Deteksi Pencilan dengan Pendekatan Bayesian pada Regresi Linear (Studi Kasus Hubungan Pengeluaran Rumah Tangga dengan PDRB di Jawa Barat Tahun 2013)

Dwiningrum Prihastiwi, Dadang Juandi, Nar Herrhyanto
Departemen Pendidikan Matematika FPMIPA Universitas Pendidikan Indonesia

Correspondent author: dwiningrumprihastiwi@gmail.com

Abstrak. Pokok pembahasan dalam skripsi ini adalah langkah-langkah mendeteksi pencilan dengan pendekatan Bayesian pada model regresi linear. Metode Bayesian memberikan hasil penaksiran yang lebih baik daripada penaksiran dengan metode klasik. Pendekatan Bayesian yang dilakukan adalah dengan mempertimbangkan distribusi awal (prior) dengan melihat fungsi likelihood data dan juga melibatkan distribusi posterior. Data pencilan merupakan salah satu masalah yang sering terjadi pada model regresi. Keberadaan data pencilan dapat mengganggu proses pengujian dan pengambilan keputusan dalam penelitian. Untuk itu, ingin dikaji lebih lanjut mengenai pencarian data pencilan dengan pendekatan Bayesian. Data pencilan ini dapat diketahui dengan membandingkan peluang prior dan peluang posterior dari data regresi yang diperoleh, jika nilai peluang posterior lebih besar dari peluang prior maka data tersebut dikatakan data pencilan.

Kata kunci: Pencilan, Bayesian.

Abstract. The main issue in this study is the steps to detect outlier by Bayesian approach in the linear regression model. Bayesian method gives better results than the assessment of assessments by classical methods. Bayesian approach is done by considering prior distribution, which is obtained by looking likelihood function, and posterior distribution. Data outliers is one of the problem that often occur in the regression model. The existence of data outliers can disturb trial processes and decision-making in research. Furthermore, would be examined on detecting outlier by Bayesian approach. The outlier can be determined by comparing prior and posterior probability which is obtained of data regression, if posterior probability value is greater than prior probability then it called by outlier.

Keyword: outlier, Bayesian.

1. PENDAHULUAN

Pada umumnya penelitian yang dilakukan untuk mengetahui seberapa besar hubungan antara variabel-variabel yang sedang diamati baik dalam bidang keilmuan matematika dan statistika maupun dalam bidang keilmuan lainnya. Steel dan Torrie dalam Rahmawati [1], menyatakan bahwa model regresi merupakan model yang cocok digunakan dalam menganalisis data penelitian yang melibatkan variabel respon (variabel terikat) dan variabel *explanatory* (variabel bebas). Suatu model regresi linear tidak akan terlepas dari permasalahan sisa. Salah satu yang dapat dilihat dari sisa adalah pencilan (*outlier*). Nazra [2] pendeteksian pencilan ini tidak hanya penting dalam rangka untuk memperbaiki model yang dicari, namun dengan diketahuinya suatu data adalah pencilan, maka seorang peneliti dapat menelusuri data tersebut untuk mengetahui dan mempelajari lebih dalam mengenai data pencilan yang diperoleh.

Azhar [3] dalam kajiannya memperoleh kesimpulan bahwa metode Bayes lebih baik daripada metode *maximum likelihood*. Hal ini ditunjukkan dengan hasil estimasi metode Bayes memiliki nilai *Mean Square Error* (MSE) serta R-*square* yang lebih kecil dari estimasi metode *maximum likelihood*.

Kesejahteraan suatu kelompok masyarakat dapat dilihat dari tingkat pendapatan masyarakatnya. Produk Domestik Regional Bruto (PDRB) merupakan salah satu indikator pertumbuhan ekonomi suatu negara/wilayah/daerah. Salah satu faktor yang dapat mempengaruhi pertumbuhan ekonomi adalah konsumsi. Dalam hal ini konsumsi dapat dilihat baik dari pengeluaran makanan maupun bukan makanan.

Dalam penelitian ini akan dikaji tentang bagaimana langkah-langkah untuk mendeteksi pencilan dengan pendekatan Bayesian pada regresi linear serta penggunaannya pada kasus PDRB. Data yang digunakan adalah pengeluaran rumah tangga makanan dan bukan makanan serta PDRB Kabupaten/Kota di Jawa Barat tahun 2013.

2. PENDEKATAN BAYESIAN PADA REGRESI LINEAR

2.1 Pendekatan Bayesian

Pada model linear harus memenuhi asumsi normalitas yakni variabel acak y dan X berdistribusi normal sehingga parameter θ juga berdistribusi normal. Model linear berbentuk y = Xθ + ε. Diberikan parameter $\theta' = (\theta_1, \ldots, \theta_n)$ vektor acak berukuran p x 1, matriks X berukuran n x p diambil dari observasi $x_1, \ldots, x_n$, vektor acak y berukuran n x 1 dan variabel acak bebas $\varepsilon' = (\varepsilon_1, \ldots, \varepsilon_n)$ yang berdistribusi normal, $N(0,\sigma^2)$.

Dalam mencari distribusi prior diperlukan fungsi *likelihood* dari parameter θ. Diketahui vektor $X\theta = E(y|\theta)$ sebagai nilai rata-rata dan matriks kovarians $D(y|\sigma^2) = \sigma^2 P^{-1}$ di mana *P* merupakan matriks definit positif.

Misalkan $(y|\theta,\ \sigma^2) \sim N(X\theta,\ \sigma^2 P^{-1})$ dan σ^2 merupakan parameter yang tidak diketahui. Didefinisikan $\tau = \sigma^{-2}$. Karena $(y|\theta,\ \sigma^2) \sim N(X\theta,\ \sigma^2 P^{-1})$ diperoleh $(\det \tau^{-1} P^{-1})^{-\frac{1}{2}} = (\det P)^{\frac{1}{2}} \tau^{\frac{n}{2}}$

Fungsi *likelihood*

$$p(y|\theta,\tau,C) = (2\pi)^{-\frac{n}{2}} (\det P)^{\frac{1}{2}} \tau^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}(y - X\ )' P (y - X\ )\right)$$

$$\propto \tau^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}(y - X\ )' P (y - X\ )\right) \qquad (3.1)$$

Distribusi prior untuk θ dan τ adalah distribusi normal-gamma

$\theta, \tau \sim NG\ (\mu, V, \alpha, b)$

dimana $\theta|\tau \sim N\ (\mu, V)$ dan $\tau \sim G\ (\alpha, b)$

Fungsi kepadatan peluang dari distribusi prior adalah :

$$p(\theta,\tau|C) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det V)^{\frac{1}{2}}} \frac{\alpha^b}{\Gamma(b)} \tau^{\frac{n}{2}+b-1}\ e\ \left(-\frac{\tau}{2}(2\alpha + (\theta - \mu)' V^{-1} (\theta - \mu)\right)$$

$$\propto \tau^{\frac{n}{2}+b-1}\ e\ \left(-\frac{\tau}{2}(2\alpha + (\theta - \mu)' V^{-1} (\theta - \mu)\right) \qquad (3.2)$$

Selanjutnya akan dicari distribusi posterior dari θ dan τ dengan mengoperasikan distribusi prior (3.2) dengan fungsi *likelihoo*d (3.1),

$$p(\theta,\tau|y,C) \propto p(\theta,\tau|C)\ x\ p(y|\theta,\tau,C)$$

$$\propto \tau^{\frac{u}{2}+b-1}\ e\ \left(-\frac{\tau}{2}(2\alpha + (\theta - \mu)' V^{-1} (\theta - \mu))\right).$$

$$\tau^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}(y - X\ )' P (y - X\ )\right)$$

$$\propto \tau^{\frac{n}{2}+\frac{u}{2}+b-1}\ e\ \left(-\frac{\tau}{2}\left(2\alpha + (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{V}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{X}\ )' \boldsymbol{P} (\boldsymbol{y} - \boldsymbol{X}\ )\right)\right) (3.3)$$

Akan diuraikan $\left(2\alpha + (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{V}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}\right) + (\boldsymbol{y} - \boldsymbol{X}\ )' \boldsymbol{P} (\boldsymbol{y} - \boldsymbol{X}\ )$

$$\left(2\alpha + (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{V}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}\right) + (\boldsymbol{y} - \boldsymbol{X}\ )' \boldsymbol{P} (\boldsymbol{y} - \boldsymbol{X}\ )$$

$$\Leftrightarrow\ 2\alpha + \theta' V^{-1} \theta - \theta' V^{-1} \mu - \mu' V^{-1} \theta + \mu' V^{-1} \mu + y' P\ \ - y' P\ \ \theta$$
$$-\theta' X' P\ \ + \theta' X' P$$

$$\Leftrightarrow\ 2\alpha + y' P\ \ + \mu' V^{-1} \mu - 2\theta' (X' P\ \ + V^{-1} \mu) + \theta' (X' P\ \ + V^{-1}) \theta$$

$$\Leftrightarrow\ 2\alpha + y' P\ \ + \mu' V^{-1} \mu - \mu_{\square}' (X' P\ \ + V^{-1}) \mu_{\square}$$
$$+ (\theta - \mu_{\square})' (X' P\ \ + V^{-1}) (\theta - \mu_{\square})$$

$$\Leftrightarrow\ 2\alpha + y' P\ \ + \mu' V^{-1} \mu - 2\mu_{\square}' (X' P\ \ + V^{-1} \mu) + \mu_{\square}' (X' P\ \ + V^{-1}) \mu_{\square}$$

$$
\begin{aligned}
&\quad +(\theta-\mu_{1})'(X'P \quad +V^{-1})(\theta-\mu_{1}) \\
\Leftrightarrow\ & 2\alpha + y'P \quad + \mu'V^{-1}\mu - 2\mu_{1}'X'P \quad - 2\mu_{1}'V^{-1}\mu + \mu_{1}'X'P \quad \mu_{1} + \mu_{1}'V^{-1}\mu_{1} \\
&\quad +(\theta-\mu_{1})'(X'P \quad +V^{-1})(\theta-\mu_{1}) \\
\Leftrightarrow\ & 2\alpha + (\mu-\mu_{1})'V^{-1}(\mu-\mu_{1}) + (y - X \quad _{1})'P(y - X \quad _{1}) \\
&\quad +(\theta-\mu_{1})'(X'P \quad +V^{-1})(\theta-\mu_{1}) \qquad (3.4)
\end{aligned}
$$

Substitusikan Persamaan (3.4) kedalam Persamaan (3.3) sehingga diperoleh

$$
\begin{aligned}
&\propto \tau^{\frac{n}{2}+\frac{u}{2}+b-1}\, e \quad \left(-\frac{\tau}{2}[2\alpha + (\mu-\mu_{1})'V^{-1}(\mu-\mu_{1}) + (y - X \quad _{1})'P(y - X \quad _{1})\right. \\
&\qquad\qquad \left. + (\theta-\mu_{1})'(X'P \quad + V^{-1})(\theta-\mu_{1})]\right) \\
&\propto \tau^{\frac{n}{2}}\, e \quad \left\{-\frac{\tau}{2}[(\theta-\mu_{1})'(X'P \quad + V^{-1})(\theta-\mu_{1})]\right\} \\
&\quad \tau^{\frac{u}{2}+b-1} \exp\left\{-\frac{\tau}{2}[2\alpha + (\mu-\mu_{1})'V^{-1}(\mu-\mu_{1})\right. \\
&\qquad\qquad \left. + \ (y - X \quad _{1})'P(y - X \quad _{1})]\right\} \qquad (3.5)
\end{aligned}
$$

Distribusi posterior untuk θ dan τ pada (3.5) merupakan distribusi normal-gamma.

$\theta, \tau \sim$ NG $(\mu_{1}, V_{1}, \alpha_{1}, b_{1})$

dimana $\theta|\tau \sim$ N (μ_{1}, V_{1}) dan $\tau \sim$ G (α_{1}, b_{1})

dengan

$$
\begin{aligned}
\mu_{1} &= \ (X'P \quad + V^{-1})^{-1}(X'P \quad + V^{-1}\mu) \\
V_{1} &= \ (X'P \quad + V^{-1})^{-1} \\
\alpha_{1} &= \ \alpha + \frac{1}{2}[(\mu-\mu_{1})'V^{-1}(\mu-\mu_{1}) + (y - X \quad _{1})'P(y - X \quad _{1})] \\
b_{1} &= \ \frac{1}{2}n + b
\end{aligned}
$$

Selanjutnya distribusi posterior dari ε dapat diturunkan dengan bentuk $\varepsilon = y - Xe$ dan diketahui bahwa ε merupakan fungsi linear dari parameter θ. De Groot [4] menjelaskan bahwa distribusi ε akan singular hanya pada ruang p-dimensi.

Misalkan $H = X(K + X'X)^{-1}X$, maka distribusi posterior dari ε yang bersyarat τ akan menghasilkan distribusi normal multivariat singular dengan rata-rata $\hat{\varepsilon} = y - Xm_1$ dan matrik kovarians $\tau^{-1}H$.

Dengan memisalkan h_{ij} merupakan elemen-elemen dari H, maka setiap ε_i(untuk i = 1, 2, …, n) memiliki distribusi t dengan lokasi $\hat{\varepsilon}_i$, ketelitian $\frac{a_1}{b_1 h_{ii}}$ dan $2a_1$ sebagai derajat kebebasan, DeGroot [4]. Matrik kovarians dari ε merupakan proporsi dari H.

Distribusi prior improper dari Zellner [5], $p(\beta,\tau|C) = \tau^{-1}$, misalkan $K \to 0$, $a \to -\frac{1}{2}p$ dan $b \to 0$. Misalkan

$$\hat{\beta} = (X'X)^{-1}X'y \;, s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{(n-p)}$$

Selanjutnya De Groot [4] menunjukkan bahwa distribusi posterior dari ε merupakan distribusi t-multivariate pada ruang p-dimensi dengan

$$\hat{\varepsilon} = y - X\hat{\beta} \,, H = X(X'X)^{-1}X \,, \qquad a_1 = \frac{1}{2}(n-p), \qquad b_1 = \frac{1}{2}(n-p)s^2$$

Untuk mendeteksi observasi yang merupakan pencilan, didefinisikan peluang p_i menjadi $p\ (|\varepsilon_i| > k\ |y)$, peluang posterior pada observasi ke-i yang didefinisikan sebagai sebuah pencilan.

Misalkan $\Phi(z)$ merupakan fungsi distribusi normal standar dengan

$$z_1 = \frac{(k - \hat{\varepsilon}_i\sqrt{\tau}\,)}{\sqrt{h_{ii}}} \,, \qquad z_2 = \frac{(-k - \hat{\varepsilon}_i\sqrt{\tau})}{\sqrt{h_{ii}}} \quad (3.6)$$

Selanjutnya diperoleh

$$p_i = p\ (|\varepsilon_i| > k\ |y) = \int \{1 - \Phi(z_1) + \Phi(z_2)\}\, p(\tau|y)d \quad (3.7)$$

Hasil dari p_i dapat dibandingkan dengan peluang prior 2Φ(-k). Nilai dengan peluang posterior p_i yang tinggi yang menjadi pencilan akan memiliki $|\hat{\varepsilon}_i|$ yang besar atau h_{ii} yang besar ataupun keduanya. Ketika $|\hat{\varepsilon}_i|$ membesar akibatnya $|\varepsilon_i|$ juga ikut membesar, namun ketika h_{ii} membesar ada ketidakpastian untuk ε_i. Banyaknya h_{ii} sering menunjukkan sebagai *leverage*.

2.2 Penentuan Nilai k

Nilai k dapat dipilih sehingga peluang prior yang tidak memiliki pencilan besar, misalkan 0,95. Ini memberikan k = $\Phi^{-1}\{0{,}5 + \frac{1}{2}(0{,}95^{1/n})\}$, dengan mensubstitusikan nilai n dan melihat daftar tabel distribusi normal baku.

Cara lain yang dapat digunakan apabila dalam pertimbangannya model diperlukan untuk menggambarkan data bukan terhadap model yang stokastik, maka nilai k = 2 dapat digunakan untuk menemukan observasi yang tidak baik dijelaskan oleh data maupun ukuran sampel.

Chaloner [6] observasi dinamakan sebagai pencilan ketika peluang posterior (3.7) lebih besar dari peluang prior 2Φ(-k).

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam pembahasan ini telah memenuhi semua asumsi klasik sehingga dapat dilanjutkan pada langkah berikutnya. Sebelum masuk pada langkah selanjutnya akan dicari model regresi dengan SPSS yakni

PDRB = -41.863.708 + 0,0021 P. Makanan + 0,0005 P. Bukan Makanan

3.1 Nilai k dan Peluang Prior

Untuk mencari pencilan diperlukan peluang prior untuk membandingkan peluang distribusi posterior. Sebelumnya akan dicari nilai k, untuk n = 26 sehingga

$$k = \Phi^{-1}\{0,5 + \tfrac{1}{2}\,(0,95^{\,1/n})\}$$
$$k = \Phi^{-1}\{0,5 + \tfrac{1}{2}\,(0,95^{\,1/26})\}$$
$$k = \Phi^{-1}\{0,5 + \tfrac{1}{2}\,(0,9980)\}$$
$$k = \Phi^{-1}\{0,5 + 0,4990\}$$
$$k = 3,08$$

Selanjutnya akan dicari nilai peluang prior dengan mensubstitusikan k = 3,08 kedalam $2\Phi(-k)$ menjadi $2\Phi(-3,08)$ sehingga diperoleh peluang prior sebesar 2(0,001) = 0,002.

3.2 Nilai Residual, Leverage, Peluang Posterior, Student Residual dan Deleted Student Residual

Dengan menggunakan *software* SPSS 18 dan Maple 13 diperoleh hasil output sebagai berikut:

Tabel Residual ($\hat{\varepsilon}_i$), *Leverage* (h_{ii}), Peluang Posterior p_i, Student Residual (r_i) dan Deleted Student Residual (t_i)

Obs	$\hat{\varepsilon}_i$	h_{ii}	$p_i(3,08)$	r_i	t_i
1	14.299.455,12677	0,097193		1,22937	1,24391
2	-2.934.982,26229	0,042589		-0,24503	-0,23996
3	-1.636.484,42847	0,042069		-0,13659	-0,13364
4	3.972.828,78386	0,071652		0,33683	0,33024
5	959.958,16598	0,087992		0,08211	0,08032
6	-11.942.876,09265	0,095478		-1,02579	-1,02702
7	1.892.978,20044	0,064433		0,15987	0,15644
8	-5.045.067,45729	0,088581		-0,43169	-0,42392

Obs	$\hat{\varepsilon}_i$	h_{ii}	$p_i(3{,}08)$	r_i	t_i
9	-21.272,82631	0,106004		-0,00184	-0,00180
10	-1.207.976,54512	0,068314		-0,10223	-0,10001
11	-5.845.026,75401	0,074907		-0,49643	-0,48814
12	998.637,71247	0,123354		0,08713	0,08523
13	-8.646.546,11840	0,239163		-0,80976	-0,80350
14	-8.449.708,87415	0,064604		-0,71368	-0,70586
15	11.157.299,91890	0,058199		0,93916	0,93665
16	23.512.723,84663	0,482533	0,005335	2,67007	3,14366
17	-3.684.714,57788	0,140641		-0,32470	-0,31829
18	-8.137.149,32082	0,049492		-0,68180	-0,67365
19	-12.057.033,46284	0,045205		-1,00797	-1,00833
20	37.510.207,52728	0,157599	0,010496	3,33850	4,54802
21	2.493.743,69882	0,106230		0,21548	0,21095
22	-1.790.698,78815	0,117241		-0,15569	-0,15235
23	-18.593.480,37562	0,249166	0,000275	-1,75287	-1,84177
24	-5.797.508,79642	0,075311		-0,49250	-0,48423
25	8.469.149,06269	0,200280		0,77363	0,76666
26	-9.476.455,36343	0,051765		-0,79497	-0,78840

Nilai peluang posterior kurang dari 10^{-6} tidak dicantumkan

Dari hasil output pada tabel di atas dapat dilihat bahwa pada observasi ke-20 memiliki nilai residual paling besar yakni 37.510.207,53 dan juga student residual yang paling besar yakni 3,33850 serta memiliki leverage sebesar 0,157599. Observasi ke-16 memiliki *leverage* yang paling besar yakni 0,482533 namun memiliki student residual yang terbesar kedua yakni 2,67007. Ketika student residual besar maka akan mempengaruhi nilai leverage h_i menjadi besar pula, begitupun sebaliknya, sehingga observasi ke-16 dan ke-20 dicurigai sebagai pencilan.

Pada tabel di atas juga dapat dilihat bahwa nilai peluang posterior observasi ke-16 sebesar 0,005335 ternyata lebih besar dari 0,002 nilai peluang prior sehingga observasi ke-16 termasuk pencilan. Untuk observasi ke-20 dapat dilihat bahwa peluang posteriornya sebesar 0,010496 ternyata lebih besar dari 0,002 nilai peluang prior sehingga observasi ke-20 termasuk pencilan. Sedangkan untuk observasi lainnya memiliki nilai peluang posterior kurang dari 0,002, artinya dalam model regresi ini terdapat observasi yang menjadi pencilan yakni observasi ke-16 dan ke-20.

Dengan adanya pencilan pada observasi ke-16 yakni Kabupaten Bekasi dan observasi ke-20 yakni Kota Bandung merupakan daerah yang memberikan kontribusi PDRB yang paling baik, dapat dilihat bahwa nilai PDRB kedua daerah ini besar. Kabupaten Bekasi mempunyai pengeluaran konsumsi makanan yang paling besar dan pengeluaran konsumsi bukan makanan yang relatif besar serta Kota Bandung yang juga mempunyai pengeluaran konsumsi makanan dan bukan makanan yang relatif besar. Oleh karena itu, hal ini bisa menjadi salah satu pertimbangan untuk meningkatkan PDRB di Jawa Barat tahun selanjutnya dengan meningkatkan pengeluaran konsumsi baik makanan maupun bukan makanan.

4. KESIMPULAN

Kesimpulan yang dapat dipaparkan adalah

1. Langkah-langkah yang diperlukan untuk mendeteksi pencilan dengan pendekatan Bayesian adalah:
   a. Distribusi improper prior dari ε yakni

      $p(\beta, \tau|C) = \tau^{-1}$, misalkan $K \to 0$, $a \to -\frac{1}{2}p$ dan $b \to 0$. Misalkan

      $$\hat{\beta} = (X'X)^{-1}X'y \ , \qquad s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{(n-p)}$$

   b. Distribusi posterior dari ε merupakan distribusi t multivariat pada ruang p-dimensi dengan

      $\hat{\varepsilon} = y - X\hat{\beta}$, $H = X(X'X)^{-1}X$,

      $a_1 = \frac{1}{2}(n-p), \qquad b_1 = \frac{1}{2}(n-p)s^2$

   c. Selanjutnya mencari distribusi peluang posterior dengan memisalkan $\Phi(z)$ merupakan fungsi distribusi normal standar dengan

      $$z_1 = \frac{(k - \hat{\varepsilon}_i\sqrt{\tau}\,)}{\sqrt{h_{ii}}} \ , \qquad z_2 = \frac{(-k - \hat{\varepsilon}_i\sqrt{\tau})}{\sqrt{h_{ii}}}$$

      sehingga diperoleh

      $$p_i = p \ \ (|\varepsilon_i| > k \ \ |y) = \int \{1 - \Phi(z_1) + \Phi(z_2)\}\, p(\tau|y)d$$

   d. Menentukan nilai k = $\Phi^{-1}\{0{,}5 + \frac{1}{2}\,(0{,}95^{\ 1/n})\}$, dengan mensubstitusikan nilai n dan melihat pada tabel distribusi normal baku.

e. Selanjutnya menentukan peluang distribusi prior dengan mensubstitusikan k kedalam $2\Phi(-k)$. Kemudian bandingkan antara nilai peluang distribusi prior dengan nilai peluang distribusi posterior. Jika nilai peluang prior lebih kecil dari nilai peluang posterior maka pengamatan tersebut dikatakan sebagai pencilan.

2. Penerapan deteksi pencilan dengan pendekatan Bayesian pada regresi linear dalam studi kasus hubungan antara pengeluaran rumah tangga dengan PDRB di provinsi Jawa Barat tahun 2013 diperoleh bahwa observasi yang menjadi pencilan adalah Kabupaten Bekasi dan Kota Bandung. Kabupaten Bekasi dan Kota Bandung merupakan daerah yang memberikan kontribusi PDRB yang paling baik, dapat dilihat bahwa nilai PDRB kedua daerah ini besar, Kabupaten Bekasi mempunyai pengeluaran konsumsi makanan yang paling besar dan pengeluaran konsumsi bukan makanan yang relatif besar serta Kota Bandung yang juga mempunyai pengeluaran konsumsi makanan dan bukan makanan yang relatif besar. Oleh karena itu, hal ini bisa menjadi salah satu pertimbangan untuk meningkatkan PDRB di Jawa Barat tahun selanjutnya dengan meningkatkan pengeluaran konsumsi baik makanan maupun bukan makanan.

REFERENSI

[1] Rahmawati, D. (2011). *Estimasi Model Regresi Linier dengan Pendekatan Bayes.* [Online]. Tersedia: http://lib.uin-malang.ac.id/?mod=th_detail&id=07610049 [25 Februari 2014]

[2] Nazra, A. (2010). *Analisis Perbandingan Berbagai Uji Pencilan Pada Analisis Regresi.* [Online]. Tersedia: *repository.unand.ac.id/2254/1/Admi_Nazra.pdf* [13 Februari 2014]

[3] Azhar, JA. (2012). Perbandingan Metode Bayes dan Metode *Likelihood* dalam Mengestimasi Parameter Model Regresi Linear. [Online]. Tersedia: https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwjyy6Gg2NrKAhWQCY4KHZJnBRkQFggoMAI&url=http%3A%2F%2Fdigilib.uin-suka.ac.id%2F10874%2F1%2FBAB%2520I%2C%2520VI%2C%2520DAFTAR%2520PUSTAKA.pdf&usg=AFQjCNERE2-lVEJ8tN37ORsjOp2XZU9jxw [3 Februari 2016]

[4] DeGroot, M.H. (2004). *Optimal Statistical Decisions*. Canada: John Wiley & Sons, Inc.

[5] Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics.* United States of America: John Wiley & Sons, Inc.

[6] Chaloner, K and Brant, R. (1988). A Bayesian Approach to Outlier Detection and Residual Analysis. *Biometrika, 75 (4),* hlm. 651-659.