TRANSFORMASI BOX COX
DALAM ANALISIS REGRESI LINIER SEDERHANA

Welly Fransiska[1], Sigit Nugroho[2], dan fachri Faisal[2]
[1]Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu
[2]Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

ABSTRAK

Penelitian ini bertujuan untuk mempelajari bagaimana mencari parameter λ dengan menggunakan Metode Kemungkinan Maksimum, mengkaji Transformasi Box Cox dalam peubah respon Y, serta untuk mengetahui apakah model yang diperoleh setelah transformasi memenuhi asumsi normalitas pada model regresi. Metode penelitian yang digunakan adalah studi literatur dan teladan penerapan. Transformasi Box Cox adalah transformasi pangkat pada respon Y. Box Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu lamda yang dipangkatkan pada variabel respon Y. Penelitian ini menggunakan teladan penerapan yang diperoleh dari buku Neter (1994) dan juga situs internet. Pengolahan data dalam penelitian ini menggunakan bantuan SPSS dan Excel. Berdasarkan teladan penerapan yang digunakan diperoleh hasil bahwa, untuk teladan 1 dan 2 asumsi kenormalan terpenuhi setelah transformasi dilakukan, transformasi yang digunakan pada teladan 1 adalah $\sqrt{Y}$, koefisien determinasi mengalami peningkatan dari 75,17% menjadi 86,4%. Sedangkan transformasi yang digunakan pada teladan 2 adalah $Ln\,Y$, serta koefisien determinasi mengalami peningkatan dari 50% menjadi 85%.

Kata kunci : *Transformasi Box Cox, Metode Kemungkinan Maksimum, Asumsi Normalitas, Koefisien determinasi.*

PENDAHULUAN

Analisis regresi merupakan salah satu cabang statistika yang paling banyak dipelajari oleh ilmuwan, baik ilmuwan bidang sosial maupun eksakta. Melalui analisis regresi, model hubungan antar variabel dapat diketahui. Secara umum, model merupakan penyederhanaan dan abstraksi dari keadaan yang sebenarnya. Model juga menolong peneliti dalam menentukan hubungan kausal (sebab akibat) antara dua atau lebih variabel bebas. Variabel dalam analisis regresi dikenal dengan nama variabel terikat (Y) dan variabel penjelas (X), atau juga lebih dikenal dengan variabel bebas (X) dan satu variabel tak bebas (Y) (Sembiring, 2003).

Dalam melakukan analisis regresi ada beberapa asumsi yang harus dipenuhi yaitu model regresi linier, galat menyebar normal dengan rataan nol dan memiliki varian tertentu, homoskedastisitas, artinya varian galat sama untuk setiap periode (homo = sama, skedastisitas = sebaran), tidak ada autokorelasi antar galat (antara μ_i dan μ_j tidak ada korelasinya), tidak terjadi multikolinieritas antar variabel bebas, jumlah observasi *n* harus lebih besar daripada jumlah variabel bebas (Naftali, 2007). Pada beberapa kasus, mentransformasi data akan membuat kecocokan model terhadap asumsi menjadi lebih baik. Transformasi data merupakan salah satu usaha untuk memperbaiki asumsi normalitas, linieritas dan homoskedastisitas. Analisis dengan data hasil transformasi masih tetap sah (Kutner *et. al*., 2005). Transformasi yang ideal harus memenuhi beberapa kriteria antara lain : ragam dari variabel yang baru tidak dipengaruhi oleh perubahan rata-rata, variabel yang baru hendaknya menyebar normal, skala pengukuran variabel yang baru hendaknya sedemikian sehingga pengaruh sesungguhnya bersifat linier dan aditif dan skala pengukuran variabel yang baru hendaknya sedemikian sehingga nilai tengah perhitungan dari contoh merupakan penduga yang efisien terhadap nilai tengah yang sesungguhnya (Draper dan Smith, 1999). Salah satu cara yang dapat digunakan untuk mengatasi asumsi kehomogenan ragam, linieritas dan kenormalan adalah dengan menggunakan Transformasi Box Cox. Transformasi Box Cox yaitu transformasi pangkat berparameter tunggal, katakanlah parameter λ terhadap Y sehingga menjadi Y^{λ}. Pada transformasi ini untuk pendugaan parameter λ akan menggunakan Metode Kemungkinan Maksimum (*Maximum Likelihood Method*).

REGRESI LINIER SEDERHANA

Istilah regresi pertama kali dikemukakan oleh seorang antropolog dan ahli meteorologi terkenal dari Inggris yang bernama Sir Francis Galton pada tahun 1855. Istilah regresi muncul dalam pidatonya di depan *Section H of The British Association* di Aberdeen, 1855, yang dimuat di majalah *Nature* September 1885 dan dalam sebuah makalah *"Regression Towards Mediocrity in Hereditary Stature"*, yang dimuat dalam *Journal of The Anthropological Institute* (Draper dan Smith, 1998). Analisis regresi pada dasarnya adalah studi mengenai ketergantungan satu variabel tak bebas dengan satu atau lebih variabel bebas, tujuannya adalah untuk mengestimasi dan atau memprediksi rata-rata populasi atau nilai rata-rata variabel tak bebas berdasarkan nilai variabel bebas yang diketahui (Gujarati, 2003).

Menurut Draper dan Smith (1998) analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel terhadap variabel lainnya.Regresi sebagai suatu teknik analisis telah dipergunakan secara luas, tidak hanya terbatas dalam bidang statistik namun juga di bidang-bidang lain seperti ekonomi, pertanian, sosial, tehnik riset dan bidang-bidang lainnya. Dalam perkembangannya terdapat dua jenis regresi yang sangat terkenal, yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana digunakan untuk menggambarkan hubungan antara satu variabel bebas *(X)* dengan satu variabel tak bebas *(Y)*, sedangkan jika terdapat lebih dari satu variabel bebas dan variabel bebasnya berpangkat satu maka persamaan regresinya disebut regresi linier berganda. Hubungan antara variabel bebas *(X)* dan variabel tak bebas *(Y)* yang bersifat linier dan sederhana dapat dituliskan sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,2,3,\ldots,n \qquad (1)$$

Parameter β_0 dan β_1 merupakan parameter yang nilainya belum diketahui. Parameter β_0 biasanya dikenal dengan intersep, yaitu jarak dari titik asal (titik 0) ke titik perpotongan antara garis regresi dengan sumbu *Y*. Interprestasi dari β_0 adalah nilai rata-rata dari penduga *Y* jika nilai *X* sama dengan nol. Parameter β_1 merupakan koefisien arah *(slope)* atau koefisien regresi. Parameter β_0 dan β_1 pada persamaan regresi linier diduga dengan $\hat{\beta}_0$ dan $\hat{\beta}_1$. Penduga parameter-parameter tersebut dapat diperoleh dengan menggunakan Metode *Maximum Likelihood.*

2.1 Asumsi-Asumsi dalam Regresi Linier

Setelah didapatkan model regresi, lakukan interpretasi terhadap hasil yang diperoleh. Hal ini disebabkan karena model regresi harus diuji terlebih dahulu apakah sudah memenuhi asumsi klasik. Gauss Markov telah membuktikan penduga dalam regresi mempunyai sifat *BLUE (Best Linier Unbiased).*

Beberapa asumsi yang harus dipenuhi dalam analisis regresi adalah:

a. Model regresi linier, artinya linier dalam parameter.
b. Galat menyebar normal dengan rataan nol dan memiliki suatu varian tertentu.
c. Homoskedastisitas, artinya varian kesalahan sama untuk setiap periode (homo = sama, skedastisitas = sebaran) dalam bentuk matematis: $Var\ (\mu/Xi) = 0$
d. Tidak ada autokorelasi antar kesalahan (antara μ_i dan μ_j tidak ada korelasinya).
e. Tidak terjadi multikolinieritas.
f. Jumlah observasi *n* harus lebih besar daripada jumlah parameter yang diestimasi (jumlah variabel bebas)

Apabila ada satu syarat saja yang tidak terpenuhi, maka hasil analisis regresi tidak dapat dikatakan bersifat *BLUE* (Naftali, 2007).

Metode *Maximum Likelihood*

Penaksiran *Maximum Likelihood* merupakan salah satu pendekatan terpenting pada penaksiran dalam semua statistik inferensia yang diperkenalkan pertama kali oleh Ronald Fisher pada tahun 1920 (Wannacott, 1990). Bila diketahui pengamatan bebas $(X_1, X_2, \dots, X_n)$ dari fungsi kepekatan peluang (kasus kontinu) dan fungsi masa peluang (kasus diskrit) $f(X,\theta)$, maka penaksir *Maximum Likelihood* $\hat{\theta}$ ialah yang memaksimumkan fungsi Kemungkinan *Maximum Likelihood* yaitu:

$$L(X_1, X_2, \dots, X_n; \theta) = f(X_1;\theta) \cdot f(X_2;\theta) \cdot \cdots \cdot f(X_n;\theta) \tag{2}$$

(Walpole and Myers.1995).

Dalam regresi linier sederhana, fungsi kemungkinan maksimum dapat dituliskan :

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

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

untuk menentukan dugaan dari β_0 dan β_1 dan σ^2 , yaitu b_0 , b_1 dan $\hat{\sigma}^2$, maka persamaan (2.12) ekuivalen :

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

$$\frac{\partial \ln L}{\partial \beta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0 \tag{3}$$

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)X_i = 0 \tag{4}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2 = 0 \tag{5}$$

penyelesaian persamaan (3) , (4) dan (5) didapat penduga dari β_0, β_1 dan σ berturut-turut adalah b_0, b_1 dan $\hat{\sigma}^2$:

$$b_0 = \bar{Y} - b_1 \bar{X} \tag{6}$$

$$b_1 = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \tag{7}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - b_o - b_1 X)^2}{n} \tag{8}$$

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

TRANSFORMASI BOX-COX

a. Pengertian Transformasi Box-Cox

Transformasi Box-Cox adalah transformasi pangkat pada respon. Transformasi Box-Cox dibahas oleh Box-Cox dalam makalah mereka pada tahun 1964. 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 1 Nilai λ dan Transformasinya.

λ	*Transformasi*
2	Y^2
1,0	Y^1
0,5	$\sqrt{Y}$
0	ln Y
-0,5	$1/\sqrt{Y}$
-1,0	1 / Y

Menurut Drapers dan Smith (1992) transformasi Box-Cox didefinisikan :

$$W = \begin{cases} (Y^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \ln Y, & \lambda = 0 \end{cases} \tag{9}$$

Transformasi yang kontinu ini bergantung pada satu parameter yaitu λ.

b. Metode kemungkinan maksimum Likelihood dalam Transformasi Box Cox

Salah satu cara untuk menduga parameter λ pada persamaan (1) adalah dengan menggunakan *Maximum Likelihood Method* dengan asumsi $\varepsilon \sim$ N(0, σ^2) untuk pilihan λ yang sesuai. Fungsi Kepekatan Peluang W didefinsikan sebagai berikut :

$$f(w|\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(W_i - \beta_0 - \beta_1 X_i)^2\right\} \tag{10}$$

Oleh karena itu fungsi kepekatan peluang Y menjadi:

$$f(y|\beta, \sigma^2, \lambda) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(W - \beta_0 - \beta_1 X_i)^2\right\} \prod_{i=1}^{n} y_i^{\lambda-1} \tag{11}$$

dimana

$$J(\lambda, Y) = \prod_{1}^{n} \frac{\partial W_i}{\partial Y} = \prod_{1}^{n} Y_i^{\lambda-1} \quad \text{untuk semua } \lambda \tag{12}$$

Berdasarkan model W= X $\beta + \varepsilon$ dan persamaan (10) diperoleh fungsi kemungkinan maksimumnya yaitu:

$$(\ln(L(\beta, \sigma^2, \lambda|y)) = \ln[(2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}(W_i - \beta_0 - \beta_1 X_i)^2\} \prod_{i=1}^{n} y_i^{\lambda-1}]$$

$$= -(\frac{n}{2})\ln 2\pi - (\frac{n}{2})\ln\sigma^2 - (\frac{1}{2\sigma^2})\sum_{i=1}^{n} (W_i - \beta_0 - \beta_1 X_i)^2\} + (\lambda - 1)\sum_{i=1}^{n} \ln y_i \tag{13}$$

Persamaan (13) menyatakan fungsi maksimum log-likelihood secara parsial yang hanya tergantung pada λ, untuk σ^2 diduga dengan $\hat{\sigma}^2(\lambda)$ sehingga diperoleh :

$$\text{L maks}(\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2(\lambda) + (\lambda-1)\sum_{i=1}^{n}\ln Y_i \tag{14}$$

dengan :

n = banyak amatan

$\hat{\sigma}^2(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i(\lambda) - \bar{y}(\lambda))^2$

$\bar{y}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} y_i(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{\lambda}-1}{\lambda}\right)$

Dengan kata lain, memaksimalkan nilai λ yang ditetapkan adalah sama dengan meminimalkan $\hat{\sigma}^2$,yaitu meminimalkan Jumlah kuadrat Galat (SSE).

c. Metode Kemungkinan Maksimum Likelihood untuk pendugaan λ

Adapun Langkah-langkah untuk menentukan λ. Langkah-langkah tersebut adalah sebagai berikut :

a) Pilih λ dari kisaran yang ditetapkan (biasanya diambil λ dari kisaran (-2,2) atau bahkan (-1,1) pada mulanya, dan kemudian memperlebar kisaran bila diperlukan).
b) Transformasikan variabel Y menggunakan persamaan (1) dan hitung $\hat{\sigma}^2$. Untuk λ yang terpilih hitunglah :

$$L_{\text{maks}}(\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2(\lambda) + (\lambda-1)\sum_{i=1}^{n}\ln Y_i$$

c) Setelah menghitung $L_{\text{maks}}(\lambda)$ untuk beberapa nilai λ dalam kisaran yang ditentukan, pasangkan $L_{\text{maks}}(\lambda)$ terhadap λ yang memaksimumkan $L_{\text{maks}}(\lambda)$. Inilah penduga kemungkinan maksimum $\hat{\lambda}$ *(maximum likelihood estimator)* bagi parameter λ.

Biasanya, nilai λ ini tidak dipergunakan dalam perhitungan selanjutnya, namun menggunakan salah satu nilai dalam barisan …,-2, -1$\frac{1}{2}$, -1, 0, $\frac{1}{2}$, 1, 1$\frac{1}{2}$,2,…yang paling dekat dengan nilai dugaan kemungkinan maksimum tersebut, tentu saja setelah memeriksa apakah nilai ini berada dalam kisaran yang ditentukan. Misal jika $\hat{\lambda}$ = 0.11, mungkin akan digunakan λ= 0, jika $\hat{\lambda}$ = 0.94 mungkin akan digunakan λ = 1 (namun pemilihan λ bisa menggunakan $\hat{\lambda}$ atau membulatkan sampai nilai perempatan terdekat).

d. Selang Kepercayaan bagi λ

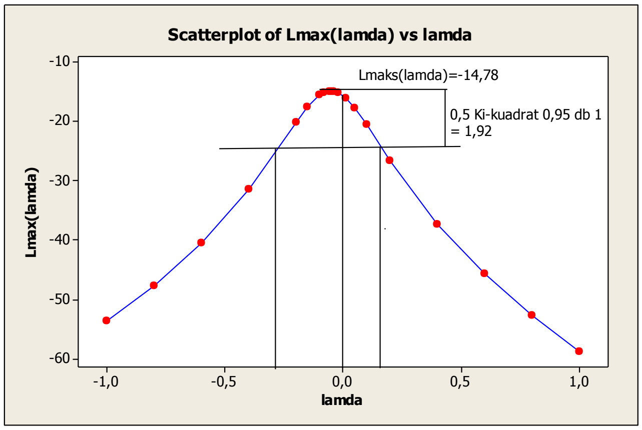
Suatu selang kepercayaan untuk λ merupakan himpunan nilai-nilai λ yang memenuhi pertidaksamaan :

$$L_{maks}(\hat{\lambda}) - L_{maks}(\lambda) \le 0{,}5\,\chi^2_{1,1-\alpha}$$

Dimana $\chi^2_{1;(1-\alpha)}$ adalah titik persentase sebaran Khi-Kuadrat dengan satu derajat bebas yang luas wilayah disebelah kananya sebesar α. Pertidaksamaan diatas dapat digambarkan dengan menarik garis mendatar setinggi

$$L_{maks}(\hat{\lambda}) - \frac{1}{2}\chi^2_{1;(1-\alpha)}$$

Berikut visualisasi dari pertidaksamaan diatas:



gambar 1 menyatakan posisi selang kepercayaan 95% untuk λ pada tebaran L_{maks} dengan λ. Garis ini memotong kurva pada 2 nilai λ yang merupakan titik-titik ujung selang kepercayaan λ.

HASIL DAN PEMBAHASAN

Teladan yang diambil berdistribusi tidak normal, hal ini dilakukan untuk melihat apakah setelah data ditransformasi, asumsi kenormalan data yang merupakan salah satu alasan dilakukannya transformasi terpenuhi atau tidak. Model yang dibentuk adalah model regresi linier sederhana yang terdiri dari variabel bebas *(X)* dan variabel tak bebas *(Y)*. Data yang dijadikan teladan akan dianalisis kenormalannya, selanjutnya data ditranformasi dengan menggunakan transformasi Box Cox.

Teladan 1

Suatu penelitian dilakukan untuk menentukan model hubungan antara umur *(X)* dan tingkat plasma pada Polyamine *(Y)*, dengan data sebagai berikut :

Tabel 2 Hubungan antara umur *(X)* dan tingkat plasma pada Polyamine *(Y)*

X	*Y*	*X*	*Y*
0,00	13,44	2,00	7,85
0,00	12,84	2,00	8,88
0,00	11,91	3,00	7,94
0,00	20,09	3,00	6,01
0,00	15,60	3,00	5,14
1,00	10,11	3,00	6,90
1,00	11,38	3,00	6,77
1,00	10,28	4,00	4,86
1,00	8,86	4,00	5,10
1,00	8,59	4,00	5,67
2,00	9,83	4,00	5,75
2,00	9,00	4,00	6,23
2,00	8,65		

Sumber data : Kutner, 2005

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. Dari data tersebut diatas dapat didefinisikan bahwa sebagai peubah respon adalah tingkat plasma pada Polyamine *(Y)* dan sebagai peubah penjelas adalah umur *(X)*, banyaknya data adalah 25, dengan Metode kemungkinan maksimum Likelihood diperoleh :

$$b_1 = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{-109}{50} = -2{,}18$$

$$b_0 = \bar{Y} - b_1\bar{X} = 9{,}1072\text{-}(\text{-}2{,}18).(2) = 13{,}467$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2}{n}$$

$$= \frac{(13{,}14 - 13{,}467 - ((-2{,}18).(0)))^2 + (12{,}84 - 13{,}467 - ((-2{,}18).(0)))^2 \ldots (6{,}23 - 13{,}467 - ((-2{,}18).(4)))^2}{25}$$

$$= \frac{78{,}45}{25} = 3{,}13$$

Sehingga model regresi yang didapatkan adalah: $Y_i = 13{,}647 - 2{,}180\ X_i$. Selanjutnya akan dihitung koefisien determinasi : $R^2 = 0{,}7517$. Artinya sebesar 75,17% dari seluruh variasi total *Y* diterangkan oleh *X*, dan 24,83% diterangkan oleh faktor lainnya. Dengan demikian model regresi sudah cukup baik karena variabel *X* dapat menerangkan variabel cukup baik karena makin dekat R^2 dengan 1 makin baik kecocokan data dengan model.

Pengujian Asumsi kenormalan

Berdasarkan model tersebut, dilakukan pengecekan asumsi kenormalannya dengan bantuan SPSS. Ketentuan analisisnya adalah jika variabel (bebas dan terikat) memiliki Z hitung < Z tabel maka data tersebut berdistribusi normal, dalam hal ini digunakan Z tabel yaitu 1,96, pada alpha 5% didapat Z tabel 1,96. Dibandingkan dengan $Z\alpha 3$ ($Skewness$) dan $Z\alpha 4$ ($Kurtosis$) dari variabel X dan Y disimpulkan bahwa data pada teladan 1 tidak berdistribusi normal karena Z hitung < Z tabel dikatakan bahwa asumsi kenormalan tidak dipenuhi.

Transformasi Box Cox

Oleh karena asumsi kenormalan tidak terpenuhi maka dilakukan transformasi terhadap variabel respon *Y* dengan menggunakan transformasi Box Cox. Setelah melakukan langkah-langkah transformasi Box Cox didapatkan L*maks* untuk λ pada kisaran (-2,2) untuk teladan 1 adalah sebagai berikut :

Tabel 3 Lmaks untuk λ pada kisaran (-2,2)

λ	-2	-1,75	-1,5	-1,25	-1	-0,75	-0,5	-0,25
L_{maks}	-31,19	-30,2	-29,34	-28,65	-28,13	-27,79	-27,65	-27,72

λ	0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
L_{maks}	-28,02	-28,57	-29,36	-30,4	-31,71	-33,27	-35,09	-37,16	-39,46

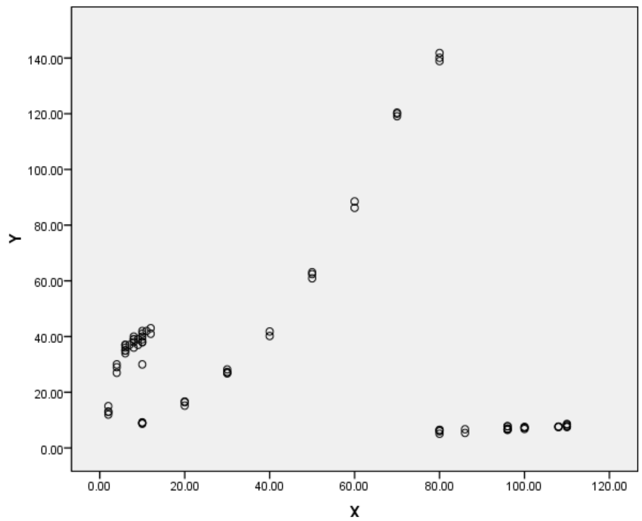
Dari nilai–nilai λ diatas, dapat dilihat bahwa λ = -0.5 menghasilkan nilai l_{maks} paling maksimum pada kisaran (2,-2) yaitu sebesar -27,65 sehingga tranformasi yang digunakan adalah $Y^{-0.5}$, artinya data awal *Y* dipangkatkan dengan -0.5 yang diberi simbol dengan *Y2*, sehingga didapat data baru setelah transformasi yaitu :

Tabel 4 Hubungan antara variabel *X* dan variabel *Y*

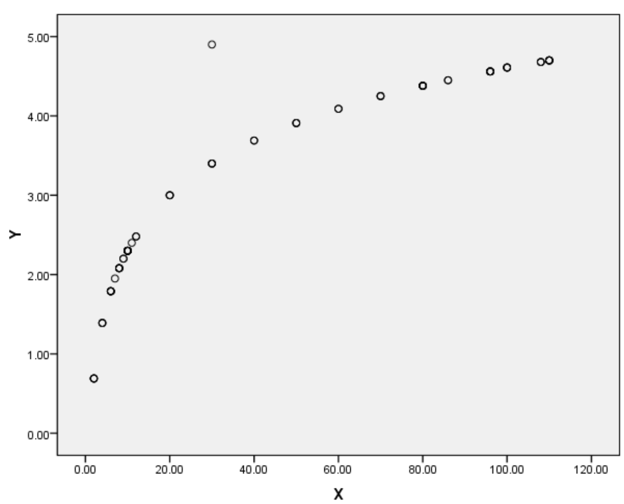
X	*Y2*
0,00	0,272772
0,00	0,279073
0,00	0,289764
0,00	0,223105
0,00	0,253185
1,00	0,314503
1,00	0,296435
1,00	0,311891
1,00	0,335957
1,00	0,341196
2,00	0,31895
2,00	0,333333
2,00	0,34001

X	*Y2*
2,00	0,356915
2,00	0,335578
3,00	0,354887
3,00	0,407909
3,00	0,441081
3,00	0,380693
3,00	0,384331
4,00	0,453609
4,00	0,442807
4,00	0,419961
4,00	0,417029
4,00	0,400642

Setelah didapatkan data baru maka akan dilakukan perhitungan untuk menduga model yaitu dengan menggunakan metode kemungkinan maksimum likelihood, didapatkan: $Y_i = 0{,}268 + 0{,}04X_i$. Sehingga diperoleh nilai $R^2 = 0{,}864 = 86{,}4\%$, dengan SPSS dapat di katakan bahwa asumsi kenormalan terpenuhi karena nilai Skewness dan Kurtosis setelah data ditransformasi adalah kurang dari Z tabel (dengan alpha 5% didapat Z tabel 1,96). Berdasarkan model yang didapat, untuk menentukan bila X =5, maka $\hat{Y}'$ = 0.468, selain itu untuk nilai determinasinya juga mengalami peningkatan yaitu dari 75,17% setelah ditransformasi menjadi 86,4% yang artinya model baru yang didapatkan lebih baik karena dapat menerangkan 86,4% dari variabel Y serta data yang semula tidak memenuhi asumsi kenormalan setelah ditransformasi juga memenuhi asumsi kenormalan, artinya transformasi yang dilakukan dapat memperbaiki model yang ada. Gambar berikut adalah scatter plot data sebelum dan sesudah transformasi data dilakukan.



Gambar 2 Scatterrplot antara variabel X dan Y sebelum transformasi



Gambar 3 Scatter plot antara variabel X dan Y setelah transformasi

Teladan 2

Teladan ini terdiri dari 79 data yang berasal dari sebuah penelitian, di mana data terdiri atas 2 variabel yaitu varianbel bebas *X* dan variabel terikat Y

Tabel 5 Hubungan antara variabel *X* dan *Y*

X	Y	X	Y	X	Y	X	Y
2	15	80	6	110	7,9	70	119,1
2	12	86	5,4	10	9,2	70	120,4
2	13	86	6,7	10	8,7	80	141,8
2	13	96	6,5	10	9	80	140,1
4	27	96	7	10	8,9	80	138,9
4	29	96	6,7	20	16,4	9	37
4	30	96	7,8	20	15,2	12	41
6	37	96	6,6	20	16,7	6	34
6	37	96	7,8	30	27,3	10	39
6	36	100	7,2	30	28,2	9	39
6	35	100	7,5	30	26,8	10	40
8	38	100	6,8	30	27	7	37
8	36	100	7,5	40	41,8	8	39
8	39	108	7,7	40	40,2	11	42
10	30	108	7,6	50	62,4	6	35
10	38	108	7,5	50	63,1	10	41
10	42	110	7,7	50	60,9	8	40
80	5,1	110	8,6	60	88,5	12	43
80	6,5	110	7,5	60	86,2	10	38
80	6,4	110	8,4	70	120		

(Anonim, 2004)

dengan Metode kemungkinan maksimum Likelihood akan dicari model untuk teladan 2, dan dengan cara yang sama pada teladan 1 didapatkan :

$$b_1 = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{-8185{,}05}{127141{,}22} = -0{,}06$$

$b_0 = \bar{Y} - b_1 \bar{X}$ = 33,4-(0,06).(45,41) =36,33

sehingga model regresi yang didapatkan adalah: $Y_i = 36{,}33 + 0{,}06\, X_i$. Selanjutnya akan dihitung koefisien determinasi : $R^2 = \frac{32280{,}45}{67915{,}} = 0{,}5$

artinya hanya sebesar 50% dari seluruh varians total Y diterangkan oleh X. Makin dekat R^2 dengan 1 makin baik kecocokan data dengan model, untuk mengatasi hal ini akan dilakukan transformasi.

Pengujian Asumsi kenormalan

Berdasarkan model tersebut, dilakukan pengecekan asumsi kenormalan dengan bantuan SPSS, dari program SPSS Didapatkan pada taraf nyata 5% didapat Z tabel 1,96. Dibandingkan dengan $Z\alpha 3(\,Skewness)$ dan $Z\alpha 4(Kurtosis)$ dari variabel X dan Y disimpulkan bahwa data pada teladan 2 tidak berdistribusi normal, dikatakan bahwa asumsi kenormalan tidak dipenuhi.

Transformasi Box Cox

Setelah melakukan langkah-langkah transformasi Box Cox didapatkan Lmaks untuk λ pada kisaran (-2,2) sebagai berikut :

Tabel 6. Lmaks untuk λ pada kisaran (-2,2)

λ	-2	-1,75	-1,5	-1,25	-1	-0,75	-0,5	-0,25
L_{maks}	-306,7	-292,1	-306,7	-267	-256,7	-248,3	-242,1	-238,6

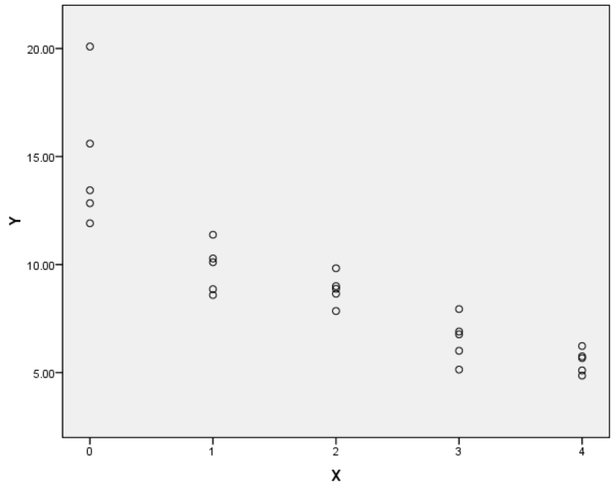
λ	0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
L_{maks}	-25	-241,7	-249,2	-261,1	-276,9	-296,1	-317,8	-341,8	-367,3

Dari nilai–nilai λ tersebut diatas, dapat dilihat bahwa λ = 0 menghasilkan nilai l_{maks} sebesar -25 sehingga tranformasi yang digunakan adalah *Ln Y*, artinya data awal *Y* di *Ln*, sehingga didapat data setelah transformasi yaitu :

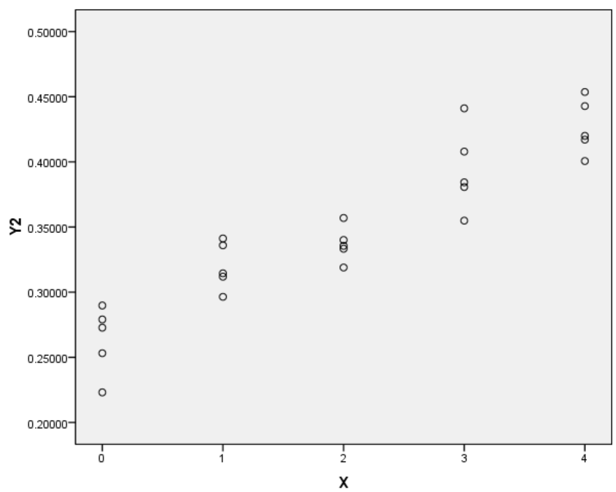
Tabel 7 Hubungan variabel X dan Y2

X	Y2	X	Y2	X	Y2	X	Y2
2	2,71	80	1,79	110	2,07	70	4,78
2	2,48	86	1,69	10	2,22	70	4,79
2	2,56	86	1,90	10	2,16	80	4,95
2	2,56	96	1,87	10	2,20	80	4,94
4	3,30	96	1,95	10	2,19	80	4,93
4	3,37	96	1,90	20	2,80	9	3,61
4	3,40	96	2,05	20	2,72	12	3,71
6	3,61	96	1,89	20	2,82	6	3,53
6	3,61	96	2,05	30	3,31	10	3,66
6	3,58	100	1,97	30	3,34	9	3,66
6	3,56	100	2,01	30	3,29	10	3,69
8	3,64	100	1,92	30	3,30	7	3,61
8	3,58	100	2,01	40	3,73	8	3,66
8	3,66	108	2,04	40	3,69	11	3,74
10	3,40	108	2,03	50	4,13	6	3,56
10	3,64	108	2,01	50	4,14	10	3,71
10	3,74	110	2,04	50	4,11	8	3,69
80	1,63	110	2,15	60	4,48	12	3,76
80	1,87	110	2,01	60	4,46	10	3,64
80	1,86	110	2,13	70	4,79		

Gambar Berikut adalah perbandingan antara data sebelum transformasi dan sesudah transformasi dilakukan.



Gambar 4 Scatter plot antara variabel X dan Y sebelum transformasi data



Gambar 5 Scatter plot antara variabel X dan Y sesudah transformasi data

Setelah didapatkan data baru maka akan dilakukan perhitungan untuk menduga model yaitu dengan menggunakan metode kemungkinan maksimum likelihood, didapatkan: $Y_i = 1{,}87 + 0{,}3\ X_i$, serta didapatkan nilai $R^2 = 0{,}85 = 85\%$.

Berdasarkan model yang didapat, untuk menentukan bila X =10, maka $\hat{Y}'$ = 4,87 karena transformasi yang cocok adalah $Ln\, Y$ maka $\hat{Y} = 1{,}58$, selain itu untuk nilai determinasinya juga mengalami peningkatan yaitu dari 50% setelah ditransformasi menjadi 85% yang artinya model baru yang didapatkan lebih baik karena dapat menerangkan 85% dari variabel Y serta data yang semula tidak memenuhi asumsi kenormalan setelah ditransformasi juga memenuhi asumsi kenormalan, ini menunjukkan bahwa transformasi pada data teladan 2 memang perlu dilakukan.

KESIMPULAN DAN SARAN

a. Kesimpulan

- Transformasi Box Cox adalah teransformasi pangkat pada respon. Box Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu lamda yang dipangkatkan pada variabel respon Y.
- Pada teladan asumsi normalitas pada model regresi yang semula tidak terpenuhi, setelah ditransformasi Box Cox asumsi tersebut terpenuhi.
- Berdasarkan teladan penerapan yang digunakan diperoleh hasil bahwa, untuk teladan 1 dan 2 asumsi kenormalan terpenuhi setelah transformasi dilakukan, transformasi yang digunakan pada teladan 1 adalah $\sqrt{Y}$, koefisien determinasi mengalami peningkatan dari 75,17% menjadi 86,4%. Sedangkan transformasi yang digunakan pada teladan 2 adalah $Ln\,Y$, koefisien determinasi mengalami peningkatan dari 50% menjadi 85%.

b. Saran

Transformasi Box Cox dapat digunakan dalam analisis regresi linier sederhana, analisis regresi linier berganda, dan multivarat. Dengan demikian, penelitian ini dapat dilanjutkan dengan membahas Transformasi Box Cox dalam analisis regresi berganda dan Multivariat.

DAFTAR PUSTAKA

Anonim. 2004. *Bab C Linieritas Homoskedastisitas*.www.Dali.staff.gunadarma.ac.ai. 1 Agustus 2010

Draper, NR and H. Smith, S. 1998. *Analisis Regresi Terapan*. Gramedia Pustaka Utama. Jakarta

Gujarati, D. N. 2003. *Econometric*. Erlangga. Jakarta

Kutner, M. H, C. J. Nachtsheim, J. Neter and W. Li. 2005. *Applied Linier Statistical* McGraw-Hill. New York

Naftali, Y. 2007. *Regresi*. Http://yohanli.wordpress.com. 2 Mei 2010

Nugroho, S. 2008. *Pengantar Statistika Matematika*, edisi 1.Unib Press. Bengkulu

Santoso, S. 2002. *Buku Latihan SPSS Statistik Multivariat*. PT Elex Media Komputindo, Jakarta.

Sembiring, R. K. 2003. *Analisis Regresi*. ITB. Bandung

Walpole, R. E dan Myers, R. H. 1995. *Ilmu Peluang dan Statistik untuk Insinyur dan Ilmuwan*, Terjemahan RK Sembiring, Edisi ke-4. ITB. Bandung

Wannacott, T. H & Wonnacott, R. J. 1990. *Introduction to Statistics.* John Willey & Sons. New York

Wei, W.W.S. 1990. *Time Analysis Univariate and Multivariate Method*. Addison Wesley Publishing Company. Inc