya lebih dari satu. Batasan yang digunakan pada fungsi *hazard* hanyalah $h(t) > 0$. Berikut ini contoh grafik fungsi *hazard*.



Gambar 1. Grafik Fungsi *Hazard*

Dapat dilihat pada grafik fungsi *hazard* di atas bahwa grafik tersebut berbentuk U biasanya dapat diartikan resiko kematian pada makhluk hidup secara biologis. Pada usia muda, tingkat atau resiko kematian tinggi. Resiko berkurang setelah dewasa, namun akan kembali bertambah setelah mendekati usia tua (Danardono, 2012).

Menurut Utami (2017), ketika peubah bebas dengan nilai *hazard rate* kurang dari 1, maka peningkatan nilai variabel bebas berhubungan dengan menurunnya risiko kematian dan lebih panjangnya waktu bertahan hidup. Sedangkan, saat nilai *hazard rate* lebih besar dari 1 peningkatan nilai variabel bebas berhubungan dengan meningkatnya risiko kematian dan lebih pendeknya waktu bertahan hidup.

2.2 Uji Kolmogorov Smirnov

Pada umumnya statistika nonparametrik dapat digunakan ketika asumsi parametrik tidak terpenuhi salah satunya tentang normalitas. Pengujian untuk membuktikan normal atau tidaknya suatu data dapat dilakukan dengan menggunakan analisis uji normalitas Kolmogorov Smirnov (Quraisy, 2020).

Menurut Susandri (2016), uji Kolmogorov Smirnov merupakan pengujian normalitas yang banyak digunakan oleh para peneliti, kelebihan dari uji ini adalah sederhana dan tidak menimbulkan perbedaan persepsi di antara pengamat. Konsep dasar dari uji normalitas Kolmogorov Smirnov adalah dengan membandingkan distribusi data (yang akan diuji normalitasnya) dengan distribusi normal baku. Distribusi normal baku adalah data yang telah ditransformasikan ke dalam bentuk *Z-Score* dan diasumsikan normal.
Sebenarnya uji Kolmogorov Smirnov merupakan salah satu dari uji kecocokan (*goodness of fit test*) dalam menentukan distribusi yang mendasari suatu kumpulan data atau variabel acak. Uji Kolmogorov Smirnov hanya dapat digunakan untuk menguji kecocokan suatu data dengan distribusi peluang kontinu seperti distribusi Normal, Lognormal, Eksponensial, Gamma, Weibull, dll (Pratama, 2020).

Fungsi distribusi yang akan digunakan pada penelitian adalah distribusi gamma dan eksponensial. Berikut ini fungsi kepeketan peluang $f(x)$dari kedua distribusi (Shao, 2003).

1. Distribusi Gamma

$$f(x) = \frac{1}{\gamma^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\gamma}} \;\; , x > 0 \tag{2.8}$$

2. Distribusi Eksponensial

$$f(x) = \theta e^{-\theta x} \;\; , x > 0 \tag{2.9}$$

2.3 Estimator Kernel

Menurut Komang & Gusti (2012), terdapat beberapa teknik pemulusan yang digunakan dalam metode nonparametrik, diantaranya: estimator histogram, kernel, deret orthogonal, deret fourier, wavelet, k-NN, dan spline.

Dalam statistika nonparametrik bentuk kurva yang kurang mulus dapat dipermulus dengan menggunakan teknik pemulusan. Salah satu teknik pemulusan yang umum digunakan adalah estimator kernel pada pemanfaatannya dilakukan pada setiap titik data (Sukarsa & Srinadi, 2012).

Estimator kernel merupakan pengembangan dari estimator histogram. Estimator kernel ini lebih khusus dalam penggunaan metode *bandwidth* walaupun mirip dengan estimator regresi nonparametrik yang lain (Eubank, 1988).

Menurut Ogden (1997), fungsi kernel dapat dinyatakan sebagai berikut:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\,(x - X_i) \tag{2.10}$$

dimana $K_h(x - X_i) = \frac{1}{h}K\left(\frac{x-X_i}{h}\right)$, maka

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{2.11}$$

dengan:

K = fungsi kernel

h = *bandwidth*

n = banyaknya data

X_i = variabel independen pada data ke-i

x = sampel acak

Dan cara menyesuaikan ketepatan estimator ini yaitu dengan mengubah h yang merupakan parameter pemulusan yang untuk mengontrol kemulusan dan kurva yang diestimasi.

Estimator kernel memiliki kemampuan yang baik dalam memodelkan data yang tidak mempunyai pola tertentu. Jika h terlalu kecil maka akan menghasilkan kurva yang sangat kasar dan sangat fluktuatif. Sedangkan, jika h terlalu besar maka akan menghasilkan kurva sangat mulus (Hardle, 1994).

2.3.1 Estimator Kernel Normal

Menurut Salha (2021), misalkan $X_1, X_2, \dots, X_n$ merupakan sampel acak dari distribusi dengan fungsi kepadatan peluang f yang tidak diketahui pada batas $[0, \infty)$, seperti f terdiferensiasi dua kali secara kontinu dan memenuhi

$$\int_0^\infty \left(xf'(x)\right)^2 dx < \infty, \int_0^\infty \left(x^2 f''(x)\right)^2 dx < \infty \,. \tag{2.12}$$

Kemudian, h adalah parameter pemulusan yang memenuhi $h \to 0$, dan $nh \to \infty$ ketika $n \to \infty$. Sehingga, fungsi kepadatan peluang pada estimator kernel normal diberikan sebagai berikut:

$$\hat{f}_N(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right) = \sum_{i=1}^{n} \phi(x, X_i, h) \tag{2.13}$$

dimana,

$$\phi(x, X_i, h) = \frac{1}{nh\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right).$$

Selanjutnya definisi fungsi distribusi kumulatif $F(x)$ pada estimator kernel normal jika $x > 0$ diberikan sebagai berikut:

$$\hat{F}_N(x) = \int_0^x \hat{f}_N(x)\, dx$$

$$\hat{F}_N(x) = \frac{1}{n}\sum_{i=1}^{n} \int_0^x \phi(x, X_i, h)\, dx$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_0^x \frac{1}{h\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right) dx \tag{2.14}$$

Dan misalkan $u = \frac{x-X_i}{h}$ dan $du = \frac{1}{h}dx \leftrightarrow dx = hdu$ maka,

$$\hat{F}_N(u) = \frac{1}{n}\sum_{i=1}^{n}\int_0^x \frac{1}{h\sqrt{2\pi}} exp\left(-\frac{1}{2}(u)^2\right)hdu$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^x \frac{1}{\sqrt{2\pi}} exp\left(-\frac{u^2}{2}\right)du$$

$$= \frac{1}{n}\sum_{i=1}^{n}(K[u]) \quad (2.15)$$

sehingga,

$$\hat{F}_N(x) = \frac{1}{n}\sum_{i=1}^{n}\left(K\left[\frac{x-X_i}{h}\right]\right). \quad (2.16)$$

2.3.2 Estimator Kernel Lognormal

Menurut Salha (2021), dengan kondisi yang sama pada estimator kernel normal, maka fungsi kepadatan peluang pada estimator kernel lognormal diberikan sebagai berikut:

$$\hat{f}_{LN}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{xh\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{\ln(x)-X_i}{h}\right)^2\right) = \sum_{i=1}^{n}\lambda(x, X_i, h) \quad (2.17)$$

dimana,

$$\lambda(x, X_i, h) = \frac{1}{nxh\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{\ln(x)-X_i}{h}\right)^2\right).$$

Selanjutnya definisi fungsi distribusi kumulatif $F(x)$ pada estimator kernel lognormal jika $x > 0$ diberikan sebagai berikut:

$$\hat{F}_{LN}(x) = \int_0^x \hat{f}_{LN}(x)\,dx$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^x \lambda(x, X_i, h)\,dx$$

$$\hat{F}_{LN}(x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^x \frac{1}{xh\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{\ln(x)-X_i}{h}\right)^2\right)dx \quad (2.18)$$

Dan misalkan $z = \frac{\ln(x)-X_i}{h}$ dan $dz = \frac{1}{xh}dx \leftrightarrow dx = xhdz$ maka,

$$\hat{F}_{LN}(z) = \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{x}\frac{1}{xh\sqrt{2\pi}}exp\left(-\frac{1}{2}(z)^2\right)xhdz$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{x}\frac{1}{\sqrt{2\pi}}exp\left(-\frac{z^2}{2}\right)dz$$

$$= \frac{1}{n}\sum_{i=1}^{n}(K[z]) \tag{2.19}$$

sehingga,

$$\hat{F}_{LN}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(K\left[\frac{\ln(x)-X_i}{h}\right]\right). \tag{2.20}$$

2.4 Pemilihan *Bandwidth* Optimum

Menurut Konecna (2018), estimator kernel memiliki suatu permasalahan yaitu dalam pemilihan *bandwidth* (h) atau parameter pemulusan. Hasil perkiraan akhir dipengaruhi ketika h bernilai secara signifikan. Akan menghasilkan kurva yang mulus ketika nilai h terlalu besar, tetapi nilai standar deviasi menjadi besar dan nilai keragaman menjadi rendah. Sehingga, diperlukannya pemilihan nilai h optimum.

Menurut Galub *et al.* (1979), dalam menentukan nilai h yang optimum, terdapat beberapa metode yang digunakan salah satunya adalah dengan menggunakan kriteria *Generalized Cross Validation* (GCV) yang didefinisikan sebagai berikut:

$$GCV(h) = \frac{n\sum_{i=1}^{n}(y_i-\hat{y}_i(h))^2}{(n-v_1)^2} \tag{2.21}$$

dengan:

n = banyaknya data

h = *bandwidth*

y_i = data sebenarnya

$\hat{y}_i(h)$ = nilai penaksir untuk pengamatan ke-i

v_1 = jumlah parameter yang efektif dalam model

2.5 MSE (*Mean Square Error*)

Menurut Kurniasih (2013), dalam menentukan pengujian mana yang paling mendekati kebenaran, yaitu dengan cara mengukur kesalahan. Metode yang digunakan dalam mengukur kesalahan biasanya menggunakan MSE. Pengujian yang terpilih atau terbaik ketika menghasilkan nilai MSE terkecil. MSE merupakan kuadrat dari rata-rata kesalahan yang dinyatakan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2.22)$$

dengan:

y_i = data sebenarnya

$\hat{y}_i$ = nilai prediksi dari variabel y_i

n = banyaknya data

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2021/2022 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini menggunakan data simulasi yang masing-masing sebanyak 200 data, yaitu data bangkitan dari:

1. Distribusi Eksponensial ($X \sim EXP(3)$)

$$f(x) = 3e^{-3x}$$

2. Distribusi Gamma ($X \sim GAMMA(1,3)$)

$$f(x) = \frac{1}{3\Gamma(1)} e^{-\frac{x}{3}}$$

Selanjutnya, menggunakan data asli untuk diterapkan ke fungsi *hazard rate* dengan estimator kernel normal dan lognormal menggunakan data kanker paru-paru dari percobaan yang dilakukan oleh North Central Cancer Treatment Group (NCCTG).

3.3 Metode Penelitian

Penelitian ini dilakukan dengan cara meninjau dan mempelajari referensi-referensi dari buku-buku, skripsi, dan jurnal. Penelitian ini dibantu dengan menggunakan perangkat lunak R Studio. Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagaiberikut:

1. Membangkitkan data simulasi distribusi eksponensial ($X \sim EXP(3)$) dan distribusi gamma ($X \sim GAMMA(1,3)$) dengan ukuran masing-masing n=200.
2. Menentukan waktu secara kontinu di selang waktu tertentu.
3. Membentuk fungsi *hazard rate* menggunakan estimator kernel
   a. Menentukan fungsi kepadatan peluang ($\hat{f}_N$) dan ($\hat{f}_{LN}$).
   b. Menentukan fungsi distribusi kumulatif ($\hat{F}_N$) dan ($\hat{F}_{LN}$).
   c. Menentukan fungsi kelangsungan hidup ($\hat{S}_N$) dan ($\hat{S}_{LN}$).
   d. Membentuk fungsi kernel pada fungsi *hazard rate* ($\hat{h}_N$) dan ($\hat{h}_{LN}$).
4. Mencari besar *bandwidth* (h) optimum dengan metode GCV(*Generalized Cross Validation*) pada masing-masing data simulasi.
5. Mencari fungsi *hazard rate* menggunakan estimator kernel normal dan lognormal untuk masing-masing data simulasi.
6. Membandingkan kinerja pada estimator kernel normal dan lognormal berdasarkan:
   a. Membentuk grafik berupa gabungan dari fungsi *hazard* asli, fungsi *hazard* menggunakan estimator kernel normal dan fungsi *hazard* menggunakan estimator kernel lognormal pada masing-masing data simulasi.
   b. Membentuk grafik berupa gabungan dari fungsi densitas asli, fungsi densitas menggunakan estimator kernel normal dan fungsi densitas menggunakan estimator kernel lognormal pada masing-masing data simulasi.
7. Mencari dan membandingkan nilai MSE dari fungsi *hazard rate* menggunakan estimator kernel normal dan lognormal pada masing-masing data simulasi.
8. Menerapkan fungsi *hazard rate* menggunakan estimator kernel normal dan lognormal pada data asli sesuai data simulasi yang digunakan.
9. Membuat kesimpulan.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan pada penelitian ini, maka diperoleh kesimpulan bahwa setelah dilakukan penelitian menggunakan data simulasi dan diterapkan ke dalam data asli, yaitu metode estimator kernel lognormal merupakan metode yang lebih baik dalam menentukan fungsi *hazard rate* dibandingkan menggunakan estimator kernel normal dalam data yang dianalisis pada penelitian ini. Kemudian fungsi *hazard rate* dengan estimator kernel lognormal pada data simulasi diperoleh nilai MSE sebesar 0,573 untuk distribusi eksponensial dan sebesar 0,421 untuk distribusi gamma. Karena nilai MSE pada distribusi gamma lebih kecil dibandingkan distribusi eksponensial, maka menggunakan data asli yang mengikuti distribusi gamma yaitu data kanker paru-paru dan terbukti bahwa dengan menggunakan estimator kernel lognormal menghasilkan grafik yang lebih mulus dibandingkan estimator kernel normal.

DAFTAR PUSTAKA

Amran, F. & Farik, A. 2015. Model Survival Nonparametrik pada Data Rawat Inap Pasien Diare di Puskesmas Indralaya. *Jurnal Matematika*. **5**(2): 105-116.

Danardono. 2012. *Analisis Data Survival*. Diktat Kuliah Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gajah Mada, Yogyakarta.

Eubank, R.L. 1988. *Spline Smoothing and Nonparametric Regression*. Mercel Dekker, New York.

Galub, G.H., Heath, M., & Wahba, G. 1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*. **21**(2): 215-223.

Hardle, W. 1994. *Applied Nonparametric Regression.* Cambridge University Press, New York.

Herawati, N & Nisa, K. 2009. Testing Normality and Bandwidth Estimation Using Kernel Method for Small Sample Size. *Jurnal Ilmu Dasar*. **10**(1): 62-67.

Jakperik, D. & Ozoje, M.O. 2012. Survival Analysis of Average Recovery Time of Tuberculosis Patient in Northen Region Ghana. *International Jurnal of Current Research*. **4**(9): 123-125.

Kleibaum, D.G. & Klein, M. 2005. *Survival Analysis a Self-Learning Text.* Second Edition. Springer, New York.

Klein, J.P. & Moeschberger, M.L. 2003. *Survival Analysis Techniques for Censored and Truncated Data*. Second Edition. Springer, New York.

Komang, G. & Gusti, A. 2012. Estimator Kernel dalam Model Regresi Nonparametrik. *Jurnal Matematika*. **2**(1): 19-30.

Konecna, K. 2018. The Priestley-Chao Estimator of Conditional Density with Uniformly Distributed Random Design. *Statistika*. **98**(3): 283-294.

Kurniasih, D. 2013. Efisiensi Relatif Estimator Fungsi Kernel Gaussian Terhadap Estimator Polinomial dalam Peramalan USD Terhadap JPY (Skripsi). Jurusan Matematika FMIPA UNS, Semarang.

Lee, E.T. & Wang, J.W. 2003. Third Edition. *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, Canada.

Liu, X. 2012. *Survival Analysis: Models and Applications*. First Edition. John Wiley & Sons, USA.

Shao, J. 2003. *Mathematical Statistics*. Second Edition. Springer, USA.

Ogden, R.T. 1997. *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhauser, Boston.

Pratama, I.M.Y. 2020. Perbandingan Metode Kemungkinan Maksimum dan Metode Momen dalam Pendugaan Parameter Distribusi Pareto dengan Dua Parameter (Skripsi). Jurusan Matematika Fakultas Sains dan Teknologi Universitas Sanata Dharma, Yogyakarta.

Quraisy, A. 2020. Normalitas Data Menggunakan Uji Kolmogorov-Smirnov dan Saphiro-Wilk. *Journal of Health, Education, Economics, Science and Techonology*. **3**(1): 7-11.

Saidi, S., Herawati, N., Nisa, K., & Setiawan, E. 2021. Nonparametric Modeling Using Kernel Method for the Estimation of the Covid-19 Data in Indonesia During 2020. *International Journal of Mathematics Trends and Technology*. **67**(6): 136-144.

Salha, R.B. 2021. Lognormal Kernel Estimator of The Hazard Rate Function. *Communications in Statistics-Simulation and Computation*. **47**(10): 1:17.

Salim, P. & Salim, Y. 2002. *Kamus Bahasa Indonesia Kontemporer*. Modern English Press, Jakarta.

Sukarsa, I.K.G. & Srinadi, I.G.A.M. 2012. Estimator Kernel dalam Model Regresi Nonparametrik. *Jurnal Matematika*. **2**: 1693-1394.

Susandri. 2016. Analisis Non Parametrik dalam Menentukan Faktor-Faktor yang Mempengaruhi Mahasiswa Tidak Melanjutkan Studi pada Semester Pertama. *Jurnal Ilmiah Media Processor*. **11**(2): 871-879.

Utami, N.F. 2017. Analisis Survival dengan Pemodelan Regresi Cox Proporsional Hazard (Skripsi). Jurusan Statistika FMIPA UNM, Makassar.