

PEMODELAN KEMATIAN BAYI MENGGUNAKAN COX
PROPORTIONAL HAZARD DENGAN *MULTIVARIATE*
ADAPTIVE REGRESSION SPLINE

(Hasil Survei Penduduk Antar Sensus 2015)

Skripsi

Disusun untuk melengkapi syarat-syarat
guna memperoleh gelar Sarjana Sains



TIAS DWI NOVIANISA

3125130795

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA










2017

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

PEMODELAN KEMATIAN BAYI MENGGUNAKAN COX PROPORTIONAL HAZARD WITH MULTIVARIATE ADAPTIVE REGRESSION SPLINE (Hasil Survei Penduduk Antar Sensus 2015)

Nama : Tias Dwi Novianisa

No. Registrasi : 3125130795

| | Nama | Tanda Tangan | Tanggal |
|------------------------|---|---|------------|
| Penanggung Jawab | |  | |
| Dekan | : Prof. Dr. Suyono, M.Si. NIP. 19671218 199303 1 005 |  | 16-08-2017 |
| Wakil Penanggung Jawab | |  | |
| Wakil Dekan I | : Dr. Muktiningsih, M.Si. NIP. 19640511 198903 2 001 |  | 16-08-2017 |
| Ketua | : Dr. Lukita Ambarwati, S.Pd, M.Si. NIP. 19721026 200112 2 001 |  | 14-08-2017 |
| Sekretaris | : Eti Dwi Wiraningsih, S.Pd., M.Si 19810203 200604 2 001 |  | 14-08-2017 |
| Penguji | : Ir. Fariani Hermin, MT NIP. 19600211 198703 2 001 |  | 14-08-2017 |
| Pembimbing I | : Dra. Widyanti Rahayu, M.Si NIP. 19661103 200112 2 001 |  | 14-08-2017 |
| Pembimbing II | : Ratna Widyati, S.Si., M.Kom 19750925 200212 2 002 |  | 14-08-2017 |

Dinyatakan lulus ujian skripsi tanggal: 10 Agustus 2017

ABSTRACT

Tias Dwi Novianisa, 3125130795. *Modelling of Infant Mortality Using Cox Proportional Hazard With Multivariate Adaptive Regression Spline (The Result of Intercensal Population Survey 2015). Thesis. Faculty of Mathematics and Natural Science ,Jakarta State University. 2017.*

The high infant mortality rate (IMR) in Indonesia still has not reached Millennium Development Goals (MDGs) target in 2015. Special handling is needed to reduce infant mortality in Indonesia, among others by looking at the causal factors that are thought to affect the infant mortality rate. This research using Cox Proportional Hazard using Multivariate Adaptive Regression Spline to analyze the significant factors that affect the infant mortality rate based on live birth data in the period January 2013 - December 2014 in Banten, by using residual textit martingale on Cox PH as a response variable on MARS modeling. The causal factors that are thought to affect the infant mortality rate involve maternal age at during childbirth (X_1), mother's parity (X_2), distance between births (X_3), mother's education (X_4), baby's gender (X_5), twin births (X_6), birth helpers (X_7), and the main source of drinking water (X_8). The results of this study indicate that the best MARS model obtained is a combination of $BF=32$, $MI=3$, and $MO=1$, with GCV value is 0.00053. Based on the best MARS model is known factors that affect the infant mortality rate are maternal age at during childbirth (X_1), mother's parity (X_2), distance between births (X_3), mother's education (X_4), baby's gender (X_5), and the main source of drinking water (X_8) with the importance of each variable is 65.28%, 88.48%, 56.00%, 48.01%, 58.67%, dan 100%.

Keywords : *infant mortality, Cox Proportional Hazrad, Multivariate Adaptive Regression Spline (MARS), Generalized Cross Validation (GCV)*

ABSTRAK

TIAS DWI NOVIANISA, 3125130795. **Pemodelan Kematian Bayi Menggunakan Cox *Proportional Hazard* dengan *Multivariate Adaptive Regression Spline* (Hasil Survei Penduduk Antar Sensus 2015). Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2017.**

Tingginya angka kematian bayi (AKB) di Indonesia masih belum mencapai target *Millenium Development Goals* (MDGs) tahun 2015. Diperlukan penanganan khusus untuk menurunkan angka kematian bayi di Indonesia dengan melihat faktor-faktor penyebab yang diduga berpengaruh terhadap laju kematian bayi. Penelitian ini menggunakan pemodelan Cox *Proportional Hazard* dengan *Multivariate Adaptive Regression Spline* (MARS) untuk menganalisis faktor-faktor signifikan yang mempengaruhi laju kematian bayi berdasarkan data bayi lahir hidup pada periode Januari 2013 - Desember 2014 di Provinsi Banten, dengan menggunakan residual *martingale* pada Cox PH sebagai variabel respon pada pemodelan MARS. Faktor-faktor penyebab yang diduga berpengaruh terhadap laju kematian bayi meliputi usia ibu saat melahirkan (x_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pendidikan ibu (X_4), jenis kelamin bayi (X_5), kelahiran kembar (X_6), penolong kelahiran bayi (X_7), dan sumber utama air minum (X_8). Hasil dari penelitian ini menunjukkan bahwa model MARS terbaik yang diperoleh adalah kombinasi dari BF=32, MI=3, dan MO=1, dengan nilai GCV sebesar 0,00053. Berdasarkan model MARS terbaik tersebut dapat diketahui faktor-faktor yang berpengaruh terhadap laju kematian bayi adalah usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pendidikan ibu (X_4), jenis kelamin bayi (X_5), dan sumber utama air minum (X_8) dengan tingkat kepentingan masing-masing variabel sebesar 65,28%, 88,48%, 56,00%, 48,01%, 58,67%, dan 100%.

Kata kunci : Kematian Bayi, Cox *Proportional Hazard*, *Multivariate Adaptive Regression Spline* (MARS), *Generalized Cross Validation* (GCV).

PERSEMBAHANKU...

” Bangunlah, lalu berilah peringatan. Dan agungkanlah Tuhanmu, dan bersihkanlah pakaianmu, dan tinggalkanlah segala (perbuatan) yang keji, dan janganlah engkau memberi (dengan maksud) memperoleh (balasan) yang lebih banyak. (Al Muddassir: 2-6)”

” Everyone has their own story ”

-Hayley Williams

Skripsi ini kupersembahkan untuk Ayah, Mama, Abang, Alm. Baba, saudara-saudara dan seluruh rekan-rekan yang membantu penulis.

”Terima kasih atas dukungan, do’a, serta kasih sayang kalian”.

KATA PENGANTAR

Puji syukur kepada Tuhan YME atas pengetahuan dan kemampuan sehingga penulis dapat menyelesaikan skripsi yang berjudul "Kematian Bayi dengan Pemodelan *Cox Proportional Hazard* Berbasis *Multivariate Adaptive Regression Spline* (Hasil Survei Penduduk Antar Sensus 2015)" yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Program Studi Matematika Universitas Negeri Jakarta.

Skripsi ini berhasil diselesaikan tidak terlepas dari adanya bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih terutama kepada:

1. Ayah dan Mama yang selalu mendukung, memberi motivasi, setia dan selalu memberikan semangat ketika penulis mulai jenuh.
2. Ibu Dra. Widyanti Rahayu, M.Si selaku Dosen Pembimbing I dan Ibu Ratna Widyati, S.Si., M.Kom selaku Dosen Pembimbing II, yang telah meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta arahan sehingga skripsi ini dapat menjadi lebih baik dan terarah.
3. Ibu Dr. Lukita Ambarwati, S.Pd., M.Si., selaku Koordinator Prodi Matematika FMIPA UNJ yang telah banyak membantu penulis.
4. Ibu Dian Handayani, M.Si., selaku Pembimbing Akademik atas segala bimbingan dan kerja sama Ibu selama perkuliahan, dan seluruh Bapak/Ibu dosen atas pengajarannya yang telah diberikan, serta karyawan/karyawati FMIPA UNJ yang telah memberikan informasi yang penulis butuhkan dalam menyelesaikan skripsi.
5. Abang kandung dan kakak ipar penulis yang terus memberi semangat,

mendoakan penulis, dan selalu menghibur ketika penulis mengalami kesulitan dalam penulisan skripsi ini.

6. Sahabat penulis sejak kecil, Fira dan Santi yang selalu ada berbagi cerita dikala senang maupun susah, serta selalu memberi semangat dan perhatian kepada penulis.
7. Teman-teman MUTIARA-ku, Monicha, Sarah, Indah, dan Andhiny yang sudah memberi kepercayaan dan motivasi agar penulis tidak menyerah dalam menyusun tugas akhir ini.
8. Teman-teman DM, Hanun, Defy, Ezania, dan Rania, teman-teman Anak Malem serta teman-teman Apalah-apalah untuk semua kesenangan dan kegilaan yang diberikan selama ini.
9. Teman-teman seperjuangan penulis, Hanun, Defy, Iren, Ita, Nurul, Umam, dan Daniel, terima kasih karena sudah saling menyemangati agar bisa lulus bareng di semester ini.
10. Kak Mega, Kak Yuka, Kak Dian, dan kakak-kakak lainnya yang telah memberi motivasi dan membantu penulis, serta yang lainnya.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Masukan dan kritikan akan sangat berarti. Semoga skripsi ini dapat bermanfaat bagi pembaca sekalian.

Jakarta, Agustus 2017

Tias Dwi Novianisa

DAFTAR ISI

| | |
|--|----------|
| ABSTRACT | i |
| ABSTRAK | ii |
| KATA PENGANTAR | iv |
| DAFTAR ISI | vi |
| DAFTAR TABEL | ix |
| DAFTAR GAMBAR | x |
| I PENDAHULUAN | 1 |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Perumusan Masalah | 3 |
| 1.3 Pembatasan Masalah | 4 |
| 1.4 Tujuan Penulisan | 4 |
| 1.5 Manfaat Penulisan | 5 |
| 1.6 Metode Penelitian | 5 |
| II LANDASAN TEORI | 6 |
| 2.1 Kematian Bayi | 6 |
| 2.2 Analisis Survival | 9 |
| 2.2.1 Data Tersensor | 10 |
| 2.2.2 Fungsi Survival dan Fungsi <i>Hazard</i> | 13 |
| 2.3 <i>Cox Proportional Hazard</i> (Cox PH) | 15 |
| 2.4 Asumsi <i>Proportional Hazard</i> | 17 |
| 2.5 Metode Kaplan Meier | 18 |

| | | |
|-----------------------|---|-----------|
| 2.6 | Residual <i>Martingale</i> | 19 |
| 2.7 | Analisis Regresi | 20 |
| 2.8 | Regresi Non-parametrik | 21 |
| 2.8.1 | Regresi Spline | 22 |
| 2.8.2 | <i>Recursive Partitioning Regression</i> (RPR) | 26 |
| 2.9 | <i>Multivariate Adaptive Regression Spline</i> (MARS) | 28 |
| III PEMBAHASAN | | 31 |
| 3.1 | <i>Multivariate Adaptive Regression Spline</i> (MARS) | 31 |
| 3.1.1 | Modifikasi Friedman | 32 |
| 3.1.2 | Estimasi Parameter | 34 |
| 3.1.3 | <i>Generalized Cross Validation</i> (GCV) | 37 |
| 3.1.4 | Uji Signifikansi | 40 |
| 3.2 | <i>Cox Proportional Hazard</i> dengan MARS | 41 |
| 3.3 | Langkah Penelitian | 42 |
| 3.4 | <i>Salford Predictive Modeler</i> (SPM) 8 | 44 |
| 3.5 | Sumber Data | 46 |
| 3.6 | Variabel Penelitian | 46 |
| 3.6.1 | Variabel respon | 46 |
| 3.6.2 | Variabel prediktor | 48 |
| 3.7 | Karakteristik Kematian Bayi | 50 |
| 3.8 | Asumsi <i>Proportional Hazard</i> | 55 |
| 3.9 | Estimasi Fungsi Survival dan Fungsi <i>Hazard</i> | 57 |
| 3.10 | Estimasi Residual <i>Martingale</i> | 59 |
| 3.11 | <i>Trial and Error</i> Model MARS | 59 |
| 3.12 | Model MARS Terbaik | 61 |

| | |
|---|-----------|
| 3.12.1 Uji Simultan | 63 |
| 3.12.2 Uji Parsial | 63 |
| 3.13 Model Cox <i>Proportional Hazard</i> dengan MARS | 65 |
| 3.14 Interpretasi Cox PH dengan MARS | 65 |
| IV PENUTUP | 72 |
| 4.1 Kesimpulan | 72 |
| 4.2 Saran | 74 |
| DAFTAR PUSTAKA | 75 |
| LAMPIRAN-LAMPIRAN | 77 |

DAFTAR TABEL

| | | |
|------|--|----|
| 3.1 | Daftar variabel yang digunakan dalam penelitian | 47 |
| 3.2 | Perbandingan Status Bayi Lahir Hidup Menurut Usia Ibu Saat Melahirkan | 51 |
| 3.3 | Perbandingan Status Bayi Lahir Berdasarkan Pendidikan Ibu . . . | 53 |
| 3.4 | Perbandingan Status Bayi Berdasarkan Jenis Kelamin Bayi . . . | 53 |
| 3.5 | Perbandingan Status Bayi Berdasarkan Kelahiran Kembar . . . | 54 |
| 3.6 | Perbandingan Status Bayi Berdasarkan Penolong Kelahiran . . . | 54 |
| 3.7 | Perbandingan Status Bayi Berdasarkan Sumber Utama Air Minum | 54 |
| 3.8 | Estimasi Fungsi Survival dan Fungsi <i>Hazard</i> | 58 |
| 3.9 | Estimasi Residual <i>Martingale</i> | 59 |
| 3.10 | Hasil GCV Berdasarkan Kombinasi BF, MI, dan MO | 60 |
| 3.11 | Interaksi Variabel pada Fungsi Basis | 70 |
| 3.12 | Tingkat kepentingan variabel | 70 |

DAFTAR GAMBAR

| | | |
|-----|---|----|
| 2.1 | Angka kematian bayi di Indonesia hasil SDKI 2012 | 6 |
| 2.2 | Ilustrasi Sensor Kanan | 11 |
| 2.3 | Skema prosedur RPR | 27 |
| 2.4 | Ilustrasi kurva MARS dengan knot 3 | 28 |
| 2.5 | Grafik Fungsi Basis | 29 |
| 2.6 | Perkalian Dua Fungsi Basis MARS | 30 |
| 3.1 | Diagram alir pemodelan Cox PH dengan MARS | 43 |
| 3.2 | Model dalam SPM 8 | 45 |
| 3.3 | <i>Options and Limits</i> dalam SPM 8 | 45 |
| 3.4 | Perbandingan Amatan Data Tersensor dan Tidak Tersensor | 50 |
| 3.5 | Jumlah Paritas Ibu | 52 |
| 3.6 | Jumlah Jarak Antar Kelahiran | 52 |
| 3.7 | Asumsi <i>Proportional Hazard</i> pada Variabel X_1 dan X_4 | 56 |
| 3.8 | Asumsi <i>Proportional Hazard</i> pada Variabel X_5 dan X_6 | 56 |
| 3.9 | Asumsi <i>Proportional Hazard</i> pada Variabel X_7 dan X_8 | 57 |

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Angka Kematian Bayi (AKB) merupakan angka yang menunjukkan jumlah kematian bayi sebelum mencapai usia 1 tahun dari setiap 1000 kelahiran hidup pada tahun tertentu atau dapat dikatakan juga sebagai probabilitas bayi meninggal sebelum mencapai usia 1 tahun (dinyatakan dengan per seribu kelahiran hidup). AKB Indonesia berdasarkan hasil SDKI pada tahun 2012 sebesar 34 per 1.000 kelahiran hidup. Hal tersebut belum mencapai target *Millenium Development Goals* (MDGs) untuk tahun 2015 di Indonesia sebesar 23 per 1.000 kelahiran hidup. Pada September 2015 telah disepakati agenda 2030 untuk pembangunan berkelanjutan atau *Sustainable Development Goals* (SDGs) yang merupakan kesepakatan pembangunan baru untuk mendorong perubahan-perubahan yang menggeser ke arah pembangunan berkelanjutan berdasarkan hak asasi manusia dan kesetaraan untuk mendorong pembangunan sosial, ekonomi dan lingkungan hidup. SDGs merupakan program untuk melanjutkan upaya dan pencapaian MDGs yang berakhir pada tahun 2015. Terdapat 38 target SDGs di sektor kesehatan yang perlu diwujudkan, salah satunya mengatasi permasalahan yang belum tuntas ditangani hingga target MDGs usai seperti angka kematian bayi (AKB).

Kasus kematian bayi yang masih tinggi hingga belum mencapai target MDGs tahun 2015 tidak dapat dibiarkan begitu saja. Oleh karena itu, diperlukan intervensi yang tepat untuk mengurangi kematian bayi tersebut. Inter-

vensi akan berjalan efektif, apabila faktor-faktor signifikan penyebab kematian bayi telah diketahui. Peneliti akan melakukan penelitian empiris untuk mengetahui faktor-faktor signifikan yang mempengaruhi laju kematian bayi pada penelitian ini. Metode yang akan digunakan peneliti untuk mengetahui laju kematian bayi yaitu analisis survival.

Analisis survival merupakan salah satu metode statistik untuk menganalisa data yang berhubungan antara kejadian dengan waktu. Salah satu metode yang terdapat pada analisis survival adalah pemodelan *Cox Proportional Hazard* atau Cox PH. Cox PH umum digunakan karena dapat memperoleh *hazard rate* tanpa perlu diketahui fungsi *baseline hazard*. Selain itu Cox PH tidak mempunyai fungsi kepadatan distribusi parametrik, sehingga regresi Cox merupakan model semiparametrik yang tidak mempunyai asumsi pada waktu survival, maka metode tersebut dikatakan *robust*, artinya hasil dari Cox PH akan mendekati hasil dari distribusi parametrik yang tepat. Penentuan waktu pada Cox PH harus jelas, mulai dari waktu mulai penelitian (*start point*), waktu berakhir penelitian (*end point*) dan waktu kejadian (*event*).

Multivariate Adaptive Regression Spline (MARS) merupakan pemodelan nonparametrik yang dikembangkan oleh Friedman (1991). MARS merupakan bentuk pendekatan dari *Recursive Partition Regression* (RPR) dengan perluasan dari fungsi *spline* sehingga model yang dihasilkan kontinu terhadap knot. Regresi *spline* mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus (Hardle, 1994). Sedangkan metode RPR mampu mengatasi kesulitan dalam menentukan knot, karena knot ditentukan oleh data. Oleh karena itu MARS menghasilkan model yang kontinu pada knot yaitu garis regresi selalu menyambung, dimana tiap knot selalu menyambung dengan fungsi basisnya. Kriner (2007) menyatakan residual *mar-*

tingale pada Cox PH dapat digunakan sebagai variabel respon MARS untuk mengatasi masalah model non-linier.

Penelitian mengenai analisis Cox PH pendekatan MARS pernah dilakukan oleh Kriner pada tahun 2007 serta Nisa' dan Budiantara pada tahun 2012. Kriner (2007) telah melakukan penelitian yang berjudul *Survival Analysis with Multivariate Adaptive Regression Splines* untuk memprediksi kematian pasien setelah mengalami serangan jantung. Pada penelitian tersebut Kriner menyimpulkan bahwa penggunaan residual *martingale* regresi cox sebagai variabel respon pada MARS dapat mengatasi masalah model nonlinear dan interaksi yang lebih sulit dibandingkan dengan metode yang biasanya. Penelitian yang serupa juga telah dilakukan oleh Nisa' dan Budiantara (2012) dengan judul Analisis Survival dengan Pendekatan *Multivariate Adaptive Regression Spline* pada Kasus Demam Berdarah *Dengue* (DBD).

Oleh karena itu, peneliti akan melakukan penelitian tentang pemodelan kematian bayi menggunakan *cox proportional hazard* dengan *multivariate adaptive regression spline* pada hasil Survei Penduduk Antar Sensus 2015 dengan faktor-faktor penduga antara lain usia ibu saat melahirkan, paritas, jarak antar kelahiran, pendidikan ibu, jenis kelamin bayi, kelahiran kembar, penolong kelahiran bayi, dan sumber utama air minum.

1.2 Perumusan Masalah

Perumusan masalah yang akan dikaji adalah sebagai berikut:

1. Bagaimana karakteristik kematian bayi dengan faktor-faktor penduganya?
2. Bagaimana model kematian bayi menggunakan *cox proportional hazard* dengan MARS?

3. Apa saja faktor-faktor yang mempengaruhi laju kematian bayi menggunakan *Cox proportional hazard* dengan pendekatan MARS?

1.3 Pembatasan Masalah

Pembatasan masalah dalam penulisan ini adalah:

1. Unit analisis yang diambil adalah bayi dalam rumah tangga yang lahir antara bulan Januari 2013 sampai Desember 2014.
2. Area yang diteliti dalam penelitian ini adalah wilayah di Provinsi Banten, dengan asumsi kondisi sosial ekonomi dalam rumah tangga tidak mengalami perubahan selama periode 2013-2014.
3. *Failure event* dalam penelitian ini yaitu kematian bayi sebelum 12 bulan pertama dalam hidupnya. Sensor yang digunakan adalah sensor kanan, yang berarti jika bayi belum mengalami *failure event* hingga masa pengamatan berakhir maka bayi akan masuk dalam data tersensor.

1.4 Tujuan Penulisan

Tujuan yang ingin dicapai dalam skripsi ini adalah:

1. Mendeskripsikan karakteristik kematian bayi dengan faktor-faktor penduganya.
2. Mengkaji model kematian bayi menggunakan *Cox proportional hazard* dengan MARS.
3. Mendeskripsikan faktor-faktor yang mempengaruhi laju kematian bayi dengan pemodelan *Cox proportional hazard* dengan pendekatan MARS.

1.5 Manfaat Penulisan

Diharapkan penelitian ini bermanfaat untuk memberikan informasi bagi pemerintah terkait tentang laju kematian bayi serta faktor-faktor yang mempengaruhinya sehingga dapat menjadi bahan masukan kepada pemerintah terkait dalam upaya menekan laju kematian bayi, serta bermanfaat bagi mahasiswa sebagai tambahan informasi penerapan ilmu statistika di bidang kesehatan, khususnya penggunaan pemodelan cox PH dan metode MARS.

1.6 Metode Penelitian

Metode yang digunakan dalam skripsi ini merupakan pemodelan *cox proportional hazard* dengan MARS yang didasarkan pada literatur bacaan berupa jurnal, skripsi, internet, dan *textbook* tentang pemodelan *cox proportional hazard* dan MARS. Data yang akan digunakan dalam proposal ini merupakan data sekunder dari SUPAS 2015. Data yang diperoleh yaitu kelahiran dan kematian bayi selama periode 2013-2014 beserta faktor-faktor penduganya. Selanjutnya, akan dicari pemodelan *Cox proportional hazard* dengan MARS beserta faktor yang mempengaruhi kematian bayi pada data yang telah diperoleh berdasarkan formulasi yang didapat sebelumnya.

BAB II

LANDASAN TEORI

2.1 Kematian Bayi

Kematian bayi merupakan kematian anak yang dilahirkan menunjukkan tanda-tanda kehidupan sampai anak belum mencapai usia satu tahun (antara 0 sampai 11 bulan) (UNICEF Indonesia, 2010). Angka Kematian Bayi (AKB) di Indonesia berdasarkan hasil SDKI pada tahun 2012 mencapai 34 per 1.000 kelahiran hidup. Hal tersebut belum mencapai target *Millenium Development Goals* (MDGs) untuk tahun 2015 yaitu sebesar 23 per 1.000 kelahiran hidup. Pada Gambar 2.1 dapat terlihat bahwa hampir seluruh AKB di Indonesia belum mencapai dari target MDGs.

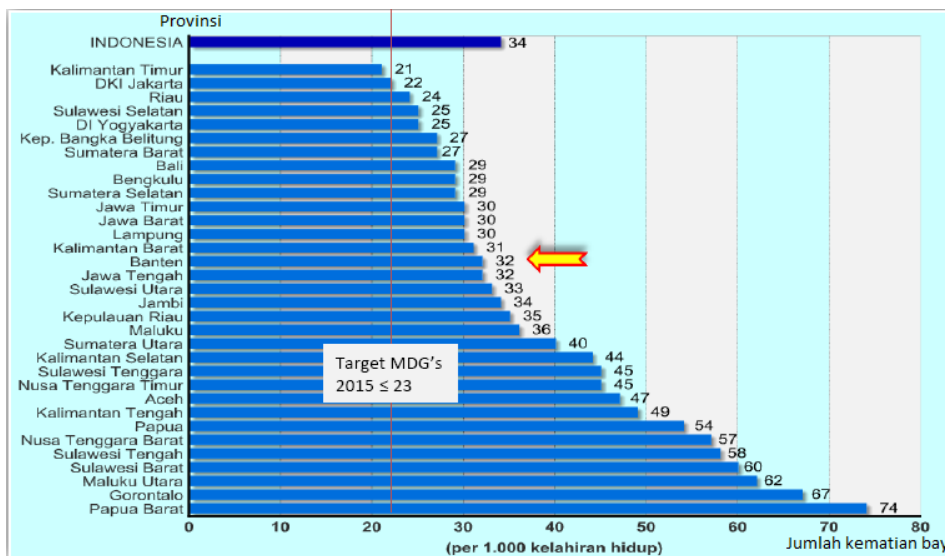


Figure 2.1: Angka kematian bayi di Indonesia hasil SDKI 2012

Kasus kematian bayi tidak dapat dibiarkan begitu saja, karena kelangsungan hidup anak sangat menentukan sumber daya manusia di masa yang akan datang. Hill (2003) dalam Ray (2015) mengemukakan bahwa kelangsungan hidup anak dipengaruhi oleh lima faktor yaitu faktor ibu, kontaminasi lingkungan, defisiensi gizi, cedera/kecelakaan, dan kontrol terhadap penyakit. Faktor ibu dan faktor lingkungan dapat diperoleh datanya melalui survei-survei BPS, salah satunya Survei Penduduk Antar Sensus (SUPAS). Perlu upaya yang tepat untuk mengurangi jumlah kematian bayi. Upaya-upaya yang dilakukan akan berjalan efektif apabila faktor-faktor signifikan penyebab kematian bayi diketahui. Oleh karena itu, faktor-faktor yang akan digunakan peneliti dalam penelitian yaitu faktor ibu, faktor bayi, faktor kontrol kesehatan, dan faktor lingkungan. Faktor-faktor tersebut digunakan sebagai variabel prediktor dari penyebab kematian bayi, antara lain:

1. Faktor ibu, meliputi:

- (a) Usia ibu saat melahirkan

Usia ibu saat melahirkan merupakan salah satu faktor kematian anak saat kelahiran. Waktu reproduksi yang sehat menurut Kemenkes (2010) yaitu usia aman untuk kehamilan dan persalinan adalah 20-35 tahun.

- (b) Paritas ibu

Paritas atau jumlah kehamilan/kelahiran yang telah dialami oleh seorang ibu menurut UNICEF Indonesia (2012) dapat mempengaruhi berat badan bayi yang dilahirkannya, hal ini terkait dengan kesuburan rahim ibu.

- (c) Jarak antar kelahiran

Resiko terhadap kematian ibu maupun anak menurut Kemenkes

(2010) dapat meningkat jika jarak antara dua kehamilan < 2 tahun atau > 4 tahun. Jarak antara dua kehamilan yang < 2 tahun berarti tubuh ibu belum kembali ke keadaan normal akibat kehamilan sebelumnya. Sebaliknya jika jarak kehamilan antara dua kehamilan > 4 tahun, akan mengakibatkan persalinan berlangsung seperti kehamilan dan persalinan pertama.

(d) Pendidikan Ibu

Status pendidikan ibu dapat mempengaruhi pola pikir ibu sebagai penunjang dalam mempermudah mencerna informasi yang diterima untuk dapat dimengerti. Tingginya pendidikan dan pengetahuan ibu tentang kesehatan akan dapat memotivasi ibu untuk menjaga dirinya agar lebih sehat dan dapat membentuk kebiasaan yang menguntungkan.

2. Faktor bayi, meliputi:

(a) Jenis kelamin

(b) Kelahiran kembar

Rata-rata berat badan anak kembar lebih rendah dari berat badan anak tunggal, hal ini terjadi karena banyaknya persalinan bayi kembar sebelum waktunya. Agustien (2016) menyatakan bahwa berat badan lahir rendah (BBLR) dapat mengakibatkan kematian pada bayi.

3. Faktor kontrol kesehatan

Faktor kontrol kesehatan pada penelitian ini yaitu penolong kelahiran bayi. Penolong kelahiran bayi dibagi menjadi dua yaitu bantuan medis (seperti dokter dan bidan) atau bantuan nonmedis (seperti dukun

dan lainnya).

4. Faktor lingkungan

Faktor lingkungan berkaitan dengan penularan penyakit kepada anggota keluarga, terutama bayi. Tingkat pencemaran lingkungan yang mencerminkan berbagai jalur penularan penyakit bisa diukur secara langsung dengan pemeriksaan mikrobiologi pada beberapa sampel, salah satunya sampel air (UNICEF Indonesia, 2010).

2.2 Analisis Survival

Analisis survival merupakan salah satu analisis yang menggambarkan suatu proses yang berhubungan dengan waktu, diawali dengan *time origin* sampai terjadinya suatu kejadian atau *end point* (Collett, 2003). Data survival adalah data tentang pengamatan jangka waktu dari awal pengamatan sampai terjadinya suatu peristiwa. Waktu survival dapat didefinisikan sebagai waktu dari awal pengamatan sampai terjadinya kejadian, waktu survival dapat berupa hari, bulan, maupun tahun. Terdapat beberapa elemen yang perlu diperhatikan dalam menentukan waktu survival, antara lain:

1. Titik awal atau waktu awal penelitian (*time origin* atau *time start point*) harus didefinisikan dengan tepat pada setiap individu, misalkan awal mula pengamatan berupa tanggal perawatan pasien.
2. Waktu akhir penelitian (*failure time* atau *end point*) didefinisikan untuk mengetahui status tersensor atau tidak tersensor, seperti kematian atau kesembuhan pasien dan lainnya.
3. Skala pengukuran waktu (*Measurement scale for the passage of time*) adalah sebagai batas waktu kejadian dari awal sampai akhir kejadian,

misalnya skala tahunan, bulanan, mingguan maupun harian.

Analisis survival berbeda dengan analisis statistik lain, dengan data tersensor sebagai pembedanya. Tipe data lama waktu tidak dapat dianalisis menggunakan regresi linier biasa karena terdapat data tersensor, dan tidak dapat pula dianalisis dengan regresi logistik karena lamanya waktu data yang diikuti tidak sama, sehingga yang paling sesuai adalah menggunakan analisis survival.

2.2.1 Data Tersensor

Millet (1998) dalam Prasetyo (2017) menyatakan bahwa data tersensor memungkinkan beberapa individu tidak dapat diobservasi secara penuh sampai terjadinya *failure event*. Collet (2003) berpendapat terdapat tiga alasan terjadinya penyensoran pada suatu data, antara lain:

1. *Lost to follow up* yaitu jika subyek pengamatan tidak terdata hingga akhir observasi dikarenakan meninggal, pindah, atau menolak untuk berpartisipasi.
2. *Drop out* yaitu jika pengamatan terhadap suatu subyek harus dihentikan karena alasan tertentu, seperti pemberian kemoterapi yang dihentikan karena efek buruknya lebih besar dibanding manfaatnya.
3. *Termination of study* yaitu jika subyek pengamatan belum mengalami *failure event* hingga observasi berakhir.

Collet (2003) menyatakan bahwa data tersensor dibagi menjadi tiga, yaitu:

1. Data tersensor kanan

Data tersensor kanan dilakukan ketika subyek pengamatan yang masuk dalam observasi dapat diamati secara penuh namun hingga akhir

penelitian belum mengalami kejadian. Sebagai ilustrasi pada Gambar 2.2 tentang kasus kematian karena kanker serviks dan ditetapkan waktu pengamatan adalah 5 tahun. Subyek penderita kanker serviks diamati sejak pertama kali didiagnosis positif kanker serviks. Subyek B dan subyek C dikatakan data tersensor kanan karena subyek B melakukan perpindahan domisili yang mengakibatkan tidak dapat diamati lagi, sedangkan subyek C masih hidup hingga akhir pengamatan.

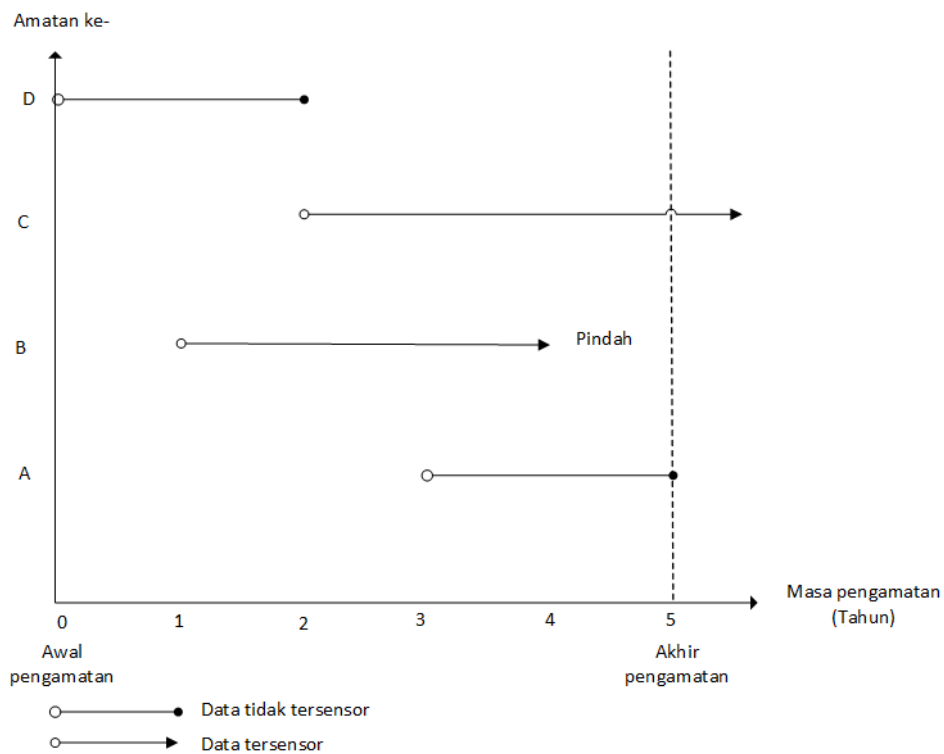


Figure 2.2: Ilustrasi Sensor Kanan

2. Data tersensor kiri

Data tersensor kiri terjadi ketika subyek pengamatan tidak teramati pada awal waktu pengamatan, akan tetapi sebelum penelitian berakhir semua *event* sudah dapat diamati secara penuh, atau dengan kata lain data

tersensor kiri terjadi apabila *event* yang ingin diperhatikan pada subyek pengamatan tersebut sudah terjadi saat subyek pengamatan tersebut masuk ke dalam penelitian. Sebagai contoh, seorang dokter ingin mengetahui diagnosis usia seseorang saat terjangkit virus HIV. Pada saat pemeriksaan terdapat seorang pasien berumur 35 tahun yang ditemukan terjangkit virus HIV, akan tetapi tidak terdapat catatan tentang waktu tepatnya pasien tersebut mulai terjangkit virus, sehingga umur pasien pada saat pemeriksaan (sekitar 35 tahun) merupakan data tersensor kiri.

3. Data sensor interval

Sensor interval terjadi ketika suatu *event* yang diamati pada subyek pengamatan terjadi pada selang waktu tertentu. Sebagai contoh, pada pemeriksaan terdapat seorang pasien berumur 35 tahun yang ditemukan terjangkit virus HIV. Apabila pada catatan medis mengindikasikan bahwa pada saat usia 30 tahun pasien tersebut belum terjangkit virus HIV, maka usia pasien yang didiagnosis terjangkit virus HIV antara 30 dan 35 tahun.

Pada model semiparametrik Cox PH, tidak terdapat distribusi parametrik khusus untuk *baseline hazard*. Akibatnya model Cox PH lebih mengacu hanya pada penggabungan informasi waktu peristiwa yang diamati dibanding dengan menentukan suatu distribusi tertentu untuk interval terjadinya suatu peristiwa. Model Cox PH mengacu pada semiparametrik karena meskipun tidak ada bentuk distribusi tertentu yang digunakan untuk *baseline hazard* tetapi risiko (*hazard*) terjadinya suatu peristiwa tetap dinyatakan sebagai fungsi dari *covariate*.

2.2.2 Fungsi Survival dan Fungsi *Hazard*

Misalkan T merupakan variabel acak non negatif dari waktu survival suatu individu. Peluang T pada analisis survival secara umum digambarkan ke dalam tiga fungsi yaitu fungsi kepadatan peluang, fungsi survival, dan fungsi *hazard*.

Definisi 2.2.1. Jika T melambangkan waktu survial dan mempunyai distribusi peluang $f(t)$, maka fungsi distribusi kumulatif dinyatakan sebagai berikut:

$$F(t) = P(T < t) = \int_0^t f(x)dx \quad ; 0 < t < \infty$$

Berdasarkan Definisi 2.2.1., diperoleh:

$$f(t) = \frac{dF(t)}{dt} = F'(t).$$

Definisi 2.2.2. Fungsi survival merupakan fungsi yang menyatakan peluang suatu subyek dapat bertahan hingga atau lebih dari waktu yang ditentukan, t . Fungsi survival, $S(t)$, dapat dinyatakan sebagai berikut:

$$S(t) = P(T \geq t)$$

Berdasarkan Definisi 2.2.1. dan Definisi 2.2.2., fungsi survival dapat dinyatakan sebagai berikut:

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(T < t) \\ &= 1 - F(t) \\ F(t) &= 1 - S(t) \\ \frac{d(F(t))}{dt} &= \frac{d(1 - S(t))}{dt} \\ f(t) &= -\frac{d(S(t))}{dt} = -S'(t). \end{aligned}$$

Definisi 2.2.3. Fungsi *hazard* merupakan probabilitas variabel acak dari waktu survival subyek pengamatan pada interval $(t, t + \Delta t)$. Fungsi *hazard* $h(t)$ dapat

ditulis dalam bentuk,

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right\}.$$

Berdasarkan Definisi 2.2.3. dapat ditarik hubungan antara fungsi survival dengan fungsi *hazard* menggunakan probabilitas bersyarat.

Definisi 2.2.4. Probabilitas bersyarat A ketika B terpenuhi didefinisikan dengan,

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

jika $P(B) \neq 0$

Berdasarkan Definisi 2.2.4., probabilitas bersyarat pada Definisi 2.2.3. yaitu

$$\frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{S(t)}$$

Karena $F(t)$ merupakan fungsi distribusi dari T , maka

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)}$$

Berdasarkan definisi turunan,

$$\begin{aligned} F'(t) &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \\ &= f(t) \end{aligned}$$

sehingga,

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

karena $f(t) = -S'(t)$ maka,

$$\begin{aligned} h(t) &= \frac{\left(-\frac{d(S(t))}{dt} \right)}{S(t)} \\ -h(t)dt &= \frac{d(S(t))}{S(t)}. \end{aligned}$$

Jika fungsi diatas diintegalkan, maka diperoleh:

$$\begin{aligned}
 - \int_0^t h(x)dx &= \int_0^t \frac{1}{S(x)}d(S(x)) \\
 &= [\ln S(x)]_0^t \\
 &= \ln S(t) - \ln S(0) \\
 &= \ln S(t)
 \end{aligned}$$

Sehingga $S(t) = \exp \left[- \int_0^t h(x)dx \right]$, dimana $H(t) = \int_0^t h(x)dx$, dengan $H(t)$ disebut fungsi *hazard* kumulatif, sehingga hubungan antara fungsi *hazard* dengan fungsi survival yaitu

$$H(t) = -\log S(t) \quad (2.1)$$

Fungsi *hazard* merupakan peluang kegagalan individu untuk bertahan selama interval waktu yang sangat pendek dengan asumsi bahwa individu tersebut telah bertahan pada awal interval. Semakin besar nilai *hazard* berarti resiko kegagalan yang dialami individu dalam penelitian semakin tinggi sehingga kemampuan bertahannya semakin kecil.

2.3 Cox Proportional Hazard (Cox PH)

Cox *Proportional Hazard* (Cox PH) merupakan model semi-parametrik yang paling sering digunakan pada analisis survival. Model Cox PH merupakan model berdistribusi semiparametrik karena tidak diperlukan informasi tentang distribusi yang mendasari waktu survival, serta dapat mengestimasi parameter Cox PH tanpa harus menggunakan fungsi *baseline hazard*. Cox PH digunakan apabila *outcome* yang diamati adalah panjang waktu suatu kejadian.

Definisi 2.3.1. Misalkan terdapat X_1, X_2, \dots, X_p variabel prediktor dan $h_0(t)$ merupakan fungsi *hazard* untuk tiap-tiap subyek pengamatan dengan semua

variabel prediktor dalam vektor x bernilai 0. Model *proportional hazard* didefinisikan sebagai berikut:

$$h_i(t) = h_0(t) \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})$$

Keterangan:

$h_i(t)$ = fungsi *hazard* subyek pengamatan ke- i

$h_0(t)$ = fungsi *baseline hazard*

x_{ji} = nilai variabel prediktor ke- j pada pengamatan ke- i , dengan
 $j = 1, 2, \dots, p$

β_j = koefisien regresi ke- j , dengan $j = 1, 2, \dots, p$.

Persamaan pada Definisi 2.3.1. dapat disebut juga sebagai regresi Cox atau model Cox PH. Model tersebut dapat dinyatakan dalam bentuk lain, yaitu:

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})$$

atau dapat dirumuskan sebagai model linier dengan log-relatif *hazard*, yaitu:

$$\ln \left[\frac{h_i(t)}{h_0(t)} \right] = (\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}).$$

Jika variabel $X_1 = X_2 = \dots = X_p = 0$, maka fungsi *hazard* tersebut merupakan fungsi *baseline hazard* yang bergantung pada waktu, sehingga diperoleh:

$$h_i(t) = h_0 \exp(0) = h_0(t).$$

Jadi yang dimaksud dengan fungsi *baseline hazard* adalah suatu fungsi *hazard* yang bergantung pada waktu dan tidak melibatkan variabel prediktornya. Perlu diketahui bahwa pada pemodelan Cox PH tidak terdapat β_0 . Hal ini dikarenakan untuk mengestimasi parameter β tidak diperlukan estimasi dari fungsi *baseline hazard*. Oleh karena itu pemodelan Cox PH disebut model semi-parametrik.

2.4 Asumsi *Proportional Hazard*

Asumsi pemodelan yang harus dipenuhi dalam Cox PH yaitu asumsi *proportional hazard* yang berarti fungsi *hazard* harus *proportional* setiap waktu karena Cox PH tidak mengakomodasi variabel yang berubah-ubah sepanjang waktu (Collet, 2003). *Proportional* artinya variabel prediktor independen terhadap waktu dan hubungan antara *hazard* kumulatif sudah *proportional* setiap waktu.

Definisi 2.4.1. Asumsi *proportional hazard* didasarkan pada fungsi survival yang didefinisikan sebagai berikut:

$$S(t, \mathbf{X}) = S_0(t) \exp \sum_{k=1}^p \beta_k X_k \quad ; 0 \leq S(t, \mathbf{X}) \leq 1$$

dengan $S_0(t)$ merupakan fungsi *baseline* survival.

Jika persamaan dari Definisi 2.4.1. diubah dalam bentuk logaritma maka fungsi tersebut menjadi:

$$\ln(S(t, \mathbf{X})) = \ln(S_0(t)) \exp \sum_{k=1}^p \beta_k X_k$$

Logaritma dari $S(t, \mathbf{X})$ dan S_0 kan menghasilkan nilai yang negatif sehingga diberikan tanda negatif di depan logaritma dan dilakukan logaritma kembali.

Secara matematis dapat ditulis sebagai berikut:

$$\begin{aligned} \ln(-\ln(S(t, \mathbf{X}))) &= \ln \left[-\ln(S_0(t)) \exp \sum_{k=1}^p \beta_k X_k \right] \\ &= \ln(-\ln(S_0(t))) + \ln \left[\exp \sum_{k=1}^p \beta_k X_k \right] \\ &= \ln(-\ln(S_0(t))) + \exp \sum_{k=1}^p \beta_k X_k \end{aligned}$$

Atau dapat dituliskan dengan,

$$-\ln(-\ln(S(t, \mathbf{X}))) = -\ln(-\ln(S_0(t))) - \exp \sum_{k=1}^p \beta_k X_k$$

Asumsi *proportional hazard* dapat diketahui dengan melihat pola dari plot $-\ln(-\ln S(t))$ atau $\ln(-\ln S(t))$ terhadap waktu survival untuk setiap variabel prediktor dengan skala kategori. Asumsi tersebut terpenuhi jika pola plot antar kategori dalam variabel prediktor membentuk pola yang sejajar atau tidak berpotongan (Kleinbaum & Klein, 2005). Pola yang saling berpotongan menunjukkan bahwa kategori antar variabel prediktor tidak memenuhi asumsi *proportional hazard*.

2.5 Metode Kaplan Meier

Telah diketahui bahwa salah satu tujuan dari analisis survival yaitu mengestimasi dan menginterpretasi fungsi survival dan fungsi *hazard*. Salah satu metode yang digunakan untuk mengestimasi fungsi survival yaitu metode Kaplan Meier. Penelitian ini menggunakan statistik nonparametrik dengan data tersensor, sehingga tepat jika menggunakan metode Kaplan-Meier. Metode Kaplan-Meier dapat mengestimasi fungsi survival dengan mengelompokkan subyek pengamatan berdasarkan pada waktu survival. Oleh karena itu dimungkinkan terdapat lebih dari satu amatan dalam satu waktu survival.

Definisi 2.5.1. Misalkan terdapat amatan yang mati dalam satu kelompok waktu survival yang saling bebas, maka penduga fungsi survival pada waktu ke t , akan menduga probabilitas ketahanan t_k . Penduga fungsi survival dengan Kaplan-Meier didefinisikan sebagai berikut:

$$\widehat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right)$$

dengan:

- k = Jumlah kelompok waktu survival
- n_j = Jumlah amatan yang bertahan hingga waktu ke t_j
- d_j = Jumlah amatan yang tidak bertahan pada waktu t_j .

Definisi 2.5.2. Misalkan terdapat sebanyak d_j amatan yang tidak bertahan pada waktu ke j (t_j), dengan $j = 1, 2, \dots, k$ dan n_j amatan yang bertahan hingga waktu ke t_j , maka fungsi *hazard* pada interval waktu t_j hingga t_{j+1} dengan metode Kaplan-Meier didefinisikan sebagai berikut:

$$\widehat{h}(t) = \frac{d_j}{n_j \tau_j}$$

dengan $t_j \leq t < t_{j+1}$ dan $\tau_j = t_{j+1} - t_j$.

Berdasarkan Definisi 2.5.1. dan Persamaan (2.1) maka penduga fungsi *hazard* kumulatif dengan metode Kaplan-Meier sebagai berikut:

$$\widehat{H}(t) = - \sum_{j=1}^k \log \left(\frac{n_j - d_j}{n_j} \right)$$

Keuntungan dari metode Kaplan-Meier adalah menganalisa data *survival* dan *failure* dengan hasil perhitungan tidak tergantung pada data yang dikelompokkan (pada sejumlah waktu interval). Estimasi fungsi *hazard* kumulatif yang diperoleh pada metode Kaplan-Meier tersebut akan digunakan untuk mengestimasi residual *martingale*.

2.6 Residual *Martingale*

Residual *martingale* merupakan salah satu residual yang terdapat pada Cox PH. Kriner (2007) menyatakan bahwa residual *martingale* pada pemodelan cox PH dapat digunakan sebagai variabel respon untuk pendekatan MARS. Persamaan residual *martingale* dirumuskan sebagai berikut (Collet, 2003):

$$\widehat{M}_i(t) = \delta_i - \widehat{H}_i(t)$$

dengan:

$\widehat{M}_i(t)$ = residual *martingale* subyek pengamatan ke- i

δ_i = 0, untuk data tersensor

= 1, untuk data tidak tersensor

$\widehat{H}_i(t)$ = Hazard Kumulatif subyek pengamatan ke- i pada waktu ke- t .

Residual *martingale* dapat menjadi gambaran mengenai perbedaan hasil pengamatan δ_i dengan angka prediksi pada kejadian $\widehat{H}_i(t)$. Ketika perbedaan antara hasil pengamatan dengan angka prediksi untuk subjek ke- i cukup besar, itu menunjukkan bahwa subyek ke- i tidak sesuai dengan model dan mengakibatkan suatu nilai yang besar pada $\widehat{M}_i(t)$. Karena interval nilai dari $\widehat{H}_i(t)$ adalah $(0, \infty)$ dan δ_j hanya bernilai 0 atau 1, maka dapat diambil kesimpulan bahwa residual *martingale* bernilai $(-\infty, 1)$ dan bernilai negatif pada data tersensor.

2.7 Analisis Regresi

Analisis regresi merupakan analisis suatu model yang digunakan untuk mengetahui hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Pada analisis regresi linier, peubah responnya memiliki skala pengukuran minimal interval. Berdasarkan banyaknya variabel prediktor yang digunakan, analisis regresi linier dibagi menjadi dua, yaitu analisis regresi linier sederhana yang hanya melibatkan satu variabel prediktor, dan analisis regresi linier berganda yang melibatkan lebih dari satu variabel prediktor. Bentuk umum regresi linier berganda adalah sebagai berikut:

$$Y = \alpha + \sum_{i=1}^p \beta_i X_i + \epsilon$$

dengan:

Y = variabel respon

α, β = koefisien regresi atau parameter yang akan ditaksir

ϵ = galat atau sisaan

Model regresi dapat diperoleh dengan melakukan estimasi terhadap parameter - parameternya menggunakan metode tertentu, diantaranya metode kuadrat terkecil (*least square*) dan metode kemungkinan maksimum (*maximum likelihood*). Analisis regresi pendekatan parametrik merupakan analisis regresi yang bentuk kurvanya diketahui. Apabila terdapat pola data yang tidak dapat diestimasi, karena menghasilkan residual dan variansi yang besar maka digunakan analisis regresi nonparametrik yang diasumsikan bentuk kurvanya tidak diketahui. Metode ini menggunakan estimasi $f(x_i)$ untuk menunjukkan fleksibilitas yang tinggi pada kurva regresi.

2.8 Regresi Non-parametrik

Regresi non-parametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva tertentu, dimana kurva regresi hanya diasumsikan *smooth*(mulus), artinya termuat di dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Eubank, 1999). Jika bentuk kurva tidak diketahui dan pola menyebar maka kurva regresi dapat diduga dengan menggunakan pendekatan model regresi nonparametrik. Selain itu, regresi non-parametrik mengabaikan asumsi-asumsi yang ada pada regresi parametrik. Secara umum, model regresi non-parametrik

dapat ditulis sebagai berikut :

$$y_i = f(x_i) + \epsilon_i \quad (2.2)$$

dengan $i = 1, 2, \dots, n$, dan

y_i = variabel respon

$f(x_i)$ = fungsi *smooth* yang tidak diketahui pada amatan ke- i

ϵ_i = error ke- i yang saling bebas.

Fungsi regresi $f(x_i)$ hanya diasumsikan termuat dalam suatu fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya berdasarkan pada kemulusan (*smoothness*) dari fungsi $f(x_i)$ tersebut. Salah satu pendekatan regresi non-parametrik adalah regresi spline. Regresi spline mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus (Hardle, 1994).

2.8.1 Regresi Spline

Spline merupakan potongan polinomial berorde q dan memiliki turunan yang kontinu dengan knot sampai orde $(q - 1)$ (Friedman, 1991). Spline memiliki dua komponen yaitu komponen polinomial dan komponen *truncated*. Salah satu kelebihan spline adalah model ini mengikuti pola data sesuai pergerakannya dengan adanya titik-titik knot. Titik knot merupakan titik yang menunjukkan perubahan data pada sub-sub interval (Nisa' & Budiantara, 2012).

Definisi 2.8.1. Misalkan terdapat t_1, t_2, \dots, t_K titik knot dengan fungsi spline berorde q . Fungsi spline didefinisikan sebagai berikut:

$$f(x) = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^K \beta_{m+k} (x_i - t_k)_+^q$$

Keterangan:

$$\begin{aligned}
i &= 1, 2, \dots, n \\
\sum_{j=0}^m \beta_j x_i^j &= \text{komponen polinomial} \\
\sum_{k=1}^K \beta_{m+k} (x_i - t_k)_+^q &= \text{komponen } \textit{truncated}
\end{aligned}$$

Jika Definisi 2.8.1. disubstitusi dengan Persamaan (2.2), maka akan terbentuk model regresi non-parametrik spline sebagai berikut:

$$y = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^K \beta_{m+k} (x_i - t_k)_+^q + \epsilon_i$$

dengan $q \geq 1$ menunjukkan orde polinomial dari fungsi spline, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, K$ dan *truncated power* fungsi spline sebagai berikut :

$$(x_i - t_k)_+^q = \begin{cases} (x_i - t_k)^q & ; (x_i - t_k) \geq 0 \\ 0 & ; (x_i - t_k) < 0 \end{cases}$$

Diharapkan pada setiap titik knot terdapat fungsi basis antar satu *region* dengan *region* lainnya. Oleh karena itu pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan $(q - 1)$ yang kontinu disetiap titik knot.

Fungsi Pemulus Spline

Berdasarkan fungsi regresi pada Persamaan (2.2) dengan f merupakan fungsi *smoothing* atau fungsi pemulus yang tidak diketahui. Oleh karena itu, dengan menggunakan data amatan sebanyak n , maka $f(x_i)$ diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS), yaitu:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \delta^2 \int_a^b (f^m(x))^2 dx \quad (2.3)$$

dengan:

δ^2 = parameter *smoothing* atau parameter pemulus

$f^m(x)$ = turunan ke- m dari $f(x)$

Penduga fungsi pemulus $\hat{f}(x_i)$ merupakan penduga fungsi yang mampu memetakan data dengan baik dan mempunyai ragam galat yang kecil serta dapat diperoleh berdasarkan data amatan, yaitu pasangan variabel prediktor dan variabel respon. Perimbangan antara fleksibilitas dan kemulusan dugaan kurva dikontrol oleh nilai parameter atau jumlah knot. Parameter pemulus yang relatif besar atau jumlah knot yang relatif kecil akan menghasilkan dugaan kurva yang sangat mulus sehingga perilaku data yang rinci tidak terlihat, sedangkan parameter pemulus yang relatif kecil atau jumlah knot yang relatif besar menghasilkan dugaan kurva yang kasar karena besarnya pengaruh variasi lokal. Pemulus *spline* mempunyai sifat fleksibel dan efektif dalam menangani sifat lokal suatu fungsi atau data.

Pemilihan Parameter Pemulus δ^2 Optimal

Pemilihan parameter pemulus δ^2 yang optimal pada Persamaan (2.3) sangat penting dilakukan untuk memperoleh model estimator kurva yang baik. Parameter δ^2 merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Pada nilai δ^2 yang kecil maka kurvanya akan kasar, dan sebaliknya pada nilai δ^2 yang besar maka kurvanya akan menjadi mulus (*smooth*), dimana fungsi yang mulus terlihat jelas secara geometrik, ketika gradien kurva pada titik-titik knot tertentu tidak berubah dengan cepat (Eubank, 1991).

Kriteria yang biasa digunakan dalam pemilihan model spline terbaik adalah *Generalized Cross Validation* (GCV). Hal itu dikarenakan nilai GCV memiliki

aspek perhitungan yang lebih sederhana dan cukup efisien. Selain GCV, kriteria model regresi yang umum digunakan masih tetap dijadikan acuan pemilihan model spline terbaik. Nilai δ^2 yang optimal berkaitan dengan nilai GCV yang minimum. Kriteria GCV didefinisikan sebagai berikut:

$$GCV(\delta^2) = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i)]^2}{\left[\frac{1}{n} \text{trace}(I - A(\delta^2)) \right]^2} \quad (2.4)$$

Keterangan:

- n = banyaknya data
- $\hat{f}(x_i)$ = penduga fungsi pemulus
- $A(\delta^2)$ = matriks Hat berukuran $n \times n$.

Matriks HAT

Matriks Hat merupakan alat pendiagnosa yang memberikan informasi titik-titik data yang berpengaruh tinggi. Matriks Hat didefinisikan sebagai berikut:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

Matriks Hat memegang peranan penting dalam mengidentifikasi pengamatan yang berpengaruh. Elemen diagonal h_{ii} dari matriks Hat didefinisikan sebagai berikut:

$$h_{ii} = x_i^T (\mathbf{X}^T \mathbf{X})^{-1} x_i \quad ; i = 1, 2, \dots, n.$$

Diagonal matriks Hat memberikan ukuran jarak yang terbakukan dari titik x_i ke pusat data dari x yaitu \bar{x} . Nilai diagonal Hat yang tinggi menunjukkan pengamatan yang ekstrim pada x .

Diagonal Hat bernilai antara 0 dan 1. Jika X memiliki rank yang penuh maka $\sum_{i=1}^n h_{ii} = p$, sehingga rata-rata dari elemen diagonal h_{ii} adalah p/n . Menurut Belsley dkk. dalam Asrani (2016), titik kritis untuk h_{ii} yaitu $2p/n$

dengan $h_{ii} > 2p/n$ berarti data memiliki potensi untuk berpengaruh pada model.

2.8.2 *Recursive Partitioning Regression (RPR)*

Recursive Partitioning Regression (RPR) merupakan suatu metode non-parametrik yang dapat dipandang sebagai suatu prosedur regresi *stepwise* untuk menduga fungsi $f(x)$ yang tidak diketahui dengan

$$\hat{f}(x) = \sum_{m=1}^S a_m B_m(x)$$

dengan $B_m(x) = I[t \in R_m]$ dan $I[.]$ merupakan fungsi indikator. Setiap fungsi indikator merupakan perkalian fungsi tangga $H(\eta)$ yang menggambarkan setiap subregion R_m . Jadi $B_m(x)$ merupakan basis fungsi yang mempunyai nilai 1 jika x merupakan anggota subregion R_m dari domain D . Keterangan:

$$\begin{aligned} a_m &= \text{koefisien fungsi basis} \\ H(\eta) &= 1 \text{ jika benar } \eta \geq 0 \\ &= 0 \text{ jika lainnya.} \end{aligned}$$

Secara umum prosedur RPR mempunyai 2 tahap yang dimulai dari region pertama $R_1 = D$. Tahap pertama atau *forward*, yaitu memilah secara iteratif daerah asal D menjadi himpunan bagian (subregion) yang saling lepas $\{R_m\}_{m=2}^M$, dengan $M \geq S$ dan M ditentukan sembarang. Tahap kedua atau *backward*, pada tahap ini yaitu menghilangkan atau memangkas (S-M) subregion dari model dengan dua kriteria yaitu evaluasi dugaan model dan jumlah subregion dalam model.

Tahapan pada RPR tersebut menghasilkan subregion yang tidak saling tumpang tindih, sehingga $\hat{f}(x)$ mendekati $f(x)$ untuk setiap subregion daerah asal. Maka fungsi basis yang dihasilkan pada prosedur RPR dapat dinyatakan

sebagai berikut:

$$B_m(x) = \prod_{k=1}^{K_m} H[s_{km}(x_{v(k,m)} - t_{km})]$$

Keterangan:

- K_m = derajat interkasi
- s_{km} = tanda pada titik knot, nilainya ± 1
- $x_{v(k,m)}$ = variabel prediktor
- t_{km} = nilai knots dari variabel prediktor $x_{v(k,m)}$.

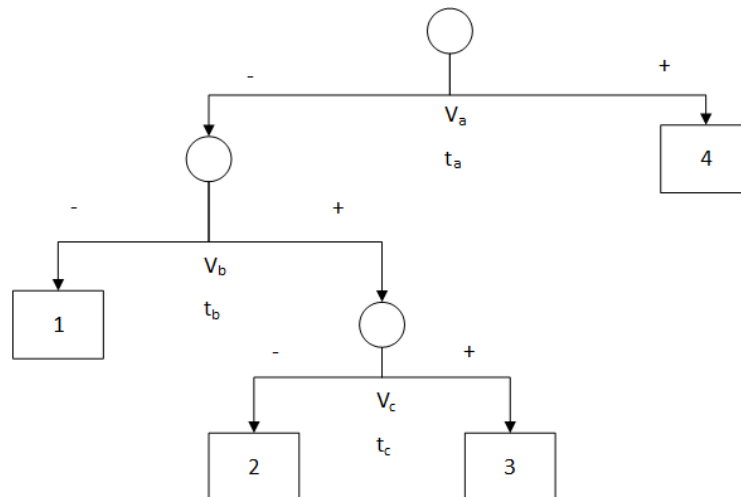


Figure 2.3: Skema prosedur RPR

RPR merupakan metode yang mampu mengatasi kesulitan dalam menentukan knot, karena knot ditentukan oleh data. Selain itu interpretasi hasilnya lebih mudah dari pada persamaan regresi biasa, karena identifikasi pengaruh dari variabel prediktor dilakukan dalam masing-masing subgrup data bukan dalam keseluruhan data seperti halnya regresi biasa. Akan tetapi RPR memiliki beberapa kekurangan diantaranya menghasilkan subregion yang saling lepas dan diskontinu pada batas subregion, serta model tidak cukup mampu menduga fungsi linier atau aditif.

2.9 *Multivariate Adaptive Regression Spline* (MARS)

Multivariate Adaptive Regression Spline (MARS) merupakan metode regresi multivariat nonparametrik yang dikembangkan oleh Friedman. MARS merupakan pengembangan model dari *Recursive Partition Regression* (RPR) yang dikombinasikan dengan metode fungsi *spline*. Friedman (1991) menyatakan bahwa terdapat beberapa hal yang perlu diperhatikan dalam membangun model MARS, antara lain:

1. *Knot*

Knot merupakan nilai variabel prediktor ketika *slope* suatu garis regresi mengalami perubahan yang dapat didefinisikan sebagai akhir dari sebuah segmen dan awal bagi segmen yang lain. Di setiap titik knot diharapkan adanya kontinuitas dari fungsi basis antar satu segmen dengan segmen lainnya. Ilustrasi mengenai knot dapat dilihat pada Gambar 3.1.

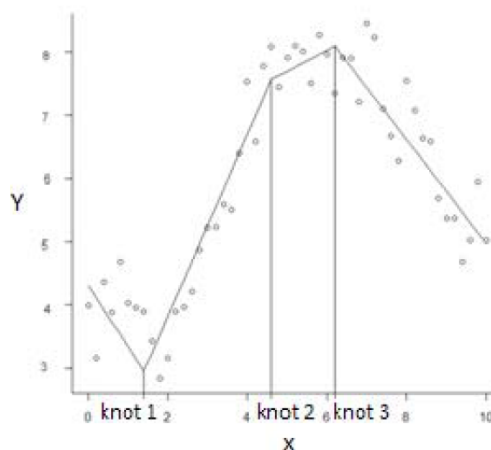


Figure 2.4: Ilustrasi kurva MARS dengan knot 3

2. Basis Function (BF)

Basis Function atau fungsi basis (BF) yaitu kumpulan dari fungsi yang digunakan untuk mewakili informasi dari satu atau beberapa variabel. Fungsi basis ini merupakan fungsi parametrik yang didefinisikan pada tiap region. Friedman menyarankan maksimum BF adalah 2 sampai dengan 4 kali jumlah variabel prediktornya. Sebagai contoh, misalkan terdapat 2 fungsi basis yang menginterpretasikan data pada Gambar 2.4. sebagai berikut:

$$BF_1 = \max(0, x - 6, 2)$$

$$BF_2 = \max(0, 6, 2 - x)$$

Berdasarkan kedua fungsi basis tersebut diketahui nilai knot untuk x sebesar 6,2. Secara matematis Gambar 2.5(a). dapat dikatakan BF_1 bernilai 0 jika $x \leq 6,2$ dan bernilai x jika $x > 6,2$, sedangkan Gambar 2.5(b) dapat dikatakan BF_2 bernilai 0 jika $x \geq 6,2$ dan bernilai x jika $x < 6,2$.

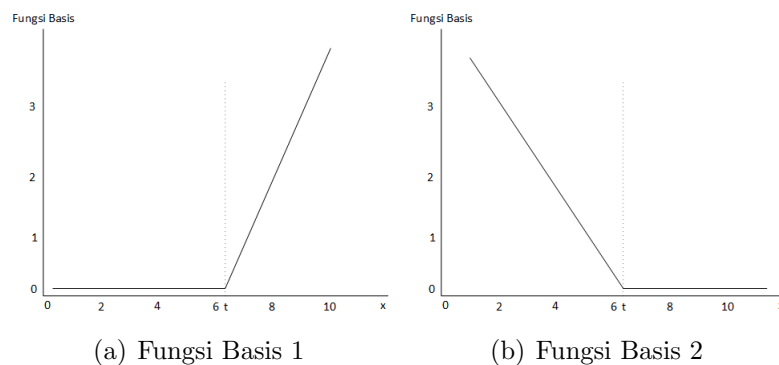


Figure 2.5: Grafik Fungsi Basis

Misalkan terdapat 2 variabel yang saling berinteraksi melalui sebuah fungsi basis yaitu $B(X_1, X_2) = (X_1 - t_1) \times (t_2 - X_2)$ maka grafik dari

perkalian fungsi basis tersebut seperti pada Gambar 2.6.

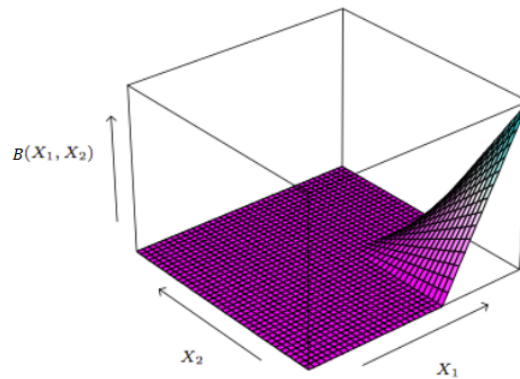


Figure 2.6: Perkalian Dua Fungsi Basis MARS

3. Maksimum Interaksi (MI)

Interaksi merupakan hasil perkalian silang antara variabel yang saling berkorelasi. Friedman membatasi jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3 dengan pertimbangan jika lebih dari 3 akan menghasilkan bentuk model yang semakin kompleks.

4. Minimum Observasi (MO)

Minimum observasi merupakan jarak antara knots atau minimum observasi antara knots sebesar 0, 1, 2, 3. $MO = 0$, artinya jarak antara titik knot dalam model adalah 0. $MO = 1$, artinya minimum jarak antara titik knot dalam model adalah 1, dan seterusnya.

Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai dari parameter pemulus yang minimum. Kelebihan utama MARS diantaranya dapat memilih variabel prediktor yang akan digunakan berdasarkan fungsi basis yang terpilih, dan mendeteksi interaksi antar variabel prediktor (Friedman, 1991).

BAB III

PEMBAHASAN

3.1 Multivariate Adaptive Regression Spline (MARS)

Regresi non-parametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva tertentu, dimana kurva regresi hanya diasumsikan *smooth*(mulus). Secara umum, model regresi nonparametrik dirumuskan sebagai berikut:

$$y_i = f(x_i) + \epsilon_i.$$

Pada model RPR penduga fungsi ($\hat{f}(x)$) dinyatakan sebagai berikut:

$$\hat{f}(x) = \sum_{m=1}^S a_m B_m(x)$$

dengan $B_m(x) = \prod_{k=1}^{K_m} H[s_{km}(x_{v(k,m)} - t_{km})]$. Keterangan:

- a_m = koefisien fungsi basis
- K_m = derajat interkasi
- s_{km} = tanda pada titik knot, nilainya ± 1
- $x_{v(k,m)}$ = variabel prediktor
- t_{km} = nilai knots dari variabel prediktor $x_{v(k,m)}$
- $H[.]$ = 1 jika $[.] > 0$
= 0 jika lainnya.

Multivariate Adaptive Regression Spline (MARS) merupakan pengembangan model nonparametrik dari *Recursive Partition Regression* (RPR) yang

dikombinasikan dengan metode fungsi *spline* sehingga model yang dihasilkan kontinu pada knot. Friedman (1991) melakukan beberapa modifikasi pada RPR untuk membentuk model MARS, modifikasi yang dilakukan oleh Friedman dapat dilihat pada subbab 3.1.1.

3.1.1 Modifikasi Friedman

MARS merupakan hasil modifikasi dari *Recursive Partition Regression* (RPR) dengan beberapa modifikasi yang dilakukan oleh Friedman (1991) antara lain:

1. Mengganti *step function* $H[\pm(x - t)]$ dengan *truncated power* fungsi spline $[\pm(x - t)]_+^q$ dan $q = 1$ untuk mengatasi diskontinu pada titik knot, sehingga fungsi basis MARS dapat dinyatakan sebagai berikut:

$$B_m(x) = \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{(k,m)})]_+$$

2. Tidak menghapus fungsi basis setelah dipilah, dengan cara demikian induk dan pilahannya masih dapat dipilah kembali sehingga diperoleh subregion yang saling tumpang tindih.
3. Membatasi perkalian pada masing-masing fungsi basis dengan variabel-variabel prediktor yang berbeda.

Pada pemodelaan MARS, pemilihan knot menggunakan tahapan *forward* dan *backward*. Pada tahap *forward*, model dibangun untuk mendapatkan fungsi dengan jumlah fungsi basis maksimum. Kriteria pemilihan fungsi basis pada tahap *forward* yaitu meminimumkan *Penalized Least Squares* (PLS) dengan metode kuadrat terkecil. Sedangkan pada tahap *backward*, setelah diperoleh model dari tahap *forward*, selanjutnya akan dikeluarkan fungsi basis spline yang paling kecil kontribusinya sampai diperoleh model yang sederhana

dengan nilai *Generalized Cross Validation* (GCV) yang paling minimum. Sehingga model MARS dinyatakan dalam persamaan berikut :

$$y_i = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{(k,m)})]_+ + \epsilon \quad (3.1)$$

dengan estimator model MARS sebagai berikut:

$$f(x_i) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{(k,m)})]_+ \quad (3.2)$$

keterangan :

- $f(x_i)$ = variabel respon
- a_0 = konstanta fungsi basis
- a_m = koefisien untuk fungsi basis ke- m
- M = maksimum fungsi basis
- K_m = derajat interaksi
- $S_{(km)}$ = +1, jika knot terletak di kanan subregion
= -1, jika knot terletak di kiri subregion
- $x_{v(k,m)}$ = variabel prediktor pada pengamatan ke- v pada (k, m)
- t_{km} = nilai knot dari variabel prediktor $x_{v(k,m)}$

Jika Persamaan (3.1) diubah dalam bentuk matriks, maka akan menjadi

$$\mathbf{Y} = \mathbf{Ba} + \epsilon \quad (3.3)$$

dengan:

$$\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$$

$$\mathbf{a} = (a_0, a_1, a_2, \dots, a_M)^T$$

dan

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} (S_{1m}(x_{1(1,m)} - t_{1m})) & \cdots & \prod_{k=1}^{K_M} (S_{Mm}(x_{1(M,m)} - t_{Mm})) \\ 1 & \prod_{k=1}^{K_1} (S_{1m}(x_{2(1,m)} - t_{1m})) & \cdots & \prod_{k=1}^{K_M} (S_{Mm}(x_{2(M,m)} - t_{Mm})) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} (S_{1m}(x_{v(1,m)} - t_{1m})) & \cdots & \prod_{k=1}^{K_M} (S_{Mm}(x_{v(M,m)} - t_{Mm})) \end{bmatrix}$$

3.1.2 Estimasi Parameter

Teorema 3.1.1. Jika \mathbf{B} matriks non singular dan parameter pemulus $\delta^2 = 0$ maka penduga parameter model MARS ($\hat{\mathbf{a}}$) adalah

$$\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}.$$

Proof. Perhatikan Persamaan (2.3) yang digunakan untuk memperoleh fungsi pemulus spline $f(x_i)$ sebagai berikut:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \delta^2 \int_a^b (f^m(x))^2 dx$$

Jika parameter pemulus $\delta^2 = 0$, maka persamaan di atas mejadi persamaan *mean square error* (MSE) sebagai berikut:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2$$

Dari Persamaan (3.3), maka dapat diketahui bahwa $f(x) = \mathbf{B}\mathbf{a}$, sehingga persamaan di atas menjadi:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{B}\mathbf{a})^2 = (\mathbf{Y} - \mathbf{B}\mathbf{a})^T (\mathbf{Y} - \mathbf{B}\mathbf{a}) = Z \quad (3.4)$$

Untuk memperoleh estimator $\hat{\mathbf{a}}$ digunakan metode kuadrat terkecil sehingga dinyatakan persamaan sebagai berikut:

$$\begin{aligned} Z &= (\mathbf{Y} - \mathbf{B}\mathbf{a})^T (\mathbf{Y} - \mathbf{B}\mathbf{a}) \\ &= (\mathbf{Y}^T \mathbf{Y} - \mathbf{a}^T \mathbf{B}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{B}\mathbf{a} + \mathbf{a}^T \mathbf{B}^T \mathbf{B}\mathbf{a}) \\ &= (\mathbf{Y}^T \mathbf{Y} - 2\mathbf{a}^T \mathbf{B}^T \mathbf{Y} + \mathbf{a}^T \mathbf{B}^T \mathbf{B}\mathbf{a}) \end{aligned}$$

Persamaan diatas diturunkan parsial terhadap \mathbf{a} untuk meminimumkan Z dengan hasil penurunan sama dengan nol. Hal tersebut dilakukan untuk mem-

peroleh estimasi parameter $\hat{\mathbf{a}}$ maka hasilnya sebagai berikut:

$$\begin{aligned}\frac{\partial Z}{\partial \mathbf{a}} &= 0 \\ -2\mathbf{B}^T\mathbf{Y} + 2\mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} &= 0 \\ -\mathbf{B}^T\mathbf{Y} + \mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} &= 0 \\ \mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} &= \mathbf{B}^T\mathbf{Y}\end{aligned}$$

Jika \mathbf{B} matriks non *singular* maka estimasi parameter $\hat{\mathbf{a}}$ menjadi:

$$\hat{\mathbf{a}} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{Y} \quad (3.5)$$

□

Teorema 3.1.2. Jika \mathbf{B} matriks non singular dan parameter pemulus $\delta^2 > 0$ maka penduga parameter model MARS ($\hat{\mathbf{a}}$)

$$\hat{\mathbf{a}} = (\mathbf{B}^T\mathbf{B} + \delta^2\mathbf{D})^{-1}\mathbf{B}^T\mathbf{Y}.$$

Proof. Perhatikan Persamaan (2.3) yang digunakan untuk memperoleh fungsi pemulus spline $f(x_i)$ sebagai berikut:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \delta^2 \int_a^b (f^m(x))^2 dx$$

Perhatikan pula

$$\begin{aligned}f(x) &= \sum_{m=0}^M a_m B_m(x) \\ f'(x) &= \sum_{m=0}^M a_m B'_m(x) \\ f''(x) &= \sum_{m=0}^M a_m B''_m(x)\end{aligned}$$

maka $f^m(x) = \sum_{m=0}^M a_m B_m^m(x)$. Misalkan $\mathbf{B}^m(x)\mathbf{B}^m(x) = S(x)$ maka

$$\begin{aligned}
\int_a^b (f^m(x))^2 dx &= \int_a^b \left(\sum_{m=0}^M a_m B_m^m(x) \right)^2 dx \\
&= \int_a^b (\mathbf{B}^m \mathbf{a})^2 dx \\
&= \int_a^b ([B^m(x)\mathbf{a}]^T [B^m(x)\mathbf{a}]) dx \\
&= \mathbf{a}^T \mathbf{a} \int_a^b \mathbf{B}^m(x)\mathbf{B}^m(x) dx \\
&= \mathbf{a}^T \mathbf{a} \int_a^b S(x) dx \\
&= \mathbf{a}^T \mathbf{a} D
\end{aligned}$$

Sehingga

$$\int_a^b (f^m(x))^2 dx = \mathbf{a}^T \mathbf{a} D \quad (3.6)$$

dengan

$$d_{ij} = \int_a^b \left(\frac{\partial^m \mathbf{B}_i(x)}{\partial x^m} \right) \left(\frac{\partial^m \mathbf{B}_j(x)}{\partial x^m} \right) dx$$

Selanjutnya substitusikan Persamaan (3.4) dan Persamaan (3.6) ke Persamaan (2.3) maka diperoleh

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \delta^2 \int_a^b (f^m(x))^2 dx = (\mathbf{Y} - \mathbf{B}\mathbf{a})^T (\mathbf{Y} - \mathbf{B}\mathbf{a}) + \delta^2 \mathbf{a}^T \mathbf{a} D = ASR(\mathbf{a})$$

sehingga penduga model MARS $\hat{\mathbf{a}}$ diperoleh dengan meminimumkan $ASR(\mathbf{a})$.

$ASR(\mathbf{a})$ minimum maka $\frac{\partial ASR(\hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}} = 0$, sehingga

$$\begin{aligned}
\frac{\partial ASR(\hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}} &= \frac{(\mathbf{Y} - \mathbf{B}\hat{\mathbf{a}})^T (\mathbf{Y} - \mathbf{B}\hat{\mathbf{a}}) + \delta^2 \hat{\mathbf{a}}^T \hat{\mathbf{a}} D}{\partial \hat{\mathbf{a}}} \\
&= \frac{(\mathbf{Y}^T \mathbf{Y} - 2\hat{\mathbf{a}}^T \mathbf{B}^T \mathbf{Y} + \hat{\mathbf{a}}^T \mathbf{B}^T \mathbf{B} \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}} \\
&= -2\mathbf{B}^T \mathbf{Y} + 2\mathbf{B}^T \mathbf{B} \hat{\mathbf{a}} + 2\delta^2 D \hat{\mathbf{a}}
\end{aligned}$$

Karena $ASR(\hat{\mathbf{a}})$ minimum maka

$$\begin{aligned}
 -2\mathbf{B}^T\mathbf{Y} + 2\mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} + 2\delta^2 D\hat{\mathbf{a}} &= 0 \\
 -\mathbf{B}^T\mathbf{Y} + \mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} + \delta^2 D\hat{\mathbf{a}} &= 0 \\
 \mathbf{B}^T\mathbf{B}\hat{\mathbf{a}} + \delta^2 D\hat{\mathbf{a}} &= \mathbf{B}^T\mathbf{Y} \\
 (\mathbf{B}^T\mathbf{B} + \delta^2 D)\hat{\mathbf{a}} &= \mathbf{B}^T\mathbf{Y} \\
 \hat{\mathbf{a}} &= (\mathbf{B}^T\mathbf{B} + \delta^2 D)^{-1}\mathbf{B}^T\mathbf{Y}
 \end{aligned}$$

□

Untuk memperoleh estimator yang optimal maka diperlukan fungsi pemulus δ^2 yang optimal pula. Pemilihan δ^2 yang optimal diperoleh dari kriteria *Generalized Cross Validation* (GCV) yang paling minimum. Model MARS terbaik atau optimal merupakan model yang memiliki nilai kriteria GCV terkecil.

3.1.3 *Generalized Cross Validation* (GCV)

Model terbaik MARS dapat dilihat dari kriteria GCV terkecil, apabila nilai GCV terkecil lebih dari satu maka dapat dilihat berdasarkan nilai *Mean Square Error* (MSE) terkecil, dan apabila memiliki nilai GCV yang sama, maka selanjutnya dilihat berdasarkan nilai R^2 yang terbesar. Diperlukan matriks Hat untuk memperoleh kriteria GCV. Untuk memperoleh matriks Hat diperlukan teorema berikut:

Teorema 3.1.3. Misalkan \mathbf{R} matriks kuadratik dengan $\mathbf{R}^{-1}(\mathbf{R}^{-1})^T = \mathbf{X}^T\mathbf{X}$, dan \mathbf{B}^{-1} merupakan faktor *cholesky* dari $\mathbf{X}^T\mathbf{X}$. Misalkan pula \mathbf{U} dan \mathbf{Q} matriks diagonal sedemikian hingga $\mathbf{U}\mathbf{Q}^{-1}\mathbf{U}^T = \mathbf{R}\mathbf{D}\mathbf{R}^T$ dan $\mathbf{Z} = \mathbf{X}(\mathbf{R}^T\mathbf{U})$, serta $\lambda = (\mathbf{R}^T\mathbf{U})^{-1}\beta$ maka penyelesaian λ sebagai berikut:

$$\begin{aligned}
 (\mathbf{I} + \delta^2\mathbf{Q})\lambda &= \mathbf{Z}^T\mathbf{Y} \\
 &= (\mathbf{U}^T\mathbf{B})\mathbf{X}^T\mathbf{Y}
 \end{aligned}$$

Selanjutnya $\mathbf{X}\beta = \mathbf{Z}\lambda$ dan matriks Hat $S(\delta^2) = \mathbf{Z}(I + \delta^2\mathbf{Q})^{-1}\mathbf{Z}^T$ dengan derajat bebas berikut:

$$\begin{aligned} \text{trace}[S(\delta^2)] &= \text{trace}[(I + \delta^2\mathbf{Q})^{-1}\mathbf{Z}^T\mathbf{Z}] \\ &= \text{trace}[(I + \delta^2\mathbf{Q})^{-1}] \\ &= \sum_i (1 + \delta^2Q_i)^{-1} \end{aligned}$$

dengan Q_i merupakan matriks diagonal ke- i dari \mathbf{Q} .

Proof.

$$\begin{aligned} I + \delta^2D &= \mathbf{X}^T\mathbf{X} + \delta^2D \\ &= \mathbf{R}^{-1}(\mathbf{R}^{-1})^T + \delta^2D \\ &= \mathbf{R}^{-1}(\mathbf{R}^{-1})^T + \mathbf{R}^{-1}\mathbf{R}\delta^2D\mathbf{R}^T(\mathbf{R}^{-1})^T \\ &= \mathbf{R}^{-1}(I + \delta^2\mathbf{R}D\mathbf{R}^T)(\mathbf{R}^{-1})^T \\ &= \mathbf{R}^{-1}\mathbf{U}(I + \delta^2Q)\mathbf{U}^T(\mathbf{R}^{-1})^T \end{aligned}$$

Sehingga

$$\mathbf{U}^T\mathbf{R}(\mathbf{X}^T\mathbf{X} + \delta^2D)\mathbf{R}^T\mathbf{U} = I + \delta^2Q$$

Hal ini berarti,

$$\mathbf{U}^T\mathbf{R}(\mathbf{X}^T\mathbf{X} + \delta^2D)\mathbf{R}^T\mathbf{U}((\mathbf{R}^T\mathbf{U})^{-1}\beta) = \mathbf{U}^T\mathbf{R}(\mathbf{X}^T\mathbf{Y})$$

atau dengan kata lain

$$\begin{aligned} (I + \delta^2R)\lambda &= \mathbf{U}^T\mathbf{R}(\mathbf{X}^T\mathbf{Y}) \\ &= \mathbf{Z}^T\mathbf{Y} \end{aligned}$$

sehingga

$$\begin{aligned} \mathbf{X} &= \mathbf{Z}\mathbf{U}^T(\mathbf{R}^{-1})^T \\ \mathbf{X}\beta &= \mathbf{Z}\mathbf{U}^T(\mathbf{R}^{-1})^T\beta \\ &= \mathbf{Z}\lambda \end{aligned}$$

□

Teorema 3.1.4. Dengan menggunakan model MARS pada Persamaan (3.1), maka δ^2 yang optimal diperoleh dengan kriteria GCV sebagai berikut:

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i)]^2}{\left[1 - \frac{C(\widetilde{M})}{n}\right]^2}$$

dengan $C(\widetilde{M}) = C(M) + d.M$, $C(M) = \text{trace}(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) + 1$ dan $2 \leq d \leq 4$.

Proof. Perhatikan Persamaan (2.4)

$$GCV = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i)]^2}{\left[\frac{1}{n} \text{trace}(I - A(\delta^2))\right]^2}$$

Jumlah parameter model MARS selain basis induk, dapat diperoleh dengan substitusi Persamaan (3.5) ke dalam Persamaan (3.3) maka diperoleh sebagai berikut:

$$\begin{aligned} \mathbf{Y} &= \mathbf{B}\mathbf{a} \\ &= \mathbf{B}[(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}] \\ &= \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \\ &= \mathbf{P}\mathbf{Y} \end{aligned}$$

Sehingga $P = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T$ matriks berukuran $(M+1) \times (M+1)$. Karena P matriks simetris dan idempoten maka trace dari matriks \mathbf{P} sama dengan rank dari \mathbf{P} yang merupakan jumlah parameter fungsi basis selain konstanta dan jumlah parameter yang diestimasi, yang dinotasikan dengan

$$C(M) = \text{trace}(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) + 1$$

Selanjutnya perhatikan penyebut GCV pada Persamaan (2.4)

$$\left[\frac{1}{n} \text{trace}(I - A(\delta^2))\right]^2$$

dengan $A(\delta^2) = S(\delta^2)$ matriks Hat yang ada pada Teorema 3.1.3. dengan $S(\delta^2) = \mathbf{B}(\mathbf{B}^T\mathbf{B} + \delta^2\mathbf{D})^{-1}\mathbf{B}^T$, sehingga

$$\begin{aligned}
\left[\frac{1}{n}\text{trace}(I - A(\delta^2))\right]^2 &= \left[\frac{1}{n}\text{trace}(I - \mathbf{B}(\mathbf{B}^T\mathbf{B} + \delta^2\mathbf{D})^{-1}\mathbf{B}^T)\right]^2 \\
&= \left[1 - \frac{1}{n}\text{trace}(\mathbf{B}(\mathbf{B}^T\mathbf{B} + \delta^2\mathbf{D})^{-1}\mathbf{B}^T)\right]^2 \\
&= \left[1 - \frac{1}{n}\text{trace}[(\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T) + \mathbf{B}\delta^2\mathbf{D}^{-1}\mathbf{B}^T]\right]^2 \\
&= \left[1 - \frac{1}{n}[\text{trace}(\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + 1) + \text{trace}(\mathbf{B}\delta^2\mathbf{D}^{-1}\mathbf{B}^T)]\right]^2 \\
&= \left[1 - \frac{1}{n}[C(M) + d.M]\right]^2 \\
&= \left[1 - \frac{1}{n}[C(\widetilde{M})]\right]^2
\end{aligned}$$

Penambahan 1 pada $\text{trace}(\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)$ dikarenakan pada model MARS selalu melibatkan basis induk a_0 , sedangkan d disarankan bernilai $2 \leq d \leq 4$. \square

3.1.4 Uji Signifikansi

Apabila telah dipilih model MARS terbaik, selanjutnya akan dilakukan uji signifikansi terhadap model MARS yang terpilih tersebut untuk mengevaluasi kecocokan model. Pengujian dilakukan dengan menguji koefisien regresi secara simultan (uji F) maupun parsial (uji t).

Uji Simultan

1. Rumusan hipotesis

H_0 : $a_1 = a_2 = \dots = a_M = 0$, model dikatakan tidak signifikan.

H_1 : minimal terdapat satu $a_m \neq 0$, dengan $m = 1, 2, \dots, M$, model dikatakan signifikan

2. Taraf signifikansi : α

3. Statistik uji:

$$F_{hitung} = \frac{\sum_{i=1}^n \frac{(\hat{y}_i - \hat{y})^2}{M}}{\sum_{i=1}^n \frac{(\hat{y}_i - \hat{y})^2}{N - M - 1}}$$

4. Daerah Kritis

tolak H_0 jika $F_{hitung} > f_{\alpha(M;N-M-1)}$ atau P-Value $< \alpha$

Uji Parsial

1. Rumusan hipotesis:

H_0 : $a_m = 0$, berarti koefisien a_m tidak berpengaruh terhadap model dengan $m = 1, 2, \dots, M$.

H_1 : $a_m \neq 0$, berarti koefisien a_m berpengaruh terhadap model dengan $m = 1, 2, \dots, M$.

2. Taraf signifikansi : α

3. Statistik uji:

$$t_{hitung} = \frac{\hat{\alpha}_m}{se(\hat{\alpha}_m)}$$

dengan

$$se(\hat{\alpha}_m) = \sqrt{var(\hat{\alpha}_m)}$$

4. Daerah kritis: Tolak H_0 jika $t_{hitung} > t_{(\frac{\alpha}{2}; N-M)}$ atau P-Value $< \alpha$ dengan N adalah jumlah pengamatan dan M adalah jumlah parameter.

3.2 Cox *Proportional Hazard* dengan MARS

Multivariate Adaptive Regression Spline (MARS) pada umumnya dapat digunakan pada 2 tipe variabel respon, yaitu biner atau kontinu. Menurut

Kriner (2007), residual *martingale* pada pemodelan *Cox Proportional Hazard* (Cox PH) dapat digunakan sebagai variabel respon kontinu pada pemodelan MARS. Residual *martingale* diperoleh dengan mengestimasi terlebih dahulu fungsi survival dan fungsi *hazard* kumulatif menggunakan metode Kaplan-Meier. Selanjutnya dilakukan pemodelan MARS dengan variabel responnya yaitu residual *martingale* yang telah diperoleh sebelumnya. Oleh karena itu, pemodelan Cox PH dengan MARS dapat diartikan sebagai pemodelan MARS dengan variabel responnya adalah residual *martingale* dari Cox PH.

3.3 Langkah Penelitian

Berdasarkan diagram alir pada Gambar 3.1, maka langkah-langkah yang digunakan dalam penelitian ini disesuaikan dengan permasalahan yang diambil, yaitu sebagai berikut:

1. Menentukan status data pada waktu survival dengan nilai $\delta = 1$ jika bayi mengalami *failure event* dan $\delta = 0$ jika bayi tidak mengalami *failure event* selama periode penelitian atau berhasil melewati usia 12 bulan.
2. Melakukan uji asumsi *proportional hazard* untuk data yang berkategori berdasarkan pada plot $\ln(-\ln(S(t)))$. Jika asumsi *proportional hazard* terpenuhi maka lanjut ke langkah berikutnya.
3. Mengestimasi fungsi survival dan fungsi *hazard* kumulatif untuk mengetahui laju kematian bayi dan tingkat kematian bayi pada waktu t .
4. Mengestimasi residual *martingale* untuk data yang tidak tersensor dengan fungsi *hazard* kumulatif yang telah diperoleh pada metode Kaplan-Meier.

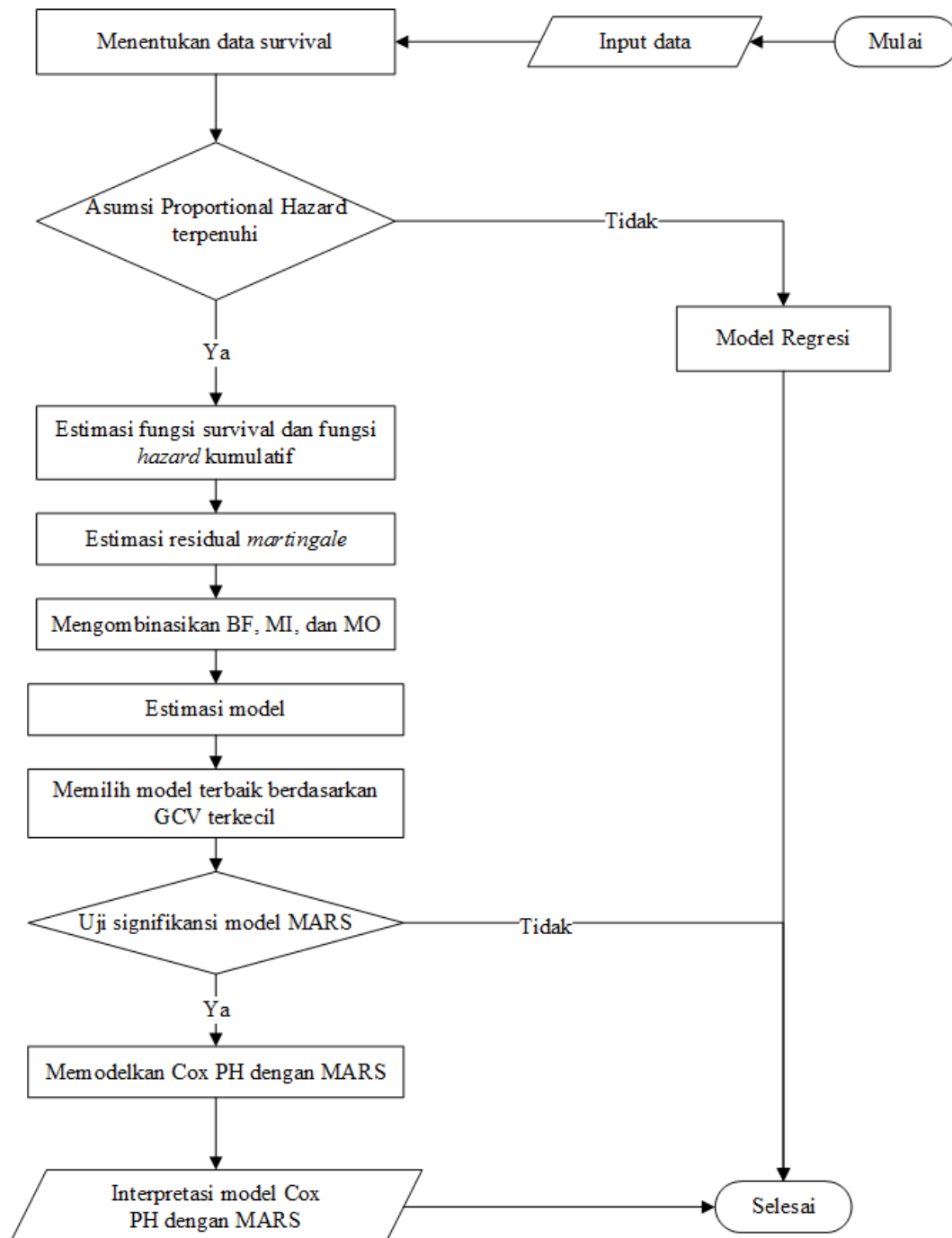


Figure 3.1: Diagram alir pemodelan Cox PH dengan MARS

5. Melakukan pemodelan MARS dengan residual *martingale* sebagai variabel respon dan kombinasi antara Basis Fungsi (16, 24, dan 32), Maksimum Interaksi (1, 2, dan 3) dan Minimum Observasi (0, 1, 2, dan 3).
6. Mengestimasi model MARS dengan masing-masing kombinasi BF, MI, dan MO yang terbentuk.
7. Memilih model MARS terbaik dengan nilai GCV yang minimum.
8. Melakukan uji signifikansi terhadap model MARS yang terpilih untuk mengevaluasi kecocokan model dengan uji koefisien regresi secara simultan (Uji F) maupun secara parsial (Uji t). Apabila kecocokan model terpenuhi maka dilakukan langkah selanjutnya.
9. Memodelkan *Cox Proportional Hazard* dengan pendekatan model MARS yang telah diketahui.
10. Interpretasi model yang diperoleh berdasarkan basis fungsi yang telah didapatkan.

3.4 *Salford Predictive Modeler (SPM) 8*

Software yang digunakan pada penelitian ini yaitu *Salford Predictive Modeler (SPM) 8*. SPM 8 merupakan *software* yang membantu dalam pengerjaan dalam memodelkan *Multivariate Adaptive Regression Spline (MARS)*. Adapun langkah-langkah membangun model MARS pada *software* SPM 8 sebagai berikut:

1. Buka file data yang ingin diolah dalam bentuk format .csv dengan klik **file** lalu **open** atau Ctrl+O.

2. Selanjutnya akan terdapat *record* mengenai banyaknya data, variabel penelitian, karakter dan numerik pada data. Pada tahap ini klik **Model**.
3. Tampilan Model Setup dapat dilihat pada Gambar . Selanjutnya pilih *sort* yaitu *File Order*, pilih target yaitu variabel respon yang idgunakan, centang variabel prediktor yang digunakan pada kolom *predictor*, lalu pilih *analysis type* yaitu *regression*, dan *analysis method* yaitu MARS.

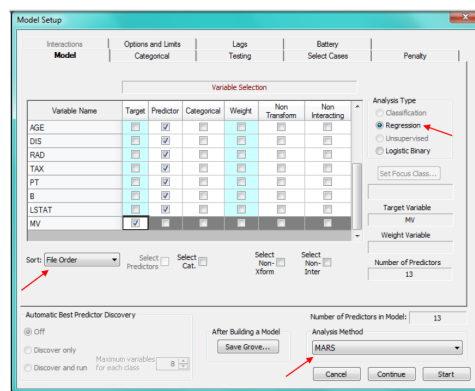
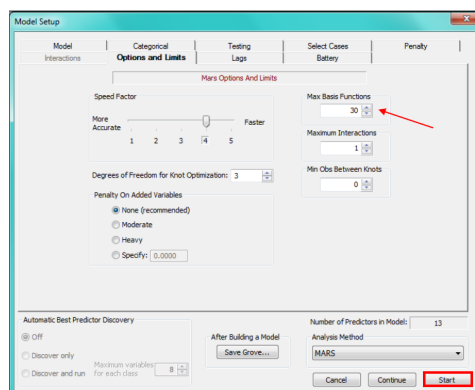


Figure 3.2: Model dalam SPM 8

4. Klik *Options and Limits* dan *input* kombinasi model BF, MI, dan MO yang sudah ditentukan, lalu klik *start*.

Figure 3.3: *Options and Limits* dalam SPM 8

3.5 Sumber Data

Sumber data yang digunakan dalam penelitian ini yaitu data sekunder yang bersumber dari Survei Penduduk Antar Sensus (SUPAS) 2015 di provinsi Banten. SUPAS 2015 merupakan survei yang dilakukan oleh BPS diantara dua waktu sensus penduduk. Badan Pusat Statistik (BPS) telah 5 kali melakukan SUPAS yaitu pada tahun 1976, 1985, 1995, 2005, dan 2015. Unit analisis yang digunakan dalam penelitian ini yaitu bayi yang lahir hidup antara Januari 2013 sampai Desember 2014 (bayi berusia 0-24 bulan). Unit analisis dianggap mengalami *failure event* apabila bayi mengalami kematian sebelum berusia genap 1 tahun atau kurang dari 12 bulan. Data kelahiran dan kematian bayi yang digunakan pada penelitian ini merupakan data yang diperoleh dari kuesioner SUPAS 2015 pertanyaan 638, 808 dan pertanyaan pada blok VII.C yaitu pertanyaan 716-723 yang dapat dilihat pada Lampiran 1.

3.6 Variabel Penelitian

Variabel-variabel penelitian yang digunakan dalam penelitian ini terdiri satu variabel respon dan 8 variabel prediktor meliputi usia ibu saat melahirkan, paritas ibu, jarak antar kelahiran, pendidikan ibu, jenis kelamin bayi, kelahiran kembar, penolong kelahiran, serta sumber utama air minum yang dapat dilihat pada Tabel 3.1.

3.6.1 Variabel respon

Variabel respon pada penelitian ini adalah usia bayi bertahan hidup (T) dari periode Januari 2013-Desember 2014 yang dicacah dalam satuan bulan dengan kategori yang dijelaskan pada halaman selanjutnya.

Table 3.1: Daftar variabel yang digunakan dalam penelitian

| Variabel | Uraian | Skala data | Kode/Nilai |
|----------|--------------------------|------------|--|
| T | Usia bayi bertahan hidup | Rasio | Bulan |
| X_1 | Usia ibu saat melahirkan | Ordinal | 1 : kurang dari 20 tahun 2 : antara 20-35 tahun 3 : lebih dari 35 tahun |
| X_2 | Paritas ibu | Rasio | kelahiran |
| X_3 | Jarak antar kelahiran | Rasio | tahun |
| X_4 | Pendidikan ibu | Ordinal | 1 : Tidak punya ijazah SD 2 : SD 3 : SMP 4 : SMA 5 : Perguruan tinggi |
| X_5 | Jenis kelamin bayi | Nominal | 1 : Laki-laki 2 : Perempuan |
| X_6 | Kelahiran kembar | Nominal | 1 : Tidak 2 : Ya |
| X_7 | Penolong kelahiran | Nominal | 1 : Bantuan non-medis 2 : Bantuan medis |
| X_8 | Sumber utama air minum | Nominal | 1 : Air kemasan 2 : Air isi ulang 3 : Leding sampai rumah 4 : Leding eceran 5 : Pompa 6 : Sumur terlindung 7 : Sumur tak terlindung 8 : Mata air terlindung 9 : Mata air tak terlindung 10 : Air sungai 11 : Air hujan |

1. $\delta = 0$

Data dikatakan tersensor jika bayi tidak mengalami *failure event*, yaitu bayi yang masih hidup hingga periode penelitian atau berhasil melewati usia 12 bulan.

2. $\delta = 1$

Data dikatakan tidak tersensor jika bayi mengalami kematian sebelum usia satu tahun atau kurang dari 12 bulan.

3.6.2 Variabel prediktor

Variabel prediktor adalah variabel yang digunakan untuk menduga variabel respon. Adapun variabel prediktor yang digunakan dalam penelitian ini sebanyak delapan variabel antara lain:

1. Usia ibu saat melahirkan (X_1) yaitu usia ibu kandung bayi saat melahirkan bayi. Kategori dari usia ibu saat melahirkan yaitu
 - 1 = usia ibu saat melahirkan bayi kurang dari 20 tahun
 - 2 = usia ibu saat melahirkan bayi antara 20-35 tahun
 - 3 = usia ibu saat melahirkan lebih dari 35 tahun
2. Paritas (X_2) yaitu jumlah kelahiran yang telah dialami ibu sebelum kelahiran bayi.
3. Jarak antar kelahiran (X_3) yaitu jarak kelahiran bayi dengan kelahiran yang sebelumnya dalam satuan tahun.
4. Pendidikan ibu (X_4) yaitu pendidikan atau ijazah terakhir yang dimiliki oleh ibu. Kategori dari pendidikan ibu yaitu:

- 1 = Tidak punya ijazah SD
 - 2 = SD
 - 3 = SMP
 - 4 = SMA
 - 5 = Perguruan tinggi
5. Jenis kelamin bayi (X_5) yaitu jenis kelamin dari unit analisis. Variabel jenis kelamin bayi dikategorikan menjadi:
- 1 = Laki-laki
 - 2 = Perempuan
6. Kelahiran kembar (X_6) yaitu bayi yang dilahirkan kembar Variabel bayi kembar dikategorikan menjadi:
- 1 = Bayi tidak kembar/Bayi tunggal
 - 2 = Bayi kembar
7. Penolong kelahiran (X_7) yaitu penolong kelahiran saat ibu melahirkan, dengan kategori berikut:
- 1 = Bantuan non-medis
 - 2 = Bantuan medis
8. Sumber utama air minum (X_8) yaitu sumber utama air yang dikonsumsi oleh keluarga dikategorikan menjadi
- 1 = Air kemasan
 - 2 = Air isi ulang
 - 3 = Leding sampai rumah
 - 4 = Leding eceran
 - 5 = Pompa
 - 6 = Sumur terlindung

- 7 = Sumur tak terlindung
- 8 = Mata air terlindung
- 9 = Mata air tak terlindung
- 10 = Air sungai
- 11 = Air hujan

3.7 Karakteristik Kematian Bayi

Penelitian ini menggunakan data sekunder yang berasal dari *Survei Penduduk Antar Sensus (SUPAS) 2015*. Wilayah yang diamati mencakup seluruh kabupaten dan kota di provinsi Banten. Data survival diperoleh dari catatan kelahiran hidup yang terjadi sejak Januari 2013 hingga Desember 2014. Selama interval waktu tersebut, tercatat 2.391 bayi lahir hidup dengan jumlah bayi yang bertahan hidup hingga berusia 12 bulan ke atas atau hingga akhir pengamatan atau yang tersensor berjumlah 2.317 bayi dan 74 bayi lainnya tidak tersensor karena mengalami kematian sebelum usia 1 tahun atau saat usia 0-11 bulan. Perbandingan amatan tersensor dan tidak tersensor di Provinsi Banten dapat dilihat pada Gambar (3.4).

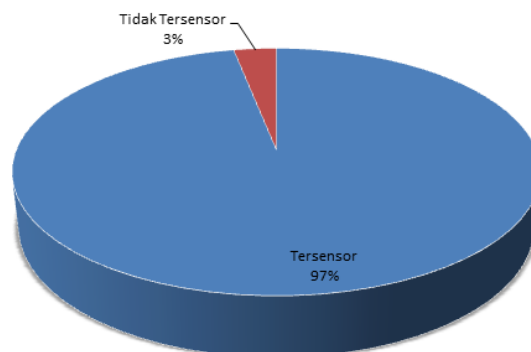


Figure 3.4: Perbandingan Amatan Data Tersensor dan Tidak Tersensor

Perhatikan Gambar 3.4., terlihat bahwa persentase bayi mati sebelum usia

1 tahun sebesar 3%. Hal tersebut masih belum mencapai target *Millenium Development Goals* (MDGs) untuk tahun 2015 yaitu 23 per 1.000 kelahiran hidup atau sebesar 2,3%. Oleh karena itu akan dilakukan penelitian mengenai faktor-faktor yang diduga menyebabkan kematian bayi.

Untuk mengetahui karakteristik bayi lahir hidup akan dilakukan analisis statistika deskriptif terhadap status bayi berdasarkan masing-masing variabel yang ada pada penelitian ini meliputi usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pendidikan ibu (X_4), jenis kelamin bayi (X_5), kelahiran kembar (X_6), penolong kelahiran bayi (X_7), dan sumber utama air minum (X_8).

Pada Table 3.2 menunjukkan jumlah dan persentase status bayi berdasarkan usia ibu saat melahirkan dengan persentase bayi yang mati sebelum 1 tahun dengan usia ibu saat melahirkan lebih dari 35 tahun lebih besar dibanding persentase bayi yang mati sebelum usia 1 tahun dengan usia ibu melahirkan antara 20-35 tahun atau kurang dari 20 tahun.

Table 3.2: Perbandingan Status Bayi Lahir Hidup Menurut Usia Ibu Saat Melahirkan

| | | Status Bayi | | Total |
|--------------------------|----------------------|------------------|---------------------------|----------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| usia ibu saat melahirkan | kurang dari 20 tahun | 216 (97,74%) | 5 (2,26%) | 221 (100%) |
| | antara 20-35 tahun | 1760 (97,62%) | 43 (2,38%) | 1803 (100%) |
| | lebih dari 35 tahun | 341 (92,92%) | 26 (7,08%) | 367 (100%) |
| Total | | | | 2391 |

Pada Gambar 3.5, diketahui bahwa nilai minimum dari paritas ibu yaitu 0 kelahiran sebanyak 925 ibu, artinya sebanyak 925 ibu melahirkan anak pertamanya sedangkan nilai maksimum dari paritas ibu yaitu 11 kelahiran sebanyak 3 ibu, artinya terdapat 3 ibu yang sebelumnya telah melahirkan sebanyak 11

kali atau dengan kata lain terdapat 3 ibu yang melahirkan anak ke-12.

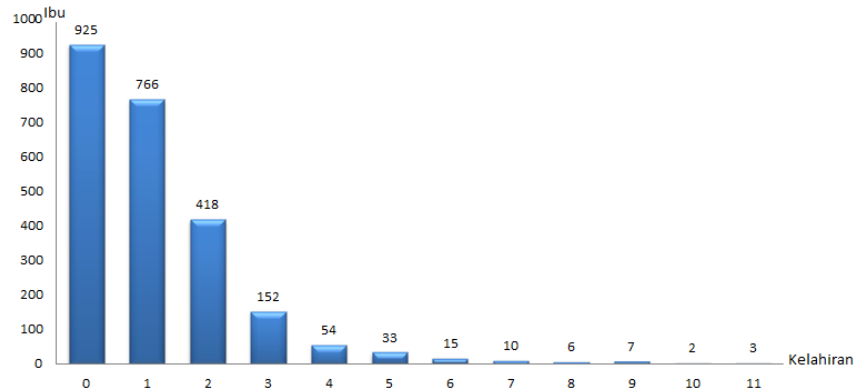


Figure 3.5: Jumlah Paritas Ibu

Pada Gambar 3.6 menunjukkan bahwa nilai minimum dari jarak antar kelahiran adalah 0 tahun sebanyak 925 ibu, artinya terdapat 925 ibu yang melahirkan anak pertama. Sedangkan nilai maksimum dari jarak antar kelahiran adalah 30 tahun sebanyak 1 ibu, artinya terdapat seorang ibu yang melahirkan bukan anak pertama dengan jarak kelahiran terakhir kali yaitu 30 tahun.

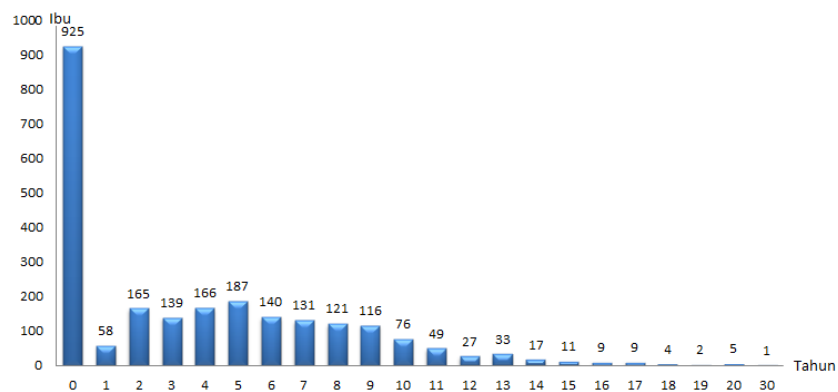


Figure 3.6: Jumlah Jarak Antar Kelahiran

Pada Tabel 3.3 dapat dilihat bahwa persentase bayi mati sebelum 1 tahun

Table 3.3: Perbandingan Status Bayi Lahir Berdasarkan Pendidikan Ibu

| | | Status Bayi | | Total |
|----------------|-------------|-----------------|---------------------------|---------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| Pendidikan ibu | Tidak punya | 176 (94,12%) | 11 (5,88%) | 187 (100%) |
| | SD | 632 (97,23%) | 18 (2,77%) | 650 (100%) |
| | SMP | 567 (97,26%) | 16 (2,74%) | 583 (100%) |
| | SMA | 680 (97,28%) | 19 (2,72%) | 699 (100%) |
| | PT | 262 (96,32%) | 10 (3,68%) | 272 (100%) |

dengan pendidikan ibu yang tidak mempunyai ijazah SD atau tidak tamat SD lebih tinggi daripada kematian bayi dengan pendidikan ibu yang lainnya.

Table 3.4: Perbandingan Status Bayi Berdasarkan Jenis Kelamin Bayi

| | | Status Bayi | | Total |
|--------------------|-----------|------------------|---------------------------|----------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| Jenis kelamin bayi | Laki-laki | 1228 (96,54%) | 44 (3,46%) | 1272 (100%) |
| | Perempuan | 1089 (97,32%) | 30 (2,68%) | 1119 (100%) |

Pada Tabel 3.4 terlihat bahwa persentase bayi laki-laki mati sebelum usia 1 tahun lebih besar dibanding persentase bayi perempuan mati sebelum usia 1 tahun.

Pada Tabel 3.5 menunjukkan bahwa persentase bayi lahir kembar yang mati sebelum usia 1 tahun lebih rendah daripada persentase bayi lahir tidak kembar atau lahir tunggal yang mati sebelum usia 1 tahun.

Table 3.5: Perbandingan Status Bayi Berdasarkan Kelahiran Kembar

| | | Status Bayi | | Total |
|------------------|-------|------------------|---------------------------|----------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| Kelahiran kembar | Tidak | 2283 (96,90%) | 73 (3,10%) | 2356 (100%) |
| | Ya | 34 (97,14%) | 1 (2,86%) | 35 (100%) |

Berdasarkan Tabel 3.6 dapat dilihat bahwa persentase bayi lahir sebelum usia 1 tahun dengan penolong kelahiran bayi berupa bantuan non-medis lebih tinggi dibanding dengan penolong kelahiran berupa bantuan medis.

Table 3.6: Perbandingan Status Bayi Berdasarkan Penolong Kelahiran

| | | Status Bayi | | Total |
|--------------------|-----------|------------------|---------------------------|----------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| Penolong kelahiran | non-medis | 525 (95,456%) | 25 (4,54%) | 550 (100%) |
| | medis | 1792 (97,34%) | 49 (2,66%) | 1841 (100%) |

Berdasarkan Tabel 3.7 dapat dilihat bahwa persentase bayi mati sebelum usia 1 tahun paling tinggi dialami oleh keluarga yang memiliki sumber utama air minum berupa leding eceran, sedangkan persentase bayi mati sebelum usia 1 tahun paling rendah dialami oleh keluarga yang menggunakan sumber utama air minum berupa air kemasan.

Table 3.7: Perbandingan Status Bayi Berdasarkan Sumber Utama Air Minum

| | | Status Bayi | | Total |
|------------------------|---------------|-----------------|---------------------------|---------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| sumber utama air minum | air kemasan | 429 (98,17%) | 8 (1,83%) | 437 (100%) |
| | air isi ulang | 726 (96,67%) | 25 (3,33%) | 751 (100%) |

Lanjutan Tabel 3.7

| | | Status Bayi | | Total |
|------------------------|----------------------------|-----------------|---------------------------|---------------|
| | | Bertahan hidup | Mati sebelum usia 1 tahun | |
| sumber utama air minum | leding sampai rumah | 57 (100%) | 0 (0%) | 57 (100%) |
| | leding eceran | 27 (90%) | 3 (10%) | 30 (100%) |
| | pompa air | 454 (97,22%) | 13 (2,78%) | 467 (100%) |
| | sumur terlindungi | 333 (96,52%) | 12 (3,48%) | 345 (100%) |
| | sumur tidak terlindungi | 113 (94,17%) | 7 (5,83%) | 120 (100%) |
| | mata air terlindungi | 77 (97,47%) | 2 (2,53%) | 79 (100%) |
| | mata air tidak terlindungi | 74 (96,10%) | 3 (3,90%) | 77 (100%) |
| | air sungai | 19 (95%) | 1 (5%) | 20 (100%) |
| | air hujan | 8 (100%) | 0 (0%) | 8 (100%) |

3.8 Asumsi *Proportional Hazard*

Asumsi yang harus dipenuhi dalam *proportional hazard* adalah fungsi *hazard* dari variabel prediktor yang bersifat kategorik harus *proportional* setiap waktu. Asumsi *proportional hazard* terpenuhi apabila garis antar kategori pada plot $\ln(-\ln S(t))$ untuk setiap variabel kategorik tidak saling berpotongan.

Berdasarkan Gambar 3.7. bagian (a) dapat dilihat plot $\ln(-\ln S(t))$ pada usia ibu saat melahirkan dengan 3 kategori, bagian (b) plot $\ln(-\ln S(t))$ pada paritas ibu dengan 5 kategorik. Kedua plot tersebut tidak saling berpotongan pada setiap kategorinya, hal itu berarti asumsi *proportional hazard* untuk dua variabel tersebut telah terpenuhi, atau dengan kata lain, dua variabel tersebut telah independen terhadap waktu serta hubungan antara kumulatif hazard sudah proporsional/ konstan setiap waktu.

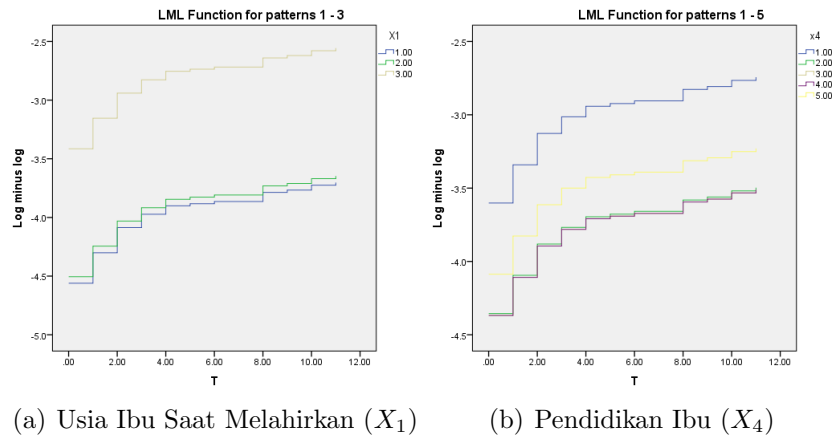


Figure 3.7: Asumsi *Proportional Hazard* pada Variabel X_1 dan X_4

Berdasarkan Gambar 3.8 pada bagian (a) terlihat plot $\ln(-\ln S(t))$ pada jenis kelamin bayi dengan 2 kategori, bagian (b) plot $\ln(-\ln S(t))$ pada kelahiran kembar dengan 2 kategori. Kedua plot tersebut tidak saling berpotongan pada setiap kategorinya, hal itu berarti asumsi *proportional hazard* untuk dua variabel tersebut telah terpenuhi, atau dengan kata lain, dua variabel tersebut telah independen terhadap waktu serta hubungan antara kumulatif hazard sudah proporsional/ konstan setiap waktu.

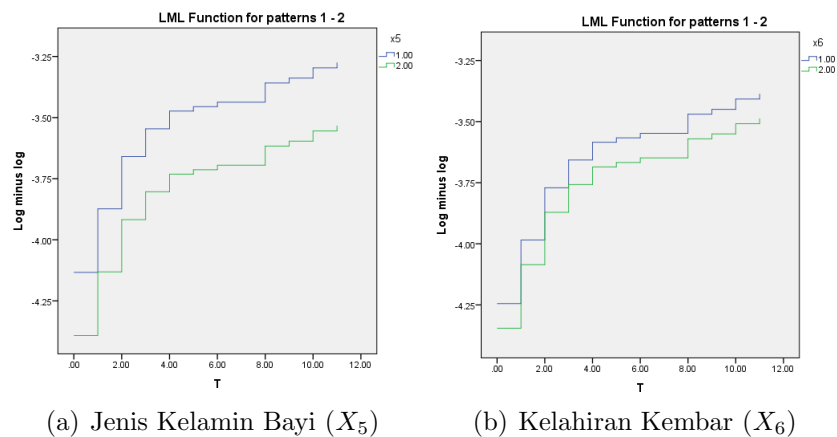


Figure 3.8: Asumsi *Proportional Hazard* pada Variabel X_5 dan X_6

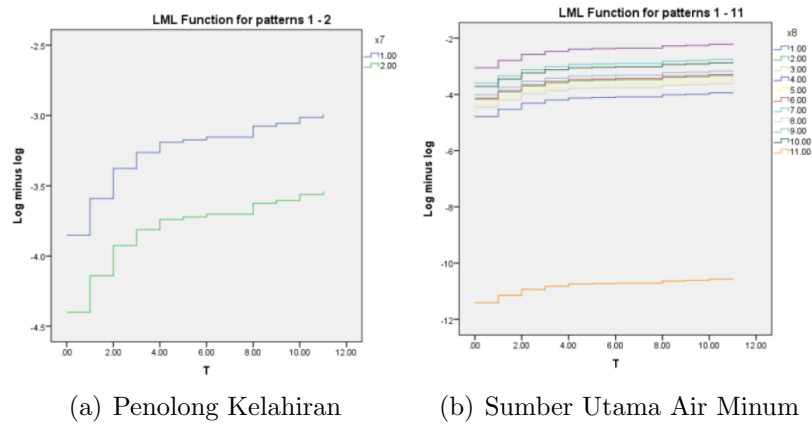


Figure 3.9: Asumsi *Proportional Hazard* pada Variabel X_7 dan X_8

Berdasarkan Gambar 3.9 pada bagian (a) merupakan plot $\ln(-\ln S(t))$ pada penolong kelahiran dengan 2 kategori, dan bagian (b) merupakan plot $\ln(-\ln S(t))$ pada sumber utama air minum dengan 11 kategori. Kedua plot tersebut tidak saling berpotongan pada setiap kategorinya, hal itu berarti asumsi *proportional hazard* untuk dua variabel tersebut telah terpenuhi, atau dengan kata lain, dua variabel tersebut telah independen terhadap waktu serta hubungan antara kumulatif hazard sudah proporsional/ konstan setiap waktu.

3.9 Estimasi Fungsi Survival dan Fungsi

Hazard

Fungsi survival digunakan untuk mengetahui probabilitas bayi untuk bertahan hidup hingga usia t bulan dan fungsi *hazard* digunakan untuk memberikan informasi tentang kegagalan bayi saat berusia t bulan untuk bertahan hidup. Estimasi fungsi survival dan fungsi *hazard* pada penelitian ini diperoleh berdasarkan metode Kaplan-Meier. Hasil estimasi dari fungsi survival dan fungsi hazard hanya dilihat berdasarkan usia kelahiran bayi hidup tanpa

dipengaruhi oleh variabel-variabel lainnya. Hasil estimasi fungsi survival dan fungsi *hazard* dapat dilihat pada Tabel 3.8.

Table 3.8: Estimasi Fungsi Survival dan Fungsi *Hazard*

| Waktu Survival (bulan) | $\widehat{S}(t)$ | $\widehat{h}(t)$ |
|---------------------------|------------------|------------------|
| 0 | 0,986 | 0,014 |
| 1 | 0,982 | 0,004 |
| 2 | 0,977 | 0,004 |
| 3 | 0,975 | 0,002 |
| 4 | 0,972 | 0,002 |
| 5 | 0,972 | 0,001 |
| 6 | 0,972 | 0,001 |
| 8 | 0,969 | 0,001 |
| 9 | 0,969 | 0,001 |
| 10 | 0,967 | 0,001 |
| 11 | 0,967 | 0,001 |

Pada Tabel 3.8 dapat terlihat bahwa nilai survival bayi usia 8 bulan sebesar 0,969, berarti peluang seorang bayi untuk bertahan hidup hingga usia 8 bulan yaitu sebesar 96,9%, atau dengan kata lain jika dalam suatu periode terdapat 1.000 orang bayi maka terdapat 969 bayi yang akan bertahan hidup hingga usia 8 bulan. Sedangkan nilai *hazard* atau laju kematian bayi pada usia 8 bulan yaitu sebesar 0,001, berarti peluang seorang bayi usia 8 bulan mengalami kematian adalah 0,1%, atau dengan kata lain jika dalam suatu periode terdapat 1.000 bayi yang mampu bertahan hingga usia 8 bulan, maka 1 bayi akan mengalami kematian sebelum berusia 9 bulan. Berdasarkan Tabel 3.8 juga dapat diketahui bahwa peluang kematian bayi paling tinggi dialami oleh usia bayi bertahan hidup 0 bulan yaitu sebanyak 14 dari 1.000 bayi yang mengalami kematian sebelum usia 1 bulan, serta semakin bertambah usia bayi bertahan hidup maka peluang bayi mengalami kematian semakin rendah. Berdasarkan hasil estimasi fungsi survival dapat diperoleh estimasi fungsi *hazard* kumulatif yang akan digunakan untuk memperoleh estimasi residual *martingale*.

3.10 Estimasi Residual *Martingale*

Pada penelitian ini, residual *martingale* digunakan sebagai variabel respon pada model MARS. Residual *martingale* diperoleh dari estimasi fungsi *hazard* kumulatif berdasarkan usia bayi bertahan hidup tanpa dipengaruhi oleh variabel-variabel lain. Hasil estimasi residual *martingale* pada data tidak tersensor dapat dilihat pada Tabel 3.9.

Table 3.9: Estimasi Residual *Martingale*

| Waktu Survival (bulan) | Fungsi Hazard Kumulatif ($\hat{H}(t)$) | Residual Martingale (RM) |
|---------------------------|---|-----------------------------|
| 0 | 0,006 | 0,994 |
| 1 | 0,008 | 0,992 |
| 2 | 0,010 | 0,990 |
| 3 | 0,011 | 0,989 |
| 4 | 0,012 | 0,988 |
| 5 | 0,012 | 0,988 |
| 6 | 0,012 | 0,988 |
| 8 | 0,013 | 0,987 |
| 9 | 0,014 | 0,986 |
| 10 | 0,014 | 0,986 |
| 11 | 0,015 | 0,985 |

Setelah diperoleh residual *martingale*, selanjutnya melakukan pemodelan MARS berdasarkan kombinasi fungsi basis (BF), maksimum interaksi (MI), dan minimum observasi (MO) dengan residual *martingale* sebagai respon.

3.11 *Trial and Error* Model MARS

Pemodelan MARS akan diterapkan pada penelitian ini dengan residual *martingale* sebagai variabel respon (Y) dengan variabel-variabel prediktor yang telah dijelaskan sebelumnya. Langkah selanjutnya yaitu mengestimasi model MARS berdasarkan kombinasi dari fungsi basis (BF), maksimum interaksi (MI), dan minimum observasi (MO) dengan nilai dari BF sebesar 16, 24, dan

32, sedangkan nilai MI sebesar 1, 2, dan 3, serta nilai MO sebesar 0, 1, 2, dan 3. Terdapat 36 kombinasi model MARS dari kombinasi BF, MI, dan MO. Model MARS terbaik merupakan model yang memiliki nilai GCV paling minimum. Hasil GCV dari 36 kombinasi tersebut dapat dilihat pada Tabel 3.10.

Table 3.10: Hasil GCV Berdasarkan Kombinasi BF, MI, dan MO

| Kombinasi ke- | Fungsi Basis (BF) | Maksimum Interaksi (MI) | Minimum Observasi (MO) | <i>Generalized Cross Validation</i> (GCV) |
|---------------|-------------------|-------------------------|------------------------|---|
| 1 | 16 | 1 | 0 | 0,00062 |
| 2 | 16 | 2 | 0 | 0,00062 |
| 3 | 16 | 3 | 0 | 0,00062 |
| 4 | 16 | 1 | 1 | 0,00062 |
| 5 | 16 | 2 | 1 | 0,00062 |
| 6 | 16 | 3 | 1 | 0,00058 |
| 7 | 16 | 1 | 2 | 0,00062 |
| 8 | 16 | 2 | 2 | 0,00062 |
| 9 | 16 | 3 | 2 | 0,00059 |
| 10 | 16 | 1 | 3 | 0,00062 |
| 11 | 16 | 2 | 3 | 0,00062 |
| 12 | 16 | 3 | 3 | 0,00055 |
| 13 | 24 | 1 | 0 | 0,00062 |
| 14 | 24 | 2 | 0 | 0,00062 |
| 15 | 24 | 3 | 0 | 0,00062 |
| 16 | 24 | 1 | 1 | 0,00062 |
| 17 | 24 | 2 | 1 | 0,00062 |
| 18 | 24 | 3 | 1 | 0,00056 |
| 19 | 24 | 1 | 2 | 0,00062 |
| 20 | 24 | 2 | 2 | 0,00062 |
| 21 | 24 | 3 | 2 | 0,00055 |
| 22 | 24 | 1 | 3 | 0,00062 |
| 23 | 24 | 2 | 3 | 0,00062 |
| 24 | 24 | 3 | 3 | 0,00058 |
| 25 | 32 | 1 | 0 | 0,00062 |
| 26 | 32 | 2 | 0 | 0,00062 |
| 27 | 32 | 3 | 0 | 0,00062 |
| 28 | 32 | 1 | 1 | 0,00062 |
| 29 | 32 | 2 | 1 | 0,00062 |
| 30 | 32 | 3 | 1 | 0,00053* |

Lanjutan Tabel 3.10

| Kombinasi ke- | Fungsi Basis (BF) | Maksimum Interaksi (MI) | Minimum Observasi (MO) | <i>Generalized Cross Validation</i> (GCV) |
|---------------|-------------------|-------------------------|------------------------|---|
| 31 | 32 | 1 | 2 | 0,00062 |
| 32 | 32 | 2 | 2 | 0,00062 |
| 33 | 32 | 3 | 2 | 0,00059 |
| 34 | 32 | 1 | 3 | 0,00062 |
| 35 | 32 | 2 | 3 | 0,00062 |
| 36 | 32 | 3 | 3 | 0,00058 |

*) GCV minimum

3.12 Model MARS Terbaik

Model MARS terbaik adalah model yang menghasilkan nilai GCV minimum. Nilai GCV dihasilkan dari mengombinasikan jumlah basis fungsi (BF), maksimum interaksi (MI), dan minimum observasi (MO) dengan total kombinasi sebanyak 36 model. Tabel 3.10 menunjukkan bahwa dari 36 kombinasi yang terbentuk, nilai GCV minimum diperoleh dari kombinasi ke-30 yaitu dengan nilai basis fungsi (BF) 32, maksimum interaksi (MI)=3, dan maksimum observasi (MO)=1 dengan nilai GCV sebesar 0,00053.

Pada tahap *forward* kombinasi ke-30 tersebut menghasilkan 32 fungsi basis. Pada fungsi basis 1 dan 2 merupakan fungsi basis untuk variabel X_4 dengan nilai knot yaitu 00011 yang berarti $BF_1 = (X_4 \in (4, 5))$ dan $BF_2 = (X_4 \in (1, 2, 3))$. Pada fungsi basis 3 dan 4 merupakan fungsi basis untuk variabel X_1 dengan nilai knot yaitu 010 dan memiliki fungsi basis induk yaitu fungsi basis 1 yang berarti $BF_3 = (X_1 \in (2)) * BF_1$ dan $BF_4 = (X_4 \in (1, 3)) * BF_1$. Pada fungsi basis 5 dan 6 merupakan fungsi basis untuk variabel X_2 dengan nilai knot yaitu 2 dan memiliki fungsi basis induk yaitu fungsi basis 4 yang berarti $BF_5 = \max(0, X_2 - 2) * BF_4$ dan $BF_6 = \max(0, 2 - X_2) * BF_4$, dan fungsi basis seterusnya hingga fungsi basis 32 dapat dilihat pada Lampiran 3.

Setelah diperoleh fungsi basis pada tahapan *forward* untuk kombinasi ke-

30 tersebut, selanjutnya dilakukan tahapan *backward* hingga diperoleh model MARS sebagai berikut:

$$\begin{aligned}
 Y = & 0.991669 - 0.111105 BF_1 - 0.11304 BF_3 + 0.06967 BF_5 + 0.145453 BF_7 \\
 & + 0.00931912 BF_9 + 0.049301 BF_{10} - 0.0650519 BF_{11} - 0.00409619 BF_{25}
 \end{aligned}
 \tag{3.7}$$

dengan:

$$\begin{aligned}
 BF_1 &= (X_4 \in (4, 5)) \\
 BF_3 &= (X_1 \in (2)) BF_1 \\
 BF_4 &= (X_1 \in (3, 1)) BF_1 \\
 BF_5 &= \max(0, X_2 - 2) BF_4 \\
 BF_7 &= (X_8 \in (6, 7, 8, 9, 10, 11)) BF_4 \\
 BF_9 &= \max(0, X_3 - 2) BF_4 \\
 BF_{10} &= \max(0, 2 - X_3) BF_4 \\
 BF_{11} &= (X_5 \in (1)) BF_4 \\
 BF_{21} &= (X_4 \in (4)) \\
 BF_{24} &= \max(0, 6 - X_3) BF_{21} \\
 BF_{25} &= (X_8 \in (2)) BF_{24}
 \end{aligned}$$

Berdasarkan model terbaik yang diperoleh, dapat disimpulkan bahwa variabel prediktor yang mempengaruhi kematian bayi menggunakan model MARS adalah usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pendidikan ibu (X_4), jenis kelamin bayi (X_5) dan sumber utama air minum (X_8). Sedangkan variabel prediktor lainnya yaitu kelahiran kembar (X_6), dan penolong kelahiran bayi (X_7) tidak mempunyai pengaruh pada model, sehingga dapat disimpulkan dua variabel tersebut tidak mempengaruhi kematian bayi. Selanjutnya akan dilakukan uji signifikansi pada model MARS terbaik dengan secara simultan (Uji F) maupun secara parsial (Uji t).

3.12.1 Uji Simultan

1. Rumusan hipotesis

$$H_0 : a_1 = a_3 = a_5 = a_7 = a_9 = a_{10} = a_{11} = a_{25} = 0.$$

$$H_1 : \text{minimal terdapat satu } a_m \neq 0, \text{ dengan } m=1, 3, 5, 7, 9, 10, \\ 11, \text{ dan } 25.$$

2. Taraf signifikansi : $\alpha = 0,05$

3. Statistik uji:

$$F_{hitung} = \frac{\sum_{i=1}^n \frac{(\hat{y}_i - \hat{y})^2}{M}}{\sum_{i=1}^n \frac{(\hat{y}_i - \hat{y})^2}{N - M - 1}}$$

4. Daerah Kritis

tolak H_0 jika $F_{hitung} > f_{0,05(8;65)}$ atau $P\text{-Value} < 0,05$

5. Keputusan:

Berdasarkan pada Lampiran 2 dengan BF=32, MI=3, dan MO=1 diperoleh $P\text{-Value}$ sebesar 0,0000, sehingga $P\text{-Value} < 0,05$ maka tolak H_0 yang berarti bahwa model signifikan.

6. Kesimpulan:

Berdasarkan taraf signifikansi 5% maka dapat disimpulkan bahwa model signifikan sehingga dapat digunakan untuk menduga kematian bayi.

3.12.2 Uji Parsial

1. Rumusan hipotesis:

$$H_0 : a_m = 0, \text{ dengan } m=1, 3, 5, 7, 9, 10, 11, \text{ dan } 25.$$

$$H_1 : a_m \neq 0, \text{ dengan } m=1, 3, 5, 7, 9, 10, 11, \text{ dan } 25.$$

2. Taraf signifikansi : $\alpha = 0,05$

3. Statistik uji:

$$t_{hitung} = \frac{\hat{\alpha}_m}{se(\hat{\alpha}_m)}$$

dengan

$$se(\hat{\alpha}_m) = \sqrt{var(\hat{\alpha}_m)}$$

4. Daerah kritis: Tolak H_0 jika $t_{hitung} > t_{(\frac{0,05}{2}, 66)}$ atau P-Value $< 0,05$.

5. Keputusan:

Berdasarkan Lampiran 2 dengan BF=32, MI=3, dan MO=1, diperoleh nilai *P-Value* dari masing-masing fungsi basis sebagai berikut:

| | | |
|------------------------------|---|---------|
| Fungsi basis 1 ($m = 1$) | = | 0,00000 |
| Fungsi basis 3 ($m = 3$) | = | 0,00000 |
| Fungsi basis 5 ($m = 5$) | = | 0,00000 |
| Fungsi basis 7 ($m = 7$) | = | 0,00000 |
| Fungsi basis 9 ($m = 9$) | = | 0,00001 |
| Fungsi basis 10 ($m = 10$) | = | 0,00000 |
| Fungsi basis 11 ($m = 11$) | = | 0,00000 |
| Fungsi basis 25 ($m = 25$) | = | 0,00008 |

Dapat dilihat nilai *P-Value* $< 0,05$ pada setiap m atau fungsi basis.

Sehingga H_0 ditolak, berarti koefisien $a_1, a_3, a_5, a_7, a_9, a_{10}, a_{11}$, dan a_{25} berpengaruh signifikan terhadap model.

6. Kesimpulan:

Berdasarkan taraf signifikansi 5% dapat disimpulkan bahwa model signifikan sehingga dapat digunakan untuk menduga kematian bayi.

Berdasarkan pengujian signifikansi pada model MARS terbaik dengan menggunakan uji simultan (uji F) dan uji parsial (uji t), diperoleh bahwa model MARS terbaik signifikan dengan variabel prediktor yang mempengaruhi kematian bayi diantaranya usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pendidikan ibu (X_4), jenis kelamin bayi (X_5) dan sumber utama air minum (X_8). Sehingga model MARS dengan BF=32, MI=3, dan MO=1 cocok digunakan untuk memodelkan kematian bayi.

3.13 Model Cox *Proportional Hazard* dengan MARS

Berdasarkan hasil pemodelan dengan menggunakan MARS pada Persamaan (3.5), maka model Cox *Proportional Hazard* atau *hazard rate* atau laju kematian pada bayi yang mati sebelum 1 tahun di Provinsi Banten dapat dimodelkan sebagai berikut:

$$\begin{aligned} \frac{h_i(t)}{h_0(t)} &= \exp(y) \\ &= \exp(0.991669 - 0.111105 BF_1 - 0.11304 BF_3 + 0.06967 BF_5 \\ &\quad + 0.145453 BF_7 + 0.00931912 BF_9 + 0.049301 BF_{10} \\ &\quad - 0.0650519 BF_{11} - 0.00409619 BF_{25}) \end{aligned}$$

3.14 Interpretasi Cox PH dengan MARS

Hasil kombinasi BF, MI, dan MO yang telah dilakukan secara berdasarkan model persamaan terbaik dengan nilai GCV minimum menunjukkan bahwa variabel-variabel prediktor yang mempengaruhi laju kematian bayi yaitu usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pen-

didikan ibu (X_4), jenis kelamin bayi (X_5), dan sumber utama air minum (X_8). Fungsi basis yang terdapat pada model serta merupakan komponen interaksi dari fungsi basis lainnya yaitu $BF_1, BF_3, BF_5, BF_7, BF_9, BF_{10}, BF_{11}$, dan BF_{25} . Interpretasi masing-masing fungsi basis pada pemodelan MARS sebagai berikut:

1. $BF_1 = (X_4 \in (4, 5))$

Fungsi basis 1 memiliki koefisien sebesar -0,111105. Hal itu berarti setiap kenaikan BF_1 sebesar satu satuan akan mengurangi laju kematian bayi sebesar 0,111105 dengan pendidikan ibu yang merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, ibu dengan pendidikan hingga lulusan SMA atau perguruan tinggi memiliki laju kematian bayi 0,111105 kali lebih rendah dibanding lainnya.

2. $BF_3 = (X_1 \in (2)) BF_1$

Fungsi basis 3 memiliki koefisien sebesar -0,11304. Hal itu berarti setiap kenaikan BF_3 sebesar satu satuan akan mengurangi laju kematian bayi sebesar 0,11304 pada usia ibu yang melahirkan antara 20-35 tahun dengan pendidikan ibu yang merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, usia ibu yang melahirkan antara 20-35 tahun dengan pendidikan ibu yang merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian bayi 0,11304 kali lebih rendah dibanding lainnya.

3. $BF_4 = (X_1 \in (3, 1)) BF_1$

Hal tersebut menunjukkan bahwa fungsi basis 4 akan bermakna apabila usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dengan pendidikan ibu yang merupakan lulusan SMA atau perguruan tinggi.

$$4. BF_5 = \max(0, X_2 - 2) BF_4$$

Fungsi basis 5 memiliki koefisien sebesar 0,06967. Hal itu berarti setiap kenaikan BF_5 sebesar satu satuan akan menambah laju kematian bayi sebesar 0,06967 pada paritas ibu atau ibu yang telah melahirkan lebih dari 2 kali dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, paritas ibu atau ibu yang telah melahirkan lebih dari 2 kali dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian bayi 0,06967 kali lebih besar dibanding lainnya.

$$5. BF_7 = (X_8 \in (6, 7, 8, 9, 10, 11)) BF_4$$

Fungsi basis 7 memiliki koefisien sebesar 0,145453. Hal itu berarti setiap kenaikan BF_7 sebesar satu satuan akan menambah laju kematian bayi sebesar 0,145453 pada rumah tangga yang menggunakan sumber utama air minum berupa sumur terlindung, sumur tak terlindung, mata air terlindung, mata air tak terlindung, air sungai, atau air hujan dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, rumah tangga yang menggunakan sumber utama air minum berupa sumur terlindung, sumur tak terlindung, mata air terlindung, mata air tak terlindung, air sungai, atau air hujan dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian bayi 0,145453 kali lebih besar dibanding lainnya.

$$6. BF_9 = \max(0, X_3 - 2) BF_4$$

Fungsi basis 9 memiliki koefisien sebesar 0,00931912. Hal itu berarti setiap kenaikan BF_9 sebesar satu satuan akan menambah laju kematian bayi sebesar 0,00931912 pada jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya lebih dari 2 tahun dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya lebih dari 2 tahun dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian bayi 0,00931912 kali lebih besar dibanding lainnya.

$$7. BF_{10} = \max(0, 2 - X_3) BF_4$$

Fungsi basis 10 memiliki koefisien sebesar 0,049301. Hal itu berarti setiap kenaikan BF_{10} sebesar satu satuan akan menambah laju kematian bayi sebesar 0,049301 pada jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya kurang dari 2 tahun dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya kurang dari 2 tahun dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian 0,049301 kali lebih besar dibanding lainnya.

$$8. BF_{11} = (X_5 \in (1)) BF_4$$

Fungsi basis 11 memiliki koefisien sebesar - 0,0650519. Hal ini berarti setiap kenaikan BF_{11} sebesar satu satuan akan mengurangi laju

kematian bayi sebesar 0,0650519 pada bayi berjenis kelamin laki-laki dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi, atau dengan kata lain, bayi berjenis kelamin laki-laki dengan usia ibu saat melahirkan kurang dari 20 tahun atau lebih dari 35 tahun dan pendidikan ibu merupakan lulusan SMA atau perguruan tinggi memiliki laju kematian 0,0650519 kali lebih rendah dibanding lainnya.

$$9. BF_{21} = (X_4 \in (4))$$

Hal tersebut menunjukkan bahwa fungsi basis 21 akan bermakna apabila pendidikan ibu merupakan lulusan SMA.

$$10. BF_{24} = \max(0, 6 - X_3) BF_{21}$$

Hal tersebut menunjukkan bahwa fungsi basis 24 akan bermakna apabila jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya kurang dari 6 tahun dan pendidikan ibu merupakan lulusan SMA.

$$11. BF_{25} = (X_8 \in (2)) BF_{24}$$

Fungsi basis 25 memiliki koefisien sebesar -0,00409619. Hal itu berarti setiap kenaikan BF_{25} satu satuan akan mengurangi laju kematian bayi sebesar 0,00409619 pada keluarga yang menggunakan sumber air minum berupa air isi ulang dengan jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya kurang dari 6 tahun dan pendidikan ibu merupakan lulusan SMA, atau dengan kata lain, keluarga yang menggunakan sumber air minum berupa air isi ulang dengan jarak antar kelahiran atau jarak ibu melahirkan dengan kelahiran sebelumnya kurang dari 6 tahun dan pendidikan ibu merupakan lulusan SMA memi-

liki laju kematian bayi 0,00409619 kali lebih rendah dibanding lainnya.

Pemodelan dengan menggunakan MARS pada penelitian ini menunjukkan bahwa terdapat variabel prediktor yang berpengaruh secara langsung terhadap laju kematian bayi, seperti pendidikan ibu serta variabel prediktor yang berpengaruh terhadap laju kematian bayi dengan adanya interaksi antar variabel prediktor lainnya.

Table 3.11: Interaksi Variabel pada Fungsi Basis

| Fungsi Basis (BF) | Interaksi |
|-------------------|--|
| 3, 4 | Usia ibu saat melahirkan dan pendidikan ibu |
| 5 | Paritas ibu, usia ibu saat melahirkan, dan pendidikan ibu |
| 7 | Sumber utama air minum, usia ibu saat melahirkan, dan pendidikan ibu |
| 9, 10 | Jarak antar kelahiran, usia ibu saat melahirkan, dan pendidikan ibu |
| 11 | Jenis kelamin bayi, usia ibu saat melahirkan, dan pendidikan ibu |
| 24 | Jarak antar kelahiran dan pendidikan ibu |
| 25 | Sumber utama air minum, jarak antar kelahiran, dan pendidikan ibu |

Tabel 3.11 merupakan interaksi antar variabel yang terdapat pada tiap fungsi basis yang signifikan terhadap model secara umum. Enam variabel prediktor yang berpengaruh tersebut menunjukkan bahwa masing-masing variabel prediktor tersebut mempunyai pengaruh terhadap laju kematian bayi.

Table 3.12: Tingkat kepentingan variabel

| Variabel | Tingkat Kepentingan |
|------------------------------------|---------------------|
| Sumber utama air minum (X_8) | 100% |
| Paritas ibu (X_2) | 88,48% |
| Usia ibu saat melahirkan (X_1) | 65,28% |
| Jenis kelamin bayi (X_5) | 58,67% |
| Jarak antar kelahiran (X_3) | 56,00% |
| Pendidikan ibu (X_4) | 48,01% |

Tabel 3.12 menunjukkan bahwa sumber utama air minum memiliki tingkat kepentingan sebesar 100% dalam mempengaruhi laju kematian bayi artinya

sumber utama air minum sangat mempengaruhi laju kematian bayi sebesar 100%. Sedangkan variabel paritas ibu, usia ibu saat melahirkan, jenis kelamin bayi, jarak antar kelahiran, dan pendidikan ibu masing-masing mempunyai tingkat kepentingan sebesar 88,48%, 65,28%, 58,67%, 56,00%, dan 48,01%.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil analisa pada pembahasan yang dilakukan pada bab sebelumnya, maka dapat diperoleh kesimpulan sebagai berikut:

1. Berdasarkan karakteristik kematian bayi yang sudah dijabarkan pada bab sebelumnya, dapat diketahui sebagai berikut:
 - Kematian bayi menurut usia ibu saat melahirkan banyak dialami oleh ibu yang melahirkan berusia lebih dari 35 tahun.
 - Menurut pendidikan ibu, kematian bayi banyak dialami oleh ibu yang tidak tamat SD.
 - Bayi berjenis kelamin laki-laki mendominasi jumlah kematian bayi sebelum usia 1 tahun dibanding bayi berjenis kelamin perempuan.
 - Kematian bayi dengan penolong kelahiran penolong berupa bantuan non medis lebih tinggi dibandingkan penolong kelahiran berupa bantuan medis.
 - Kematian bayi banyak dialami oleh keluarga yang memiliki sumber utama air minum berupa leding eceran, sedangkan jumlah kematian bayi paling rendah dialami oleh keluarga yang menggunakan sumber utama air minum berupa air kemasan.

2. Model Cox *Proportional Hazard* dengan *Multivariate Adaptive Regression Spline* pada variabel prediktor kematian bayi diperoleh sebagai berikut:

$$\begin{aligned} \frac{h_i(t)}{h_0(t)} = & \exp(0.991669 - 0.111105 BF_1 - 0.11304 BF_3 + 0.06967 BF_5 \\ & + 0.145453 BF_7 + 0.00931912 BF_9 + 0.049301 BF_{10} \\ & - 0.0650519 BF_{11} - 0.00409619 BF_{25}) \end{aligned}$$

dengan

$$\begin{aligned} BF_1 &= (X_4 \in (4, 5)) \\ BF_3 &= (X_1 \in (2)) BF_1 \\ BF_4 &= (X_1 \in (3, 1)) BF_1 \\ BF_5 &= \max(0, X_2 - 2) BF_4 \\ BF_7 &= (X_8 \in (6, 7, 8, 9, 10, 11)) BF_4 \\ BF_9 &= \max(0, X_3 - 2) BF_4 \\ BF_{10} &= \max(0, 2 - X_3) BF_4 \\ BF_{11} &= (X_5 \in (1)) BF_4 \\ BF_{21} &= (X_4 \in (4)) \\ BF_{24} &= \max(0, 6 - X_3) BF_{21} \\ BF_{25} &= (X_8 \in (2)) BF_{24} \end{aligned}$$

Model di atas diperoleh dari model MARS terbaik dengan kombinasi fungsi basis (BF), maksimum interaksi (MI) dan minimum observasi (MO) dengan nilai BF=32, MI=3, dan MO=1 yang memiliki nilai GCV terkecil daripada kombinasi lainnya yaitu sebesar 0,00053.

3. Variabel-variabel prediktor yang berpengaruh signifikan mempengaruhi laju kematian bayi pada model yang diperoleh diantaranya usia ibu saat melahirkan (X_1), paritas ibu (X_2), jarak antar kelahiran (X_3), pen-

didikan ibu (X_4), jenis kelamin bayi (X_5), dan sumber utama air minum (X_8) dengan tingkat kepentingan masing-masing variabel sebesar 65,28%, 88,48%, 56,00%, 48,01%, 58,67%, dan 100%.

4.2 Saran

Terdapat beberapa hal yang dapat disarankan pada penelitian ini, yaitu:

- Diperlukan perhatian lebih terutama pada kesehatan ibu, seperti usia ibu saat melahirkan dan jarak melahirkan dengan kelahiran sebelumnya serta kesehatan lingkungan sekitar sebagai upaya meningkatkan kualitas hidup masyarakat, seperti sumber utama air minum masyarakat.
- Diperlukan pengamatan yang lebih terperinci mengenai faktor-faktor penduga kematian bayi, seperti riwayat penyakit ibu, dan lainnya agar diperoleh model yang lebih akurat dari penelitian ini.
- Bagi peneliti selanjutnya disarankan melakukan perbandingan antara metode Cox PH pendekatan MARS, dengan metode Weibull *Proportional Hazard* pendekatan MARS agar dapat diketahui metode yang optimal untuk memodelkan laju kematian bayi.

LAMPIRAN-LAMPIRAN

Lampiran 1. Kuesioner SUPAS 2015

| VII.C. DAFTAR RIWAYAT KELAHIRAN ANAK Untuk perempuan yang pernah melahirkan | | | | | | | | | | | | | | | | | | | |
|--|--|---------------------------|------------------------------------|----------------|---|--------------------------------|--|--|--------------------------------------|--------------------------------|--------------------------------------|--------------------------------|--------------------------------------|--------------------------------|---|--------------------------------|--|------|---|
| 715. | Siapakah nama anak (pertama, kedua, dst) | 716. | Apakah (NAMA) tunggal atau kembar? | 717. | Apakah (NAMA) laki-laki atau perempuan? | 718. | Pada bulan apa dan tahun berapa (NAMA) dilahirkan? | 719. | Siapa saja yang menolong persalinan? | 720. | Apakah (NAMA) masih hidup? | 721. | JIKA MASIH HIDUP Berapa umur (NAMA)? | 722. | JIKA MASIH HIDUP Apakah (NAMA) tinggal bersama Ibu? | 723. | JIKA SUDAH MENINGGAL Berapa umur (NAMA) ketika ia meninggal? CATAT DALAM BULAN, JIKA KURANG DARI 1 HARI TULIS "00" CATAT DALAM BULAN JIKA KURANG DARI 2 TAHUN CATAT DALAM TAHUN JIKA 2 TAHUN LEBIH | 724. | Apakah ada anak lahir hidup lain antara (NAMA) ANAK SEBELUMNYA dan (NAMA)? (TERMASUK ANAK YANG SUDAH MENINGGAL) |
| 1. | _____ NAMA | TUNGGAL...1 KEMBAR...2 | LK.....1 PR.....2 | BULAN TAHUN | DOKTER A BIDAN B DUKUN C LAINNYA D | YA.....1 TIDAK...2 P.723 | UMUR DALAM TAHUN | YA.....1 TIDAK 2 ANAK BERKUT NYA | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | YA.....1 TIDAK...2 P.724 | | |
| 2. | _____ NAMA | TUNGGAL...1 KEMBAR...2 | LK.....1 PR.....2 | BULAN TAHUN | DOKTER A BIDAN B DUKUN C LAINNYA D | YA.....1 TIDAK...2 P.723 | UMUR DALAM TAHUN | YA.....1 TIDAK 2 ANAK BERKUT NYA | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | HARI.....1 BULAN...2 TAHUN...3 | YA.....1 TIDAK...2 P.724 | YA.....1 TIDAK...2 P.724 | | |

Lampiran 2. Input Data

| Amatan ke- | T | Sensor | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 |
|------------|----|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 12 | 0 | 3 | 2 | 9 | 3 | 1 | 1 | 2 | 8 |
| 2 | 23 | 0 | 2 | 0 | 0 | 3 | 2 | 1 | 2 | 3 |
| 3 | 11 | 0 | 2 | 1 | 9 | 3 | 2 | 1 | 2 | 8 |
| 4 | 12 | 0 | 3 | 2 | 7 | 2 | 1 | 1 | 2 | 10 |
| 5 | 17 | 0 | 2 | 1 | 8 | 2 | 2 | 1 | 2 | 10 |
| 6 | 10 | 0 | 3 | 1 | 16 | 3 | 2 | 1 | 1 | 10 |
| 7 | 22 | 0 | 2 | 0 | 0 | 2 | 1 | 1 | 1 | 9 |
| 8 | 12 | 0 | 1 | 0 | 0 | 2 | 2 | 1 | 1 | 3 |
| 9 | 17 | 0 | 2 | 1 | 3 | 4 | 1 | 1 | 1 | 6 |
| 10 | 22 | 0 | 2 | 1 | 9 | 2 | 2 | 1 | 1 | 7 |
| 11 | 10 | 0 | 3 | 3 | 4 | 2 | 1 | 1 | 1 | 10 |
| 12 | 22 | 0 | 2 | 2 | 8 | 2 | 2 | 1 | 1 | 10 |
| 13 | 23 | 0 | 1 | 0 | 0 | 3 | 2 | 1 | 2 | 3 |
| 14 | 4 | 0 | 1 | 0 | 0 | 3 | 2 | 1 | 2 | 5 |
| 15 | 21 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 1 | 5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2376 | 7 | 0 | 2 | 1 | 3 | 5 | 1 | 1 | 2 | 1 |
| 2377 | 17 | 0 | 2 | 1 | 4 | 5 | 2 | 1 | 2 | 1 |
| 2378 | 15 | 0 | 2 | 1 | 9 | 3 | 1 | 1 | 2 | 6 |
| 2379 | 10 | 0 | 3 | 1 | 5 | 5 | 1 | 1 | 2 | 1 |
| 2380 | 20 | 0 | 2 | 1 | 7 | 3 | 2 | 1 | 2 | 6 |
| 2381 | 19 | 0 | 2 | 1 | 9 | 4 | 1 | 1 | 2 | 5 |
| 2382 | 2 | 0 | 2 | 1 | 7 | 4 | 1 | 1 | 2 | 6 |
| 2383 | 22 | 0 | 3 | 2 | 5 | 4 | 2 | 1 | 2 | 1 |
| 2384 | 22 | 0 | 2 | 1 | 2 | 5 | 1 | 1 | 2 | 1 |
| 2385 | 3 | 0 | 2 | 1 | 2 | 5 | 2 | 1 | 2 | 1 |
| 2386 | 24 | 0 | 2 | 2 | 5 | 5 | 1 | 1 | 2 | 1 |
| 2387 | 6 | 0 | 2 | 1 | 6 | 3 | 2 | 1 | 2 | 5 |
| 2388 | 11 | 0 | 2 | 0 | 0 | 4 | 2 | 1 | 2 | 6 |
| 2389 | 24 | 0 | 2 | 2 | 6 | 4 | 1 | 1 | 2 | 6 |
| 2390 | 2 | 0 | 2 | 2 | 9 | 4 | 2 | 1 | 2 | 6 |
| 2391 | 4 | 0 | 2 | 1 | 5 | 4 | 2 | 1 | 2 | 1 |

Lampiran 3. Hasil Output Model MARS

Terbaik

Kombinasi BF=32, MI=3, MI=1

=====

MARS Results

=====

Forward Stepwise Knot Placement

=====

| BasFn(s) | GCV | IndBsFns | EfPrms | Variable | Knot Parent | BsF |
|----------|---------|----------|--------|----------|--------------|-----|
| 0 | 0.00062 | 0.0 | 1.0 | | | |
| 2 1 | 0.00064 | 1.0 | 5.0 | X4 | 00011 | |
| 4 3 | 0.00069 | 2.0 | 9.0 | X1 | 010 X4 | 1 |
| 6 5 | 0.00070 | 4.0 | 14.0 | X2 | 2.00000 X1 | 4 |
| 8 7 | 0.00067 | 5.0 | 18.0 | X8 | 000001111 X1 | 4 |
| 10 9 | 0.00071 | 7.0 | 23.0 | X3 | 2.00000 X1 | 4 |
| 12 11 | 0.00063 | 8.0 | 27.0 | X5 | 10 X1 | 4 |
| 14 13 | 0.00063 | 9.0 | 31.0 | X8 | 000001111 X1 | 4 |
| 16 15 | 0.00071 | 10.0 | 35.0 | X8 | 000001111 X1 | 3 |
| 18 17 | 0.00083 | 11.0 | 39.0 | X5 | 10 X1 | 3 |
| 20 19 | 0.00095 | 13.0 | 44.0 | X3 | 4.00000 X4 | 1 |
| 22 21 | 0.00116 | 14.0 | 48.0 | X4 | 00010 | |
| 24 23 | 0.00152 | 16.0 | 53.0 | X3 | 6.00000 X4 | 21 |
| 26 25 | 0.00221 | 17.0 | 57.0 | X8 | 010000000 X3 | 24 |
| 28 27 | 0.00371 | 18.0 | 61.0 | X8 | 100000000 X3 | 20 |
| 30 29 | 0.00765 | 19.0 | 65.0 | X3 | 3.00000 X4 | 1 |
| 32 31 | 0.02465 | 20.0 | 69.0 | X7 | 10 X4 | 1 |

=====

Final Model (After Backward Stepwise Elimination)

=====

| Basis Fun | Coefficient | Variable | Knot | Parent |
|-----------|-------------|----------|---------|--------|
| 0 | 0.99167 | | | |
| 1 | -0.11111 | X4 | 00011 | |
| 3 | -0.11304 | X1 | 010 | X4 |
| 5 | 0.06967 | X2 | 2.00000 | X1 |

```

7      0.14545 X8                000001111 X1
9      0.00932 X3                2.00000 X1
10     0.04930 X3                2.00000 X1
11     -0.06505 X5                10 X1
25     -0.00410 X8                010000000 X3

```

Piecewise Linear GCV = 0.00053, #efprms = 28.19999

```

=====
ANOVA Decomposition on 8 Basis Functions
=====

```

| fun | std. dev. | -gcv | #bsfns | #efprms | variable |
|-----|-----------|---------|--------|---------|----------------|
| 1 | 0.05424 | 0.00097 | 1 | 3.40000 | X4 |
| 2 | 0.04938 | 0.00095 | 1 | 3.40000 | X1 X4 |
| 3 | 0.02090 | 0.00082 | 1 | 3.40000 | X1 X4 X2 |
| 4 | 0.02869 | 0.00103 | 1 | 3.40000 | X1 X4 X8 |
| 5 | 0.01981 | 0.00062 | 2 | 6.80000 | X1 X4 X3 |
| 6 | 0.01471 | 0.00066 | 1 | 3.40000 | X1 X4 X5 |
| 7 | 0.00964 | 0.00058 | 1 | 3.40000 | X4 X8 X3 |

```

=====
Variable Importance
=====

```

| Variable | Importance | -gcv |
|----------|------------|---------|
| X8 | 100.00000 | 0.00090 |
| X2 | 88.48248 | 0.00082 |
| X1 | 65.28510 | 0.00069 |
| X5 | 58.66973 | 0.00066 |
| X3 | 56.00066 | 0.00065 |
| X4 | 48.00744 | 0.00062 |
| X7 | 0.00000 | 0.00053 |

X6 0.00000 0.00053

=====

MARS Regression: Training Data

=====

W: 74.00 R-SQUARED: 0.66100

MEAN DEP VAR: 0.98666 ADJ R-SQUARED: 0.61928

UNCENTERED R-SQUARED = R-0 SQUARED: 0.99979

| Parameter | Estimate | S.E. | T-Ratio | P-Value |
|-------------------|----------|---------|-----------|---------|
| Constant | 0.99167 | 0.00227 | 437.38660 | 0.00000 |
| Basis Function 1 | -0.11111 | 0.01309 | -8.48920 | 0.00000 |
| Basis Function 3 | -0.11304 | 0.01352 | -8.35806 | 0.00000 |
| Basis Function 5 | 0.06967 | 0.00975 | 7.14651 | 0.00000 |
| Basis Function 7 | 0.14545 | 0.01624 | 8.95542 | 0.00000 |
| Basis Function 9 | 0.00932 | 0.00187 | 4.97607 | 0.00001 |
| Basis Function 10 | 0.04930 | 0.00929 | 5.30970 | 0.00000 |
| Basis Function 11 | -0.06505 | 0.01228 | -5.29743 | 0.00000 |
| Basis Function 25 | -0.00410 | 0.00098 | -4.19960 | 0.00008 |

F-STATISTIC = 15.84286

S.E. OF REGRESSION = 0.01521

P-VALUE = 0.00000

RESIDUAL SUM OF SQUARES = 0.01504

[MDF,NDF] = [8, 65]

REGRESSION SUM OF SQUARES = 0.02932

=====

Basis Functions

=====

BF1 = (X4 in (4, 5));

BF3 = (X1 in (2)) * BF1;

BF4 = (X1 in (3, 1)) * BF1;

BF5 = max(0, X2 - 2) * BF4;

BF7 = (X8 in (6, 7, 8, 9, 10, 11)) * BF4;

BF9 = max(0, X3 - 2) * BF4;

BF10 = max(0, 2 - X3) * BF4;

BF11 = (X5 in (1)) * BF4;

BF21 = (X4 in (4));

BF24 = max(0, 6 - X3) * BF21;

BF25 = (X8 in (2)) * BF24;

Y = 0.991669 - 0.111105 * BF1 - 0.11304 * BF3 + 0.06967 * BF5

+ 0.145453 * BF7 + 0.00931912 * BF9 + 0.049301 * BF10

- 0.0650519 * BF11 - 0.00409619 * BF25;

MODEL RM = BF1 BF3 BF5 BF7 BF9 BF10 BF11 BF25;

SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya yang bertanda tangan di bawah ini, mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta:

Nama : Tias Dwi Novianisa
No. Registrasi : 3125130795
Program Studi : Matematika

Menyatakan bahwa skripsi ini yang saya buat dengan judul "**Pemodelan Kematian Bayi Menggunakan Cox *Proportional Hazard* dengan *Multivariate Adaptive Regression Spline* (Hasil Survei Penduduk Antar Sensus 2015)**" adalah :

1. Dibuat dan diselesaikan oleh saya sendiri.
2. Bukan merupakan duplikat skripsi yang pernah dibuat oleh orang lain atau jiplakan karya tulis orang lain.

Pernyataan ini dibuat dengan sesungguhnya dan saya bersedia menanggung segala akibat yang timbul jika pernyataan saya tidak benar.

Jakarta, Agustus 2017

Yang membuat pernyataan


Tias Dwi Novianisa

DAFTAR RIWAYAT HIDUP



TIAS DWI NOVIANISA. Lahir di Jakarta, 13 November 1995. Anak kedua dari pasangan Bapak Syamsudin dan Ibu Eni Aryani. Saat ini bertempat tinggal di Jalan Petojo Sabangan XI RT 010/004 No 36, Jakarta Pusat 10160.

Email : tias13dwi@gmail.com

Riwayat Pendidikan : Penulis mengawali pendidikan di SDN Sukabumi Utara 09 Pagi selama 2 tahun, lalu saat kenaikan kelas 3 pindah ke Tangerang dan bersekolah di SD Kunciran 07, dan kemudian pindah kembali ke Jakarta saat kenaikan kelas 4 dan melanjutkan sekolah di SDN Petojo Selatan 02 Pagi hingga lulus SD pada tahun 2007. Setelah itu, penulis melanjutkan pendidikan ke SMPN 1 Jakarta hingga lulus pada tahun 2010. Selanjutnya kembali melanjutkan pendidikan ke jenjang berikutnya yaitu SMA di SMAN 35 Jakarta dan dinyatakan lulus pada tahun 2013. Pada tahun yang sama, penulis melanjutkan pendidikan ke Perguruan Tinggi di Universitas Negeri Jakarta jurusan Matematika melalui jalur SNMPTN. Pada Agustus 2017 penulis telah memperoleh gelar Sarjana Sains untuk Program Studi Matematika, FMIPA UNJ.

Riwayat Organisasi : Selama di bangku perkuliahan, penulis aktif di berbagai organisasi kemahasiswaan. Dalam dua tahun pertama, penulis mendapat kepercayaan sebagai staff Biro Enterpreneur BEMJ Matematika. Pada tahun ketiga perkuliahan, penulis diberi kepercayaan untuk menjadi koor. Publikasi salah satu acara BEMJ Matematika yaitu Pelangi Matematika XXII, serta staff publikasi pada acara Calculus Cup VII.

Riwayat Pekerjaan : Penulis mulai menjadi pengajar privat Matematika jenjang SD, SMP dan SMA sejak tahun 2014 hingga sekarang.