Jurnal Matematika UNAND
Vol. **4** No. **2** Hal. 100 – 107
ISSN : 2303–2910

PERBANDINGAN METODE *MAXIMUM LIKELIHOOD* DAN METODE BAYES DALAM MENGESTIMASI PARAMETER MODEL REGRESI LINIER BERGANDA UNTUK DATA BERDISTRIBUSI NORMAL

CATRIN MUHARISA, FERRA YANUAR, HAZMIRA YOZZA
Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
catrinmuharisa@yahoo.com

Abstrak. Analisis regresi merupakan salah satu metode untuk melihat hubungan antara variabel bebas (*independent*) dengan variabel terikat (*dependent*) yang dinyatakan dalam model regresi. Beberapa metode yang bisa digunakan untuk mengestimasi parameter model regresi, diantaranya adalah metode klasik dan metode Bayes. Salah satu metode klasik adalah metode *maximum likelihood*. Penelitian ini membahas tentang perbandingan metode *maximum likelihood* dan metode Bayes dalam mengestimasi parameter model regresi linear berganda untuk data berdistribusi normal. Adapun rumus untuk mengestimasi parameter dengan metode *maximum likelihood* adalah $\hat{\boldsymbol{\beta}}\mathbf{=(X^TX)^{-1}X^TY}$ dan $\hat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{\infty} e_i$. Sedangkan untuk mengestimasi parameter dengan metode Bayes adalah dengan menggunakan distribusi *prior* dan fungsi *likelihood*. Distribusi *prior* yag dipilih pada kajian ini adalah $f(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} f(\beta_j|\sigma^2)f(\sigma^2)$ dengan $\beta_j \sim N(\mu_{\beta_j}, \sigma^2)$ dan $\sigma^2 \sim IG(a, b)$. Distribusi *prior* konjugat tersebut kemudian dikalikan dengan fungsi *likelihood* $L(\boldsymbol{\beta}, \sigma^2)$ sehingga membentuk distribusi *posterior* $f(\boldsymbol{\beta}|\sigma^2)$. Distribusi *posterior* inilah yang digunakan untuk mengestimasi parameter model melalui proses Markov Chain Monte Carlo (MCMC). Algoritma MCMC yang digunakan adalah algoritma Gibbs Sampler. Model regresi linear berganda yang diperoleh dengan metode maximum likelihood adalah

$$\hat{y} = -27,8210000 + 0,0307430X_1 + 0,0039211X_2 + 0,0034631X_3 + 0,6537000X_4$$

dengan kecocokan modelnya adalah sebesar 95,7 %. Sedangkan model regresi linear berganda yang diperoleh dengan metode Bayes adalah

$$\hat{y} = -26,620000 + 0,029380X_1 + 0,004204X_2 + 0,003321X_3 + 0,656200X_4$$

dengan kecocokan modelnya adalah sebesar 99,99 %. Dengan demikian dapat disimpulkan bahwa metode Bayes lebih baik dari pada metode maximum likelihood.

Kata Kunci: Model Regresi Linear Berganda, metode Maximum Likelihood, dan metode Bayes

1. Pendahuluan

Analisis regresi merupakan suatu alat dalam statistik yang mengalami perkembangan pesat dan banyak digunakan dalam berbagai bidang kehidupan. Analisis tersebut mempunyai tujuan untuk mengestimasi hubungan antara variabel *dependent* (terikat) dengan variabel *independent* (bebas). Pada penelitian ini, analisis regresi dilakukan dengan menggunakan distribusi normal.

Banyak metode yang bisa digunakan untuk mengestimasi parameter model regresi, diantaranya adalah metode klasik dan metode Bayes. Salah satu metode yang digunakan dalam metode klasik adalah metode *maximum likelihood.* Metode *maximum likelihood* merupakan metode estimasi parameter yang memaksimumkan fungsi kemungkinan. Pada metode Bayes, parameter populasi memiliki distribusi awal yang dinamakan dengan distribusi *prior.* Distribusi *prior* ini dapat berasal dari data penelitian sebelumnya atau berdasarkan intuisi seorang peneliti. Informasi *prior* dari sebaran parameter tersebut kemudian digabungkan dengan informasi dari data yang didapat dari pengambilan sampel sehingga didapat distribusi *posterior* dari parameter.

Dalam tulisan ini dilakukan pengujian mengenai pendugaan parameter dari model yang berdistribusi normal dengan menggunakan metode *maximum likelihood* dan metode Bayes. Dari uraian diatas peneliti akan membandingkan hasil estimasi dari kedua metode.

2. Model Regresi Linear Berganda

Model regresi linear berganda melibatkan lebih dari satu variabel independen $X_1, X_2, \ldots, X_n$ dengan satu variabel dependen. Bentuk model dituliskan sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \tag{2.1}$$

Apabila model dibentuk dalam matriks, maka Model (2.1) menjadi :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

dimana:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

3. Mengestimasi Parameter Model Regresi Menggunakan Metode *Maximum Likelihood*

Mengestimasi parameter dengan menggunakan metode *maximum likelihood* yaitu dengan menggunakan fungsi *likelihood* model regresi seperti berikut :

$$L(\boldsymbol{\beta}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2\right) \tag{3.1}$$

dan logaritma natural fungsi *likelihood* model regresi adalah:

$$\ln L(\boldsymbol{\beta}, \sigma^2) = \frac{n}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) - \left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2\right) \tag{3.2}$$

Dimana masing-masing turunannya adalah sebagai berikut :

$$\begin{aligned}
\frac{\partial \ln(\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)}{\partial \beta_0} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2 = 0 \\
\frac{\partial \ln(\beta_0, \beta_1, \cdots, \beta_p, \sigma^2)}{\partial \beta_1} &= \frac{x_{i1}}{\sigma^2} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})]^2 = 0 \\
&\vdots \\
\frac{\partial \ln(\beta_0, \beta_1, \cdots, \beta_p, \sigma^2)}{\partial \beta_p} &= \frac{x_{ip}}{\sigma^2} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})]^2 = 0 \qquad (3.3)
\end{aligned}$$

Persamaan (3.3) disederhanakan akan menjadi :

$$\begin{aligned}
\beta_0 n + \beta_1 \sum_{i=1}^{n} x_{i1} + \cdots + \beta_p \sum_{i=1}^{n} x_{ip} &= \sum_{i=1}^{n} y_i, \\
\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} {x_{i1}}^2 + \ldots + \beta_p \sum_{i=1}^{n} x_{i1} x_{ip} &= \sum_{i=1}^{n} y_i x_{i1}, \\
&\vdots \\
\beta_0 \sum_{i=1}^{n} x_{ip} + \beta_1 \sum_{i=1}^{n} {x_{ip}}^2 + \ldots + \beta_p \sum_{i=1}^{n} {x_{ip}}^2 &= \sum_{i=1}^{n} y_i x_{ip} \qquad (3.4)
\end{aligned}$$

Sistem Persamaan (3.4) disebut persamaan normal. Jika dinyatakan dalam bentuk matriks, maka persamaan normal diatas akan menjadi

$$\mathbf{X^T X}\boldsymbol{\beta} = \mathbf{X^T Y} \qquad (3.5)$$

$$\begin{bmatrix} n & \beta_1 \sum_{i=1}^{n} x_{11} & \cdots & \sum_{i=1}^{n} x_{1p} \\ \sum_{i=1}^{n} x_{i1} & \beta_1 \sum_{i=1}^{n} {x_{i1}}^2 & \cdots & \sum_{i=1}^{n} x_{i1} x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{i1} x_{ip} & \cdots & \sum_{i=1}^{n} {x_{ip}}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_{i1} \\ \vdots \\ \sum_{i=1}^{n} y_i x_{ip} \end{bmatrix} \qquad (3.6)$$

Dengan demikian diperoleh hasil estimasi parameter adalah sebagai berikut :

$$\hat{\boldsymbol{\beta}} = \mathbf{(X^T X)^{-1} X^T Y} \qquad (3.7)$$

Sedangkan untuk variansinya diperoleh turunan sebagai berikut :

$$\frac{\partial \ln(\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2 \qquad (3.8)$$

Penyelesaian persamaan (3.8) adalah

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2}{n} \qquad (3.9)$$

dengan σ^2 adalah standard error regresi, dapat ditulis sebagai berikut.

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} [y_i - \hat{y_i}]^2}{n} = \frac{1}{n}\sum_{i=1}^{n} {e_1}^2 \tag{3.10}$$

4. Mengestimasi Parameter Model Regresi Menggunakan Metode Bayes

Pada metode Bayes untuk mengestimasi parameter dalam model regresi linear berganda digunakan distribusi *prior* bersama fungsi *likelihood*. Fungsi *likelihood* untuk model regresi normal telah dicari pada Persamaan (3.1) karena sama-sama berdistribusi normal. Pada model regresi normal, distribusi *prior* yang digunakan seperti berikut:

$$\begin{aligned} \beta_j &\sim N(\mu_{\beta_j}\sigma^2) \text{ untuk } j = 0, 1, \cdots, p \\ \sigma^2 &\sim IG(a, b), \end{aligned}$$

maka distribusi *prior* dari parameter pada metode Bayes adalah

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^{n} f(\beta_j|\sigma^2) f(\sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{-1/2}} \exp\left(-\frac{1}{2\sigma^2}\left((\beta_0 - \mu_{\beta_0})^2 (\beta_1 - \mu_{\beta_1})^2 \cdots \left(\beta_p - \mu_{\beta_p}\right)^2\right)\right), \\ &\times \frac{b}{\Gamma(a)} \frac{1}{\sigma^{2(a+1)}} \exp\left(\frac{-b}{a}\right) \end{aligned} \tag{4.1}$$

Distribusi posterior yang diperoleh adalah sebagai berikut :

$$\begin{aligned} Posterior &\propto L(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) \\ &\propto \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})]^2\right) \\ &\times \frac{1}{(2\pi\sigma^2)^{-1/2}} \exp\left(-\frac{1}{2\sigma^2}\left((\beta_0 - \mu_{\beta_0})^2 (\beta_1 - \mu_{\beta_1})^2 \ldots \left(\beta_p - \mu_{\beta_p}\right)^2\right)\right) \\ &\times \frac{b}{\Gamma(a)} \frac{1}{\sigma^{2(a+1)}} \exp\left(\frac{-b}{a}\right) \end{aligned} \tag{4.2}$$

5. Markov Chain Monte Carlo (MCMC)

Metode MCMC merupakan suatu metode untuk menentukan nilai parameter dari suatu integrasi analitik yang sulit. Ide dasar dari metode MCMC adalah mengambil sampel acak dalam jumlah besar dari fungsi kepekatan peluang untuk membangkitkan parameter baru secara akurat.

Terdapat beberapa macam algoritma MCMC, dua diantaranya adalah yang paling populer yaitu algoritma *Metropolis-Hasting* dan algoritma *Gibbs sampler*. Pada skripsi ini digunakan algoritma *Gibbs sampler* untuk mencari nilai posterior dalam metode Bayes.

Algoritma *Gibbs sampler* adalah sebagai berikut :

(1) Menentukan nilai awal $\boldsymbol{\beta}^{(0)}, \sigma^{2\,(0)}$
(2) Untuk iterasi $t = 1, \cdots, T$

- Atur $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t-1)}$ dan $\sigma^2 = \sigma^{2\,(t-1)}$
- Untuk $j = 0, \cdots, p$
  Bangkitkan $\sigma^{2\,(t)}$ dari $f(\sigma^2|\mathbf{x})$
  Bangkitkan $\beta_0^{(t)}$ dari $f(\beta_0 \mid \beta_1^{(t-1)}, \beta_2^{(t-1)}, \cdots, \beta_p^{(t-1)}, \sigma^{2\,(t)}, \mathbf{x})$
  Bangkitkan $\beta_1^{(t)}$ dari $f(\beta_1 \mid \beta_0^{(t)}, \beta_2^{(t-1)}, \cdots, \beta_p^{(t-1)}, \sigma^{2\,(t)}, \mathbf{x})$
  Bangkitkan $\beta_2^{(t)}$ dari $f(\beta_2 \mid \beta_0^{(t)}, \beta_1^{(t)}, \cdots, \beta_p^{(t-1)}, \sigma^{2\,(t)}, \mathbf{x})$
  $\vdots$
  Bangkitkan $\beta_p^{(t)}$ dari $f(\beta_k \mid \beta_1^{(t)}, \beta_2^{(t)}, \cdots, \beta_{p-1}^{(t)}, \sigma^{2\,(t)}, \mathbf{x})$

Iterasi diatas akan berhenti sampai T, jika parameter model yang dihasilkan belum konvergen lakukan kembali langkah ke-2 sampai parameter konvergen.

6. Pemodelan Data dengan Metode *Maximum Likelihood*

Hasil masing-masing parameter dengan metode *maximum likelihood* dapat dilihat pada Tabel 1.

Tabel 1. Estimasi Parameter Metode *Maximum Likelihood*$(t_{0,05;204} = 1,96)$

Variabel Independent	Nilai Estimasi	Standar Deviasi	t_{hit}	Keterangan
Konstanta	-27,8210000	4,4220000	-	-
X_1	0,0307430	0,0093710	3,28	Berpengaruh
X_2	0,0039211	0,0009884	3,97	Berpengaruh
X_3	0,0034631	0,0003954	8,76	Berpengaruh
X_4	0,6527000	0,0305000	21,40	Berpengaruh

Setelah diperoleh nilai estimasi parameternya dan setiap variabel yang diuji telah signifikan maka dapat dibentuk model regresi linear pada metode *maximum likelihood* adalah

$$\hat{y} = -27,8210000 + 0,0307430X_1 + 0,0039211X_2 + 0,0034631X_3 + 0,6537000X_4$$

dengan kecocokan modelnya adalah sebesar 95,7 %

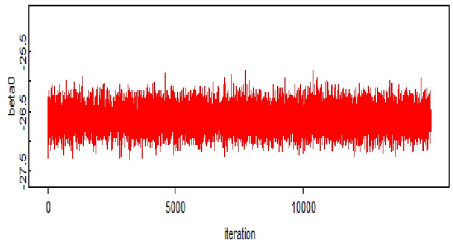
7. Pemodelan Data dengan Metode Bayes

Hasil estimasi parameter dengan metode Bayes dapat dilihat pada Tabel 2.

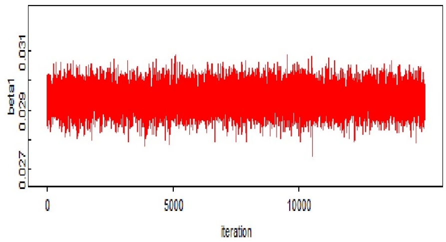
Selanjutnya akan dilakukan uji kekonvergenan terhadap parameter model yang telah dihasilkan. Uji kekonvergenan pada metode Bayes diuji dengan melihat *Trace Plot*. Pada kasus performa *hardware* komputer yang diambil sebanyak 15000 sampel diperoleh *Trace Plot* seperti pada Gambar 1 - Gambar 6.

Tabel 2. Estimasi Parameter Metode Bayes

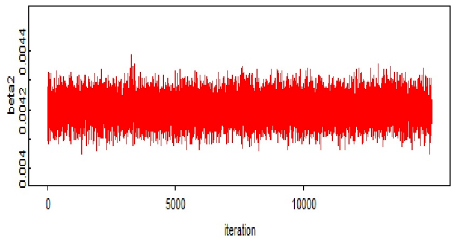
Parameter	**Hasil Estimasi**	**Standar Deviasi**	**2,5%**	**Median**	**7,5%**
β_0	-26,620000	0,195900	-27,000000	-26,620000	-26,230000
β_1	0,029380	4,200E-4	0,028580	0,029380	0,030210
β_2	0,004204	4,382E-5	0,004117	0,004204	0,004290
β_3	0,003321	1,741E-5	0,003286	0,003321	0,003355
β_4	0,656200	0,001366	0,653600	0,656200	0,658900
σ^2	2,082000	0,006659	2,069000	2,082000	2,095000



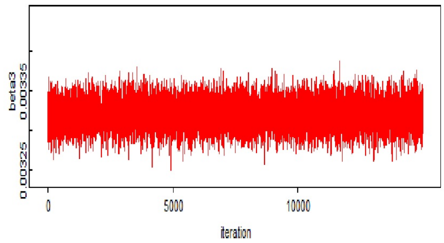
Gambar 1. *Trace Plot* β_0



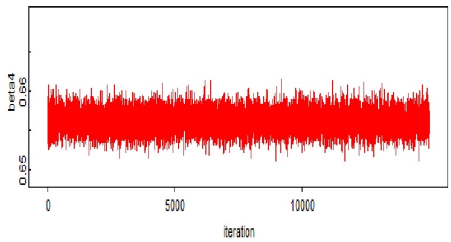
Gambar 2. *Trace Plot* β_1



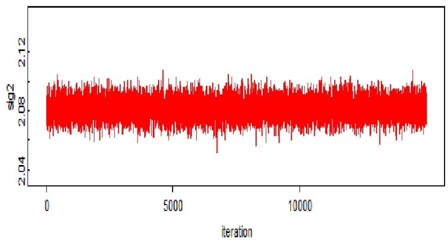
Gambar 3. *Trace Plot* β_2



Gambar 4. *Trace Plot* β_3



Gambar 5. *Trace Plot* β_4

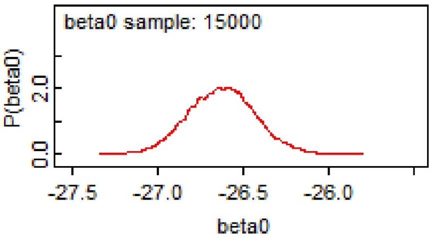


Gambar 6. *Trace Plot* σ^2

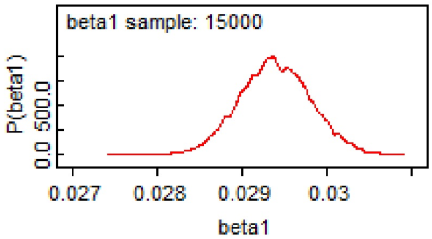
Pada Gambar $1-6$ estimasi parameter model telah tersebar disekitar dua garis horizontal yang linear maka parameter dikatakan konvergen.

Pada Gambar $7-12$ berikut ini disajikan fungsi kepekatan peluang untuk setiap parameter yang diestimasi nilainya.

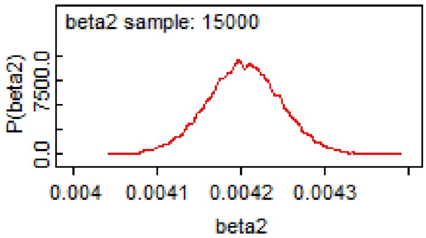
Pada Gambar $7-12$ diperlihatkan bahwa kurva fungsi kepekatan dari semua parameter model telah membentuk kurva normal. Dengan demikian dapat disim-



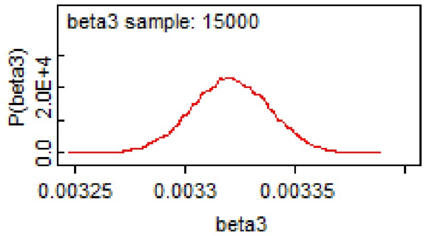
Gambar 7. Fungsi Kepekatan $\hat{\beta}_0$



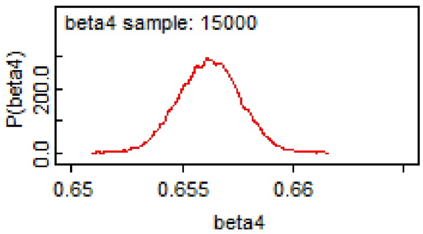
Gambar 8. Fungsi Kepekatan $\hat{\beta}_1$



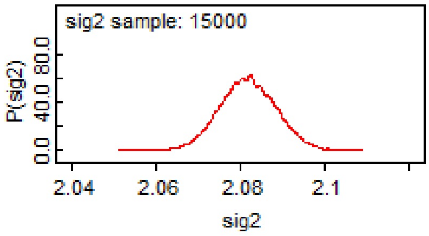
Gambar 9. Fungsi Kepekatan $\hat{\beta}_2$



Gambar 10. Fungsi Kepekatan $\hat{\beta}_3$



Gambar 11. Fungsi Kepekatan $\hat{\beta}_4$



Gambar 12. Fungsi Kepekatan $\hat{\sigma^2}$

pulkan bahwa parameter model berdistribusi normal. Dengan demikian model regresi yang terbentuk dengan metode Bayes adalah

$$\hat{y} = -26,620000 + 0,029380X_1 + 0,004204X_2 + 0,003321X_3 + 0,656200X_4$$

dengan kecocokan modelnya adalah sebesar 99,99 %.

8. Perbandingan Estimasi Parameter Model Regresi Pada Metode ***Maximum Likelihood*** dengan Metode Bayes

Mengukur kecocokan model pada kedua metode menghasilkan hasil yg berbeda. Koefisien determinasi pada metode Bayes lebih besar dari metode *maximum likelihood* maka dapat dikatakan model pada metode Bayes lebih baik daripada metode *maximum likelihood.*

9. Kesimpulan

Model regresi linear berganda yang diperoleh dengan metode *maximum likelihood* adalah

$$\hat{y} = -27,8210000 + 0,0307430X_1 + 0,0039211X_2 + 0,0034631X_3 + 0,6537000X_4,$$

dengan kecocokan modelnya adalah sebesar 95,7 %. Model regresi linear berganda yang diperoleh dengan metode Bayes adalah

$$\hat{y} = -26,620000 + 0,029380X_1 + 0,004204X_2 + 0,003321X_3 + 0,656200X_4,$$

dengan kecocokan modelnya adalah sebesar 99,99 %. Koefisien determinasi pada metode Bayes lebih besar dari metode *maximum likelihood*, sehingga dapat disimpulkan bahwa model pada metode Bayes lebih baik daripada metode *maximum likelihood*.

10. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Bapak Dr. Dodi Devianto, Ibu Dr. Maiyastri, dan Ibu Dr. Lyra Yulianti yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

[1] Azhar, J. A. 2012. *Perbandingan Metode Bayes dan Metode Likelihood dalam Mengestimasi Parameter Model Regresi Linear.* Skripsi S-1. UIN Sunan Kalijaga, tidak diterbitkan.
[2] Box, G.E.P dan Tiao,G.C. 1973. *Bayesian Inference In Statistical Analysis.* Addison Wesley Company. Inc.: Philippines.
[3] Bain, L.J and M.Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics.* Second Edition. Duxbury Press: California.
[4] Carl Hamacher, Z. Vranesic, and S. Zaky. 2004. *Organisasi Komputer.* Ed.5. Andi : Yogyakarta.
[5] Ein-Dor, P and Feldmesser, J. 1987. *Computer Hardware Data Set.* https://archive.ics.uci.edu/ml/datasets/Computer+Hardware. Diakses pada 22 September 2014 13:38.
[6] Jogiyanto. 2010. *Pengenalan Komputer.* Andi : Yogyakarta.
[7] Mutiarani, Vania. 2012. Penerapan Model Regresi Linier Bayesian Untuk Mengestimasi Parameter dan Interval Kredibel. *Prosiding Nasional Matematika dan Pendidikan Matematika MS-53.* Kontribusi Pendidikan Matematika dan Matematika dalam Membangun Karakter Guru dan Siswa. FMIPA UNY. Yogyakarta.
[8] Ntzoufras, I. 1973. *Bayessian Modelling Using WinBugs.* Department of Statistics Athens University of Economics and Bussines: Greece.
[9] Simatupang, E. R. 2008. *Model Regresi Literatur.* Skripsi S-1. Universitas Indonesia, tidak diterbitkan.
[10] Walpole, R.E. 1993. *Pengantar Statistika Edisi ke-3.* PT Gramedia PustakaUtama: Jakarta
[11] Walpole, R.E and Myers, R. H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan* Edisi ke-4. ITB : Bandung.