Jurnal Sains & Matematika (JSM) ISSN 0854-0675
Volume 15, Nomor 2, April 2007 Artikel Penelitian: 69-72

Estimasi Model Regresi Linier Dengan Metode Median Kuadrat Terkecil

Tarno[1]
[1] Staf Pengajar Program Studi Statistka FMIPA Undip

ABSTRAK---Model regresi linier merupakan model yang paling sering digunakan dalam analisis statistika. Model regresi linier ini digunakan untuk menyatakan hubungan fungsional antara satu atau beberapa variabel bebas (prediktor) terhadap satu variabel terikat (respon). Dalam analisis regresi, mengestimasi parameter secara otomatis mengestimasi model regresi. Untuk memperoleh estimasi model regresi dapat dilakukan dengan beberapa metode antara lain: metode kuadrat terkecil, metode maksimum likelihood dan sebagainya. Salah satu metode yang paling populer adalah metode kuadrat terkecil (OLS). Pada prinsipnya metode kuadrat terkecil mengestimasi model regresi dengan meminimalkan rata-rata kuadrat sesatan (MSE). Dalam tulisan ini dibahas suatu metode alternatif untuk mendapatkan estimasi model regresi yaitu metode median kuadrat terkecil (LMS). Pada metode LMS, estimasi model yang diperoleh adalah suatu model yang memiliki median kuadrat sesatan terkecil. Prosedur estimasinya adalah dengan memilih p titik sampel (dengan p: banyaknya parameter di dalam model termasuk intersept) dari n titik sampel hasil pengamatan, kemudian ditentukan suatu persamaan yang melalui p titik tersebut. Setelah diperoleh sejumlah persamaan yang melalui p titik tersebut, kemudian ditentukan median dari residual kuadrat. Persamaan atau model yang diestimasi melalui p titik yang menghasilkan nilai median kuadrat terkecil merupakan model yang terpilih.

Kata Kunci: regresi linier, estimasi parameter, sesatan kuadrat

PENDAHULUAN

Salah satu model yang sangat berguna dan sering digunakan dalam berbagai bidang aplikasi adalah model linier umum :

$$y_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + \varepsilon_i, i = 1,2,...,n \qquad (1)$$

dengan $\mathbf{y}_i$ adalah respon ke-i , $\mathbf{x}_i$: p-vektor variabel prediktor yang berkaitan dengan y_i, $\boldsymbol{\beta}$: p-vektor parameter yang tidak diketahui dan ε_i : sesatan random. Masalah regresi linier dapat diformulasikan sebagai kasus khusus dari model (1).

Apabila $\mathbf{x}_i$ dalam model tersebut tertentu (deterministik), diasumsikan bahwa ε_i independen dengan mean 0 dan variansi σ^2. Jika $\mathbf{x}_i$ tersebut random, maka model (1) dikatakan sebagai model korelasi. Dalam suatu model korelasi, $(y_i, \mathbf{x}_i^{'})$ diasumsikan independen dan berdistribusi identik dengan momen kedua berhingga dan $E(y_i \mid \mathbf{x}_i) = \mathbf{x}_i^{'}\boldsymbol{\beta}$, σ_i^2 menyatakan variansi bersyarat dari y_i diberikan $\mathbf{x}_i$.

Parameter $\boldsymbol{\beta}$ dalam model tersebut dikenal sebagai parameter regresi. Dalam model regresi dengan asumsi bahwa sesatan berdistribusi normal dengan mean 0 dan variansi konstan σ^2, serta memenuhi asumsi-asumsi dari regresi klasik lainnya, maka estimasi parameternya dapat ditentukan dengan menggunakan metode kuadrat terkecil (OLS).

Metode kuadrat terkecil merupakan suatu metode estimasi parameter regresi yang paling populer. Disamping metode tersebut, terdapat sejumlah metode alternatif yang dapat digunakan untuk mendapatkan estimasi model regresi linier, seperti: metode maksimum likelihood, metode median terkecil (least median) dan sebagainya. Dari sejumlah metode yang ada, masing-masing mempunyai kelebihan dan kekurangan, misalnya: untuk metode OLS digunakan apabila estimasi parameternya merupakan penyelesaian eksak dari persamaan normal, sedangkan MLE merupakan metode pendekatan apabila penyelesaian untuk parameter regresinya tidak eksak. Terlepas dari kelebihan dan kekurangan masing-masing metode, dalam tulisan ini dibahas tentang Metode Median Terkecil (Least Median Squre) yang dapat digunakan sebagai suatu metode alternatif untuk estimasi parameter regresi linier.

REGRESI LINIER DAN ESTIMASI PARAMETER

Struktur probabilitas dari model linier biasanya dinyatakan seperti persamaan (1), yaitu:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, i = 1,2,...,n$$

Suku sesatan ε_i dalam persamaan (1) diasumsikan sebagai suatu sampel random dari suatu distribusi F yang tidak diketahui yang mempunyai nilai harapan 0,

$$F \to (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) = \boldsymbol{\varepsilon} \text{ dan } E_F(\boldsymbol{\varepsilon}) = 0 \quad (2)$$

Dari (1) dan (2) berakibat:

$E(y_i \mid \mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$, dan harus memenuhi asumsi linieritas.

Jika diinginkan untuk mengestimasi parameter regresi $\boldsymbol{\beta}$ dari data pengamatan $((x_1, y_1), (x_2, y_2), ..., (x_n, y_n))$, maka nilai estimasi dari $\boldsymbol{\beta}$, katakanlah **b** memberikan nilai residual kuadrat:

$$\text{RSE}(\mathbf{b}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\mathbf{b})^2 \quad (3)$$

Estimasi kuadrat terkecil dari $\boldsymbol{\beta}$ merupakan nilai $\hat{\boldsymbol{\beta}}$ dari **b** yang meminimalkan RSE(**b**)

$$\text{RSE}(\hat{\beta}) = \min_{b}[RSE(\mathbf{b})] \quad (4)$$

Jika $\mathbf{X}$ menyatakan matriks n x p dengan baris ke-i adalah $\mathbf{x_{i\cdot}}$ dan $\mathbf{y} = (y_1, y_2, ..., y_n)'$. Maka estimasi kuadrat terkecilnya merupakan solusi persamaan normal:

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{y},$$

yang diberikan oleh rumus:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (5)$$

METODE MEDIAN KUADRAT TERKECIL

Regresi median kuadrat terkecil disingkat LMS merupakan suatu teknik estimasi model yang kuarang sensitive dibandingkan dengan metode kuadrat terkecil (OLS). Perbedaan antara keduanya terletak pada pemilihan kriteria kecocokan model. Pada metode kuadrat terkecil (OLS) kriteia yang digunakan adalah meminimalkan rata-rata kuadrat residual:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\mathbf{b})^2 \quad (6)$$

Rata-rata sample sensitive terhadap nilai yang mempengaruhinya, tetapi untuk median tidak. Oleh karena itu agar persamaan (6) kurang sensitive, rata-rata tersebut diganti dengan median, dengan memberikan median residual kuadrat:

$$\text{MSR(b)=median}(y_i - \mathbf{x}_i'\mathbf{b})^2 \quad (7)$$

Estimasi LMS dari $\boldsymbol{\beta}$ adalah nilai $\hat{\boldsymbol{\beta}}$ yang meminimalkan MSR(**b**),

$$\mathbf{MSR}(\hat{\beta}) = \min_{b}[\text{MSR}(\mathbf{b})] \quad (8)$$

Perbedaan antara OLS dan LMS adalah bukan pada pemilihan model tetapi bagaimana mengukur kecocokan antara model dengan data observasi. MSR(b) kurang sensitive dibandingkan RSE(b) dalam hal menjabarkan titik-titik data.

SIMULASI

Dalam bagian ini diberikan contoh penerapan metode median kuadrat terkecil (LMS) serta perbandingannya terhadap metode kuadrat terkecil (OLS). Data yang digunakan untuk simulasi adalah data tentang: jumlah sisa hormone anti-inflammatory (dalam miligram) sebagai respon dan lamanya pemakaian (dalam jam) sebagai variabel prediktor, diambil dari: Efron & Tibshirani, 1993, Introduction to Bootstrap, Chapman & Hall Inc., USA. Data lengkap yang digunakan untuk simulasi dapat dilihat pada Tabel 1.

Tabel 1. Data jumlah sisa hormone anti-inflammatory (dalam miligram) sebagai respon (Y) dan lamanya pemakaian (dalam jam) sebagai variabel prediktor (X).

No	X	Y	No	X	Y	No	X	Y
1	99	25.8	10	376	16.3	19	119	28.8
2	152	20.5	11	385	11.6	20	188	22.0
3	293	14.3	12	402	11.8	21	115	29.7
4	155	23.2	13	29	32..5	22	88	28.9
5	196	20.6	14	76	32.0	23	58	32.8
6	53	31.1	15	296	18.0	24	49	32.5
7	184	20.9	16	151	24.1	25	150	25.4
8	171	20.8	17	177	26.5	26	107	31.7
9	52	30.4	18	209	25.8	27	125	28.5

Listing Program menggunakan software ”R” untuk mendapatkan estimasi model regresi dengan Metode Median Kuadrat Terkecil (LMS) adalah sebagai berikut:

```
Regresi_median<-function (B,x,y)
 {
      I<-seq(1:length(x))
      C<-rep(1,length(x))
      Is<-matrix(0,2*B, nrow=2)
      yy<-matrix(0,2*B, nrow=2)
      xx<-matrix(0,2*B, nrow=2)
      beta<-matrix(rep(0,2*B), nrow=2)
      res<-matrix(rep(0,length(x)*B), nrow=length(x))
      sisa2<-matrix(rep(0,length(x)*B), nrow=length(x))
      nilaitengah<-rep(0,B)
      for(j in 1:B)
            {
            Is[,j] <- sample(I, replace = F, size=2)
            yy[,j] <-y[Is[,j]]
            xx[,j] <-x[Is[,j]]
            beta[,j] <- glm(yy[,j] ~ xx[,j])$coef
            X<-matrix(c(C,x),nrow=length(x))
            y<-matrix(y,nrow=length(y))
            res[,j]<-y-(X%*%beta[,j])
            sisa2[,j] <- (res[,j])*(res[,j])
            nilaitengah[j]<-median(sisa2[,j])
  }
  rbind(nilaitengah, beta)
 }
```

Hasil estimasi parameter dengan menggunakan metode kuadrat terkecil maupun metode Median kuadrat terkecil dapat dilihat pada Tabel 2 berikut:

Tabel 2. Estimasi parameter dengan OLS dan LMS

No	Metode	Intersept (β_o)	Gradien (β_1)	MSE atau MSR
1	OLS	34.16417	-0.05745	5.262578
2	LMS	37.37337	-0.05864	2.267882

Dari tabel 1 diperoleh estimasi model untuk masing-masing metode sebagai berikut:

$\hat{y} =$ 34.16417 – 0.05745 x, untuk metode OLS dengan MSE 5.262578; dan

$\hat{y} =$ 37.37337 – 0.05864 x, untuk metode LMS dengan MSR 2.267882.

Dengan demikian baik menggunakan metode OLS maupun LMS diperoleh estimasi model regresi yang tidak memiliki perbedaan yang mencolok atau kedua model yang diperoleh mempunyai keakuratan yang hampir sama.

KESIMPULAN

Pada metode LMS, estimasi model yang diperoleh adalah suatu model yang memiliki median kuadrat sesatan terkecil. Prosedur estimasinya adalah dengan memilih p titik sampel (dengan p: banyaknya parameter di dalam model termasuk intersept) dari n titik sampel hasil pengamatan, kemudian ditentukan suatu persamaan yang melalui p titik tersebut. Setelah diperoleh sejumlah model yang melalui p titik tersebut (banyaknya model yang terbentuk maksimal ada C_p^n), kemudian ditentukan median dari residual kuadrat. Persamaan atau model yang diestimasi melalui p titik yang menghasilkan nilai median kuadrat terkecil merupakan model yang terpilih.

DAFTAR PUSTAKA

1. Asshish Sen, Muni Srivastava, 1994, Regression Analysis: Theory, Methods, and Analysis, Springer-Verlag, New York Inc.
2. Efron and Tibshirani, 1993, An Introduction to Bootstrap, Chapman & Hall Inc., USA.
3. Hjorth, J.S.U, (1994), Computer Intensive Statistical Methods, Validation Model Selection and Bootstrap, Chapman and Hall, New York.
4. Montgomery & Peck, 1982, Intoduction to Linear Regression Analysis, John Wiley and Sons, New York.