

## PERBANDINGAN METODE KUADRAT TERKECIL DAN METODE BAYES DALAM MENGESTIMASI MODEL REGRESI LINIER SEDERHANA DENGAN GALAT HETEROSKEDASTISITAS

RYAN HANDANI, MAIYASTRI, RITA DIANA

*Program Studi Matematika,  
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,  
Kampus UNAND Limau Manis Padang, Indonesia,  
email : dhan7545@gmail.com*

**Abstrak.** Analisis regresi merupakan salah satu metode dalam statistik yang digunakan untuk melihat hubungan antara variabel bebas dengan variabel tak bebas, salah satunya yang dinyatakan dalam model regresi linier sederhana, dimana dalam pendugaan parameternya terdapat dua pendekatan yaitu, pendekatan klasik pendekatan Bayesian. Salah satu metode estimasi dengan pendekatan klasik adalah Metode Kuadrat Terkecil (MKT), dimana dalam MKT terdapat asumsi-asumsi yang harus dipenuhi diantaranya adalah *error* berdistribusi normal, *independent* dan identik (homogen). Jika salah satu asumsi dilanggar seperti *error*nya tidak identik (heteros) maka metode Bayes dapat digunakan untuk mengatasi masalah tersebut. Dengan metode Bayes diperoleh hasil bias yang lebih kecil dibandingkan dengan MKT.

### 1. Pendahuluan

Analisis regresi merupakan suatu kajian terhadap hubungan antara variabel bebas dan variabel tidak bebas. Model regresi dapat dituliskan sebagai  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  dimana  $y_i$  menyatakan variabel tak bebas,  $x_i$  menyatakan variabel bebas,  $\beta_0, \beta_1$  merupakan parameter dan  $\varepsilon_i$  merupakan faktor galat. Dari model regresi ini dapat diestimasi besarnya pengaruh secara kuantitatif dari masing-masing variabel bebas terhadap variabel tak bebas.

Ada beberapa metode penyusunan model regresi, tetapi sejauh yang menyangkut analisis regresi, metode yang paling luas digunakan adalah metode kuadrat terkecil atau *Ordinary Least Square*. Estimasi parameter model regresi yang di peroleh dengan OLS merupakan estimator yang baik bila model regresi memenuhi asumsi-asumsi tertentu yang sering disebut dengan asumsi model regresi linear klasik. Satu asumsi kritis dari model regresi linear klasik adalah bahwa galat  $\varepsilon_i$  adalah galat yang diasumsikan semuanya mempunyai varians yang sama. yang mana apabila asumsi ini dilanggar kita mempunyai kasus Heteroskedastisitas.

Sedangkan pada metode Bayes memperkenalkan suatu metode yang diperlukan untuk mengetahui bentuk distribusi awal (*prior*) dari populasi yang dikenal dengan metode bayesian. Informasi ini kemudian digabungkan dengan informasi dari sampel untuk digunakan dalam mengestimasi parameter populasi. Setelah informasi dari data yang didapat dari pengambilan sampel digabungkan dengan informasi

*prior* dari parameter, akan didapat distribusi *posterior* dari parameter. Rataan dari distribusi *posterior* ini yang akan menjadi parameter regresi dengan metode Bayes.

## 2. Kajian Pustaka

Pada bagian ini akan diterangkan mengenai teori-teori terkait metode estimasi dengan Metode Kuadrat Terkecil dan metode Bayes.

### 2.1. Metode Kuadrat Terkecil

regresi linier sederhana melibatkan satu variabel bebas dengan satu variabel tak bebas. Bentuk umum model regresi linear sederhana dapat dituliskan sebagai

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \text{ dimana}$$

$y_i$  : variabel tak bebas yang bersifat acak

$x_i$  : variabel bebas yang bersifat tetap

$\beta_0, \beta_1$  : parameter regresi

$\varepsilon_i$  : variabel galat

Salah satu metode estimasi yang digunakan untuk menduga parameter model regresi linear sederhana dalam menduga nilai  $\beta_0, \beta_1$  adalah Metode Kuadrat Terkecil dengan model dugaan:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

dengan :

$$\hat{\beta}_0 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2},$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Agar pendugaan parameter pada model regresi linier sederhana menghasilkan estimator linier tidak bias terbaik, mesti memenuhi beberapa asumsi yang disebut asumsi klasik, sebagai berikut:

- (1)  $\varepsilon_i$  merupakan peubah acak menyebar menurut sebaran normal.
- (2) Nilai tengah  $\varepsilon_i = 0$ , dituliskan  $E(\varepsilon_i) = 0$ .
- (3) Ragam  $\varepsilon_i$  konstan yang nilainya tidak diketahui, dituliskan  $Var(\varepsilon_i) = \sigma^2$ .
- (4) Galat pengamatan-pengamatan yang berbeda saling bebas

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j.$$

Secara ringkas dapat dituliskan sebagai berikut

$$\varepsilon_i \sim N(0, \sigma^2).$$

Salah satu asumsi model klasik adalah asumsi homoskedastisitas yaitu varian tiap unsur galat merupakan suatu bilangan konstan  $Var(\varepsilon_i) = \sigma^2$ , Jika terjadi

pelanggaran terhadap asumsi ini, maka model regresinya dikatakan berada dalam keadaan heterokedastisitas, yaitu varian tiap unsur galat tidak konstan

$$\text{Var}(\varepsilon_i) = \sigma_i^2.$$

## 2.2. Metode Bayes

Pada metode Bayes untuk menduga parameter nilai estimatornya diperoleh dari mean distribusi posterior. Misalkan data observasi  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  mempunyai distribusi tertentu dengan himpunan parameter  $\theta = (\theta_1, \theta_2, \dots, \theta_n)'$  yang merupakan variabel random, maka distribusi *posterior* dari parameter  $\theta$  atau  $f(\theta|\mathbf{y})$  dapat dinyatakan dalam bentuk:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})}$$

dengan:

$$f(\mathbf{y}) = \int_{\Omega_{\theta_1}} \cdots \int_{\Omega_{\theta_p}} f(\mathbf{y}|\theta)f(\theta)d\theta_1 \cdots d\theta_p : \text{normalizing constant,}$$

$$f(\theta) : \text{distribusi prior dari parameter } \theta,$$

$$f(\mathbf{y}|\theta) = \prod f(y_i|\theta) : \text{fungsi Likelihood data.}$$

Metode MCMC merupakan suatu metode untuk menentukan nilai parameter dengan menggunakan teknik simulasi numerik yang digunakan untuk menyelesaikan masalah pemodelan kompleks. Ide dasar dari MCMC adalah membangkitkan data sampel dari distribusi *posterior* setiap parameter model sesuai proses Markov Chain dengan menggunakan simulasi Monte Carlo secara iteratif hingga diperoleh kondisi yang konvergen.

Markov Chain merupakan proses iterasi sekumpulan variabel random  $\theta$  sedemikian sehingga  $f(\theta^{t+1}|\theta^t, \dots, \theta^1) = f(\theta^{t+1}|\theta^t)$ . Hasil estimasi *posterior* parameter dikatakan baik jika terpenuhinya sifat dari Markov Chain yang *strongly ergodic* berupa:

- (1) *Irreducible*, artinya sampel parameter yang dibangkitkan melalui proses MCMC adalah bersifat random.
- (2) *Aperiodic*, artinya sampel parameter yang dibangkitkan tersebut tidak memiliki pola yang periodik dalam domain nilai tertentu.
- (3) *Recurrent*, artinya perubahan sampel parameter terjadi secara stabil dalam domain nilai tertentu.

Penggunaan metode MCMC dalam analisis Bayesian memerlukan algoritma *sampling* yang sesuai untuk mendapatkan sampel dari suatu distribusi. Salah satu algoritma yang sering digunakan adalah *Gibbs Sampling*, Proses *Gibbs Sampling* dilakukan dengan cara membangkitkan rangkaian *Gibbs* variabel random berdasarkan sifat-sifat dasar proses Markov Chain yang menggunakan konsep distribusi unidimensional yang terstruktur dalam bentuk *full conditional*.

### 3. Data dan Hasil

Data yang digunakan pada penelitian ini adalah data yang dibangkitkan secara acak dengan menggunakan software R, dimana nilai  $x_i, \beta_0, \beta_1$  ditetapkan kemudian dengan galat  $\varepsilon_i$  mengalami heteroskedastisitas kemudian dihitung nilai  $y_i$ . lalu akan diuji apakah data yang kita bangkitkan sebelumnya telah mengalami heteroskedastisitas, yang mana kita akan menggunakan uji-*glejser* dengan taraf nyata  $\alpha = 0,05$ , Berdasarkan data yang dibangkitkan dengan jumlah sampel  $n=30$  dan ragam  $\sigma^2 = 1$ , dengan iterasi sebanyak 1000 diperoleh nilai *p-value* sebesar  $0.028 < 0.05$  yang artinya bahwa data yang kita bangkitkan mengalami heteroskedastisitas.

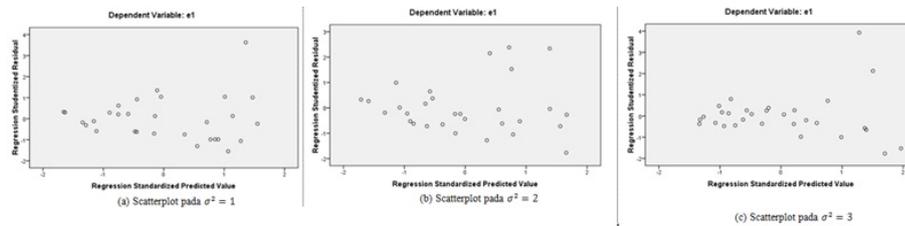
#### 3.1. Estimasi Parameter Model Regresi Linier Sederhana dengan Metode Kuadrat Terkecil

Dengan menggunakan data yang telah dibangkitkan akan dilakukan estimasi dengan metode kuadrat terkecil untuk menduga  $\hat{\beta}_0, \hat{\beta}_1$  dan serta nilai dari Absolut bias yang disajikan pada Tabel 1.

Tabel 1. Koefisien regresi model regresi linier sederhana dengan metode Kuadrat terkecil

No	Parameter	Nilai Estimasi	Absolut Bias	Standar deviasi
1	$\hat{\beta}_0$	2.254	3,07496	0.29
2	$\hat{\beta}_1$	2.981	0,31151	0.012

Untuk melihat pengaruh galat Heteroskedastisitas dapat dilihat pada Scatterplot pada Gambar 1 bahwa varians galat tidak konstan dimana tampak bahwa data memiliki pola.



Gambar 1. Scatterplot dengan ukuran sample  $n=30$  dengan iterasi 1000

### 3.2. Estimasi Parameter Model Regresi Linier Sederhana dengan Pendekatan Bayes

Jika  $y_i$  adalah variabel *random* yang merupakan respon dengan galat heteroskedastisitas  $\varepsilon_i \sim N(0, \sigma_i^2)$  maka  $y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 x_i, \sigma_i^2)$  dan  $\sigma_i^2 = \frac{1}{\tau_i}$  adalah parameter persisi, maka fkp dari  $y_1$  adalah

$$f(y|\beta_0, \beta_1, \tau_i) = \left(\frac{\tau_i}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\tau_i}{2}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

sehingga bentuk Likelihood dari  $\mathbf{y}$  sebagai berikut:

$$f(y|\beta_0, \beta_1, \tau) \propto \prod \sqrt{\tau_i} \exp\left\{-\frac{1}{2} \sum \tau_i ((y_i - \beta_0 - \beta_1 x_i)^2)\right\},$$

dengan menggunakan distribusi *prior* konjugat yang berdistribusi normal untuk parameter  $\beta_0, \beta_1$  yang dinyatakan sebagai berikut.

$$\beta_j \sim N(\mu_{\beta_j}, \sigma_{\beta_j}^2),$$

dimana  $j = 0, 1$ , dan distribusi *prior* konjugat yang berdistribusi gamma untuk parameter  $\tau_i$  yang dinyatakan sebagai berikut.

$$\tau_i \sim \gamma(a, b),$$

sehingga bentuk proposional distribusi *posterior* diperoleh sebagai berikut:

$$f(\beta_0, \beta_1, \tau|y) \propto \prod \sqrt{\tau_i} \exp\left\{-\frac{1}{2} \sum \tau_i ((y_i - \beta_0 - \beta_1 x_i)^2)\right\} \\ x \sqrt{\tau_{\beta_j}} \exp\left(-\frac{\tau_{\beta_j}}{2} (\beta_j - \mu_{\beta_j})^2\right) \tau_i^{a-1} \exp(-b\tau_i).$$

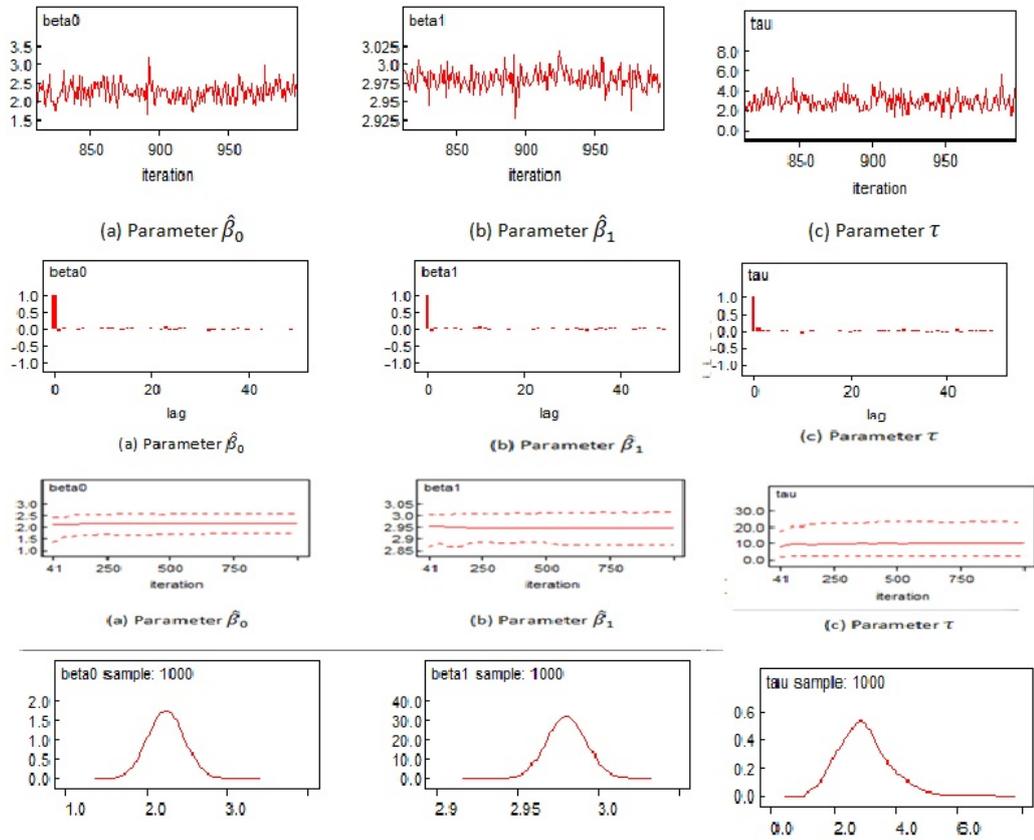
Dari distribusi *posterior* diatas dengan menggunakan data yang dibangkitkan dengan jumlah sampel  $n=30$  dan ragam  $\sigma^2 = 1$ , dengan iterasi sebanyak 1000, dilakukan proses MCMC dengan algoritma *Gibbs Sampling* menggunakan bantuan *software* WinBUGS dan diperoleh hasil seperti pada Gambar 2.

Trace plot parameter menunjukkan nilai acak dengan pola yang stasioner dan bersifat random. Trace plot untuk estimasi parameter beberapa koefisien  $\hat{\beta}_0, \hat{\beta}_1$  dan  $\tau$ , Autocorrelation plot untuk parameter yang diestimasi memperlihatkan nilai autokorelasi yang tidak signifikan karena berada dalam 95% *confidence interval*. Hal ini menunjukkan bahwa sampel yang dibangkitkan dalam proses MCMC memiliki sifat yang random. Quantiles plot menunjukkan ergodic mean hasil estimasi parameter  $\hat{\beta}_0, \hat{\beta}_1$  dan  $\tau$  yang dihasilkan telah mencapai nilai yang stabil. Kondisi ini merupakan salah satu indikasi bahwa proses iterasi estimasi parameter  $\hat{\beta}_0, \hat{\beta}_1$  dan  $\tau$  Telah mencapai kondisi equilibrium atau konvergen.

Plot fungsi kepadatan distribusi posterior koefisien regresi model regresi linier sederhana dengan pendekatan Bayes, Terlihat bahwa plot fungsi kepadatan *posterior* koefisien regresi model regresi linier sederhana dengan pendekatan Bayes yang diperoleh sesuai dengan distribusi *prior* yang dipilih.

Selanjutnya, ringkasan posterior parameter yang diperoleh dari proses estimasi parameter model regresi linier sederhana dengan pendekatan Bayes pada Gambar 3 berikut.

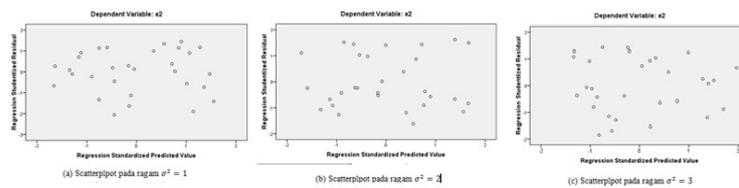
Dapat dilihat pada scatterplot pada Gambar 4 bahwa varians galat telah konstan dan data menyebar acak yang artinya tidak terjadi heteroskedastisitas



Gambar 2. Diagnostic plot pada proses Markov Chain Monte Carlo

No	Parameter	Posterior mean	Absolut Bias	Standar deviasi
1	$\hat{\beta}_0$	2.251	0,2781	0.2235
2	$\hat{\beta}_1$	2.981	0,01986	0.01211
3	$\tau$	2.958		0.7936

Gambar 3. Tabel Koefisien regresi model regresi linier sederhana dengan metode Bayes



Gambar 4. Scatterplot dengan ukuran sample  $n=30$  dengan iterasi 1000

Hal yang sama juga dilakukan pada data dengan berbagai simulasi yang berbeda dan diperoleh hasil yang sama.

**4. Perbandingan Estimasi Parameter metode OLS dan Metode Bayes**

Setelah dilakukan estimasi model regresi linier terhadap parameter  $\hat{\beta}_0, \hat{\beta}_1$  dan  $\tau$  dengan metode OLS dan Metode Bayes, berikutnya akan kita lakukan perbandingan terhadap kedua metode yaitu salah satu kriteria uji kita adalah dengan membandingkan nilai absolut bias. dimana dari nilai absolut bias ini dapat dilihat metode mana yang lebih baik, apabila nilai absolut bias semakin kecil berarti model dugaan parameter kita sudah bagus.

Pada Gambar 5 diberikan tabel nilai dugaan dan nilai absolut bias dari kedua metode pada iterasi (T) 1000 kali dengan  $n = 10, 20$ , dan  $30$  dengan berbagai macam ragam ( $\sigma^2 = 1, 2$  dan  $3$ ).

		Nilai Absolut Bias											
		Ragam = 1				Ragam = 2				Ragam = 3			
		Ols		Bayes		Ols		Bayes		Ols		Bayes	
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
10		1,765976	0,565351	0,2171	0,05595	2,980654	1,194804	0,2079	0,05309	4,594361	1,669646	0,36	0,06252
20		2,532187	0,409666	0,118	0,0133	4,990034	0,865684	0,2982	0,02461	8,392009	1,193094	0,4686	0,06151
30		3,074963	0,311514	0,2781	0,01986	6,271104	0,622440	0,2781	0,01986	9,018307	1,037985	0,401	0,03501

Gambar 5. Tabel Nilai dugaan dan nilai absolut bias estimasi parameter metode OLS dan metode Bayes

Dari Gambar 5 dapat dilihat bahwa nilai absolut bias dengan metode Bayes lebih kecil dan stabil ketimbang dengan Metode Kuadrat Terkecil.

**5. Kesimpulan**

Berdasarkan uraian di atas dapat disimpulkan bahwa

- (1) Penduga dengan metode Bayes memiliki nilai Absolut Bias yang lebih kecil dan stabil dari penduga metode kuadrat terkecil pada data simulasi yang dibangkitkan.
- (2) Penduga dengan metode Bayes lebih baik dalam menangani kasus dengan galat Heteroskedastisitas dari penduga metode kuadrat terkecil itu dikarenakan pada metode Bayes tetap memenuhi sifat MCMC.

**6. Ucapan Terima Kasih**

Penulis mengucapkan terima kasih kepada Bapak Dodi Devianto, Bapak Yudiantri Asdi, dan Bapak Narwen yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

**Daftar Pustaka**

- [1] Walpole, Ronald E. 1993. *Pengantar Statistika* Edisi Ke-3. Jakarta: PT Gramedia Pustaka Utama
- [2] Box, G.E.P and Tiao, G.C 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company, Inc; Philippines
- [3] Gujarati, N. 1995. *Basic Econometrics*. McGraw-Hill, California
- [4] Bolstad, W.M. 2007. *Introduction to Bayesian Statistical* Second Edition. John Wiley & Sons. Inc; America