S-6

PENERAPAN MODEL REGRESI LINIER BAYESIAN UNTUK MENGESTIMASI PARAMETER DAN INTERVAL KREDIBEL

Vania Mutiarani[1], Adi Setiawan[2], Hanna Arini Parhusip[3]

[1] Mahasiswa Program Studi Matematika FSM UKSW
[2, 3] Dosen Program Studi Matematika FSM UKSW

e-mail: [1]vania.mutiarani@yahoo.com, [2]adi_setia_03@yahoo.com, [3]hannaariniparhusip@yahoo.co.id

Abstrak

Dalam penelitian statistik, sering diselidiki hubungan antara dua atau lebih variabel, apakah ada hubungan sebab-akibat atau tidak. Pada paper ini, data yang diambil adalah data SUSENAS tahun 2011 dari BPS Salatiga yaitu pendapatan dan pengeluaran masyarakat Salatiga dengan sampel $n = 30$ sehingga akan diselidiki hubungan antara pendapatan sebagai variabel Y dengan pengeluaran sebagai variabel X. Hubungan sebab-akibat tersebut dapat membentuk garis regresi berbentuk linier yang tidak dapat ditentukan secara tepat, sehingga diperlukan taksiran parameter untuk model regresi linier. Untuk mengestimasi parameter tersebut digunakan model regresi linier Bayesian yaitu dengan distribusi prior konjugat $p(\boldsymbol{\beta}, \sigma^2) = p(\sigma^2)p(\boldsymbol{\beta}|\sigma^2)$ dengan $p(\sigma^2) \sim \text{Inv} - \text{Gamma}(a_0, b_0)$ dan $p(\boldsymbol{\beta}|\sigma^2) \sim N(\boldsymbol{\mu}_0, \sigma^2 \boldsymbol{\Lambda}_0^{-1})$ dan distribusi posterior $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\sigma^2|\mathbf{y}, \mathbf{X})p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X})$ dengan $p(\sigma^2|\mathbf{y}, \mathbf{X}) \sim \text{Inv} - \text{Gamma}(a_n, b_n)$ dan $p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X}) \sim N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1})$. Kemudian dirancang rantai Markov dari distribusi posterior dengan Gibbs sampling sebanyak 5000 iterasi dan diperoleh taksiran parameter yang merupakan rata-rata dari nilai Gibbs sampler yaitu $\sigma^2 = 0.005718$, $\beta_0 = 2.101$ dan $\beta_1 = 0.708$. Dari nilai-nilai Gibbs sampler yang telah didapatkan, dihasilkan fungsi densitas untuk masing-masing parameter sehingga interval kepercayaan Bayesian (interval kredibel) 95% untuk taksiran parameter σ^2 adalah (0.003384, 0.009659), untuk β_0 yaitu (1.607, 2.601) dan (0.6282, 0.7879) untuk parameter β_1.

Kata kunci: model regresi linier Bayesian, estimasi parameter, interval kredibel

I. PENDAHULUAN

I.1. Latar Belakang

Dalam penelitian statistik, sering diselidiki hubungan antara dua atau lebih variabel, apakah ada hubungan sebab-akibat atau tidak. Analisis regresi linier adalah salah satu bagian dari statistik yang merupakan suatu analisis terhadap persamaan regresi dimana hubungan variabel independen dan dependen berbentuk garis lurus (Sembiring, 2003). Garis lurus tersebut disebut juga garis regresi yang tidak dapat ditentukan secara tepat, sehingga diperlukan taksiran parameter untuk model regresi linier. Metode yang biasa digunakan untuk mengestimasi parameter adalah metode kuadrat terkecil. Namun ada cara lain, yaitu menggunakan model regresi linier Bayesian.

Dalam statistik, regresi linier Bayesian merupakan pendekatan untuk regresi linier dimana analisis statistik yang dilakukan dalam konteks inferensi Bayesian (Web 1).

 ISBN : 978-979-16353-8-7

Untuk mendapatkan taksiran parameter, dirancang rantai Markov dari distribusi posterior dengan bantuan Gibbs sampling.

Interval kepercayaan adalah suatu kisaran nilai yang dianggap mengandung nilai parameter populasi yang sebenarnya (Fauzy, 2008). Interval kepercayaan dapat digunakan sebagai taksiran suatu parameter dan dapat pula dipandang sebagai pengujian hipotesis, yaitu apakah suatu parameter sama dengan suatu nilai tertentu (Sembiring, 2003). Dalam statistik Bayesian, interval kepercayaan Bayesian (interval kredibel) merupakan interval di dalam domain dari distribusi probabilitas posterior yang digunakan untuk penaksiran interval (Web 2).

Dalam penelitian ini akan dijelaskan bagaimana penerapan model regresi linier Bayesian untuk mengestimasi parameter dan interval kepercayaan Bayesian (interval kredibel) dengan mengambil data SUSENAS (Survey Sosial Ekonomi Nasional) tahun 2011 dari BPS Salatiga yaitu pendapatan dan pengeluaran masyarakat Salatiga dengan sampel $n = 30$.

II. DASAR TEORI

II.1. Regresi Linier Bayesian

Dalam statistik, regresi linier Bayesian merupakan pendekatan untuk regresi linier dimana analisis statistik yang dilakukan dalam konteks inferensi Bayesian (Web 1). Saat model regresi memiliki error yang berdistribusi normal, dan jika bentuk khusus dari distribusi prior diasumsikan, hasil eksplisit tersedia untuk distribusi probabilitas posterior dari parameter model.

II.2. Model

Dalam konteks model regresi linier, **y** disebut sebagai variabel dependen, dan **X** variabel independen. Dalam notasi matriks, sistem ini dapat ditulis sebagai (Lancaster, 2003)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \qquad (1)$$

dengan **y** adalah vektor kolom, untuk pengamatan $i = 1, \dots, n$ dipunyai nilai y; **X** adalah matriks $n \times k$, untuk $j = 0, \dots, k$; $\boldsymbol{\beta}$ adalah vektor kolom dari turunan parsial k yaitu $\{\beta_j\}$, dan vektor kolom ε berisi sebanyak n error sehingga

$$\mathbf{y} = [y_i] = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix},$$

$$\boldsymbol{\beta} = [\beta_j] = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = [\varepsilon_i] = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Berdasar pada persamaan (1), model regresi linier dengan intersep dapat ditulis dalam bentuk matriks sebagai

$$\mathbf{y} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i \qquad (2)$$

dengan vektor $\mathbf{x}_i^{\mathrm{T}}$ merupakan tiap baris dari matriks $\mathbf{X}$ yang berdimensi $n \times k$. Pada paper ini, $k = 1$ karena $\mathbf{X}$ hanya 1 variabel. Oleh karena itu, $\mathbf{x}_i^{\mathrm{T}} = [1 \quad x_i] = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ untuk $i = 1, \dots, n$ dan vektor $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ dan ε_i saling bebas dan berdistribusi identik, yaitu normal dengan mean nol dan variansi σ^2, dinotasikan dengan $\varepsilon_i \sim N(0, \sigma^2)$. Persamaan (2) dapat ditulis ulang menjadi :

$$y_i = [1 \quad x_i] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \tag{3}$$

Persamaan (3) merupakan model regresi linier dengan intersep. Untuk memudahkan dalam penulisan, $\mathbf{x}_i^{\mathrm{T}}$ selanjutnya ditulis sebagai **X.** Galat (*error*) ε_i berdistribusi normal, sehingga $(y_i|\mathbf{X}, \boldsymbol{\beta}, \sigma^2)$ juga berdistribusi normal (Puspaningrum, 2008) dengan $\mathrm{E}(y_i) = \beta_0 + \beta_1 x_i$ dan $V(y_i) = \sigma^2$.

Jadi $(y_i|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ dengan fungsi kepadatan probabilitas :

$$p(y_i|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^2\right]$$

Kemudian fungsi likelihood didefinisikan sebagai :

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} p(y_i|\mathbf{X}, \boldsymbol{\beta}, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

$$= (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)$$

Jadi, fungsi likelihood :

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right). \tag{4}$$

II.3. Distribusi Prior Konjugat

Prior konjugat adalah suatu prior yang jika dikombinasikan dengan fungsi likelihood akan menghasilkan suatu posterior dengan distribusi yang sama dengan distribusi prior (Gelman, 2006). Karena log-likelihood kuadratik dalam $\boldsymbol{\beta}$, log-likelihood ditulis ulang sedemikian rupa sehingga likelihood menjadi normal dalam $(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$ dengan $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$. Ditulis

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) + \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}).$$

Likelihood ditulis ulang menjadi

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-v/2} \exp\left(-\frac{vs^2}{2\sigma^2}\right)$$

$$\times (\sigma^2)^{-(n-v)/2} \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T(\boldsymbol{X}^T\boldsymbol{X})\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)\right) \tag{5}$$

dengan $vs^2 = \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$, dan $v = n - k$, k merupakan jumlah koefisien regresi. Dilihat dari persamaan (5), hal ini menunjukkan suatu bentuk untuk prior :

$$p(\boldsymbol{\beta}, \sigma^2) = p(\sigma^2)p(\boldsymbol{\beta}|\sigma^2) \qquad (6)$$

dengan σ^2 berdistribusi $\text{Inv} - \text{Gamma}(a_0, b_0)$ dengan $a_0 = v_0/2$ dan $b_0 = v_0 s_0^2$ dengan v_0 dan s_0^2 diperoleh dari informasi awal atau ditentukan secara subyektif. Kepadatan prior ditulis sebagai

$$p(\sigma^2) \propto (\sigma^2)^{-(v_0/2+1)} \exp\left(-\frac{v_0 s_0^2}{2\sigma^2}\right).$$

Lebih lanjut, prior bersyarat $\boldsymbol{\beta}|\sigma^2$ berdistribusi normal yang dinotasikan sebagai $N(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Lambda}_0^{-1})$ dengan $\boldsymbol{\mu}_0$ ditentukan secara subyektif, dan memiliki kepadatan prior bersyarat :

$$p(\boldsymbol{\beta}|\sigma^2) \propto (\sigma^2)^{-k/2}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Lambda}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right).$$

II.4. Distribusi Posterior

Untuk menyatakan distribusi posterior, digunakan teorema Bayes yaitu dapat dinyatakan sebagai berikut (Lancaster, 2003)

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

sehingga dengan likelihood (persamaan (4)) dan prior yang telah ditentukan pada persamaan (6), distribusi posterior dapat dinyatakan sebagai

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2)p(\sigma^2)p(\boldsymbol{\beta}|\sigma^2) \qquad (7)$$

$$\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)$$

$$\times (\sigma^2)^{-(a_0+1)} \exp\left(-\frac{b_0}{2\sigma^2}\right)$$

$$\times (\sigma^2)^{-k/2}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Lambda}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right).$$

Dengan beberapa pengaturan ulang, posterior pada persamaan (7) dapat ditulis ulang sehingga mean posterior $\boldsymbol{\mu}_n$ dari vektor parameter $\boldsymbol{\beta}$ dapat dinyatakan dalam estimator kuadrat terkecil $\widehat{\boldsymbol{\beta}}$ dan mean prior $\boldsymbol{\mu}_0$ dengan kekuatan dari prior ditunjukkan oleh matriks prior presisi $\boldsymbol{\Lambda}_0 = 1/\sigma^2$

$$\boldsymbol{\mu}_n = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{\Lambda}_0\boldsymbol{\mu}_0). \qquad (8)$$

Untuk membenarkan bahwa $\boldsymbol{\mu}_n$ memang mean posterior, istilah kuadrat dalam eksponensial dapat diatur kembali sebagai bentuk kuadrat dalam $\boldsymbol{\beta} - \boldsymbol{\mu}_n$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Lambda}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0)$$

$$= (\boldsymbol{\beta} - \boldsymbol{\mu}_n)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)(\boldsymbol{\beta} - \boldsymbol{\mu}_n) + \mathbf{y}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\mu}_n^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)\boldsymbol{\mu}_n + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0.$$

Selanjutnya, posterior dapat dinyatakan sebagai distribusi normal dikalikan dengan distribusi invers-gamma :

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-k/2}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\mu}_n)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)(\boldsymbol{\beta} - \boldsymbol{\mu}_n)\right)$$

$$\times (\sigma^2)^{-(n+v_0)/2-1} \exp\left(-\frac{b_0 + \mathbf{y}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\mu}_n^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)\boldsymbol{\mu}_n + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0}{2\sigma^2}\right)$$

Oleh karena itu, distribusi posterior dapat diparameterisasi sebagai berikut :

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X})p(\sigma^2|\mathbf{y}, \mathbf{X}) \qquad (9)$$

dengan kedua faktor sesuai dengan kepadatan dari distribusi $N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1})$ dan $\text{Inv} - \text{Gamma}(a_n, b_n)$ dengan parameternya diberikan oleh

$\mathbf{\Lambda}_n = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{\Lambda}_0),\ a_n = \frac{1}{2}(n + v_0),\ b_n = b_0 + \frac{1}{2}(\mathbf{y}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\mu}_0^{\mathrm{T}}\mathbf{\Lambda}_0\boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{\Lambda}_n\boldsymbol{\mu}_n).$

Berdasarkan parameter di atas, persamaan (8) diperbarui sesuai konteks Bayesian menjadi

$$\boldsymbol{\mu}_n = (\mathbf{\Lambda}_n)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{\Lambda}_0\boldsymbol{\mu}_0\right) = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{\Lambda}_0)^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{y} + \mathbf{\Lambda}_0\boldsymbol{\mu}_0). \qquad (10)$$

II.5. MCMC (Markov Chain Monte Carlo)

Salah satu cara untuk merancang rantai Markov yaitu dari distribusi posterior dengan $p(\sigma^2|\mathbf{y},\mathbf{X}) \sim \mathrm{Inv-Gamma}(a_n, b_n)$ dan $p(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}) \sim N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{\Lambda}_0)^{-1})$ yaitu dengan Gibbs Sampling yang menghasilkan rantai Markov oleh sampling dari distribusi bersyarat.

Sebelumnya, disusun distribusi prior konjugat dengan $p(\sigma^2) \sim \mathrm{Inv-Gamma}(a_0, b_0)$ dengan $a_0 = v_0/2$ dan $b_0 = v_0 s_0^2$ dengan v_0 dan s_0^2 ditentukan secara subyektif dan $p(\boldsymbol{\beta}|\sigma^2) \sim N(\boldsymbol{\mu}_0, \sigma^2\mathbf{\Lambda}_0^{-1})$ dengan $\boldsymbol{\mu}_0$ ditentukan secara subyektif dan prior presisi $\mathbf{\Lambda}_0 = 1/\sigma^2$ dengan memilih nilai σ^2.

Jika $\sigma^2 \sim \mathrm{Inv-Gamma}(a_n, b_n)$, maka :

$$\sigma^2|\mathbf{y},\mathbf{X} \sim \mathrm{Inv-Gamma}\left[\frac{1}{2}(n + v_0), b_0 + \frac{1}{2}(\mathbf{y}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\mu}_0^{\mathrm{T}}\mathbf{\Lambda}_0\boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^{\mathrm{T}}\mathbf{\Lambda}_n\boldsymbol{\mu}_n)\right] \quad (*)$$

Jika $(\beta_0, \beta_1) \sim BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, (Bain et al., 1991) maka :

1. Distribusi dari β_0 bersyarat pada $\beta_1 = \beta_1^{(0)}$

   $$\beta_0|\beta_1^{(0)} \sim N\left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(\beta_1^{(0)} - \mu_2\right), \sigma_1^2(1 - \rho^2)\right]. \qquad (**)$$

2. Distribusi dari β_1 bersyarat pada $\beta_0 = \beta_0^{(0)}$

   $$\beta_1|\beta_0^{(0)} \sim N\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(\beta_0^{(0)} - \mu_1\right), \sigma_2^2(1 - \rho^2)\right], \qquad (***)$$

Diberikan σ^2 dan vektor $\boldsymbol{\beta}$ yang tidak diketahui : $\boldsymbol{\beta} = (\beta_0, \beta_1)$

1. Dipilih nilai awal ${\sigma^2}^{(0)},\ \beta_0^{(0)},\ \beta_1^{(0)}$
2. Sampel ${\sigma^2}^{(1)}$ dari $p\left({\sigma^2}^{(1)}\middle|\mathbf{y},\mathbf{X}\right)$ sehingga ${\sigma^2}^{(1)}|\mathbf{y},\mathbf{X}$ memenuhi (*)

   Sampel $\beta_0^{(1)}$ dari $p\left(\beta_0^{(1)}\middle|{\sigma^2}^{(1)}, \beta_1^{(0)}, \mathbf{y}, \mathbf{X}\right)$ sehingga $\beta_0^{(1)}|{\sigma^2}^{(1)}, \beta_1^{(0)}$ memenuhi (**)

   Sampel $\beta_1^{(1)}$ dari $p\left(\beta_1^{(1)}\middle|{\sigma^2}^{(1)}, \beta_0^{(1)}, \mathbf{y}, \mathbf{X}\right)$ sehingga $\beta_1^{(1)}|{\sigma^2}^{(1)}, \beta_0^{(1)}$ memenuhi (***)
3. Langkah 2 diulangi sebanyak 1000 kali (untuk mendapat hasil yang lebih akurat, dilakukan iterasi sebanyak lebih dari 1000 kali)
4. Akhirnya didapatkan sampel dari $p(\sigma^2|\mathbf{y},\mathbf{X})$ dan $p(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X})$ dalam bentuk rantai Markov.

II.6. Interval Kredibel (Interval Kepercayaan Bayesian)

Dalam statistik Bayesian, interval kredibel $(1 - \alpha)100\%$ merupakan interval di dalam domain dari distribusi probabilitas posterior yang digunakan untuk penaksiran interval (Web 2).

Salah satu metode untuk mengestimasi interval kredibel yang paling mudah digunakan adalah interval kredibel dua ekor (Johnson, 2009). Interval kredibel dua ekor disusun dengan menemukan kuantil $\alpha/2$ dan $1 - \alpha/2$ dengan tingkat signifikansi α.

ISBN : 978-979-16353-8-7

III. METODE PENELITIAN

Data yang diambil dalam penelitian adalah data SUSENAS (Survey Sosial Ekonomi Nasional) tahun 2011 dari BPS Salatiga yaitu pendapatan dan pengeluaran masyarakat Salatiga dengan sampel $n = 30$. Dalam melakukan perhitungan, digunakan alat bantu program WinBUGS 1.4.3.

Langkah-langkah penyelesaian untuk mengestimasi parameter dan interval kredibel menggunakan model regresi linier Bayesian sebagai berikut :

1. Merancang rantai Markov dari distribusi posterior $p(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}) \propto p(\sigma^2|\mathbf{y},\mathbf{X})p(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X})$ dengan $p(\sigma^2|\mathbf{y},\mathbf{X}) \sim \text{Inv} - \text{Gamma}(a_n, b_n)$ dan $p(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}) \sim N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1})$ yaitu dengan Gibbs Sampling yang menghasilkan 3 rantai Markov dengan iterasi sebanyak 5000 yaitu untuk taksiran parameter σ^2, β_0 dan β_1.
2. Taksiran parameter σ^2, β_0 dan β_1 diperoleh dengan mencari rata-rata dari nilai Gibbs sampler.
3. Dari nilai-nilai Gibbs sampler tersebut, dihasilkan fungsi densitas untuk σ^2 berdistribusi invers-gamma dan β_0, β_1 berdistribusi normal.
4. Mencari interval kredibel 95% untuk masing-masing taksiran parameter dengan tingkat signifikansi $\alpha = 0.05$ berdasar pada fungsi densitas.

IV. HASIL PENELITIAN DAN PEMBAHASAN

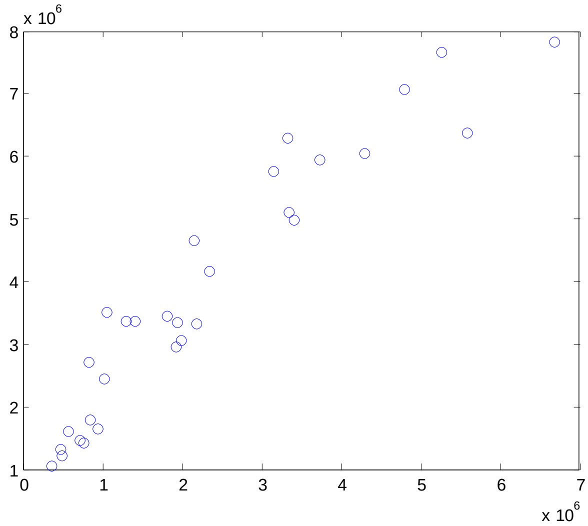
Data yang digunakan dalam penelitian merupakan nilai logaritma dari data asli dengan pendapatan sebagai variabel dependen $\mathbf{y}$ terhadap pengeluaran sebagai variabel independen $\mathbf{X}$ dengan pengamatan $n = 30$ dinyatakan pada Tabel 1.

Tabel 1. Data Pendapatan dan Pengeluaran

Y	X	log(Y)	log(X)
2705000	824382,6667	6,4322	5,9161
3364333,333	1418616	6,5269	6,1519
6040000	4297649,667	6,7810	6,6332
3353000	1936732,667	6,5254	6,2871
1426333,333	773483	6,1542	5,8885
7664400	5270866	6,8845	6,7219
6285666,667	3323499,667	6,7984	6,5216
5950000	3734666,667	6,7745	6,5723
5750000	3154333,333	6,7597	6,4989
1323333,333	472666,3333	6,1217	5,6746
1650333,333	937132,6667	6,2176	5,9718
1051400	359666,3333	6,0218	5,5559
7826616,667	6677783,333	6,8936	6,8246
1466666,667	719666,3333	6,1663	5,8571
3500000	1048666	6,5441	6,0206
3450000	1820000	6,5378	6,2601
6366666,667	5580666,333	6,8039	6,7467
4661666,667	2147499,333	6,6685	6,3319
4976666,667	3411166	6,6969	6,5329

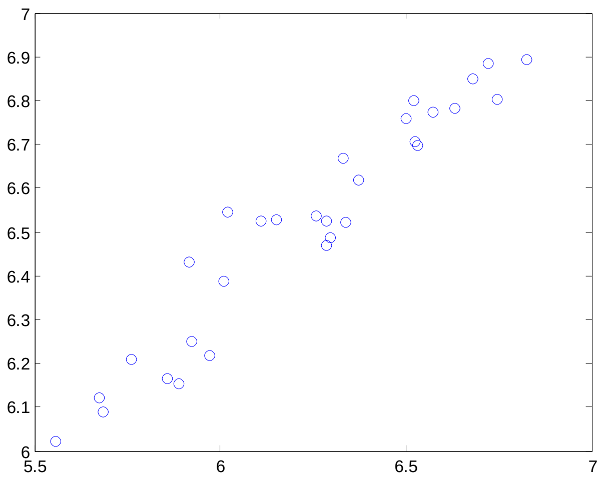
7068000	4795400	6,8493	6,6808
5098666,667	3340066	6,7075	6,5238
1225000	485166	6,0881	5,6859
3316666,667	2182400	6,5207	6,3389
2439166,667	1024266	6,3872	6,0104
1780000	841666,6667	6,2504	5,9251
4156333,333	2352566	6,6187	6,3715
3057666,667	1987100	6,4854	6,2982
2950833	1934774,333	6,4699	6,2866
1613666,667	578949,6667	6,2078	5,7626
3356666,667	1292166	6,5259	6,1113

Diagram pencar data asli pada Tabel 1 ditampilkan oleh Gambar 1.



Gambar 1. Diagram Pencar Data Asli

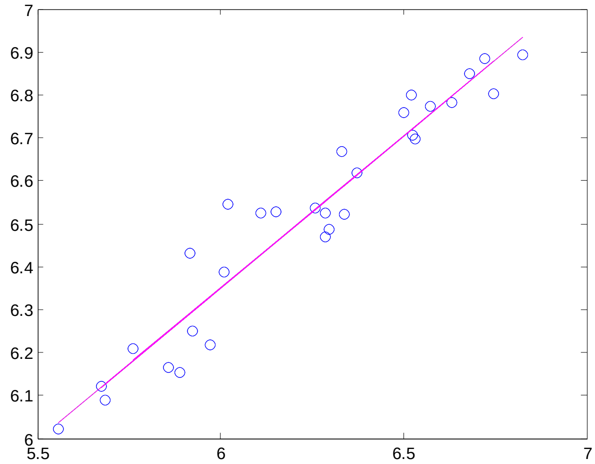
Sedangkan diagram pencar untuk data logaritma dari data asli ditunjukkan oleh Gambar 2.



Gambar 2. Diagram Pencar Data Logaritma Dari Data Asli

Dari Gambar 2 terlihat sebaran data cenderung berada di sekitar garis lurus maka dapat dikatakan hubungan antara kedua peubah tersebut membentuk pola linier. Karena data-data tersebut menunjukkan hubungan linier maka selanjutnya dapat ditentukan persamaan regresi dugaannya.

Untuk mendapatkan taksiran $\widehat{\boldsymbol{\beta}} = (\beta_0, \beta_1)$ berdasarkan metode kuadrat terkecil, nilai-nilai log(X) dan log(Y) dimasukkan ke dalam persamaan $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ dengan $\mathbf{X} = [\mathbf{1}; \log(\mathrm{X})]$ dan $\mathbf{y} = \log(\mathrm{Y})$ sehingga diperoleh nilai $\beta_0 = 2.105302$ dan $\beta_1 = 0.707418$. Jadi persamaan garis regresi dugaan adalah $\hat{y}_i = 2.105302 + 0.707418\, x_i$. Diagram pencar data dan persamaan garis regresi $\hat{y}_i$ ditampilkan pada Gambar 3.



Gambar 3. Garis Regresi dengan Metode Kuadrat Terkecil

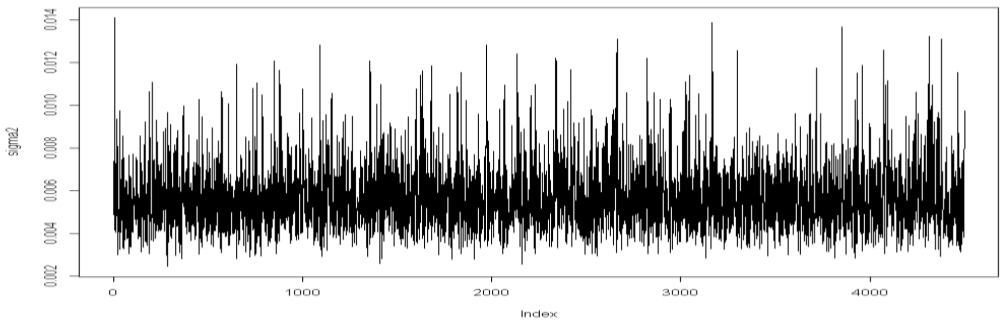
Selanjutnya untuk mendapatkan estimasi parameter σ^2 dan $\widetilde{\boldsymbol{\beta}} = (\beta_0, \beta_1)$ dengan model regresi linier Bayesian, dirancang rantai Markov dari distribusi posterior $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\sigma^2|\mathbf{y}, \mathbf{X})p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X})$ dengan $p(\sigma^2|\mathbf{y}, \mathbf{X}) \sim \mathrm{Inv-Gamma}(a_n, b_n)$ dan $p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X}) \sim N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1})$ yaitu dengan Gibbs sampling sebanyak 5000 iterasi.

1. Taksiran Parameter σ^2 dan Interval Kredibel σ^2

   Dilakukan Gibbs sampling sebanyak 5000 iterasi untuk mendapatkan taksiran parameter σ^2 yang berdistribusi $\mathrm{Inv-Gamma}(a_n, b_n)$ dengan memilih nilai awal ${\sigma^2}^{(0)} = 1$. Kemudian dengan memotong 500 iterasi pertama, diperoleh hasil:

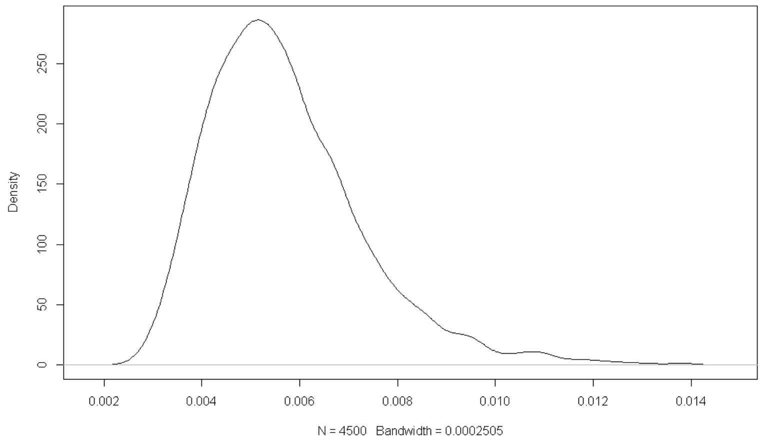
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
sigma2	0.005718	0.001617	3.245E-5	0.003384	0.005455	0.009659	501	5000

Rantai Markov untuk taksiran parameter σ^2 sebagai berikut:



Gambar 9. Rantai Markov untuk Taksiran Parameter σ^2

Gambar 9 menunjukkan nilai-nilai Gibbs sampler sebanyak 4500 nilai yang membentuk rantai Markov. Dengan mencari rata-rata dari 4500 nilai Gibbs sampler tersebut, maka diperoleh hasil taksiran parameter $\sigma^2 = 0.005718$. berdasarkan nilai-nilai Gibbs sampler tersebut, dihasilkan fungsi densitas pada Gambar 10 sehingga interval kredibel 95% untuk taksiran σ^2 adalah (0.003384, 0.009659).



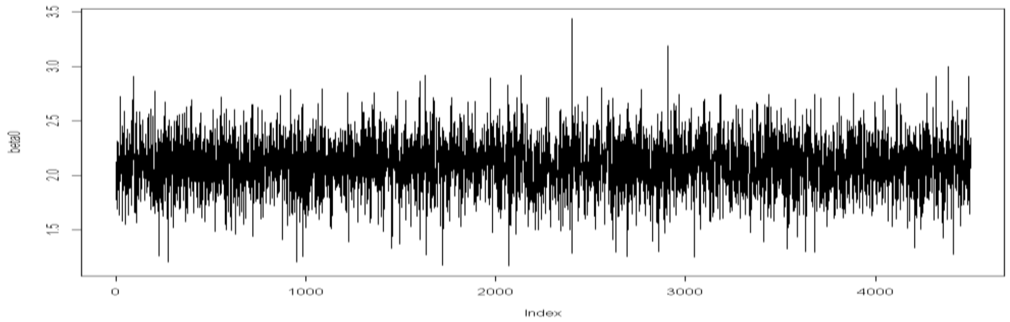
Gambar 10. Fungsi Densitas Parameter σ^2

2. Taksiran Parameter β_0 dan Interval Kredibel β_0

Dipilih nilai awal $\beta_0^{(0)} = 0$. Dengan Gibbs sampling sebanyak 5000 iterasi dan memotong 500 iterasi pertama untuk taksiran β_0 yang berdistribusi normal, didapatkan hasil:

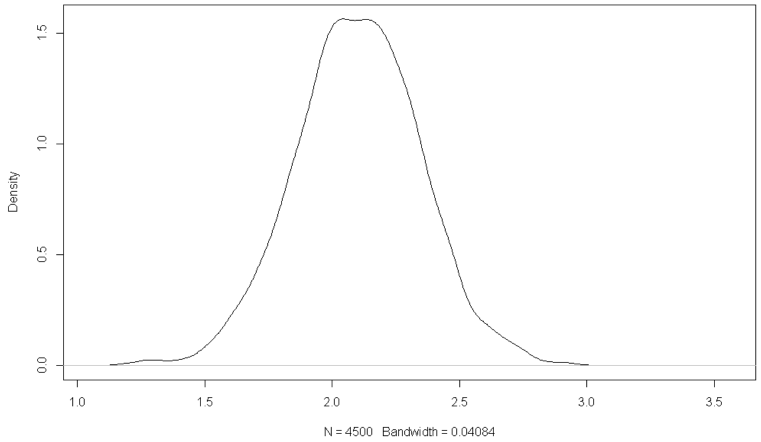
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	2.101	0.2505	0.00362	1.607	2.102	2.601	501	5000

Rantai Markov untuk taksiran β_0 ditunjukkan pada gambar di bawah ini :



Gambar 4.Rantai Markov untuk Taksiran Parameter β_0

Berdasarkan Gambar 4, taksiran parameter β_0 didapat dengan mencari rata-rata dari nilai Gibbs sampler sebanyak 4500 dan menghasilkan nilai taksiran $\beta_0 = 2.101$. Dari nilai-nilai Gibbs sampler yang telah dihitung, menghasilkan fungsi densitas pada Gambar 5, sehingga diperoleh interval kredibel 95% yaitu (1.607, 2.601).



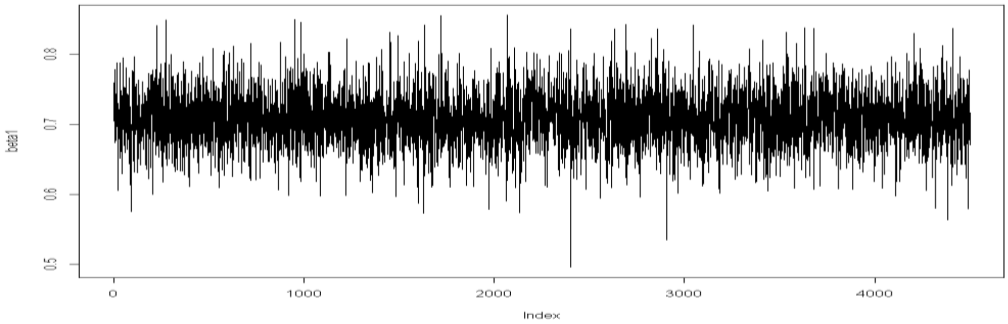
Gambar 5. Fungsi Densitas Parameter β_0

3. Taksiran Parameter β_1 dan Interval Kredibel β_1

Untuk memperoleh taksiran parameter β_1 yang berdistribusi normal, dipilih nilai awal $\beta_1^{(0)} = 0$ kemudian dengan melakukan Gibbs sampling sebanyak 5000 iterasi dan memotong 500 iterasi pertama didapatkan hasil :

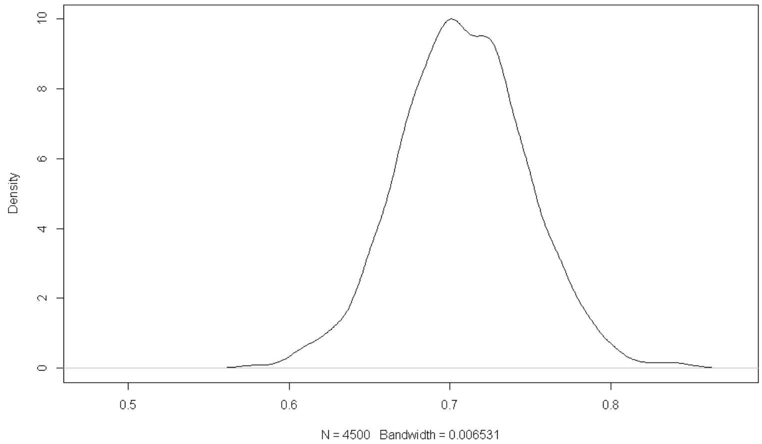
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	0.708	0.04012	5.815E-4	0.6282	0.7074	0.7879	501	5000

Rantai Markov untuk taksiran β_1 sebagai berikut:



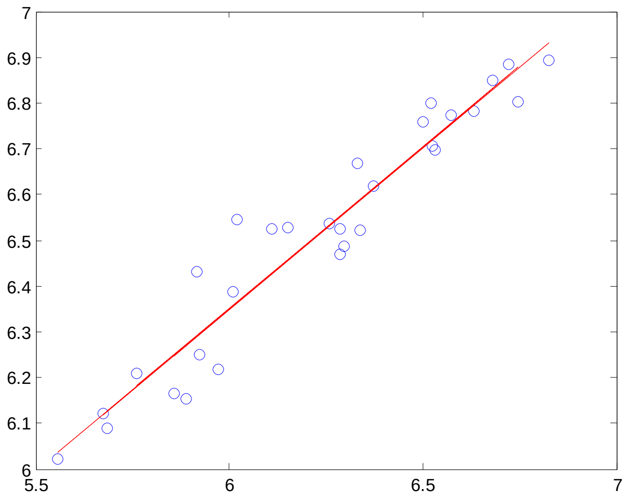
Gambar 6. Rantai Markov untuk Taksiran Parameter β_1

Dengan mencari rata-rata dari nilai Gibbs sampler sebanyak 4500 iterasi yang ditunjukkan pada Gambar 6, didapat nilai taksiran $\beta_1 = 0.708$ dan menghasilkan fungsi densitas pada Gambar 7. Jadi, interval kredibel 95% untuk parameter β_1 yaitu (0.6282, 0.7879).



Gambar 7. Fungsi Densitas Parameter β_1

Selanjutnya, dengan estimasi parameter $\widetilde{\boldsymbol{\beta}} = (2.101, 0.708)$ dibentuk persamaan garis regresi dugaan: $\widetilde{y}_i = 2.101 + 0.708\ x_i$. Diagram pencar data dan persamaan garis regresi ditampilkan pada Gambar 8.



Gambar 8. Garis Regresi dengan Metode Bayesian

V. KESIMPULAN

Dari hasil penelitian tentang data SUSENAS (Survey Sosial Ekonomi Nasional) tahun 2011 dari BPS Salatiga yaitu data logaritma dari data pendapatan dan pengeluaran masyarakat Salatiga dengan sampel $n = 30$ yang telah diperoleh pada pembahasan, dapat disimpulkan bahwa :

- Untuk mendapatkan estimasi parameter σ^2, β_0, dan β_1 dirancang rantai Markov dari distribusi posterior $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\sigma^2|\mathbf{y}, \mathbf{X})p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X})$ dengan $p(\sigma^2|\mathbf{y}, \mathbf{X}) \sim \text{Inv} - \text{Gamma}(a_n, b_n)$ dan $p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X}) \sim N(\boldsymbol{\mu}_n, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0)^{-1})$ yaitu dengan Gibbs sampling sebanyak 5000 iterasi. Diperoleh nilai-nilai Gibbs sampler sebanyak 5000 (untuk masing-masing parameter), dan setelah dirata-rata, didapat nilai taksiran $\sigma^2 = 0.005718$, $\beta_0 = 2.101$, dan $\beta_1 = 0.708$.
- Dengan tingkat signifikansi $\alpha = 0.05$, interval kredibel 95% untuk masing-masing taksiran σ^2, β_0, dan β_1 merupakan interval dimana terdapat taksiran parameternya. Dari nilai-nilai Gibbs sampler yang ada, dihasilkan fungsi densitas untuk σ^2 yang berdistribusi invers-gamma, sehingga diperoleh interval kredibel 95% yaitu (0.003384, 0.009659). Sedangkan fungsi densitas untuk β_0 dan β_1 yang berdistribusi normal, diperoleh interval kredibel 95% yaitu (1.607, 2.601) untuk β_0 dan (0.6282, 0.7879) untuk β_1.

VI. DAFTAR PUSTAKA

Bain, Lee J. dan Engelhardt, Max. 1991. *Introduction To Probability and Mathematical Statistics Second Edition*. USA : Duxbury.

Fall, Roger Levy. 2007. *Lecture 6: Bayesian Parameter Estimation; Confidence Intervals (Bayesian and Frequentistic).*

Fauzy, Akhmad. 2008. *Statistik Industri.* Jakarta : Erlangga.

Gelman, Andrew. 2006. *Bayesian Analysis*. Department of Statistics and Department of Political Science : Columbia University.

Johnson, Matthew S. 2009. *Introduction to Bayesian Statistics with WinBUGS*. New York : Columbia University.

Lancaster, Tony. 2003. *An Introduction to Modern Bayesian Econometrics*.

Puspaningrum, Dessy. 2008. *Penerapan Metode Bayesian Untuk Mengestimasi Parameter Pada Model Regresi Linier Sederhana*. UKSW : FSM.

Sembiring, R. K. 2003. *Analisis Regresi Edisi Kedua*. Bandung : ITB.

Web 1 : http://en.wikipedia.org/wiki/Bayesian_linear_regression
Bayesian Linear Regression
Diunduh pada 28 Agustus 2012

Web 2 : http://en.wikipedia.org/wiki/Credible_interval
Credible Interval
Diunduh pada 5 September 2012