Prosiding Seminar Matematika dan Pendidikan Matematika *ISBN: 978-602-6122-20-9*
hal 927-934 November 2016 *http://jurnal.fkip.uns.ac.id*

MASALAH NILAI AWAL ITERASI NEWTON RAPHSON UNTUK ESTIMASI PARAMETER MODEL REGRESI LOGISTIK ORDINAL TERBOBOTI GEOGRAFIS (RLOTG)

Shaifudin Zuhdi, Dewi Retno Sari Saputro

Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
shaifudinzuhdi@gmail.com

Abstrak: Model RLOTG merupakan pengembangan dari model regresi logistik dengan variabel respon memiliki lebih dari dua kategori dan memiliki tingkatan. Model tersebut merepresentasikan hubungan antara variabel bebas dengan variabel respon dan peluang kejadian yang diakibatkan oleh variabel bebas yang memperhatikan pengaruh lokasi pengamatan. Parameter untuk model RLOTG menunjukkan karakteristik populasi pengamatan, umumnya estimasi parameter menggunakaan metode *maximum likelihood*. Metode tersebut menghasilkan suatu sistem persamaan nonlinier yang sulit ditentukan penyelesaian eksaknya maka digunakan pendekatan numerik. Pendekatan numerik yang digunakan adalah iterasi Newton Raphson. Iterasi tersebut dimulai dengan menentukan nilai awal (*initial value*). Perhitungan dengan Newton Raphson dapat konvergen atau divergen sesuai pada nilai awal yang diberikan. Tujuan penelitian ini adalah melakukan estimasi parameter pada model RLOTG dengan iterasi Newton Raphson menggunakan dua nilai awal yaitu 0 dan nilai awal dari estimator parameter model regresi logistik ordinal (RLO). Penerapan dilakukan pada pemodelan tingkat banyak penderita demam berdarah dengue (DBD) di Kota Semarang dengan model RLO untuk mendapatkan estimator parameter sebagai nilai awal, selanjutnya ditentukan nilai estimasi parameter model RLOTG. Hasil penelitian menunjukkan bahwa dengan nilai awal 0 dicapai konvergen pada seluruh lokasi, namun iterasi yang diperlukan lebih banyak daripada nilai awal dari estimasi parameter model RLO. Pada penerapan, ditunjukkan bahwa estimasi parameter dengan nilai awal dari estimator parameter model RLO membutuhkan iterasi yang lebih sedikit daripada nilai awal 0, namun untuk beberapa lokasi, estimasi parameter modelnya tidak mencapai konvergen.

Kata kunci: Estimasi Parameter, Newton Raphson, Nilai Awal, Model RLOTG

PENDAHULUAN

Metode numerik adalah suatu algoritme, menyangkut langkah-langkah untuk menyelesaikan masalah numerik. Dalam penyelesaian masalah numerik digunakan suatu pendekatan dengan teknik kalkulasi berulang (teknik iterasi) dalam menentukan solusi hampiran. Karena teknik-teknik iterasi ini merupakan model pendekatan, maka terdapat kesalahan atau tidak menghasilkan solusi eksak. Teknik-teknik yang digunakan memiliki potensi membuat suatu kesalahan yang di evaluasi secara bertahap untuk mendapatkan nilai kesalahan yang sangat kecil. Kesalahan (error) terjadi pada penyelesaian numerik seperti penyederhanaan hasil numerik, data-data yang diperoleh pada hasil pengukuran yang kurang akurat atau pembulatan, nilai-nilai pendekatan pada numerik dan sebagainya. Menurut Butt (2009), ada tiga jenis sumber kesalahan dalam perhitungan numerik yaitu kesalahan manusia (*human error*), kesalahan pembulatan (*round off error*), dan kesalahan pemotongan (*truncation error*).

Penyelesaian masalah numeric dalam menentukan solusi persamaan nonlinier yaitu pencarian akar $f(x) = 0$ dilakukan secara iterasi dengan beberapa metode. Metode pencarian akar tersebut salah satunya adalah metode Newton Raphson. Kinerja metode ini lebih cepat dalam mencapai konvergensi, karena memiliki laju konvergensi kuadrat. Metode Newton Raphson merupakan metode terbaik yang terkenal untuk menentukan akar suatu fungsi persamaan nonlinier, yang ditemukan oleh Isaac Newton (1669) & Josep Raphson (1690) (Meza, 2010). Namun penggunaan metode Newton Rapshon masih terdapat kelemahan dalam mencapai konvergensi, salah satu penyebabnya adalah kesalahan pembulatan pada barisan iterasi yang dihasilkan komputer sehingga berbeda dengan barisan iterasi yang dihasilkan secara teori (Ypma, 1983 &Argyros, 2000).

Model RLO dengan data bertipe ordinal pada variabel respon dan bertipe interval pada variabel bebasnya (McCullagh & Nelder, 1983), estimasi parameter RLO dengan metode *maximumlikelihood* yang menghasilkan sistem persamaan nonlinier, yang sulit ditentukan penyelesaian eksaknya.Model RLOTG merupakan representasi hubungan antara variabel respon dengan variabel bebas yang mempertimbangkan lokasi secara geografis (Atkinson *et al.*, 2003). Menurut Rifada & Purhadi (2011), faktor letak secara geografis merupakan faktor pembobot pada model RLOTG. Seperti halnya model RLO, estimasi parameter model RLOTG menggunakaan metode *maximum likelihood*. Metode tersebut menghasilkan suatu sistem persamaan nonlinier yang sulit ditentukan penyelesaian eksaknya.

Kedua estimasi parameter pada model RLO dan model RLOTG memerlukan penyelesaian pendekatan, salah satunya adalah metode Newton Raphson (NR). Perhitungan tersebut dapat menggunakan pemrograman rekursif berbasis *software*. Menurut Darnius (2002), *software* R merupakan *software* terintegrasi yang memiliki fasilitas untuk merekayasa data, perhitungan, dan penampilan grafik. Estimasi parameter dilakukan pada *software* R dengan membuat program berdasarkan algoritme.

Penyelesaian dengan metode NRkonvergen dengan cepat apabila iterasi dimulai cukup dekat dengan akar-akar yang diinginkan. Sehingga metode NR untuk penyelesaian pendekatan untuk estimasi parameter model RLO dan model RLOTG. Pada penelitian ini penulis tertarik untuk melakukan simulasi menyangkut masalah nilai awal iterasi dengan metode NR dengan program R yang telah dibuat oleh Zuhdi & Saputro (2016).

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan yakni melakukan running program dan simulasinya untuk estimasi parameter model RLOTG. Langkah-langkah dalam penelitian adalah mengeksekusi program yang telah disusun. Uraian langkahnya dinyatakan sebagai berikut.

A. *Mengestimasi parameter model RLO dengan menjalankan program yang telah dibuat.*

Berikut langkah-langkah yang dilakukan untuk mengestimasi parameter model RLO.

1. Memberikan nilai awal parameter model berupa vektor dengan ukuran 7×1 pada proses iterasi dengan *software* R.
2. Memberikan batas eror pada susunan sintaks program, sebagai batas berhentinya iterasi dan diperoleh nilai hasil estimasi parameter.
3. Melakukan pengecekan nilai hasil estimasi parameter merupakan nilai yang maksimum berdasarkan syarat metode *maximum likelihood*.

B. *Mengestimasi parameter model RLOTG dengan menjalankan program yang telah dibuat.*

Berikut langkah-langkah yang dilakukan untuk mengestimasi parameter model RLOTG.

1. Memberikan nilai awal parameter model berupa vektor dengan ukuran 7×1pada proses iterasi dengan nilai awal 0 dan nilai awal hasil dari estimasi parameter model RLO pada langkah A dengan *software* R.
2. Memberikan batas eror pada susunan sintaks program, sebagai batasberhentinya iterasi dan diperoleh nilai hasil estimasi parameter.
3. Menentukan lokasi yang akan diestimasi parameter. Karena model RLOTG merupakan model lokal, maka setiap lokasi pengamatan memiliki model.
4. Melakukan pengecekan nilai hasil estimasi parameter merupakan nilai yang maksimum berdasarkan syarat metode *maximum likelihood*.
5. Melakukan analisis hasil iterasi berdasarkan hasil running program pada *software* R.

HASIL DAN PEMBAHASAN

Hasil dan pembahasan meliputi model RLO yang diperlukan dan untuk estimasi parameter model RLOTG serta hasil estimasi parameter model RLOTG. Berikut adalah uraiannya.

A. Model RLO

Model RLO merupakan model yang merepresentasikan hubungan antara variabel respon yang mempunyai skala ordinal dengan variabel bebas yang bersifat kategori dan/atau kontinu (Hosmer & Lemeshow, 2000). Model regresi logistik dapat juga disebut sebagai model logit. Jika variabel respon memiliki G kategori berskala ordinal dan x_i menyatakan vektor variabel bebas pada pengamatan ke-i, maka model RLO dapat dinyatakan sebagai

$$\text{logit}[P(Y_a \leq g|x_a)] = \alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i},$$

dengan $g = 1,2,\dots,G-1$, $a = 1,2,\dots,n$, dan $i = 1,2,\dots,p$. $P(Y_a \leq g|x_i)$menyatakan peluang kumulatif kurang dari atau sama dengan kategori ke-g terhadap x_i. Parameter α_g merupakan *intercept* yang memenuhi kondisi $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{g-1}$ dan β_i merupakan vektor koefisien regresi yang bersesuaian dengan x_i. Dengan peluang masing-masing G kategori respon adalah

$$P_g(x_i) = \frac{\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}{1+\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})} - \frac{\exp\left(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i}\right)}{1+\exp\left(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i}\right)}, g = 1,2,\dots,G.$$

Nilai peluang untuk masing-masing kategori respon digunakan sebagai pedoman untuk pengklasifikasian.

Estimasi parameter model RLO dilakukan dengan metode *maximum likelihood estimation* (*MLE*). Metode *MLE* digunakan untuk mendapatkan penyelesaian maksimum dari fungsi *likelihood*, namun hanya dapat diperoleh sistem persamaan nonlinier. Metode Newton Raphson digunakan untuk menyelesaikan sistem persamaan nonlinier hingga diperoleh nilai parameter yang maksimum. Pada model RLO, misalkan diambil n sampel vektor variabel random $Y_1, Y_2, \dots, Y_n$, dengan $Y_a = [y_{a1} y_{a2} \cdots y_{a,G-1}]^T$ dan peluang hasil kategori ke- g adalah $p_g(x_i)$, fungsi *likelihood*nya adalah

$$L = \prod_{a=1}^{n}\prod_{g=1}^{G}(P_g(x_i))^{y_{ag}}$$

$$= \prod_{a=1}^{n}\prod_{g=1}^{G}\left(\frac{\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}{1+\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})} - \frac{\exp(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}{1+\exp(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}\right)^{y_{ag}}.$$

Prinsip dari *MLE* adalah mengestimasi vektor parameter $V = [\alpha_1\alpha_2 \ldots\ \alpha_{G-1}\beta_1\beta_2 \ldots\ \beta_p]^T$ dengan cara memaksimumkan fungsi *likelihood*. Untuk memudahkan perhitungan, fungsi *likelihood* ditransformasi menjadi *ln-likelihood* yang dinyatakan sebagai

$$\ell = \sum_{a=1}^{n}\sum_{g=1}^{G} y_{ag}\ ln\left(\frac{\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}{1+\exp(\alpha_g + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})} - \frac{\exp(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}{1+\exp(\alpha_{g-1} + \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta_i})}\right).$$

B. Model RLOTG

Peluang yang dapat digunakan untuk RLOTGadalah logit kumulatif. Pada logit kumulatif, masing-masing parameter dipengaruhi lokasi geografis data. Purhadi *et al.*(2012) menyatakan model RLOTG dengan variabel respon G buah kategori berskala ordinal adalah

$$\text{Logit}[P(Y_a \le g|x_a)] = \boldsymbol{\alpha_g}(u_a, v_a) + \boldsymbol{x}_i^{T}\boldsymbol{\beta_i}(u_a, v_a)$$

dengan $P(Y_a \le g|x_a)$ adalah peluang kumulatif kategori ke-g terhadap x_a , $g = 1,2,\ldots,G$ adalah banyaknya kategori variabel respon, $\alpha_g(u_a, v_a)$ adalah parameter *intercept* dan memenuhi kondisi $\alpha_1(u_a, v_a) \le \alpha_2(u_a, v_a) \le \cdots \le \alpha_{G-1}(u_a, v_a)$, (u_a, v_a) adalah titik koordinat (*lattitude*, *longitude*) untuk lokasi a, dan $\beta_i(u_a, v_a) = [\beta_1(u_a, v_a)\beta_2(u_a, v_a) \ldots\ \beta_p(u_a, v_a)]$ merupakan vektor parameter regresi untuk lokasi ke-a. Menurut Purhadi *et al.*(2012), peluang masing-masing G kategori respon adalah

$$p_g(x_a) = \frac{\exp(\boldsymbol{\alpha_g}(u_a,v_a) + \boldsymbol{x}_i^{T}\boldsymbol{\beta_i}(u_a,v_a))}{1+\exp(\boldsymbol{\alpha_g}(u_a,v_a) + \boldsymbol{x}_i^{T}\boldsymbol{\beta_i}(u_a,v_a))} - \frac{\exp(\boldsymbol{\alpha_{g-1}}(u_a,v_a) + \boldsymbol{x}_i^{T}\boldsymbol{\beta_i}(u_a,v_a))}{1+\exp(\boldsymbol{\alpha_{g-1}}(u_a,v_a) + \boldsymbol{x}_i^{T}\boldsymbol{\beta_i}(u_a,v_a))}.$$

Estimasi parameter model RLOTG menggunakan metode *maximum likelihood* terboboti. Misalkan diambil n sampel acak $Y_1, Y_2, \ldots, Y_n$ dengan peluang hasil pada kategori ke-g adalah $p_g(x_a)$. Menurut Hosmer & Lemeshow (2000), fungsi *likelihood* untuk n pengamatan adalah

$$L = \prod_{a=1}^{n}[p_1(x_a)^{y_{a1}}p_2(x_a)^{y_{a2}} \ldots p_G(x_a)^{y_{aG}}]$$

dengan $y_{ag} = 1$ jika $Y_a = g$ dan $y_{ag} = 0$ jika $Y_a \ne g$.

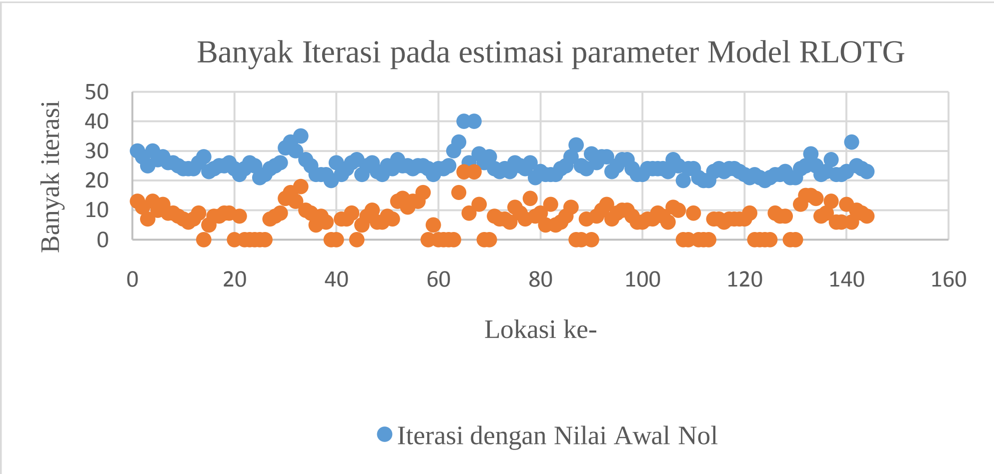
Pembobot pada setiap lokasi memiliki nilai berbeda yang dipengaruhi lokasi lain yang berdekatan. Pembobot diberikan pada bentuk *likelihood* untuk model lokal RLOTG. Untuk memudahkan perhitungan, *likelihood* terboboti ditransformasi menjadi *ln-likelihood* terboboti. Misalkan pembobot untuk setiap lokasi (u_a, v_a) adalah $w_{ab}\,(u_a, v_a)$ fungsi *ln-likelihood* terboboti adalah

$$l = \sum_{a=1}^{n} [y_{a1} Ln\big(p_1(x_a)\big) + y_{a2} Ln\big(p_2(x_a)\big) + \ldots + y_{aG}\, Ln\big(p_G(x_a)\big)]\, w_{ab}\,(u_a, v_a).$$

C. *Hasil Iterasi Model RLO dan RLOTG*

Iterasi memiliki satu arti yang penting yaitu pengulangan, namun pengertian ini dikembangkan lebih lanjut sesuai dengan bidang yang bersangkutan seperti pemrograman yang menyatakan bahwa iterasi merupakan suatu langkah atau urutan yang terdiri atas satu langkah algoritme dan dilakukan dalam program *loop* yang sering diartikan sebagai program yang berulang.

Penentuan masalah nilai awal akan menentukan banyaknya iterasi, pada penelitian ini, menggunakan dua nilai awal yang berbeda. Nilai awal yang digunakan adalah nol dan nilai estimator parameter model RLO. Untuk menghentikan proses iterasi digunakan batas eror (e) dengan nilai 0.00001. Dari proses iterasi yang dilakukan yakni dilakukan pada 144 model RLOTG, terdapat 31 model yang tidak diperoleh estimasi parameternya karena belum dicapai konvergen sementara 113 model dapat diperoleh estimasi parameternya. Plot banyaknya iterasi Newton Raphson untuk estimasi parameter model RLOTG dengan dua nilai awal berbeda ditunjukkan pada Gambar 1.



GAMBAR 1. PLOT BANYAKNYA ITERASI PADA NILAI AWAL NOL DAN NILAI AWAL HASIL ESTIMASI MODEL RLO PADA 144 KELURAHAN DI KABUPATEN SEMARANG

Pada Gambar 1 tampak bahwa dengan digunakan nilai awal hasil estimasi model RLO lebih cepat mencapai konvergensi (jumlah iterasi lebih sedikit dibandingkan dengan nilai awal nol).

SIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa dengan nilai awal 0 dicapai konvergensinya pada estimasi parameter model RLOTG, namun iterasi yang diperlukan lebih banyak daripada nilai awal hasil estimasi parameter model RLO. Selainnya, pada beberapa lokasi dengan nilai awal dari estimasi model RLO estimasi parameter modelnya tidak mencapai konvergen.

Meskipun dengan nilai awal hasil estimasi parameter model RLO lebih baik daripada nilai awal nol, namun masih terdapat estimasi parameter model yang belum dicapai konvergen. Hal tersebut berbeda apabila digunakan nilai awal nol, meskipun banyaknya iterasi yang dipergunakan jauh lebih banyak namun semua model RLOTG mencapai konvergen. Bagi yang tertarik dengan metode numerik dapat melakukan kajian tentang hal ini lebih lanjut.

DAFTAR PUSTAKA

Atkinson, P. M., S. E. German, D. A. Sear, & M. J. Clark. (2003). *Exploring The Relations Between Riverbank Erison and Geomorphological Control Using Geographically Weighted Logistic Regression*. Ohio: Ohio State University. Vol. 35. Issue: 1. Pages: 58-82.

Butt, R. (2009). *Introduction to Numerical Analysis Using Matlab*, U.S.A (page 10-11).

Darnius, O. (2002). *Aplikasi Software R dalam Analisis Regresi*. Sumatra Utara: Jurusan Matematika. FMIPA USU.

Hosmer, D. W. & S. Lemeshow. (2000). *Applied Logistic Regression*. USA: John Willey and Sons Inc.

Meza, C. J. (2010). *Newton's Method*. Lawrence Berkeley National Laboratory. 94720. California.

McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models*. 2^{nd} Ed. Chapman and Hall.

Purhadi, Rifada, M. & Wulandari, P. (2012). *Geographically Weighted Ordinal Logistic Regression Model*: International Journal of Mathematics and Computation, Vol. 16, Issue: 3.

Rifada, M. & Purhadi. (2011). *Pemodelan Tingkat Kerawanan Demam Berdarah Dengue di Kabupaten Lamongan dengan Pendekatan Geographically Weighted Ordinal Logistic Regression*. Prosiding Seminar Nasional Statistika. Universitas Diponegoro. Semarang.

Ypma, T. J. (1982). *Numerical Solution of Systems of Nonlinear Algebraic Equations*. Thesis. Oxford.

Zuhdi, S. & Saputro, D. R. S. (2016). *R Programming for Parameters Estimation of Geographically Weighted Ordinal Logistic Regression (GWOLR) Model based on Newton Raphson*. The 2nd International Conference and Applied Statistics. Bandung: Universitas Padjajaran. (proses penerbitan)