

**TESIS**

**MODEL REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DENGAN  
METODE *GENERALIZED CROSS VALIDATION*, *UNBIASED RISK*, DAN  
*GENERALIZED MAXIMUM LIKELIHOOD* PADA FAKTOR  
YANG MEMPENGARUHI JUMLAH KEMATIAN BAYI  
DI KALIMANTAN TIMUR TAHUN 2018**



**E. GINA MUSTIKAWATI**

**UNIVERSITAS AIRLANGGA  
FAKULTAS KESEHATAN MASYARAKAT  
PROGRAM MAGISTER  
PROGRAM STUDI KESEHATAN MASYARAKAT  
SURABAYA  
2020**

**TESIS**

**MODEL REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DENGAN  
METODE *GENERALIZED CROSS VALIDATION*, *UNBIASED RISK*, DAN  
*GENERALIZED MAXIMUM LIKELIHOOD* PADA FAKTOR  
YANG MEMPENGARUHI JUMLAH KEMATIAN BAYI  
DI KALIMANTAN TIMUR TAHUN 2018**



**E. GINA MUSTIKAWATI  
NIM 101814153018**

**UNIVERSITAS AIRLANGGA  
FAKULTAS KESEHATAN MASYARAKAT  
PROGRAM MAGISTER  
PROGRAM STUDI KESEHATAN MASYARAKAT  
SURABAYA  
2020**

**MODEL REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DENGAN  
METODE *GENERALIZED CROSS VALIDATION*, *UNBIASED RISK*, DAN  
*GENERALIZED MAXIMUM LIKELIHOOD* PADA FAKTOR  
YANG MEMPENGARUHI JUMLAH KEMATIAN BAYI  
DI KALIMANTAN TIMUR TAHUN 2018**

**TESIS**

**Untuk memperoleh gelar Magister Kesehatan (M.Kes)  
Minat Studi Biostatistik  
Program Studi Kesehatan Masyarakat  
Fakultas Kesehatan Masyarakat  
Universitas Airlangga**

**Oleh :**

**E. GINA MUSTIKAWATI  
NIM 101814153018**

**UNIVERSITAS AIRLANGGA  
FAKULTAS KESEHATAN MASYARAKAT  
PROGRAM MAGISTER  
PROGRAM STUDI KESEHATAN MASYARAKAT  
SURABAYA  
2020**

**PENGESAHAN**

**Dipertahankan di depan Tim Penguji Tesis  
Minat Studi Biostatistik  
Program Studi Kesehatan Masyarakat  
Fakultas Kesehatan Masyarakat Universitas Airlangga  
dan diterima untuk memenuhi persyaratan guna memperoleh gelar  
Magister Kesehatan (M.Kes)  
Pada Tanggal 18 November 2020**

**Mengesahkan**

**Universitas Airlangga  
Fakultas Kesehatan Masyarakat**

**Dekan,**



**Dr. Santi Martini, dr., M.Kes  
NIP 196609271997022001**

**Tim Penguji**

**Ketua : 1. Prof. Kuntoro, dr., M.PH., Dr.PH**  
**Anggota : 2. Dr. Arief Wibowo, dr., M.S**  
**3. Dr. Mahmudah, Ir., M.Kes**  
**4. Dr. M. Atoillah Isfandiari, dr., M.Kes**  
**5. Dr. Pipit Festi Wiliyanarti, S.KM., M.Kes**

**PERSETUJUAN**

**TESIS**

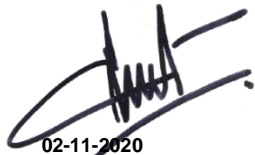
**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Magister Kesehatan (M.Kes)  
Minat Studi Biostatistik  
Program Studi Kesehatan Masyarakat  
Fakultas Kesehatan Masyarakat  
Universitas Airlangga**

Oleh :

**E. GINA MUSTIKAWATI  
NIM 101814153018**

**Menyetujui,  
Surabaya, 18 November 2020**

**Pembimbing Ketua,**



**Dr. Arief Wibowo, dr., MS  
NIP 195903101986011001**

**Pembimbing,**



**Dr. Mahmudah, Ir, M.Kes  
NIP 196901101993032002**

**Mengetahui,  
Koordinator Program Studi Kesehatan Masyarakat**



**Dr. M. Bagus Qomaruddin, Drs., M.Sc  
NIP 196502161990021001**

### PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini, saya :

Nama : E. Gina Mustikawati  
NIM : 101814153018  
Program Studi : Kesehatan Masyarakat  
Minat Studi : Biostatistik  
Angkatan : 2018/2019  
Jenjang : Magister

Menyatakan bahwa saya tidak melakukan kegiatan plagiat dalam penulisan tesis saya yang berjudul:

**MODEL REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DENGAN METODE *GENERALIZED CROSS VALIDATION*, *UNBIASED RISK*, DAN *GENERALIZED MAXIMUM LIKELIHOOD* PADA FAKTOR YANG MEMPENGARUHI JUMLAH KEMATIAN BAYI DI KALIMANTAN TIMUR TAHUN 2018.**

Apabila suatu saat nanti terbukti saya meakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 18 November 2020

METERAI  
TEMPEL  
6000  
E. Gina Mustikawati

A 6000 Rupiah postage stamp is affixed to the document. The stamp features the Garuda Pancasila emblem and the text 'METERAI TEMPEL' and '6000 RUPIAH'. A handwritten signature in black ink is written over the stamp. Below the stamp, the name 'E. Gina Mustikawati' is printed.

## KATA PENGANTAR

Puji syukur kehadirat ALLH SWT atas Karunia dan Hidayah-nya penyusunan tesis dengan judul “**Model Regresi Nonparametrik *Spline Truncated* dengan Metode *Generalized Cross Validation, Unbiased Risk, dan Generalized Maximum Likelihood* pada Faktor yang Memengaruhi Jumlah Kematian Bayi di Kalimantan Timur Tahun 2018**” ini dapat terselesaikan.

Tesis ini berisikan tentang model regresi nonparametrik spline truncated dan faktor yang mempengaruhi kematian bayi di Kalimantan Timur Tahun 2018. Hasil temuan diharapkan dapat membantu para pengambil kebijakan khususnya terkait program-program kesehatan dalam penurunan Angka Kematian Bayi (AKB) di provinsi Kalimantan Timur.

Penghargaan setinggi-tingginya penulis sampaikan kepada Dr. Arief Wibowo, dr., MS selaku pembimbing ketua yang dengan kesabaran dan perhatiannya dalam memberikan bimbingan, semangat dan saran hingga tesis ini bisa terselesaikan dengan baik. Ucapan terima kasih yang tak terhingga juga saya sampaikan kepada Dr. Mahmudah, Ir, M.Kes selaku pembimbing kedua yang telah banyak meuangkan waktu untuk memberikan bimbingan, motivasi dan saran demi kesempurnaan tesis ini.

Terwujudnya tesis ini adalah berkat dukungan dari berbagai pihak, oleh karena itu rasa terima kasih yang sebesar-besarnya disampaikan kepada yang terhormat:

1. Prof. Dr. Mohammad Nasih, SE., Mt., Ak., CMA selaku Rektor Universitas Airlangga
2. Dr. Santi Martini, dr.,M.Kes selaku Dekan Fakultas Kesehatan Masyarakat Universitas Airlangga
3. Dr. M. Bagus Qomaruddin, Drs., M.Sc selaku Ketua Program Studi Program Magister Kesehatan Masyarakat Fakultas Kesehatan Masyarakat Universitas Airlangga
4. Dr. Arief Wibowo, dr., MS selaku Ketua Minat Studi Biostatistik Fakultas Kesehatan Masyarakat Universitas Airlangga
5. Ketua Penguji dan anggota penguji atas kesediaan waktu dan bimbingan dalam perbaikan tesis ini, yang meliputi:
  - a. Dr. M. Atoillah Isfandiari, dr.,M.Kes selaku ketua penguji
  - b. Dr. Arief Wibowo, dr.,M.S selaku anggota penguji
  - c. Dr. Mahmudah, Ir.,M.Kes selaku anggota penguji
  - d. Prof. Kuntoro, dr.,M.PH.,Dr.PH selaku anggota penguji
  - e. Dr. Pipit Festi Wiliyanarti,S.KM.,M.Kes selaku anggota penguji
6. Dinas Kesehatan Provinsi Kalimantan Timur atas ketersediaan data yang digunakan dalam penelitian.
7. Para Dosen Program Studi Ilmu Kesehatan Masyarakat, secara khusus Departemen Biostatistik yang telah memberikan bimbingan, petunjuk dan dorongan kepada penulis selama menempuh proses pendidikan.

8. H. E. Achmad Fariyal, B.Sc dan Hj. Herawati, S.Pd atas dukungan doa, moril dan materil dalam mendorong penulis menyelesaikan berbagai jenjang pendidikan.
9. E. Novita Fahriza, SH dan E. Siska Handayani, SKM, yang selalu memberikan semangat dan doa.
10. Rekan-rekan mahasiswa Magister Ilmu Kesehatan Masyarakat angkatan 2018/2019, khususnya minat Biostatistika.
11. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan dalam bentuk apapun kepada penulis selama proses pendidikan.

Kritik dan saran yang bersifat konstruktif dari para pembaca sangat diharapkan untuk perbaikan tesis ini, sehingga bisa diperoleh hasil yang berkualitas. Akhir kata semoga tesis ini bermanfaat untuk semua pihak.

Surabaya, 18 November 2020

Penulis



*SUMMARY*

**SPLINE TRUNCATED NONPARAMETRIC REGRESSION MODEL  
USING GENERALIZED CROSS VALIDATION, UNBIASED RISK, AND  
GENERALIZED MAXIMUM LIKELIHOOD METHODS ON FACTORS  
THAT INFLUENCES THE NUMBER OF INFANT MORTALITY  
EAST KALIMANTAN AT 2018**

Regression analysis is a data analysis method that describes the relationship between response and predictor variables (Hosmer and Lemeshow, 2000). The form of the relationship pattern between predictor and the response variables can be identified based on past information or by using a scatter plot (Hardle, 1990). If there is no information about the shape of the regression curve whether it is linear, quadratic or otherwise, then the approach used is nonparametric regression. Among these approaches, nonparametric regression with the spline approach is the method that is often used.

Spline Truncated is a function where change in the behavior of different curves at different intervals. There are various advantages using a spline, including: having a good visual interpretation, being flexible, and being able to handle smooth functional characters. In addition, splines can handle data patterns and can describe changes in data behavior that fluctuate in certain sub-intervals. Spline also has advantage of overcoming data patterns that show sharp rises / falls with the help of knot points and the resulting curve is relatively smooth (Budiantara, 2009).

There are several methods for selecting the optimal knot points in the nonparametric spline regression, are Unbiased Risk (UBR), Generalized Cross Validation (GCV), and Generalized Maximum Likelihood (GML) method.

One of the health problems that is an indicator of health development is the Infant Mortality Rate (IMR) which is still high nationally. The number of infant deaths in East Kalimantan Province has fluctuated from 2014 to 2018. Data conditions with an unknown regression curve can be found in the data on infant mortality. Like previous studies which state that data on infant mortality and the variables that influence it do not form a certain pattern so that it uses a nonparemetric spline regression approach to model infant mortality data (Megasari and Budiantara, 2013) (Pratiwi, 2017). Therefore, research on comparison of the Spline Truncated nonparametric regression model using the Generalized Cross Validation (GCV), Unbiassed Risk (UBR) and Generalized Maximum Likelihood (GML) method on factors that affect the number of infant mortality in East Kalimantan Province at 2018.

This is a non-reactive study using secondary data. The population of study included 187 health centers in East Kalimantan Province, while the sample was the total population. dependent variable of this research is the number of infant deaths. As independent variables are several factors that affect the number of infant deaths are athe percentage of visits from pregnant women K4, thepercentage of delivery assistance by health personnel, percentage of neonatal complications handled, percentage of infant health services, percentage of

complete basic immunization and percentage of low birth weight (LBW). The initial analysis of the spline nonparametric regression is to create a scatter plot and normality test to find out whether it can be continued with nonparametric analysis. Then proceed with determining the optimal knot point at 1, 2 and 3 knot points using the GCV, UBR and GML methods. After obtaining the optimal knot point from each method, it is continued with the determination of the best model selected based on the criteria of the smallest score of MSE and the largest score of  $R^2_{adjusted}$ .

The best spline nonparametric regression model is obtained using the GML method at 1 knot point, with MSE score is 12.41 and  $e R^2_{adj}$  score is 0.657. Compared with the GCV method at 3 knots points with MSE score is 12.45 and  $R^2_{adj}$  score is 0.589 and the UBR method at 2 knots points with MSE score is 12.60 and  $R^2_{adj}$  score is 0.592.

$$\begin{aligned} \text{Infant mortality} = & 0,0027 - 0,031(\text{persentase K4}) + 0,068(\text{persentase K4} - 115,9)_+ \\ & - 0,003(\text{persentase persalinan oleh nakes}) - \\ & 0,003(\text{persentase persalinan oleh nakes} - 118,9)_+ \\ & - 0,009(\text{persentase komplikasi neonatal}) + \\ & 0,134(\text{persentase komplikasi neonatal} - 335,71)_+ \\ & - 0,058(\text{persentase yankes bayi}) \\ & - 0,008(\text{persentase yankes bayi} - 327,08)_+ \\ & - 0,003(\text{persentase imunisasi dasar lengkap}) - \\ & 0,003(\text{persentase imunisasi dasar lengkap} - 277,62)_+ \\ & - 0,009\text{persentase BBLR} + 0,134(\text{persentase BBLR} - 14,48)_+ \end{aligned}$$

The spline truncated nonparametric regression analysis using the GML method produced the best model compared to the GCV and UBR methods on factors that affect the number of infant mortality in East Kalimantan Province in 2018. The factors that affect the number of infant deaths include the percentage of visits from pregnant women K4, the percentage of delivery assistance by health personnel, percentage of neonatal complications handled, percentage of infant health services, percentage of complete basic immunization and percentage of low birth weight (LBW).

## RINGKASAN

**MODEL REGRESI NONPARAMETRIK *SPLINE TRUNCATED* DENGAN  
METODE *GENERALIZED CROSS VALIDATION*, *UNBIASED RISK*, DAN  
*GENERALIZED MAXIMUM LIKELIHOOD* PADA FAKTOR  
YANG MEMPENGARUHI JUMLAH KEMATIAN BAYI  
DI KALIMANTAN TIMUR TAHUN 2018**

Analisis regresi merupakan metode analisis data yang menggambarkan hubungan antara variabel respons dan variabel prediktor (Hosmer dan Lemeshow, 2000). Bentuk pola hubungan antara variabel prediktor dengan variabel respon dapat diidentifikasi berdasarkan informasi masa lalu atau dengan menggunakan *scatter plot* (Hardle, 1990). Apabila tidak ada informasi apapun mengenai bentuk kurva regresi apakah berbentuk linier, kuadratik atau yang lainnya, maka pendekatan yang digunakan adalah regresi nonparametrik. Diantara metode pendekatan tersebut, regresi nonparametrik dengan pendekatan *spline* merupakan metode yang sering digunakan.

*Spline Truncated* merupakan fungsi dimana terdapat perubahan pola perilaku kurva yang berbeda pada interval-interval yang berlainan. Berbagai macam kelebihan jika menggunakan *spline* antara lain: mempunyai interpretasi visual yang baik, bersifat fleksibel, serta mampu menangani karakter fungsi yang bersifat mulus. Selain itu, *spline* dapat mengatasi pola data dan dapat menggambarkan perubahan perilaku data yang berubah-ubah pada sub-subinterval tertentu. *Spline* juga mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik/turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan relatif mulus (Budiantara, 2009).

Terdapat beberapa metode untuk memilih titik knot optimal dalam regresi nonparametrik *Spline* antara lain metode *Unbiased Risk* (UBR), *Generalized Cross Validation* (GCV), dan *Generalized Maximum Likelihood* (GML).

Salah satu permasalahan kesehatan yang menjadi indikator pembangunan kesehatan adalah Angka Kematian Bayi (AKB) yang masih tinggi secara nasional. Jumlah kematian bayi di Provinsi Kalimantan Timur mengalami fluktuasi dari tahun 2014 – 2018. Kondisi data dengan bentuk kurva regresi yang tidak diketahui dapat ditemukan pada data kematian bayi. Seperti penelitian-penelitian terdahulu yang menyebutkan bahwa data kematian bayi dan variabel yang mempengaruhinya tidak membentuk pola tertentu sehingga menggunakan pendekatan regresi nonparametrik *spline* untuk memodelkan data kematian bayi (Megasari dan Budiantara, 2013) (Pratiwi, 2017). Oleh karena itu perlu dilakukan penelitian tentang perbandingan model regresi nonparametrik *Spline Truncated* dengan menggunakan metode *Generalized Cross Validation* (GCV), *Unbiased Risk* (UBR) dan *Generalized Maximum Likelihood* (GML) pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018.

Penelitian ini adalah penelitian *non reaktif* dengan menggunakan data sekunder. Populasi penelitian mencakup 187 puskesmas di Provinsi Kalimantan

Timur, sedangkan sampel adalah keseluruhan jumlah populasi. Variabel dependen penelitian adalah jumlah kematian bayi. Sebagai variabel independen adalah beberapa faktor yang mempengaruhi jumlah kematian bayi meliputi persentase kunjungan ibu hamil K4, persentase pertolongan persalinan oleh tenaga kesehatan, persentase komplikasi neonatal yang ditangani, persentase pelayanan kesehatan bayi, Persentase imunisasi dasar lengkap dan persentase berat bayi lahir rendah (BBLR). Analisis awal regresi nonparametrik *spline* adalah dengan membuat *scatter plot* dan uji normalitas untuk mengetahui apakah dapat dilanjutkan dengan analisis nonparametrik. Kemudian dilanjutkan dengan penentuan titik knot optimal pada 1, 2 dan 3 titik knot dengan menggunakan metode GCV, UBR dan GML. Setelah diperoleh titik knot optimal dari masing masing metode, dilanjutkan dengan penentuan model terbaik yang dipilih berdasarkan kriteria nilai MSE terkecil dan  $R^2_{adjusted}$  terbesar.

Hasil penelitian menunjukkan bahwa model regresi nonparametrik *spline* terbaik diperoleh dengan menggunakan metode GML pada 1 titik knot, dengan nilai MSE yaitu 12,41 dan nilai  $R^2_{adj}$  yaitu 0,657. Dibandingkan dengan metode GCV pada 3 titik knot dengan nilai MSE sebesar 12,45 dan nilai  $R^2_{adj}$  sebesar 0,589 serta metode UBR pada 2 titik knot dengan nilai MSE sebesar 12,60 dan nilai  $R^2_{adj}$  sebesar 0,592.

$$\begin{aligned} \text{Kematian Bayi} = & 0,0027 - 0,031(\text{persentase K4}) + 0,068(\text{persentase K4} - 115,9)_+ \\ & - 0,003(\text{persentase persalinan oleh nakes}) - \\ & 0,003(\text{persentase persalinan oleh nakes} - 118,9)_+ \\ & - 0,009(\text{persentase komplikasi neonatal}) + \\ & 0,134(\text{persentase komplikasi neonatal} - 335,71)_+ \\ & - 0,058(\text{persentase yankes bayi}) \\ & - 0,008(\text{persentase yankes bayi} - 327,08)_+ \\ & - 0,003(\text{persentase imunisasi dasar lengkap}) - \\ & 0,003(\text{persentase imunisasi dasar lengkap} - 277,62)_+ \\ & - 0,009\text{persentase BBLR} + 0,134(\text{persentase BBLR} - 14,48)_+ \end{aligned}$$

Analisis regresi nonparametrik *spline truncated* dengan menggunakan metode GML menghasilkan model terbaik dibandingkan dengan metode GCV dan UBR pada faktor yang mempengaruhi jumlah kematian bayi di Provinsi Kalimantan Timur tahun 2018. Adapun faktor yang mempengaruhi jumlah kematian bayi meliputi persentase kunjungan ibu hamil K4, persentase pertolongan persalinan oleh tenaga kesehatan, persentase komplikasi neonatal yang ditangani, persentase pelayanan kesehatan bayi, Persentase imunisasi dasar lengkap dan persentase berat bayi lahir rendah (BBLR).