ESTIMASI MODEL LINIER TERGENERALISASI LOG-LOGISTIK PADA DATA UJI HIDUP TERSENSOR II MENGGUNAKAN ALGORITMA *FISHER-SCORING*

Skripsi

disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Sri Resmiasih
4111414006

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2019

PERNYATAAN KEASLIAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila dikemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Semarang,

Yang membuat pernyataan

Sri Resmiasih

NIM. 4111414006

PENGESAHAN

Skripsi dengan judul

Estimasi Model Linier Tergeneralisasi Log-Logistik Pada Data Uji Hidup Tersensor II Menggunakan Algoritma *Fisher-Scoring*

disusun oleh

Sri Resmiasih

4111414006

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada

tanggal

Panitia

Ketua

Prof. Dr. Sudarmin, M.Si
NIP. 196601231992031003

Sekretaris

Drs. Arief Agoestanto, M.Si
NIP. 196807221993031005

Ketua Penguji

Dr. Scolastika Mariani, M.Si
NIP. 196502101991022001

Anggota Penguji/
Pembimbing I

Dr. Nur Karomah Dwidayati, M.Si
NIP. 196605041990022001

Anggota Penguji/
Pembimbing II

Drs. Arief Agoestanto, M.Si
NIP. 196807221993031005

MOTTO

"Jika kamu tidak keras kepala, kamu akan mudah untuk menyerah. Jika kamu tidak fleksibel, kamu akan sulit menemukan solusi untuk masalah yang sedang kamu hadapi"

(Jeff Bezos)

"Saya tidak pernah gagal. Saya hanya menemukan 10.000 cara yang tidak tepat."

(Thomas A. Edison)

PERSEMBAHAN

Skripsi ini saya persembahkan kepada :

1. Kedua orang tua saya, dan adikku Giandra Sakha Wiratama yang telah memberikan doa, dan dukungannya.
2. Selamet Widodo yang selalu memberikan motivasi dan dukungan.
3. Rekan jurusan Matematika Angkatan 2014 yang selalu memberikan semangat.
4. Almamaterku tercinta

PRAKATA

Puji syukur kehadirat Allah SWT atas segala rahmat dan nikmat-Nya yang senantiasa tercurah sehingga penyusunan skripsi yang berjudul "Estimasi Model Linier Tergeneralisasi Log-Logistik Pada Data Uji Hidup Tersensor II Menggunakan Algoritma *Fisher-Scoring*" sebagai syarat untuk memperoleh gelar Sarjana Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang dapat terselesaikan dengan baik.

Penyusunan skripsi ini tidak lepas dari bantuan dan dukungan dari berbagai pihak berupa saran, bimbingan maupun petunjuk. Oleh karena itu, dengan penuh kerendahan hati penulis menguucapkan terimakasih kepada :

1. Prof. Dr. Fathur Rokhman, M.Hum, Rektor Universitas Negeri Semarang yang telah memberikan kesempatan kepada penulis untuk menyelesaikan studi Strata 1 di Universitas Negeri Semarang;
2. Prof. Dr. Sudarmin, M.Si., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang yang telah mengizinkan penulis melakukan penelitian dan telah mengesahkan skripsi ini;
3. Drs. Arief Agoestanto, M.Si, Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang yang telah memberikan dukungan dan motivasi kepada penulis untuk menyelesaikan skripsi ini;

4. Dr. Nur Karomah D, M.Si, Dosen Pembimbing I yang telah memberikan waktu, bimbingan, dukungan, arahan, ide, kritik dan saran selama menyelesaikan skripsi;
5. Drs. Arief Agoestanto, M.Si, Dosen Pembimbing II yang dengan sabar memberikan motivasi, bimbingan, saran, arahan dan dukungan selama penyusunan skripsi;
6. Dr. Scolastika Mariani, M.Si, Dosen Wali sekaligus dosen penguji yang telah memberikan arahan, ide, kritik dan saran selama menyelesaikan skripsi;
7. Seluruh dosen dan staf karyawan di lingkungan Universitas Negeri Semarang terkhusus Jurusan Matematika yang telah berkenan mendidik, memberi banyak ilmu, pengalaman dan inspirasi selama penulis belajar di kampus;
8. Kedua orang tua, Bapak Sumadi dan Ibu Puni’ah yang selalu memberikan do’a, dukungan, bimbingan, kasih sayang, motivasi dan semangat tiada henti dan adikku tersayang Giandra Sakha Wiratama sebagai penyemangat ketika lelah;
9. Selamet Widodo, teman yang telah memberikan dukungan berupa semangat, motivasi, serta berbagi keluh dan kesah kepada penulis dalam menyelesaikan skripsi;
10. Teman-teman kost Mekarsari, yang selalu menemani menjalani sisa hari dengan penuh canda dan ceria;

11. Keluarga Matematika 2014 yang telah berbagi banyak suka dan duka serta pengalaman yang tidak bisa diulang;
12. Semua pihak yang telah membantu dan mendukung penulis dalam penelitian dan penyusunan skripsi yang tidak dapat disebutkan satu persatu.

Penulis menyadari skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan agar dapat menghasilkan karya yang lebih baik lagi. Semoga skripsi ini dapat memberikan manfaat bagi penulis dan para pembaca.

Semarang,

Penulis

ABSTRAK

Resmiasih,S.2019. *Estimasi Model Linier Tergeneralisasi Log-logistik Pada Data Uji Hidup Tersensor II Menggunakan Algoritma Fisher-Scoring.* Skripsi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama Dr. Nur Karomah D, M.Si. dan Pembimbing Pendamping Drs. Arief Agoestanto, M.Si.

Kata kunci: Model Linier Tergeneralisasi, Distribusi Log-Logistik, Data Uji hidup,Algoritma Fisher-Scoring.

Tujuan dari penelitian ini adalah untuk memperoleh nilai estimasi model linier tergeneralisasi Log-logistik berdasarkan maximum likelihood estimation dengan menggunakan algoritma Fisher-Scoring dan melakukan uji kecocokan model. Metode untuk mengestimasi parameter model dapat diperoleh dengan menyelesaikan persamaan maksimum likelihood yang berbentuk

$$\frac{-r\hat{\beta}}{\hat{\alpha}} + \left(\frac{2\hat{\beta}}{\hat{\alpha}}\right)\sum_{i=1}^{r}\left[1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\right]^{-1}\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}} - \frac{(r-n)}{1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}}\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)\left(\frac{t_r}{\hat{\alpha}}\right)^{\hat{\beta}} = 0$$

dan

$$-\frac{r}{\hat{\beta}} + \sum_{i=1}^{r}\ln\frac{t_i}{\hat{\alpha}} - 2\sum_{i=1}^{r}\left[1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\right]^{-1}\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\ln\left(\frac{t_i}{\hat{\alpha}}\right) + \frac{(r-n)}{1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}}\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)\ln\left(\frac{t_r}{\hat{\alpha}}\right) = 0$$

dan dalam perhitungannya digunakan algoritma Fisher-Scoring melalui software R. untuk menguji kecocokan model digunakan uji Residual Deviance. Hasil penerapan model linier tergeneralisasi Log-logistik pada data uji hidup pasien kanker paru-paru di RSUP Dr. Kariadi Semarang diperoleh bentuk model dugaan

$$y_i = -2.05650 + 1.62366x_1 + 1.65555x_2 - 0.06665x_3 + 1.19602x_4$$

Berdasarkan model tersebut diperoleh kesimpulan bahwa Semakin bertambahnya umur, skor performance, dan perawatan maka semakin berkurang waktu tahan hidup pasien kanker paru-paru tersebut. Sedangkan dalam uji kecocokan model pada data uji hidup pasien kanker paru-paru di RSUP Dr. Kariadi Semarang diperoleh bahwa model dugaan sesuai dengan model sebenarnya.

DAFTAR ISI

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

Halaman

DAFTAR SIMBOL

α : Location Parameter

β : Scale Parameter

n : Banyak Data yang diamati

r : Kegagalan Data yang diamati

$f(y_i)$: Fungsi Densitas Distribusi Log-Logistik

$F(t)$: Fungsi Komulatif

$S(t)$: Fungsi Survivor

$H(t)$: Fungsi Hazard atau Fungsi Kegagalan

ε : Standart Eror Distribusi Logistik

$G(\eta_i)$: Fungsi Link Distribusi

L : Fungsi Likelihood

H : Matriks Hessian

Wj : Uji Wald

SE : Nilai Standart Eror

$A_{n,p}^2$: Uji Anderson Darling

$D_{n,p}^{\alpha}$: Nilai pada Tabel Kaziol and Byar.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Model statistika yang memegang peranan sangat penting dalam suatu penelitian untuk melakukan suatu analisis hubungan antar beberapa variabel adalah model linier. Bentuk model linier yang paling sederhana dan sering digunakan adalah model regresi sederhana yang hanya melibatkan satu variabel respon dan satu variabel prediktor. Seiring perkembangan jaman dan ilmu pengetahuan beberapa ahli matematika telah mengembangkan model linier yang sederhana menjadi model linier tergeneralisasi. Model linier tergeneralisasi ini menggunakan asumsi bahwa variabel respon memiliki distribusi keluarga eksponensial. Distribusi eksponensial merupakan distribusi yang mempunyai sifat-sifat dan karakteristik tertentu untuk menentukan estimasi menurut kecukupan statistik. Beberapa distribusi yang termasuk keluarga eksponensial adalah distribusi Gaussian, Poisson, Binomial dan Log-logistik.

Salah satu distribusi keluarga eksponensial yang dapat digunakan untuk menganalisa data ketahanan hidup adalah distribusi Log-Logistik karena logaritma natural dari variabel tahan hidupnya berdistribusi secara logistik. Distribusi Log-Logistik juga mempunyai bentuk fungsi kegagalan (hazard) yang tidak monoton naik atau monoton turun dan juga tidak konstan seperti pada distribusi Weibull.

Analisa data uji hidup adalah suatu metode untuk menganalisis data yang berhubungan dengan waktu, mulai dari *start-point* sampai dengan terjadinya suatu kejadian khusus atau *end-point.* Jika semua individu diuji sampai terjadinya kematian atau kegagalan tertentu maka disebut sampel lengkap. Pengujian untuk sampel lengkap mempunyai keuntungan yaitu semua individu dapat teramati secara keseluruhan. Namun metode sampel lengkap membutuhkan waktu yang sangat lama dalam penelitian dan biaya yang cukup besar. Bagi suatu instansi yang melakukan penelitian tentu hal tersebut sangat merugikan,untuk itu diperlukan penyensoran data agar lebih efisien dari segi waktu maupun biaya.

Dalam statistika ada beberapa jenis penyensoran data yang dapat digunakan untuk mempercepat suatu penelitian. Sensor tipe I dibatasi oleh waktu dan setiap individu masuk kedalam pengamatan secara bersama. Sensor tipe II dibatasi oleh banyaknya r individu yang gagal dari sebanyak n individu dalam pengamatan $(1 \leq r \leq n)$ Sensor tipe III dibatasi oleh waktu dan setiap individu masuk kedalam pengamatan dalam waktu yang berbeda. Pada penelitian ini penulis menggunakan penyensoran tipe II dengan alasan waktu yang dibutuhkan untuk melakukan penelitian relatif lebih cepat, dengan cara mengambil sebagian data untuk tidak diamati. Hal ini menyebabkan biaya yang dikeluarkan relatif sedikit.

Berdasarkan uraian diatas, diperlukan suatu metode yang dapat digunakan dalam penarikan kesimpulan mengenai suatu populasi yang biasa disebut inferensi statistik. Didalam inferensi statistika estimasi parameter suatu populasi merupakan salah satu hal penting. Salah satu metode yang sering digunakan adalah *Maximum Likelihood Estimation* (MLE) yang menggunakan pendekatan distribusi dengan memaksimumkan fungsi *likelihood* dan *least square method* (metode kuadrat terkecil) yang menggunakan pendekatan geometris dengan meminimumkan galatnya. Namun pada kenyataannya sering kali estimasi menggunakan MLE maupun metode kuadrat terkecil mendapatkan kendala. Ketika diperoleh dalam bentuk implisit dan nonlinear MLE kurang tepat menghasilkan estimasi parameter sehingga diperlukan bantuan algoritma Fisher-scoring dengan pendekatan metode numerik Newton Raphson untuk iterasinya.

Algorima *Fisher-scoring* merupakan modifikasi algoritma *Newton Raphson* yang menggunakan vektor *score* dan matriks insformasi Fisher. Algoritma Newton Raphson secara iteratif menggunakan matriks Hessian yang elemennya merupakan turunan kedua suatu fungsi *likelihood.* Hal ini menyebabkan hasil iterasinya tidak selalu konvergen. Berdasarkan hal tersebut, algoritma Newton Raphson dimodifikasi menggantikan matriks Hessian-nya dengan menggunakan matriks informasi fisher yang selanjutnya lebih dikenal dengan algoritma *fisher-scoring.* Beberapa penelitian telah dilakukan oleh para ahli mengenai algoritma *fisher-scoring* antara lain adalah "Membandingan Keunggulan Pada Algoritma Newton Raphson dan Algoritma *fisher-scoring* dalam Perhitungan Estimasi Maksimum *Likelihood" "(* Schworer, A. and P. Hovey, 2004). Pada penelitian tersebut ditunjukkan bahwa algoritma *fisher-scoring* lebih baik daripada algoritma Newton Raphson karena algoritma *fisher-scoring* tetap konvergen ketika algoritma Newton Raphson tidak konvergen. Selanjutnya "Menerapkan Estimasi Parameter dengan Algoritme Fisher *scoring* pada Model Regresi Logistik biner"(Marius, O. U. and O. I. C. Anaene,2013). Penelitian selanjutnya tentang algoritma *Fisher-scoring* adalah “Maksimum *Likelihood* berdasarkan Newton Raphson, *Fisher-scoring dan* algoritma Ekspektasi Maksimasi pada Data Khusus” (Purba, S. A. and Sutarman, 2017). Penelitian tersebut menjelaskan tentang ketiga algoritma yang memiliki hasil estimasi yang sama pada kasus data normal. Namun pada estimasi yang menggunakan algoritma Newton Raphson memiliki jumlah iterasi yang lebih banyak dari algoritma *Fisher-scoring* dan Ekspektasi Maksimasi. Keunggulan menggunakan algoritma *Fisher-scoring* adalah lebih dijamin konvergensinya daripada algoritma Newton Raphson.

Berdasarkan penelitian-penelitian sebelumnya penulis tertarik untuk mengembangkan penelitian mengenai **“ESTIMASI MODEL LINIER TERGENERALISASI LOG-LOGISTIK PADA DATA UJI HIDUP TERSENSORE II MENGGUNAKAN ALGORITMA *FISHER-SCORING*”**. Pada penelitian kali ini data statistik yang digunakan penulis diasumsikan mengikuti pola distribusi *Log-Logistic*. Distribusi *Log-Logistic* hampir sama dengan distribusi Log-Normal.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas dapat dirumuskan permasalahan inti sebagai berikut.

1. Bagaimana hasil uji kesesuaian model linier tergeneralisasi Log-logistik dengan menggunakan statistic uji *Residual Deviance*?
2. Bagaimana estimator model linier tergeneralisasi log logistik pada data tahan hidup pasien kanker paru-paru dengan algoritma fisher scoring pada program R?

1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas,batasan masalah yang diperlukan pada penelitian ini sebagai berikut.

1. Model linier tergeneralisasi menggunakan distribusi log-logistik
2. Estimasi parameter yang digunakan dalam menentukan model linier tergeneralisasi adalah algoritma *Fisher-scoring*.
3. Program dalam penelitian ini menggunakan program R
4. Data yang digunakan merupakan data tersensor tipe II.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas didapatkan tujuan sebagai berikut.

1. Menguji kesesuaian model linier tergeneralisasi Log-logistik dengan menggunakan statistic uji *Residual Deviance.*

2. Mendapatkan estimasi model linier tergeneralisasi Log-logistic pada data tahan hidup pasien kanker paru-paru dengan algoritma fisher scoring pada program R.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penulisan ini sebagai berikut.

1.5.1 Bagi Penulis

1. Penulis dapat mengembangkan dan mengaplikasikan pengetahuan dan keilmuan di bidang matematika.
2. Penulis dapat menentukan model linier tergeneralisasi distribusi log-logistik pada data uji hidup tersensor tipe II.
3. Penulis dapat mengetahui proses dan hasil estimasi parameter distribusi log-logistik dengan menggunakan algoritma *fisher scoring*.

1.5.2 Bagi Pembaca

1. Memberikan informasi tentang bagaimana memilih model linear tergeneralisasi log-logistic yang tepat dengan mempertimbangkan nilai estimasi parameter dengan algoritma *fisher-scoring*.
2. Untuk penelitian selanjutnya bisa menggunakan algoritma *fisher scoring* pada distribusi selain log-logistic, bisa distribusi Weibull atau kelompok distribusi eksponensial lainnya.

BAB II

TINJAUAN PUSTAKA

2.1 Konsep Dasar Distribusi Survival

2.1.1 Fungsi Padat Peluang

Fungsi padat peluang pada analisis survival adalah peluang suatu individu mati atau gagal dalam interval waktu t sampai t + Δt , dengan waktu T merupakan variabel random. Fungsi densitas peluang dari T dapat dinyatakan sebagai $f(t)$,

$$f(t) = \lim_{\Delta t \to 0} \left[\frac{P\big(t \leq T < (t + \Delta t)\big)}{\Delta t} \right] \tag{2.1}$$

yang mempunyai sifat sebagai berikut.

a. $f(t) \geq 0, t \geq 0$

b. $\int_0^\infty f(t)dt = 1.$

Jika T merupakan variabel random dari waktu hidup suatu individu dalam interval [0, ∞) , maka fungsi distribusi kumulatif *F(t)* untuk distribusi kontinu dengan fungsi densitas peluang *f(t)* dinyatakan sebagi berikut.

$F(t) = P(T \leq t)$ atau $F(t) = \int_0^t f(x)dx \, , untuk\ t > 0$ $\quad$ $(\text{Lawless}, 1982{:}\,8)$.

2.1.2 Fungsi Tahan Hidup

fungsi *survivor* didefinisikan sebagai peluang suatu individu dapat bertahan hidup sampai waktu t. Jika T merupakan variabel random dari waktu hidup suatu individu dalam interval [0, ∞) ,maka fungsi survivor S(t) dapat dinyatakan dalam persamaan:

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

$$= \int_t^{\infty} f(x)dx$$

Dengan demikian diperoleh persamaan yang menyatakan hubungan antara fungsi survivor dan fungsi distribusi kumulatif, yaitu $S(t) = 1 - F(t)$.

2.1.3 Fungsi Kegagalan

Fungsi *hazard* menyatakan tingkat kegagalan suatu komponen pada waktu t, jika diketahui bahwa komponen tersebut tetap hidup hingga waktu t. Menurut (Lawless,1982:8) fungsi hazard adalah peluang suatu individu mati dalam interval waktu t sampai t + Δt , jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t, yang dinyatakan sebagai berikut.

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right] \tag{2.2}$$

Jika f(t) adalah fungsi densitas peluang pada waktu t, maka diperoleh:

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

$$= \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < (t + \Delta t)) \cap (T \geq t)}{P(T \geq t)\Delta t} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < (t + \Delta t))}{P(T \geq t)\Delta t} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{1}{\Delta t} \frac{F(t + \Delta t) - F(t)}{1 - F(t)} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)} \right]$$

$$= \frac{F'(t)}{S(t)}$$

$$h(t) = \frac{f(t)}{S(t)}. \tag{2.3}$$

2.2 Data Penyensoran

Dalam penelitian uji hidup, data waktu hidup dapat berbentuk data lengkap, data tersensor tipe I dan data tersensor tipe II. Pada pengambilan data lengkap , percobaan akan dihentikan jika semua komponen atau individu yang diteliti gagal atau mati (Lawless,1982:31). Metode menggunakan data lengkap memerlukan waktu yang lama sehingga jarang digunakan.

Data tersensor adalah data yang diperoleh sebelum semua data teramati waktu hidupnya, sedangkan dalam waktu pengamatan telah berakhir atau oleh sebab lain. Data tersensor tipe I merupakan data uji hidup yang dihasilkan setelah penelitian berjalan selama waktu yang telah ditentukan. Sedangkan data tersensor tipe II merupakan data hasil penelitian dimana penelitian dihentikan setelah kematian atau kegagalan tertentu telah terjadi (Lawless,1982:31).

Data tersensor tipe II merupakan data kematian atau kegagalan yang tidak lengkap yaitu data waktu kematian atau kegagalan dari r observasi terkecil dalam sampel random yang berukuran n dengan $1 \leq r \leq n$. Dalam suatu penelitian , penyensoran tipe II lebih sering digunakan yaitu dalam uji hidup yang terdapat observasi sebanyak n tetapi penelitian dihentikan ketika observasi mengalami kegagalan ke-r sehingga dapat menghemat waktu dan biaya. Dalam penyensoran ini r ditentukan terlebih dahulu sebelum data dikumpulkan.

2.3 Keluarga Eksponensial

Suatu variabel acak *y* dengan *probability density function f* dan parameter θ dikatakan menjadi anggota keluarga eksponensial, jika *f* dapat dinyatakan sebagai :

$$f(y_i; \theta_i, \psi) = exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\psi)} + c(y_i, \psi)\right\} \tag{2.4}$$

dengan $a(.)$,$b(.)$ dan $c(.)$ masing-masing adalah fungsi khusus yang diketahui, θ adalah parameter kanonik (parameter lokasi) dan ψ adalah parameter disperse (parameter skala) (Hardle, dkk, 2004:151).

Salah satu distribusi yang termasuk keluarga eksponensial adalah: distribusi Weibull, distribusi Gamma,distribusi Binomial dan distribusi log-logistik.

2.4 Distribusi Log-Logistik

Dalam statistika distribusi log-logistik merupakan salah satu distribusi peluang kontinu untuk variable random non-negatif. Distribusi ini digunakan dalam analisis tahan hidup sebagai model parametrik. Misalkan untuk meneliti waktu penyembuhan suatu penyakit. Distribusi log-logistik juga telah dikembangkan di bidang industry untuk menganalisis tahan hidup komponen dari suatu produk.

Distribusi log-logistik merupakan distribusi yang berasal dari distribusi logistik dengan variable random Y yang mempunyai fungsi densitas peluang

$$f(y) = \frac{\sigma^{-1} exp\left[\frac{(y-\mu)}{\sigma}\right]}{\{1 + exp\left[\frac{(y-\mu)}{\sigma}\right]\}^2}, t > 0, \alpha > 0, \beta > 0 \tag{2.5}$$

Dengan $\mu = \log \alpha,\ \ \sigma = \beta^{-1}\ , -\infty < \mu < \infty$ dan $\sigma > 0.$

Fungsi distribusi komulatifnya adalah

$$F(t) = \frac{t^{\beta}}{a^{\beta} + t^{\beta}}, t > 0, \alpha > 0, \beta > 0 \tag{2.6}$$

Fungsi tahan hidupnya adalah

$$S(t) = \frac{1}{1 + (\frac{t}{a})^{\beta}}, t > 0, \alpha > 0, \beta > 0 \tag{2.7}$$

Dalam statistika distribusi log-logistik merupakan salah satu distribusi peluang kontinu untuk variable random non-negatif. Distribusi ini digunakan dalam analisis tahan hidup sebagai model parametrik. Misalnya untuk meneliti waktu penyembuhan suatu penyakit. Distribusi log-logistik juga telah dikembang dalam bidang industri untuk mengalisis tahan hidup komponen dari suatu produk.

Variable random T dikatakan mengikuti distribusi log-logistik dengan parameter α dan parameter shape β, jika mempunyai fungsi densitas :

$$f(t;\alpha,\beta)=\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{2}},\qquad t>0,\text{dimana }\alpha>0,\text{dan }\beta>0$$

untuk selanjutnya dinotasikan sebagai $T-L_L(\alpha,\beta)$. Nilai parameter shape yaitu β menyatakan suatu bentuk yang bermacam-macam dari kurva fungsi densitas,yaitu naik,turun atau mendatar. Sehingga kondisi ini sangat cocok digunakan untuk berbagai model data survival.

2.4.1 Simulasi Data Berdistribusi Log-logistik

Sebanyak 15 data tahan hidup $t=(1{,}4{,}1{,}13{,}8{,}3{,}1{,}2{,}1{,}3{,}2{,}1{,}1{,}3{,}2)$ merupakan hasil bangkitan dengan menggunakan bantuan software Minitab. Dari data tersebut diperoleh nilai $AD=0.359$ P-value >0.250.

Hasil uji kecocokan distribusi Log-Logistik untuk data bangkitan sebanyak 10 data adalah sebagai berikut.

Hipotesis :

$H_0: F(t)= F_0(t)$ (Data berdistribusi Log-Logistik)

$H_1: F(t)\neq F_0(t)$ (Data tidak berdistribusi Log-Logistik)

Statistic Uji :

$$A_{n,p}^2=-\sum_{i=1}^{r}\left(\frac{2_i-1}{n}\log F_0\big(t_{(i)}\big)-\frac{2n-2_i+1}{n}\log[1-F_0\big(t_{(i)}\big)]\right)$$

$$+\frac{r^2}{n}\log F_0(L)-\frac{(n-r)^2}{n}lo\,g[1-F_0(L)]-nF_0(L)$$

Kriteria penolakan :

Tolak H_0 jika $A_{n,p}^2 > D_{n,p}^{\alpha}$

Keputusan :

Berdasarkan hasil output Minitab didapat nilai uji Anderson Darling sebesar 0.505 dan $P = F_0(L) = 0.161 \approx 0.2$, maka berdasarkan tabel Koziol dan Byar diperoleh $D_{n,p}^{\alpha} = D_{10,0.2}^{5\%} = 0.9268$. diketahui bahwa $A_{n,p}^2 = 0.615 < D_{50,0.2}^{5\%} = 0.9268$, sehingga H_0 diterima. Artinya bahwa data waktu ketahanan pasien kanker paru-paru berdistribusi Log-Logistik.

2.5 Model Linier Tergeneralisasi

Model linier tergeneralisasi merupakan perluasan konsep model regresi linier yang telah banyak digunakan. Model linier tergeneralisasi mengasumsikan bahwa respon $Y(\text{variabel dependen})$ sama dengan kombinasi linier $\boldsymbol{X^T\beta}$ dengan nilai error berdistribusi normal.

$E(y_i|x_i) = G(x_i{}^{\tau}\beta)$ (Hardle, 2004:151). (2.8)

dengan y_i adalah variabel respon ke-i, x_i adalah variabel prediktor ke-i pada variabel prediktor ke-j, $\boldsymbol{\beta}$ adalah vektor parameter yang diestimasi, dan G adalah fungsi link.

Model linear tergeneralisasi termasuk kasus yang khusus, regresi linear dan analisis model varian,logit dan probit model untuk respon quantal. Log linear model dan multinomial respon merupakan model yang dapat digunakan untuk menghitung dan menentukan model yang biasa digunakan untuk analisis data uji hidup.

Analisis regresi yang responnya termasuk salah satu keluarga eksponensial disebut Generalisasi Model Linier atau lebih dikenal dengan GLM (*Generalized Linear Models*). *Generalized Linear Models* (GLM) memperluas model regresi biasa yang mencakup variable respon berdistribusi tidak normal dan fungsi model mean. Ada tiga komponen utama dalam analisis GLM seperti berikut ini.

1. Komponen Random

   Komponen random dari GLM terdiri dari variable respon Y dengan observasi bebas $(y_1, ..., y_n)$ dari sebuah distribusi dalam keluarga eksponensial. Bentuk fungsi densitas probabilitas dari distribusi keluarga eksponensial, yaitu sebagai berkut.

$$f(y_i;\theta_i,\psi) = exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\psi)} + c(y_i,\psi)\right\}$$

$$\text{(McCullagh and Nelder, 1989)}$$

2. Komponen Sistematis

   Komponen sistenmatis dari GLM adalah hubungan dari vektor $(\eta_{1,} \dots, \eta_n)$ untuk menjelaskan variable-variabel yang berhubungan dalam sebuah model linier.

$$\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

(2.9)

Atau dalam matriks dituliskan dalam bentuk $\boldsymbol{\eta} = \boldsymbol{X\beta}$

$$\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix},\ \beta^0 = \begin{bmatrix} {\beta_0}^0 \\ {\beta_1}^0 \\ \vdots \\ {\beta_p}^0 \end{bmatrix},$$

Kombinasi linier dari variable yang dijelaskan disebut predictor linier. Dengan $\boldsymbol{\eta}$ adalah vektor (n x 1) dari observasi, X adalah matriks (n x c) dari variable bebas, β adalah matriks (c x 1) dari koefisien regresi dengan $c = p + l$.

3. Fungsi *Link*

Fungsi *link* adalah fungsi yang menghubungkan ekspektasi variable respon dengan komponen sistematis.

Diberikan fungsi $\mu_i = E(Y_i), i = 1, 2, 3, \ldots, n$ model link μ_i untuk $\boldsymbol{\eta_i}$ adalah $\boldsymbol{\eta_i} = g(\mu_i)$.

$\boldsymbol{\eta_i} = g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}$ dengan $i = 1,2, \ldots, n$

Fungsi link $g(\mu_i) = \mu_i$ dinamakan identitas link dengan $\boldsymbol{\eta_i} = \boldsymbol{\mu_i}$. Fungsi link yang menstransformasikan nilai meannya ke parameter *natural* dinamakan *kanonikal link.*

2.6 Estimasi Parameter

Pada regresi logistik, untuk mencari penaksir titik, terdapat tiga metode yang umumnya digunakan, yaitu : metode momen, *maximum likelihood estimation*, dan metode *bayes*. Namun, dari beberapa metode tersebut, metode *maximum* *likelihood estimation* merupakan metode yang kemungkinan menghasilkan penaksir yang baik.

Dalam menentukan penaksir *maximum likelihood*, diperlukan adanya fungsi yang disebut fungsi likelihood. Fungsi ini menyatakan kemungkinan dari data yang diamati sebagai fungsi dari parameter yang tidak diketahui. Untuk menaksir parameter β, dengan cara memaksimumkan fungsi likelihood dengan syarat bahwa data tersebut mengikuti suatu distribusi tertentu. Pada regresi logistik biner, setiap pengamatan mengikuti distribusi Bernoulli sehingga didapat fungsi likelihoodnya (Hosmer dan Lemenshow, 2000 : 9).

Misalkan $y_1, y_2, \dots, y_n$ merupakan variable acak independen dari suatu distribusi dengan pdf $f(y_1, \beta)$, untuk $\beta \in \Omega$ dengan Ω ruang parameter, maka pdf bersama antara $y_1, y_2, \dots, y_n$ adalah $f(y_1, \beta), f(y_2, \beta), \dots . f(y_n, \beta)$. Jika pdf bersama tersebut dinyatakan sebagai fungsi terhadap β maka dinamakan fungsi likelihood yang dinotasikan L (Hogg dan Craig,1995) dan ditulis :

$$L(\beta; y_1, y_2, \dots, y_n) = f(y_1, \beta), f(y_2, \beta), \dots . f(y_n, \beta) = \Pi_{i=1}^{n} f(y_i, \beta) \quad (2.10)$$

Jika statistic $\hat{\beta} = t(y_1, y_2, \dots, y_n)$ memaksimumkan $L(\beta; y_1, y_2, \dots, y_n)$ maka statistik $\hat{\beta} = t(y_1, y_2, \dots, y_n$) adalah maximum likelihood estimator dari β.

2.7 Algoritma *Fisher-Scoring*

Menurut Thomas,dkk(2011) algoritma adalah prosedur komputasi dimana mengambil sebuah nilai atau menentukan nilai sebagai input dan menghasilkan beberapa nilai sebagai output. Sebuah algoritma adalah sebuah urutan langkah-langkah komputasi yang dapat merubah sebuah input menjadi output. Menuru Lawless(2003) misalkan terdapat fungsi-fungsi $\frac{\partial l(\beta)}{\partial \beta j} = 0$ dengan j=1,2,….p maka nilai-nilai βj yang memenuhi fungsi implisit tersebut dapat diperoleh melalui iterasi Newton-Raphson sebagai berikut.

$$\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} - H\left(\beta^{(r)}\right)^{-1} D\left(\beta^{(r)}\right) \quad , r = 0,1,2, \ldots \tag{2.11}$$

dengan

$\beta = (\beta_{1,}\beta_{2,}, \ldots, \beta_{p,})^T$, $D(\beta) = (\frac{\partial l(\beta)}{\partial \beta j}, \ldots., \frac{\partial l(\beta)}{\partial \beta j})^T$, $H(\beta) = [\frac{\partial^{(2)} l(\beta)}{\partial \beta j \partial \beta k}]^{(\ \ )}$ dan j,k-0,1,2,…, p

Adapun langkah-langkah dalam algoritma Newton Raphson adalah sebagai berikut:

1) Menentukan nilai awal β_0
2) Menentukan $D(\beta_0) dan\ H(\beta_0)$
3) Menghitung estimator parameter untuk r=0,1,2,.. dengan menggunakan persamaan (2.8)
4) Mengulangi iterasi sampai diperoleh nilai yang konvergen yaitu max $|\hat{\beta}^{(r+1)} - \hat{\beta}^{(r)}| \leq \varepsilon$ dengan ε adalah konstanta positif yang ditentukan.

2.8 Uji Signifikansi Parameter

Pengujian parameter dalam model regresi logistik biner bertujuan untuk mengetahui apakah variabel prediktor dalam model yang diperoleh berpengaruh secara signifikan terhadap variabel respon. Pengujian tersebut dapat dilakukan baik secara simultan maupun secara parsial. Untuk menguji signifikansi dari parameter dalam model dapat dilakukan dengan uji rasio likelihood, dan uji wald (Hosmer dan Lemeshow, 2000 : 11).

2.8.1 Uji Individu

Uji signifikansi parameter secara individu digunakan untuk memeriksa signifikansi parameter β terhadap variabel prediktor secara parsial atau individu. Proses pengujian yang dilakukan pada uji ini menggunakan uji Wald (Hosmer dan lemeshow, 2000 : 16) dan dapat dituliskan sebagai berikut:

$H_o : \beta_j = 0$ (parameter β tidak berpengaruh secara signifikan terhadap variabel respon secara individu)

$H_1 : \beta_j \neq 0 ; j = 1,2, \dots , p$ (parameter β berpengaruh secara signifikan terhadap variabel respon secara individu)

Bentuk statistik uji Wald :

$$W_j = \frac{(\hat{\beta}_j)}{[SE(\hat{\beta}_j)]} \tag{2.12}$$

dimana

β_j = nilai koefisien regresi logistik untuk variabel ke- j

$\hat{\beta}_j$ = penduga dari parameter β

$SE(\hat{\beta}_j)$ = nilai standar error untuk variabel ke -i

Hasil rasio diperoleh dari uji statistik dibawah hipotesis H_0, akan mengikuti Z atau sebaran normal baku. Kriteria penolakan yaitu Tolak Ho, jika $W_j > Z_{(\propto;db)}$yang berarti parameter β berpengaruh secara signifikan terhadap variabel respon secara individu.

2.8.2 Uji Simultan

Uji simultan atau serentak merupakan pengujian yang dilakukan untuk memeriksa signifikansi parameter β terhadap variabel prediktor secara keseluruhan. Pada uji ini, statistik yang digunakan adalah statistik uji G atau *Likelihood Ratio Test* (Hosmer dan Lemeshow, 2000 : 14).

Hipotesis :

H_o : $\beta_1 = \beta_2 = \cdots = \beta_j = 0$ (variabel prediktor tidak berpengaruh secara signifikan terhadap variabel respon)

H_1 : paling sedikit ada satu $\beta_j \neq 0$ *($i = 1,2, \dots, j$)* (variabel prediktor berpengaruh secara signifikan terhadap variabel respon)

Bentuk statistik uji *Likelihood Ratio Test*:

$$G = -2\, ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n} \hat{\pi}_i{}^{y_i} \, (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \tag{2.13}$$

atau $G = 2\{\sum_{i=1}^{n} [y_i \, ln\, (\hat{\pi}_i) + (1 - y_i)\, ln\, (1 - \hat{\pi}_i)] - [n_1\, ln\, (n_1) + n_0\, ln(n_0) - n\, ln(n)]\}$

dimana :

n_1 = banyaknya observasi yang berkategori 1 (individu mati selama penelitian)

n_0 = banyaknya observasi yang berkategori 0 (individu hidup sampai akhir penelitian)

n = banyaknya observasi ($n_1 + n_0$)

Dalam hipotesis H_0, bahwa β_1 sama dengan nol, statistik uji G mengikuti distribusi *Chi-Square*. Hasil uji statistik diatas bahwa H_0 ditolak jika $G > \chi^2{}_{(db,\propto)}$ atau *p-value* $< \propto$.

2.8.3 Uji Kesesuaian Model

Jika T adalah variable random dengan distribusi kontinu $F(t)$ dan mempertimbangkan rumusan hipotesis:

$H_0: F(t) = F_0(t)$ (Data berdistribusi Log-Logistik)

$H_1: F(t) \neq F_0(t)$ (Data tidak berdistribusi Log-Logistik)

Statistic Anderson Darling untuk data tersensor tipe II adalah sebagai berikut.

$$A_{n,p}^2 = -\sum_{i=1}^{r}\left(\frac{2_i - 1}{n}\log F_0\left(t_{(i)}\right) - \frac{2n - 2_i + 1}{n}\log\left[1 - F_0\left(t_{(i)}\right)\right]\right)$$
$$+\frac{r^2}{n}\log F_0(L) - \frac{(n-r)^2}{n}lo\,g[1 - F_0(L)] - nF_0(L) \tag{2.14}$$

Koziol dan Byar (1975) menetapkan bahwa $A_{n,p}^2$ mendekati distribusi dari $D_{n,p}^{\alpha}$ dalam kasus data tersensor tipe II, $p = F_0(L)$. Kriteria penolakan : Tolak H_0 jika $A_{n,p}^2 > D_{n,p}^{\alpha}$.

2.9 Program R

Pada saat ini, banyak paket perangkat lunak yang digunakan dalam membantu perhitungan estimasi parameter dan mengolah data sampai mendapatkan nilai hasil estimasi. Menurut (Rosadi,2010:1), R merupakan suatu sistem analisis statistika yang relatif lengkap, sebagai hasil dari kolaborasi riset berbagai statistikawan di seluruh dunia. Saat ini R dapat dikatakan *lingua franca* (bahasa standar) untuk keperluan komputasi statistika modern. Versi awal R dibuat tahun 1992 di Universitas Auckland, New Zaeland oleh Ross Ihka dan Robert Gentleman (yang menjadi asal muasal akronim nama R untuk perangkat lunak ini).

Kelebihan program R dapat dituliskan sebagai berikut.

1. Probabilitas, jika memilih perangkat lunak ini, pengguna (*user*) bebas untuk mempelajari dan menggunakannya sampai kapan pun (berbeda, misalnya dengan lisensi perangkat berversi pelajar)
2. Multiptalform, R merupakan sistem operasi multiplatform, lebih kompatibel dari pada perangkat lunakstatistika manapun yang pernah ada. Dengan demikian, jika pengguna memutuskan untuk berpindah ke sistem operasi, penyesuaiannya akan relatif lebih mudah untuk dilakukan.

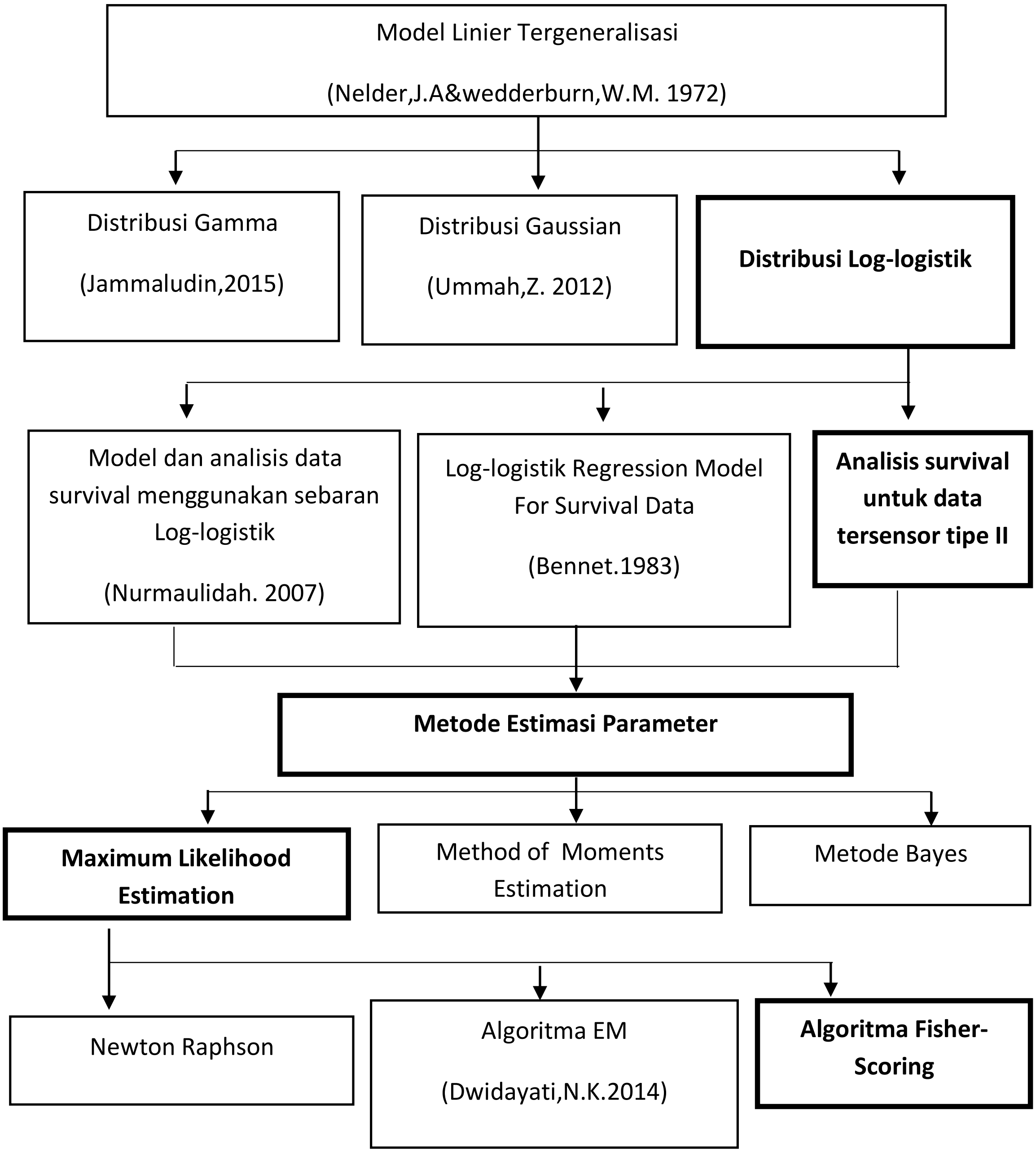
3. Umum dan berada dibarisan terdepan, berbagai metode analisis statistika (metode klasik maupun metode baru) telah diprogramkan ke bahasa R. Dengan demikian, perangkat lunak ini dapat digunakan untuk berbagai macam analisis statistika, baik pendekatan klasik maupun pendekatan statistika modern.
4. Bisa diprogram, pengguna dapat memprogramkan metode baru atau mengembangkan modifikasi dari fungsi-fungsi analisis statistika yang telah ada dalam sistem R.
5. Bahasa berbasis analisis matriks, bahasa R sangat baik untuk melakukan pemrograman dengan baris matriks (seperti halnya dengan bahasa MATLAB dan GAUSS).
6. Fasilitas grafik yang relatif baik.

2.10 Kerangka Berpikir

Menurut Nelder,J.A&wedderburn,W.M. 1972 model linier merupakan model statistika yang digunakan untuk menjelaskan hubungan antara dua variabel atau lebih. Ketika distribusi dari variabel respon merupakan anggota keluarga eksponensial maka disebut model linier tergeneralisasi. Model linier tergeneralisasi termasuk kasus yang khusus yang dapat digunakan untuk menghitung dan menentukan model yang biasa digunakan untuk analisis data uji hidup. Salah satu distribusi anggota keluarga eksponensial yang dapat digunakan untuk menganalisis data uji hidup adalah distribusi Gamma, distribusi Gaussian dan distribusi Log-logistik. Pada tahun 2015 Jammaludin telah meneliti tentang model linier tergeneralisasi Gamma pada data uji hidup lampu, kemudian Zahrotul Ummah meneliti tentang model linier tergeneralisasi Gaussian pada data perkembangan nilai ekpor dan impor hasil pertanian. Penelitian selanjutnya mengenai model regresi Log-logistik pada data uji hidup oleh Bennet,1983 yang hasilnya adalah distribusi log-logistik merupakan distribusi yang memiliki nilai hazard atau nilai kegagalan yang tidak monoton naik maupun tidak monoton turun. Distribusi log-logistik sangat cocok digunakan untuk menganalisis ketahanan suatu individu maupun suatu unit. Dalam menentukan estimasi parameter suatu data uji hidup metode yang biasa digunakan adalah metode bayes dan estimasi dengan metode likelihood. Estimasi dengan menggunakan metode likelihood lebih sering digunakan untuk penelitian, Estimasi dengan likelihood merupakan metode pendekatan geometris dengan meminimumkan galatnya. Namun nilai yang dihasilkan masih kurang emplisit sehingga perlu juga digunakan pendekatan dengan metode numerik *Newton Raphson*. Pada penelitian yang dilakukan oleh Dwidayati,N.K.2014 mengenai "Estimasi Parameter dalam Model Weibull untuk Analisis Cre rate Penderita Kanker Payudara" estimasi parameter dalam model Mixture menggunakan algoritma EM namun mengalami kendala dalam penyelesaian persamaan yang diperoleh dari turunan pertama pada ekspektasi fungsi log-likelihood dalam bentuk not close form sehingga untuk menyelesaikan persamaan ini diperlukan iterasi yang tidak dapat dikerjakan secara manual dan menggunakan algoritma yang sesuai. Kus,C & Kaya,M.F.2006 mengenai algoritma EM dengan hasil algoritma akan konvergen dengan andal tetapi agak lambat (dibandingkan dengan Newton {metode Raphson) ketika jumlah informasi dalam data yang hilang relatif besar. Oleh karena itu, dalam penelitian ini, perkiraan kemungkinan maksimum untuk parameter distribusi Loglogistic adalah diperoleh dengan menggunakan algoritma EM berdasarkan hak Tipe-II progresif sampel yang disensor.

Pada perkembangannya metode *Newton Raphson* telah dikembangkan menjadi algoritma *Fisher scoring*. Alasan penulis menggunakan algoritma *Fisher scoring* adalah algoritma *fisher scoring* merupakan pengembangan dari *Newton Raphson*, dan apabila *Newton Raphson* tidak konvergen algoritma *Fisher scoring* tetap konvergen.

Skema :



Gambar 2.1 Kerangka Berfikir

BAB V

PENUTUP

5.1 Simpulan

Dari hasil pengolahan data dan interpretasi hasil yang didapatkan pada bab sebelumnya, dapat ditarik kesimpulan sebagai berikut.

1. Hasil uji kesesuaian model linier tergeneralisasi log-logistik diperoleh melalui uji *Residual Deviance* adalah $Y_i = \alpha + \beta\varepsilon_i$ dengan, $\hat{\alpha}$ dan $\hat{\beta}$

$$\frac{-r\hat{\beta}}{\hat{\alpha}} + \left(\frac{2\hat{\beta}}{\hat{\alpha}}\right)\sum_{i=1}^{r}\left[1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\right]^{-1}\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}} - \frac{(r-n)}{1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}}\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)\left(\frac{t_r}{\hat{\alpha}}\right)^{\hat{\beta}} = 0$$

   Dan

$$-\frac{r}{\hat{\beta}} + \sum_{i=1}^{r}\ln\frac{t_i}{\hat{\alpha}} - 2\sum_{i=1}^{r}\left[1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\right]^{-1}\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\ln\left(\frac{t_i}{\hat{\alpha}}\right)$$
$$+\frac{(r-n)}{1+\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}}\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)\ln\left(\frac{t_r}{\hat{\alpha}}\right) = 0$$

   Karena kedua persamaan memiliki bentuk yang kompleks dan tidak dapat diselesaikan secara analitik sehingga untuk menyelesaikan persamaan tersebut digunakan metode iterasi dengan menggunakan algoritma *Fisher-scoring.*

2. Hasil penerapan estimasi model linier tergeneralisasi log-logistik pada data tahan hidup tersensor tipe II menggunakan algoritma *Fisher-*

scoring pada data kanker paru-paru dengan iterasi sebanyak empat kali diperoleh bentuk model dugaan sebagai berikut.

$$y_i = -2.05650 + 1.62366x_1 + 1.65555x_2 - 0.06665x_3 + 1.19602x_4$$

Nilai estimasi parameter model linier tergeneralisasi log-logistik dengan menggunakan *Maximum Likelihood Estimator* $\hat{\alpha}$ adalah 21.95577 dan $\hat{\beta}$ adalah 1.507590.

5.2 Saran

Saran yang dapat diberikan agar dapat diperhatikan oleh peneliti selanjutnya adalah sebagai berikut.

1. Untuk mencari estimator model linier tergeneralisasi yang berdistribusi Log-logistik dapat menggunakan metode *maximum likelihood estimation.*
2. Jika metode *maximum likelihood estimation* menghasilkan estimator estimator model linier tergeneralisasi yang berdistribusi Log-logistik berbentuk implisit maka bisa menggunakan bantuan pendekatan numerik menggunakan algoritma *Fisher-scoring.*
3. Bukan hanya distribusi Log-logistik yang dapat dicari estimatornya dengan menggunakan algoritma Fisher-scoring, untuk penelitian selanjutnya bisa menggunakan distribusi Gamma, Poisson maupun Weibull.

DAFTAR PUSTAKA

Alakus,K & Erilli, N. 2011. *Confidence Intervals Estimation for survival function in log-logistic Distribution and Proportional odds Regression Based on Cencored Survival time Data*. Biometrics and Biostatistics, 11: 2-3.

Bain, L.J. and Engelhart, M. 1992. *Introduction to Promodulity and Mathematical Statistics, 2nd ed.* California: Duxbury Press, Belmont.

Bennet, S. 1983. *Log-logistic Regression Models for Survival Data*. Applied Statistics, 32. 165-171.

Dwidayanti,N.K.& Abidin,Z. 2014. *Estimasi Parameter dalam Model Weibull untuk Analisis Cre rate Penderita Kanker Payudara*. Jarlitbangnov: Penelitian Fundamental RISTEKDIKTI.

Everitt, S.B. 1994. *A Handbook of Statistical Analysis Using S-PLUS*. London: Chapman and Hall.

Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. Canada: John Wiley & Sons, Inc.

Lawless, J.F. 2003. *Statistical Models and Methods for Lifetime Data*. Canada: John Wiley & Sons, Inc.

Lee, E.T. 2003. *Statistical Methods for Survival Data Analysis 3rd Edition*. Canada: John Wiley & Sons, Inc.

Marius, O. U. and O. I. C. *Anaene, Estimating the Fisher's Scoring Matrix Formula from Logistic Model.* American Journal of Theoretical and Applied Statistics,vol. 2. pp. 221-227. 2013.

McCullagh, P. dan Nelder, J.A, 1989. *Generalized Linear Models Second Edition*. London: Chapman and Hall.

Nelder,J.A & Wedderburn,W.M. 1972. *Generalized Linear Models*. Journal Of The Royal Statistical Society,vol.135. 3(1972):370-384.

Nurmaulidah. 2007. *Model dan Analisis Data Survival menggunakan Sebaran Log logistic.* Skripsi. Bogor: Institut Pertanian Bogor.

Ojo,M.O & Olapade,A.K. 2003. *On The Generalized Logistic and Log-Logistic Distributions*. Kragujevac J.Math, 25(2003): 65-73.

Putri,A.N,dkk. 2016. *Informasi Fisher pada Algoritme Fisher Scoring untuk Estimasi Parameter Model Regresi Logistik Ordinal Terboboti Geografis (RLOTG)*. Seminar Nasional Matematika dan Pendidikan Matematika UNY, hlm. 59-64.

Sari,Dewi.2011. *Analisis survival untuk data tersensor tipe II menggunakan model distribusi log-logistik*. Skripsi. Surabaya: Unniversitas Negeri Yogyakarta.

Schworer, A. and P. Hovey. 2004. *Newton Raphson versus Fisher Scoring Algorithms in Calculating Maximum Likelihood Estimates*. Dayton.

Smyth, G.K, 2002, *Optimation*, Encyclopedia of Environmetrics, Volume 3, pp1481-14687 Edited:Abdel H. El Shaarawi and Walter W.Piegorsch,Chicester

Sudarno,dkk.2012.*Model Regresi Data Tahan Hidup Tersensor Tipe III Berdistribusi Log-logistik.*volume 1,hal 83-92.

Thomas, H. C. dkk. 2001. *Introduction To Algorithms*. USA: MIT Press.

Tirta, I.M., 2008, *Model Statistika Linier (Versi Elektronik)*, Jurusan Matematika FMIPA, Universitas Jember

Utami, D.T. 2015. *Analisis Data Uji Hidup Pasien Knaker Paru-Paru di RSUP Dr. Kariadi Semarang Dengan Model Regresi*. Skripsi. Semarang: Unniversitas Negeri Semarang.

Ummah,Z. 2012. *Estimasi Model Linear Tergeneralisasi Gaussian Berdasarkan Maximum Likelihood Estimator dengan Menggunakan Algoritma Fisher Scoring*. Skripsi. Surabaya: Universitas Airlangga.