JURNAL MATEMATIKA DAN KOMPUTER
Vol. 7. No. 3, 8 - 17, Desember 2004, ISSN : 1410-8518

PEMODELAN STATISTIKA DENGAN TRANSFORMASI BOX COX

Dwi Ispriyanti
Staf Pengajar Jurusan Matematika
Fakultas MIPA UNDIP
Semarang

Abstrak

Analisis regresi adalah salah satu teknik statistika yang digunakan untuk menentukan model hubungan satu variabel respon (Y) dengan satu atau lebih variabel penjelas (X).Dalam melakukan analisisnya harus dipenuhi bahwa $\varepsilon_i \sim NID(0, \sigma^2)$ *, jika asumsi tersebut tidak dipenuhi maka dapat dilakukan transfomasi terhadap Y yang dipangkatkan dengan parameter* λ*,sehingga menjadi* Y^{λ} *. Pendugaan parameter* λ *dapat dicari dengan menggunakan Metode Kemungkinan Maksimum (Maximum Likehood Method).* λ *dipilih sedemikian, sehingga didapat jumlah kuadrat sisaan yang paling minimum.*

Kata Kunci : *Transformasi Box Cox, Maximum Likelihood Method*

1. PENDAHULUAN

Analisis Regresi merupakan salah satu cabang statistika yang paling banyak dipelajari oleh ilmuwan, khususnya para peneliti, baik ilmuwan bidang sosial maupun eksakta. Melalui analisis regresi model hubungan antar variabel dapat diketahui. Variabel dalam analisis regresi dikenal dengan nama variabel respon(Y) dan dan variabel penjelas (X).

Dalam melakukan analisis regresi ada beberapa asumsi yang harus dipenuhi antara lain ε_i (galat) bebas satu sama lain, mempunyai nilai tengah nol, ragam kontan, dan mengikuti sebaran normal, yang lebih umum ditulis $\varepsilon_i \sim NID(0, \sigma^2)$ .Apabila kenormalan data , kehomoginan ragam dan linieritas tak dipenuhi , maka dapat dilakukan transformasi terhadap variabel respon. Salah satu cara untuk mengatsi kehomoginan ragam dengan menggunakan transformasi Box Cox, yaitu transformasi pangkat berparameter tunggal, katakanlah λ terhadap Y, yang menjadi Y^{λ} . Pendugaan parameter λ dapat dicari dengan menggunakan

Metode Kemungkinan Maksimum (aximum Likehood Methods). λ dipilih sedemikian, sehingga didapat jumlah kuadrat sisaan yang paling minimum.

Dalam penulisan ini, diambil contoh data tentang tingkat plasma pada polyamine yang diambil dari sampel 25 anak sehat yang berumur 0 (baru lahir), 1 th,2 th 3 th dan 4 th, penelitian dilakukan untuk mengetahui model hubungan antara umur dan tingkat plasma, dan analisisnya dibantu dengan menggunakan sofware SPSS versi 10.

2. METODE KEMUNGKINAN MAKSIMUM

Metode kemungkinan maksimum merupakan metode untuk memperoleh estimator.Misalkan X variabel random dengan distribusi probabilitas f(x,θ) , dimana parameter θ tidak diketahui, maka fungsi kemungkinan maksimumnya :

$L//(\theta /X) = \prod_{i}^{n} f(x_i, \theta)$

Pada model regresi , metode kemungkinan maksimum adalah sebagai berikut :

Pandang model regresi dalam matriks :

$$Y = X\beta + \varepsilon \quad ; \varepsilon \sim NID(0, \sigma^2)$$

X Fixed & β konstan , sehingga Var (Y) = σ^2

Dalam ragresi linier sederhana, fungsi kemungkinan maksimum dapat dituliskan :

$$L(Y_i, x_i, \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2} \sum_{1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2\}$$

(1)

Untuk menentukan dugaan dari $\beta_0 dan \beta_1$ dan σ^2 , yaitu b_0 , b_1 dan $\hat{\sigma}^2$, maka persamaan (1) eqivalent :

Ln

$$(Y_i, X_i, \beta_0, \beta_1, \sigma^2) = -(\frac{n}{2}) \ln 2\pi - (\frac{n}{2}) \ln \sigma^2 - (\frac{1}{2\sigma^2}) \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2\}$$

$$\frac{\partial \ln L}{\partial \beta_0} = 0 \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0$$

(2)

$$\frac{\partial \ln L}{\partial \beta_1} = 0 \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i = 0 \text{, dan}$$

(3)

$$\frac{\partial \ln L}{\partial \sigma^2} = 0 \Rightarrow -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 = 0$$

(4)

Penyelesaian persamaan (2) , (3) dan (4) didapat :

$b_0 = \bar{Y} - b_1 \bar{X}$

$$b_1 = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - b_o - b_1)^2}{n}$$

b_0 dan b_1 adalah intersep dan slope, $\hat{\sigma}$ adalah standard error dari regresi.

Secara analog pada model $Y = X\beta + \varepsilon$

$$\text{Ln } L = -(\frac{n}{2})\ln 2\pi - (\frac{n}{2})\ln \sigma^2 - (\frac{1}{2\sigma^2})(Y - X\beta)'(Y - X\beta)$$

$$\frac{\partial \ln L}{\partial \beta i} = 0 \Rightarrow b = (X'X)^{-1}(X'Y)$$

3. TRANSFORMASI BOX COX

Transformasi Box Cox adalah transformasi pangkat pada respon. Box Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu λ yang dipangkatkan pada variabel respon Y, sehingga transformasinya menjadi Y^{λ}, λ adalah parameter yang perlu diduga. Tabel dibawah adalah beberapa nilai λ dengan transformasinya .

Tabel Nilai λ dan Taransformasinya

λ	Transformasi
2	Y^2
0.5	$\sqrt{Y}$
0	log Y / ln Y
-0.5	$1/\sqrt{Y}$
-1.0	1 / Y

Prosedure Box Cox adalah secara simultan, menduga λ dalam model :

$W= X\beta + \varepsilon$; $\varepsilon_i \sim NID(0, \tau^2)$

$W=(w_1 ,w_2 ,\ldots\ldots, w_n)^T$

Menurut Drapers S, Harry S , 1992 W didefinisikan :

$$W = \begin{cases} (Y^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \ln Y, & \lambda = 0 \end{cases}$$

(5)

Pendugaan parameter λ dapat dicari dengan menggunakan metode kemungkinan maksimum . Dari model $W= X\beta + \varepsilon$, maka

$$L(\beta, \lambda, \tau^2) = = (2\pi\tau^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\tau^2}(w_i - \beta_0 - \beta_1 x_i)^2\}$$

$$\text{Ln } L= -(\frac{n}{2})\ln 2\pi - (\frac{n}{2})\ln\sigma^2 - (\frac{1}{2\tau^2})\sum_{i=1}^{n} (W_i - \beta_0 - \beta_1 X_i)^2\}$$

$$\text{Ln } L = -(\frac{n}{2})\ln 2\pi - (\frac{n}{2})\ln\sigma^2 - (\frac{1}{2\tau^2})(W - X\beta)'(W - X\beta)$$

$$\text{L maks } (\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2(\lambda) + \ln J(\lambda, Y)$$

n adalah banyaknya amatan , dan $\hat{\sigma}^2$ (λ) adalah 1/n x JKS setelah menduga model regresi dengan λ yang ditentukan.

$$J(\lambda, Y) = \prod_1^n \frac{\partial W_i}{\partial Y} = \prod_1^n Y_i^{\lambda - 1}$$, untuk semua λ

(6)

ln $J(\lambda,Y) = (\lambda-1)\sum_{1}^{n} \ln Y_i$, sehingga L maks(

$\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2(\lambda) + (\lambda-1)\sum \ln Y_i$

Jika kita reduksi terhadap konstante, maka L maks($\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2(\lambda)$

Sehingga memaksimalkan dengan nilai λ yang ditetapkan adalah identik dengan meminimalkan $\hat{\sigma}^2$,yaitu meminimalkan dari Jumlah kuadrat Sesatan(JKS) yang diperoleh dari pengepasan model regresi.

Dari uraian tentang metode Box Cox diatas, maka dapat disimpulkan langkah-langkah untuk menentukan λ . Langkah-langkah tersebut adalah sebagai berikut :

1. Pilih λ dari kisaran yang ditetapkan (biasanya (-2,2) atau (-1,1)
2. untuk masing-masing λ, buat model W= X $\beta + \varepsilon$, W adalah seperti ditetapkan pada persamaan (5) . Dari model regresi ini didapat JKS, langkah 2 ini dilakukan terus menerus pada setiap λ yang ditetapkan sehingga diperoleh beberapa JKS.
3. Plot antara JKS dan λ
4. Pilih sebagai penduga λ adalah λ yang meminumkan JKS.

Bentuk alternatif lain, yang lebih disukai pemakai adalah :

$$V_i = \frac{W_i}{\{J(\lambda,Y)\}^{\frac{1}{n}}} \qquad (7)$$

dan L maks (λ) = $= -\frac{n}{2}\ln\hat{\sigma}^2(\lambda,V)$

dengan V = (V1,V2,V3,…..)' dan $\hat{\sigma}^2(\lambda,V)$ =S(λ,V)/n adalah jumlah kuadrat sisa yang diperoleh berdasarkan model V= X $\beta + \varepsilon$. Singkatnya adalah meminumkan fungsi S(λ,V) Dengan menggunakan persamaan (6) , maka persamaan (7) menjadi :

$$V_i = \frac{W_i}{\hat{Y}^{\lambda-1}} \text{ , dimana } \hat{Y} = (Y1, Y2, \ldots.)^{\frac{1}{n}}$$

Suatu pendekatan 100(1-α) % selang kepercayaan untuk λ dapat dihitung :

$$JKS^{*} = JKS(\lambda) \{ 1 + \frac{t^2 \frac{\alpha}{2}, v)}{v} \}$$

(8)

JKS(λ) adalah JKS minimum dan v aadalah derajad bebas dari galat.Dengan membuat grafik, dengan sumbu x adalah harga-harga λ dan sumbu y adalah harga-harga JKS, akan diperoleh suatu kurva dan dengan menarik garis horizontal pada nilai JKS * pada sumbu y ,maka akan diperoleh batas –batas selang untuk λ yang sesuai.

CONTOH TERAPAN

Data

Suatu penelitian dilakukan untuk menentukan model hubungan antara umur (X) dan tingkat plasma pada Polyamine (Y), data diambil pada anak balita yang sehat, sejumlah 25 anak yang berumur 0 (baru lahir) , 1 th , 2 th , 3 th dan 4 th., masing –masing diambil 5 anak.Data sebagai berikut :

Y	13.44	12.84	11.91	20.09	15.60	10.11	11.38	10.28	8.86	8.59	9.83	9.00	
X	0	0	0	0	0	1	1	1	1	1	2	2	
Y	8.65	7.85	8.88	7.94	6.01	5.14	6.90	6.77	4.86	5.10	5.67	5.75	6.23
X	2	2	2	3	3	3	3	3	4	4	4	4	4

Sumber data : Neter J, 1990

Hasil dan Pembahasan

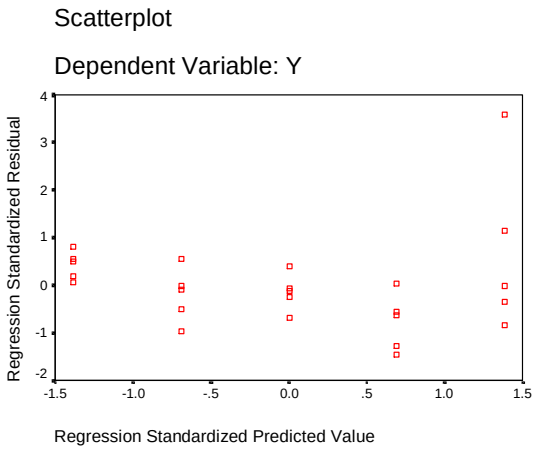
Dari data tersebut diatas dapat kita definisikan bahwa sebagai peubah respon adalah tingkat plasma pada Polyamine (Y) dan sebagai peubah penjelas adalah Umur (X), banyaknya data adalah 25 , Karena hanya ada 1 peubah penjelas, maka model yang digunakan :

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

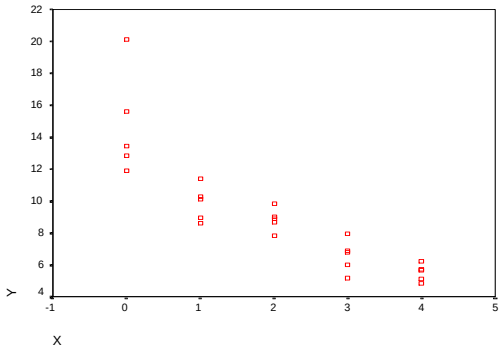
Dengan bantuan SPSS, didapat model di atas :

$$Y = 13{,}475 - 2{,}182\ X\ ,\ R^2 \text{ adj} = 0{,}743$$

Dari model tersebut, dilakukan pengecekan asumsi[2] nya.



Gambar 1: Predict Value terhadap Standard Residual



Gambar 2 : Plot antara Y dan X

Ternyata dari gambar 1 di atas asumsi kehomoginan ragam tak dipenuhi , artinya ada pola tertentu (tidak acak), dilihat gambar plot Y dan X (gambar 2) juga menunjukkan curve yang tak linier dan dengan uji kolmogorow _smirnov didapat sig=0.016 < 0.05 artinya kenormalan juga tak dipenuhi, Sehingga perlu dilakukan ditransformasi terhadap peubah respon Y, yaitu Y^{λ} dengan langkah –langkah sebagai berikut :

1. λ diambil range (2,-2)
2. Hitung $\hat{Y}$ = (Y1 x Y2 x Yn25)$^{1/25}$

3. Untuk tiap harga λ , hitung $\hat{Y}^{\lambda-1}$
4. Hitung $V_i = \frac{W_i}{\hat{Y}^{\lambda-1}}$
5. Regresikan antara V dan X , sehingga didapat JKS
6. Lanjutkan pada nilai λ yang lain
7. Tentukkan λ yang mempunyai JKS terkecil.

Dari langkah-langkah diatas didapat nilai nilai λ dan JKS sebagai berikut :

λ	2	1	0.9	0.7	0.5	0.3	0.1	0	-0.1	
JKS	1278459	77.983	70.348	57.801	48.391	41.424	36.444	34.519	33.101	
λ	-0.3	-0.4	**-0.5**	-0.6	-0.7	-0.9	-1	-1.2	-1.5	-2
JKS	31.223	30.732	**30.621**	30.721	31.104	32.703	33.909	37.118	44.014	673000000

Dari nilai –nilai λ tersebut diatas, dapat dilihat bahwa λ = -0.5 mempunyai JKS paling kecil , sehingga tranformasi yang digunakan adalah $Y^{-0.5}$, artinya data awal Y dipangkatkan dengan -0.5 yang diberi simbul dengan Y2, kemudian Y2 dengan X diregresikan dan model yang didapat adalah sebagai berikut :

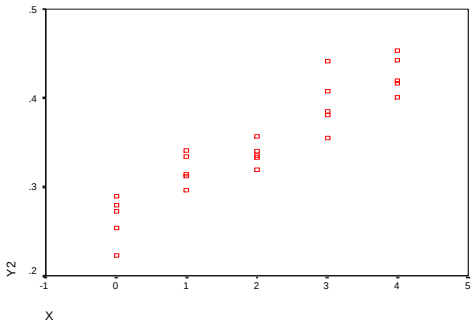
$\hat{Y}' = 0.268 + 0.04\ X$

R^2=0.866, R^2_{adj} =0.861

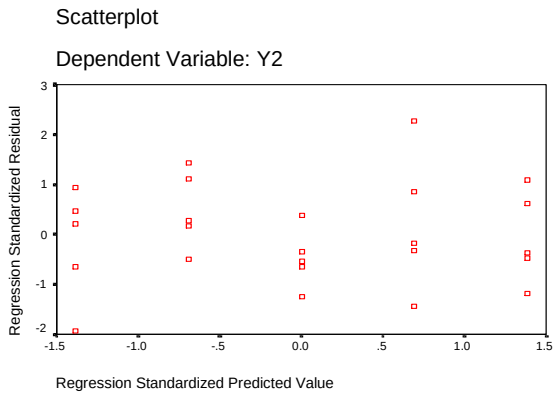
F= 149.221,sig=0.000

Dengan sig =0.00 berarti model tersebut sangat significant, kemudian dilakukan pengecekan terhadap asumsi-asumsi :

Gambar di bawah ini adalah Gambar –gambar setelah Y ditranformasi diperoleh sbb:



Gambar 3: Plot Y2 dan X



Gambar 4 : Predict Value terhadap Standard Residual setelah ditranformasi

Dilihat dari gambar diatas, maka asumsi –asumsi yang semula tidak dipenuhi, dengan melakukan tranformasi Boc Cox menjadi dipenuhii, yaitu linieritas (gambar 3), kehomoginan ragam(gambar 4) , kenormalan (sig= 0.200 >0.05). Sehingga model yang cocok untuk hubungan tingkat plasma pada polyamine dan umur adalah $\hat{Y}'$ **= 0.268 + 0.04 X .** Dari model ini dapat digunakan untuk menentukan bila X =5 , maka $\hat{Y}'$ = 0.468, karena $\hat{Y}' = \frac{1}{\sqrt{\hat{Y}}}$, maka $\hat{Y} = \frac{1}{\hat{Y}'^2}$, sehingga untuk X=5 , $\hat{Y}$ = 4.566 , artinya jika anak berumur 5 tahun, maka rata-rata tingkat plasma pada polyamine adalah 4,566.

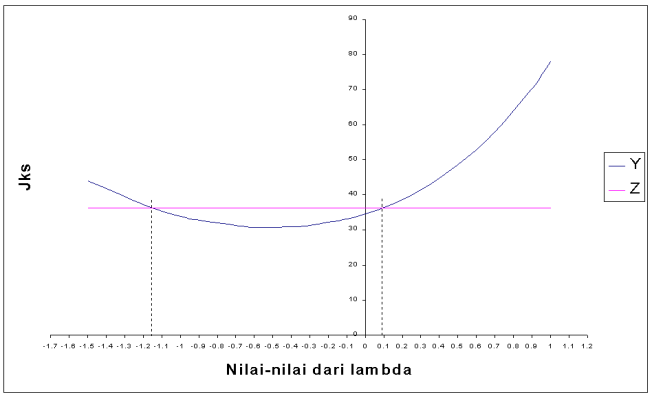
Dari nilai – nilai λ diatas, dihitung Selang kepercayaan dengan tingkat kesalahan 5 % sebagai berikut :

$$\text{Dari persamaan (8) : } JKS^* = JKS(\lambda)\left(1 + \frac{t^2_{0.25,23}}{23}\right)$$

$$= 30.621\left(! + \frac{(2.07)^2}{23}\right)$$

$$= 36{,}326$$

Dengan menarik garis horizontal pada nilai 36,326 pada sumbu y ' maka akan diperoleh nilai bawah dan atas untuk λ . Sehingga SK untuk λ adalah -1,16<λ <0.095



Gambar 5 : Plot λ dan JKS

4. KESIMPULAN

Transformasi Box COX adalah pendugaan parameter λ yang dipangkatkan pada variabel Y sehingga mendapat JKS yang minimum. Jika didapat JKS minimum dengan λ=1, maka data tak perlu dilakukan transformasi karena sudah terjadi kebebasan antara ragam respon dengan rata- rata respon. Dengan transformasi Box Cox secara simultan dapat dicapai kenormalan dari sebaran, kekonstanan ragam dari galat dan linieritas dari struktur model.

Bila dalam menduga parameter λ kurang tepat, maka akan dicirikan oleh lebarnya selang kepercayaan.

DAFTAR PUSTAKA

Drapper, NR and Harry Smith, S, *Analisis Regresi Terapan,* edisi kedua, Gramedia Pustaka Utama, Jakarta,1992

Montgomery DC and Elizabeth A.P, *Introduction to linier Regression Analysis* Second Edition, John Wiley & Sons, New York,1982

Neter, J and W.Wasserman, *Apllied Linier Statistical Models*, Richard D, Irwin, Japan,1994

Weisberg S, *Apllied Linier Regression,* Second edition, John Wiley & Sons, New York,1985

JURNAL MATEMATIKA DAN KOMPUTER
Vol. 7. No. 3, 8 - 17, Desember 2004, ISSN : 1410-8518