

Pendekatan Baru dari Transformasi Box-Cox untuk Mengatasi Masalah Ketidaknormalan pada Variabel Respon

Dwi Andini Juniawati*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*dwiandini556@gmail.com, nusarhajarisman@unisba.ac.id

Abstract. Regression analysis is used to model between two quantitative variables. When doing regression modeling it is often found that the data is not normally distributed, the way to overcome this is to do a transformation. One of the transformations used is the Box-Cox transformation. Box and Cox consider a single-parameter transformation class, namely the parameter λ with respect to Y so that Y^λ with λ is the parameter to be estimated. Estimation of the parameter λ uses a maximum likelihood estimator. In general, the value of λ is assumed to be known, but some values of λ (called λ_0) the assumption of normality are not met. This has an impact on the problem of parameter estimation based on maximum likelihood to be inappropriate so that another alternative is needed to obtain the value of λ , namely using a new approach method which is carried out by means of a grid search combined with normality testing. In this study a new approach to the Box-Cox transformation method will involve the response variable, namely the amount of poverty and the predictor variables, namely the Open Unemployment Rate, Human Development Index, and Minimum Wage. In this study, we wanted to find out the factors that influence the amount of poverty in West Java Province in 2022. The data used is secondary data sourced from the Central Statistics Agency and the West Java Province Website. From the analysis results, the optimum λ estimation is 0.5 and the regression model is $y^{3/5} = (0.0168^{5/3}) + (1.2410^{5/3} X_1) - (2.3160^{5/3} X_2) + (0.000066^{5/3} X_3)$.

Keywords: *Regression Analysis, Approach from Box-Cox Transformation, Maximum Likelihood Estimator.*

Abstrak. Analisis regresi digunakan untuk memodelkan di antara dua variabel kuantitatif. Ketika melakukan pemodelan regresi sering kali ditemukan data tidak berdistribusi normal, cara untuk mengatasinya dapat melakukan transformasi. Salah satu transformasi yang digunakan adalah transformasi Box-Cox. Box dan Cox mempertimbangkan kelas transformasi berparameter tunggal yaitu parameter λ terhadap Y sehingga menjadi Y^λ dengan λ adalah parameter yang harus diduga. Pendugaan parameter λ menggunakan penduga kemungkinan maksimum. Pada umumnya nilai λ diasumsikan sudah diketahui, namun beberapa nilai dari λ (disebut λ_0) asumsi kenormalan tidak terpenuhi. Hal ini berdampak pada masalah pendugaan parameter yang didasarkan pada kemungkinan maksimum menjadi tidak tepat sehingga perlu alternatif lain untuk memperoleh nilai λ yaitu menggunakan metode pendekatan baru yang dilakukan dengan cara pencarian secara grid yang dikombinasikan dengan pengujian kenormalan. Dalam penelitian ini pendekatan baru metode transformasi Box-Cox akan melibatkan variabel respon yaitu Jumlah kemiskinan dan variabel prediktor yaitu Tingkat Pengangguran Terbuka, Indeks Pembangunan Manusia, dan Upah Minimum Kerja. Pada penelitian ini ingin mengetahui faktor-faktor yang mempengaruhi jumlah kemiskinan di Provinsi Jawa Barat tahun 2022. Data yang digunakan adalah data sekunder yang bersumber dari Badan Pusat Statistika dan Website Provinsi Jawa Barat. Dari hasil analisis didapat pendugaan λ yang optimum yaitu 0.5 dan model regresi $y^{3/5} = (0.0168^{5/3}) + (1.2410^{5/3} X_1) - (2.3160^{5/3} X_2) + (0.000066^{5/3} X_3)$.

Kata Kunci: *Analisis Regresi, Pendekatan dari Transformasi Box-Cox, Penduga Kemungkinan Maksimum.*

A. Pendahuluan

Analisis regresi digunakan untuk memodelkan dan meringkas hubungan di antara dua variabel kuantitatif. Dalam melakukan model analisis regresi terlebih dahulu melakukan pengujian asumsi klasik. Gauss Markov telah membuktikan penduga dalam regresi mempunyai sifat BLUE (*Best Linier Unbiased*). Salah satu asumsi yang perlu terpenuhi yaitu data mengikuti distribusi normal atau kenormalan terpenuhi. Ketika melakukan pemodelan regresi sering kali ditemukan data tidak berdistribusi normal atau asumsi kenormalan data tidak terpenuhi.

Pada masalah ketidaknormalan dapat diselesaikan dengan beberapa cara, termasuk menggunakan transformasi Box-Cox. Pada transformasi Box-Cox dapat memberikan model alternatif yang mungkin dapat memberikan lebih baik untuk keperluan pendugaan maupun untuk keperluan peramalan. Box dan Cox mempertimbangkan kelas transformasi berparameter tunggal yaitu parameter λ sehingga menjadi Y^λ dengan λ merupakan parameter yang harus diduga. Pada transformasi ini, pendugaan parameter λ menggunakan pendugaan kemungkinan maksimum (PKM) untuk melakukan transformasi pada variabel respon (Draper & Smith, 1992).

Pada umumnya nilai λ diasumsikan sudah diketahui, akan tetapi untuk beberapa nilai dari λ (yang disebut λ_0) asumsi kenormalan tidak terpenuhi. Hal ini berdampak pada masalah pendugaan parameter yang didasarkan pada kemungkinan maksimum menjadi tidak tepat (Velez, 2015). Dengan demikian inferensi yang didasarkan pada nilai λ yang diperoleh dari hasil pendugaan kemungkinan maksimum juga menjadi tidak tepat sehingga diperlukan alternatif lain untuk memperoleh nilai λ . Dalam skripsi ini akan diperkenalkan pendekatan baru dari metode transformasi Box-Cox yang digunakan untuk menduga λ sebagai alternatif dari penduga kemungkinan maksimum. Metode pendekatan baru ini dilakukan dengan cara pencarian secara grid yang dikombinasikan dengan pengujian kenormalan dalam rangka mencari nilai λ . Statistik uji untuk menguji kenormalan yang digunakan dalam skripsi ini adalah Shapiro wilk.

Pada pengujian pendekatan lama dari transformasi Box-Cox dengan pengujian pendekatan baru dari transformasi Box-Cox memiliki perbedaan. Pada pendekatan lama umumnya hanya menghasilkan satu nilai λ dan belum dapat dipastikan dapat menunjukkan hasil data berdistribusi normal. Sedangkan pada pendekatan baru ini akan menghasilkan beberapa nilai λ dan sudah melakukan pengujian normalitas sehingga dapat dipastikan hasil data tersebut berdistribusi normal. Pendekatan baru dari metode transformasi Box-Cox akan diterapkan pada data jumlah kemiskinan di Provinsi Jawa Barat pada tahun 2022.

Jumlah kemiskinan disebabkan oleh beberapa faktor seperti tingkat pengangguran yang tinggi, tingkat pendapatan yang dibawah standar dan kualitas sumber daya manusia yang tercermin dari indeks pembangunan manusia yang kurang (Kurniawan, 2018).

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana penerapan baru dari metode transformasi Box-Cox dalam pencarian nilai λ yang optimal dengan dikombinasikan dengan uji normalitas dari variabel respon?” serta “faktor-faktor apa saja yang mempengaruhi jumlah kemiskinan di Provinsi Jawa Barat tahun 2022?”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Untuk menerapkan pendekatan baru dari metode transformasi Box-Cox dalam pencarian nilai λ yang optimal dengan dikombinasikan dengan uji normalitas dari variabel respon.
2. Untuk mengetahui faktor-faktor yang mempengaruhi jumlah kemiskinan di Provinsi Jawa Barat tahun 2022 menggunakan analisis regresi

B. Metodologi Penelitian

Metode Pengumpulan Data

Pada penelitian ini menggunakan data yang bersifat sekunder. Data sekunder pada penelitian ini yaitu bersumber dari Badan Pusat Statistika (BPS) dan situs website provinsi jawa barat yang terdiri atas 18 Kabupaten dan 9 Kota. Data yang digunakan pada penelitian ini terdiri atas variabel respon (Y) yaitu data Jumlah kemiskinan dan variabel prediktor (X) memiliki 3 (tiga) kategori yaitu Tingkat Pengangguran Terbuka (TPT) (X_1), Indeks Pembangunan Manusia (IPM) (X_2), dan Upah Minimum Kerja (UMK) (X_3).

Metode Analisis Data

Analisis Regresi

Analisis regresi merupakan salah satu teknik analisis data dalam statistika yang seringkali digunakan untuk mengkaji hubungan antara beberapa variabel dan meramal suatu variabel. Salah satu tujuan dari regresi yaitu menentukan persamaan regresi yang baik sehingga dapat digunakan untuk menaksir nilai variabel respon (Kutner, Nachtsheim dan Neter, 2004).

Pendugaan Parameter

Pendugaan merupakan proses yang digunakan untuk menghasilkan suatu nilai tertentu terhadap suatu parameter. Penduga parameter pada mulanya akan digunakan untuk menduga suatu populasi dari sampel. Penduga merupakan suatu tahapan yang terpenting dalam menentukan model peluang yang tepat dari sekumpulan data (Yendra & Noviadi, 2015).

Penduga parameter dapat dilakukan dengan beberapa metode, namun metode kemungkinan maksimum adalah yang paling populer atau yang paling sering digunakan dalam penelitian. Hal ini dikarenakan metode tersebut sangat berhubungan dengan kemampuan numerik, terutama dalam menghasilkan titik penyelesaian dalam suatu persamaan. Pendugaan nilai parameter menggunakan metode kemungkinan maksimum merupakan teknik digunakan untuk menaksir parameter yang distribusi populasi diketahui. Dalam bentuk matriks persamaan pendugaan parameter β adalah:

$$b = (X^tX)^{-1}(X^tY) \quad \dots(1)$$

Pengujian Model Regresi

Uji Simultan adalah uji pada semua variabel prediktor secara bersamaan pada suatu model. Uji ini digunakan untuk melihat apakah variabel prediktor secara bersama-sama berpengaruh terhadap variabel respon (Kurniawan & Yuniarto, 2016).

Uji Parsial digunakan untuk mengetahui pengaruh dari setiap variabel prediktor terhadap variabel respon (Kurniawan & Yuniarto, 2016).

Uji Diagnostik Model

Uji normalitas digunakan untuk mengetahui apakah variabel respon berdistribusi normal atau tidak. Uji Shapiro Wilk adalah salah satu metode yang dapat digunakan pada uji normalitas (Alam & Mutaqin, 2023).

Uji heterokedastisitas bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan varians dari residual satu pengamatan ke pengamatan lain. Untuk model regresi yang baik yaitu terdapat homoskedastisitas atau dapat dikatakan tidak terjadi heteroskedastisitas. Uji Glejser dapat digunakan untuk mendeteksi heteroskedastisitas. Uji Glejser mengusulkan untuk meregresi nilai absolut residual sebagai variabel terikat dengan variabel bebas yang digunakan.

Uji multikolinearitas bertujuan untuk mengetahui apakah dalam model regresi terdapat korelasi antar variabel bebas. Cara mendeteksi multikolinearitas dapat dilihat dari nilai Variance Inflation Factor (VIF).

Transformasi Box-Cox

Transformasi Box-Cox adalah transformasi pangkat pada respon. Box-Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu λ yang dipangkatkan dengan variabel respon Y sehingga transformasinya menjadi y^λ . λ adalah parameter yang diperlukan diduga (Fransiska, 2012). Tujuan dari transformasi adalah untuk memberikan suatu model alternatif yang dapat memberikan lebih sukses baik untuk keperluan pendugaan maupun untuk keperluan peramalan. Transformasi memiliki manfaat karena tidak mengarah pada penghapusan pengamatan dan juga mengakomodasi semua pengamatan ke dalam distribusi yang cenderung kurang condong daripada data asli (Hajarisman, 2014). Pada transformasi Box-Cox variabel respon misalkan $y = (y_1, y_2, \dots, y_k)$ adalah data yang akan diterapkan transformasi Box-Cox. Maka mendefinisikan transformasi Box-Cox sebagai berikut:

$$y_i^{(\lambda)} = \begin{cases} \{\lambda^{-1}(y_i^{(\lambda)} - 1)\} & \text{if } \lambda \neq 0 \\ \log(y_i) & \text{if } \lambda = 0 \end{cases} \quad \dots(2)$$

Sehingga untuk λ tidak diketahui maka:

$$y^\lambda = X\beta + \varepsilon \quad \dots(3)$$

Pendekatan Baru Pada Transformasi Box-Cox

Transformasi ini dihasilkan dari modifikasi keluarga transformasi yang didefinisikan oleh Tukey untuk menjelaskan ketidakcontinuan pada $\lambda=0$. Karena sifat terkenal dari penduga kemungkinan maksimum (PKM) biasanya memperkirakan λ dengan $\hat{\lambda}_{MLE}$. Hernandez dan Johnson (1980) memperoleh hasil asimtotik untuk kasus ini yaitu menggunakan pendekatan baru dengan pencarian nilai λ yang dikombinasikan dengan uji normalitas yang dilakukan pada data yang di transformasi untuk menentukan λ^* nilai optimal dari λ , sehingga nilai λ dari uji normalitas yang optimal adalah hasil p-value tertinggi.

Misalkan $y = (y_1, y_2, \dots, y_n)$ adalah data yang akan diterapkan transformasi Box-Cox. Maka mendefinisikan transformasi Box-Cox dengan menggunakan persamaan (2).

C. Hasil Penelitian dan Pembahasan

Deskripsi Data

Kondisi jumlah kemiskinan dan kondisi faktor-faktor yang mempengaruhi meningkatnya jumlah kemiskinan pada penelitian ini dijelaskan melalui statistik deskriptif seperti berikut:

Tabel 1. Deskripsi Data Penelitian

Variabel	Rata-rata	Minimum	Maksimum	Simpangan Baku
Y (Kemiskinan)	150.778	12.7	474.7	102.4137
X_1 (TPT)	7.8011	1.56	10.78	2.33196
X_2 (IPM)	72.6130	65.94	82.50	4.72093
X_3 (UMK)	3.072.179,43	1.852.099,52	4.816.921,17	1.009.196,75

Sumber: Data Penelitian yang Sudah Diolah, 2023.

Dari tabel di atas, dapat diketahui bahwa jumlah kasus kemiskinan (Y) memiliki rata-rata 150.778 jika dibulatkan sebesar 160 jiwa dengan nilai maksimum sebesar 474.7 dan jika dibulatkan menjadi 475 jika berada di Kabupaten Bogor, dan wilayah dengan jumlah kemiskinan (Y) minimum adalah Kota Banjar sebesar 12.7 jika dibulatkan 13 jiwa. Jumlah Tingkat Pengangguran Terbuka (TPT) memiliki rata-rata 7.8 persen dengan nilai maksimum sebesar 10.78 persen berada di Kota Cimahi dan nilai minimum berada di Kabupaten Pangandaran dengan nilai 1.56 persen. Indeks Pembangunan Manusia (IPM) memiliki rata-rata 72.6 persen dengan nilai minimum sebesar 65.94 persen berada di Kabupaten Cianjur dan nilai maksimum berada di Kota Bandung dengan nilai 82.5 persen, dan Upah Minimum Kerja (UMK) memiliki rata-rata 3.072.179,43 rupiah dengan nilai minimum sebesar 1.852.099,52 rupiah berada di Kota Banjar dan nilai maksimum berada di Kota Bekasi sebesar 4.816.921,17 rupiah.

Analisis Regresi

Dalam analisis regresi salah satu metode untuk pendugaan model regresi adalah dengan menggunakan kemungkinan maksimum. Pendugaan koefisien regresi dilakukan dengan bantuan software R-Studio menggunakan data jumlah kemiskinan, didapat hasil sebagai berikut:

$$\hat{y} = 0.0011 + 0.1090X_1 - 0.1649X_2 + 0.000043X_3$$

Langkah selanjutnya ialah melakukan pengujian parameter secara simultan untuk mengetahui apakah semua variabel prediktor berpengaruh secara bersama-sama terhadap

variabel respon dengan hasil sebagai berikut:

Tabel 2 Hasil Uji Parsial

Parameter	Koefisien	Galat Baku	t-value	p-value
intersep	0.0011	0.026	4.271	0.000
X1	0.0109	8.895	1.225	0.2329
X2	0.1649	4.026	-4.096	0.0004
X3	0.00004	0.000	1.911	0.0686
F-statistic: 6.25, p-value: 0.002918				

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$; Tidak ada pengaruh secara signifikan antara variabel prediktor terhadap variabel respon

$H_1 : \text{minimal ada satu } \beta_k \neq 0 ; k = 1,2,3$; Minimal ada satu variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon

Berdasarkan hasil output menggunakan software *R-Studio* didapat bahwa nilai p-value sebesar 0.002918 untuk menguji hipotesis uji simultan, kriteria uji adalah tolak H_0 jika $p\text{-value} < \alpha$. Dengan taraf nyata $\alpha = 5\%$ maka diperoleh hasil bahwa $p\text{-value} > \alpha$ atau $0.0029 < 0.05$ maka H_0 ditolak. Artinya variabel prediktor secara bersamaan berpengaruh terhadap variabel respon. Kemudian langkah selanjutnya melakukan pengujian parameter secara parsial untuk mengetahui apakah ada pengaruh dari setiap variabel prediktor terhadap variabel respon.

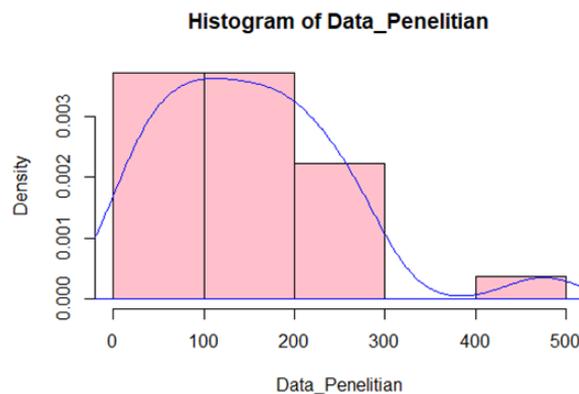
$H_0 : \beta_k = 0$; Variabel prediktor ke-k tidak berpengaruh secara signifikan terhadap variabel respon

$H_1 : \beta_k \neq 0$; Variabel prediktor ke-k berpengaruh secara signifikan terhadap variabel respon

Berdasarkan hasil pengujian secara parsial diatas dengan menggunakan taraf nyata $\alpha = 5\%$, kriteria uji dari pengujian secara parsial yaitu $\text{sig} < \alpha$ maka H_0 ditolak terlihat bahwa terdapat 1 variabel yang berpengaruh terhadap Jumlah Kemiskinan (Y) yaitu Indeks Pembangunan Manusia (X_2).

Uji Normalitas

Uji normalitas dilakukan sebagai syarat untuk analisis regresi, berguna untuk melihat apakah data yang telah dikumpulkan mempunyai distribusi normal atau tidak. Uji normalitas menggunakan analisis grafik dan analisis statistik. Analisis grafik yaitu berupa grafik histogram. Adapun grafik histogram dapat dilihat pada Gambar 1.



Gambar 1 Histogram Data Kemiskinan di Jawa Barat tahun 2022

Gambar diatas merupakan grafik histogram. Grafik histogram dikatakan normal jika distribusi data membentuk lonceng, tidak condong ke kiri atau tidak condong ke kanan. Grafik histogram diatas tidak membentuk lonceng dan condong ke kanan sehingga grafik histogram

tersebut dinyatakan tidak normal. Uji normalitas dengan grafik histogram perlu diperkuat dengan uji statistik. Uji statistik menggunakan uji *Shapiro Wilk*. Hasil uji normalitas dengan grafik terkadang berbeda dengan hasil uji statistik. Grafik tampak normal, namun secara statistik bisa terjadi sebaliknya. Maka dari itu untuk memastikan hasil grafik berdistribusi normal dilakukan uji statistic dengan hipotesis sebagai berikut:

H_0 : Sisaan berdistribusi normal

H_1 : Sisaan tidak berdistribusi normal

Didapat hasil *p-value* menggunakan *Shapiro Wilk* yaitu 0.068 Maka kriteria uji pada uji normalitas adalah tolak H_0 jika *p-value* < α . Digunakan taraf nyata $\alpha = 10\%$ dapat disimpulkan bahwa pada *Shapiro Wilk* didapat H_0 ditolak, artinya bahwa variabel respon yakni variabel jumlah kemiskinan tidak berdistribusi normal.

Uji Heterokedastisitas

Untuk mengetahui data sampel berasal dari populasi yang memiliki varians yang sama atau tidak malah dilakukan pengujian heterokedastisitas dengan menggunakan uji Glejser. Adapun hasil pengujian dengan menggunakan software SPSS, dengan hipotesis sebagai berikut:

H_0 : Sisaan tidak terjadi heterokedastisitas

H_1 : Sisaan terjadi heterokedastisitas

Tabel 2. Hasil Uji Heterokedastisitas

Variabel	VIF
X_1	0.323
X_2	0.360
X_3	0.391

Dengan kriteria uji tolak H_0 jika *p-value* < α digunakan taraf nyata $\alpha = 5\%$ didapat bahwa variabel X_1, X_2, X_3 memiliki nilai *p-value* > 0.05. Hal ini dapat disimpulkan bahwa pada sisaan dari variabel X_1, X_2, X_3 tidak terjadi heterokedastisitas.

Uji Multikolinieritas

Untuk mengetahui apakah ada korelasi antara variabel prediktor yang satu dengan variabel prediktor yang lain dalam model regresi yang digunakan. Metode yang digunakan untuk mendeteksi ada atau tidaknya menggunakan Variance Inflation Factor (VIF). Adapun hasil pengujian dengan menggunakan software SPSS didapat hasil sebagai berikut:

Tabel 3. Nilai VIF Pada Variabel Prediktor

Variabel	VIF
X_1	1.713
X_2	1.438
X_3	2.051

Berdasarkan tabel di atas dapat diketahui bahwa nilai VIF pada variabel X_1, X_2, X_3 memiliki nilai kurang dari 10. Hal ini dapat disimpulkan bahwa pada variabel X_1, X_2, X_3 tidak terjadi masalah multikolinieritas dalam model regresi tersebut, berarti dalam penelitian ini tidak terdapat pola hubungan linier antara variabel respon dan variabel prediktor dalam model regresi dan layak digunakan untuk analisis lebih lanjut.

Pendekatan Baru Pada Transformasi Box-Cox

Pada penelitian ini diketahui bahwa data kemiskinan di Provinsi Jawa Barat tahun 2022 setelah melakukan pengujian normalitas pada variabel respon didapat bahwa data tidak berdistribusi normal sehingga cara untuk mengatasi yaitu menggunakan transformasi. Transformasi Box-Cox

ini melakukan pendekatan berbasis p-value untuk λ . Nilai λ digunakan sebagai parameter yang diperlukan untuk diduga λ kemudian akan dipasangkan dengan p-value hasil pengujian normalitas. Karakteristik dari metode ini yaitu nilai λ yang dipilih dipasangkan dengan nilai p-value yang tertinggi. Menurut Kutner (2005) nilai λ antara -2 sampai 2. Maka menerapkan nilai transformasi Box-Cox terhadap variabel respon berdasarkan persamaan (2) Dilakukan dengan software excel dan diperoleh hasil sebagai berikut:

Tabel 4. Hasil Nilai Transformasi Box-Cox

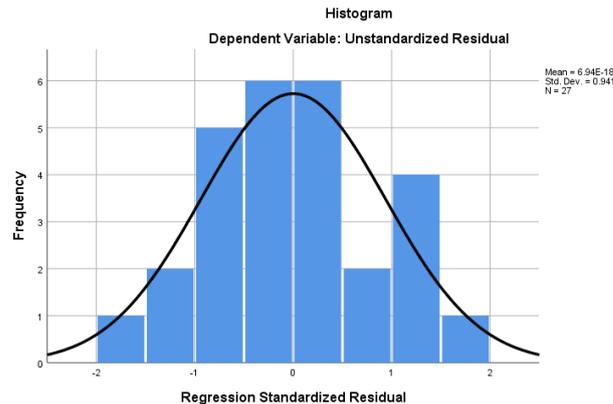
Wilayah Jawa Barat	Jumlah Kemiskinan	λ						
		-2	-1.9	1.9	2
Bogor	474.7	0.499998	0.526311	64038.24	112669.50
Sukabumi	186.3	0.499986	0.526290	10830.03	17353.35
Cianjur	246.8	0.499992	0.526301	18479.48	30454.62
Bandung	258.6	0.499993	0.526302	20194.32	33436.48
Garut	276.7	0.499993	0.526304	22964.33	38280.95
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kota Bekasi	137.4	0.499974	0.526270	6072.72	9438.88
Kota Depok	64.4	0.499879	0.526123	1438.70	2073.18
Kota Cimahi	31.2	0.499486	0.525553	362.67	486.22
Kota Tasikmalaya	87.1	0.499934	0.526207	2553.83	3792.71
Kota Banjar	12.7	0.496900	0.522108	65.31	80.15

Untuk mengetahui data berdistribusi normal dengan menggunakan data yang sudah di transformasi dilakukan pengujian normalitas Shapiro Wilk untuk nilai λ tertentu, pengujian dilakukan menggunakan software R-studio. Adapun hasil pengujian normalitas untuk variabel respon dapat dilihat seperti tabel berikut:

Tabel 5. Hasil Pengujian Normalitas Untuk Variabel Respon

No	λ	p-value	Statistik Uji
1	-2	0.0000	0.5670
2	-1.9	0.0000	0.5826
3	-1.8	0.0000	0.5990
4	-1.7	0.0000	0.6168
5	-1.6	0.0000	0.6369
6	-1.5	0.0000	0.6597
7	-1.4	0.0000	0.6834
8	-1.3	0.0000	0.7077
9	-1.2	0.0000	0.7324
10	-1.1	0.0000	0.7581
11	-1	0.0000	0.7828
12	0	0.0000	0.6958
13	0.1	0.4574	0.9642
14	0.2	0.4687	0.9647
15	0.3	0.4077	0.9619
16	0.4	0.6568	0.9721
17	0.5	0.8367	0.9789
18	0.6	0.8570	0.9798
19	0.7	0.7402	0.9751
20	0.8	0.4964	0.9658
21	0.9	0.2724	0.9543
22	1	0.1124	0.9386
23	1.1	0.0432	0.9216
24	1.2	0.0174	0.9049
25	1.3	0.0734	0.8884
26	1.4	0.0032	0.8716
27	1.5	0.0013	0.8535
28	1.6	0.0006	0.8342
29	1.7	0.0003	0.8159
30	1.8	0.0001	0.7970
31	1.9	0.0001	0.7781
32	2	0.0000	0.7590

Berdasarkan tabel diatas dapat dilihat bahwa pada data kemiskinan ketika λ bernilai -2 sampai 0 memberikan hasil p-value yang kecil, dan pada saat λ bernilai 1 sampai 2 juga memberikan hasil p-value yang kecil sehingga dengan menggunakan taraf nyata 5% menghasilkan kesimpulan bahwa p-value $< \alpha$ artinya bahwa variabel respon tidak berdistribusi normal. Sebaliknya ketika λ bernilai 0.1 sampai 0.9 memberikan hasil p-value yang besar. Pada rentang λ bernilai 0.1 sampai 0.9 didapat bahwa nilai λ yang tertinggi pada saat λ sebesar 0.6 dengan nilai p-value = 0.8570. sehingga dengan menggunakan taraf nyata 10% menghasilkan nilai 0.8570 > 0.05 sehingga dengan kriteria uji Tolak H_0 p-value $< \alpha$ menggunakan taraf nyata 5% menghasilkan nilai 0.8570 > 0.05 artinya bahwa H_0 diterima atau dengan kata lain variabel respon berdistribusi normal. Sehingga pendugaan nilai λ yang optimal yaitu 0.6.



Gambar 2 Histogram Data Hasil Transformasi untuk $\lambda=0.6$

Analisis Regresi Setelah Transformasi

Dalam analisis regresi salah satu metode untuk mengestimasi model regresi adalah dengan menggunakan penduga kemungkinan maksimum (PKM). Pendugaan koefisien regresi dilakukan dengan bantuan software R-studio menggunakan data hasil pengujian dari pendekatan baru metode transformasi Box-Cox dengan $\lambda = 0.6$, didapat hasil sebagai berikut:

$$y^{\frac{3}{5}} = (0.0168^{\frac{5}{3}}) + (1.2410^{\frac{5}{3}}X_1) - (2.3160^{\frac{5}{3}}X_2) + (0.0000066^{\frac{5}{3}}X_3)$$

Berdasarkan hasil pengujian yang telah dilakukan dengan menggunakan pendekatan baru dari metode transformasi Box-Cox selanjutnya akan dilakukan pengujian parameter secara simultan dan pengujian parameter secara parsial. Adapun hasil yang diperoleh menggunakan software R-Studio, dengan hasil sebagai berikut:

Tabel 6. Hasil Uji Parsial Setelah Transformasi

Variabel	Koefisien	Std. Error	t-value	p-value
intersep	0.01684	0.3466	4.859	0.0000
X1	1.2410	1.1640	1.066	0.2976
X2	-2.3160	0.5268	-4.396	0.0002
X3	0.0000066	0.0000	2.250	0.0343
F-statistic: 7.119, p-value: 0.001497				

Langkah selanjutnya ialah melakukan pengujian parameter secara simultan yaitu untuk mengetahui apakah semua variabel prediktor berpengaruh secara bersama-sama terhadap variabel respon, dilakukan pengujian simultan parameter dengan hasil sebagai berikut:

Berdasarkan hasil output menggunakan software R-Studio didapat bahwa nilai p-value sebesar 0.0014. Kriteria uji ialah tolak H_0 jika p-value $< \alpha$. Dengan taraf nyata $\alpha = 5\%$ maka

diperoleh hasil bahwa $0.0014 < 0.05$ maka H_0 ditolak. Artinya variabel prediktor secara bersamaan berpengaruh terhadap variabel respon.

Kemudian selanjutnya melakukan pengujian parameter secara parsial untuk mengetahui apakah ada pengaruh dari setiap variabel prediktor terhadap variabel respon, dengan hasil sebagai berikut:

Berdasarkan tabel 6 diatas dengan menggunakan taraf nyata $\alpha = 5\%$ maka terlihat bahwa terdapat 2 variabel yang berpengaruh terhadap jumlah kemiskinan (Y) yaitu Indeks Pembangunan Manusia (X_2), Upah Minimum Kerja (X_3).

Berdasarkan hasil pengujian secara parsial dari data kemiskinan terdapat perbedaan dari variabel prediktor yang mempengaruhi terhadap variabel respon, pada tabel 2 menyatakan bahwa terdapat 1 variabel yang berpengaruh terhadap jumlah kemiskinan (Y) yaitu indeks pembangunan manusia (X_2). Sedangkan setelah dilakukan transformasi melalui pendekatan baru dari metode transformasi Box-Cox terdapat 2 variabel yang berpengaruh terhadap jumlah kemiskinan (Y) yaitu indeks pembangunan manusia (X_2) dan upah minimum kerja (X_3)

D. Kesimpulan

Kesimpulan yang diperoleh berdasarkan analisis yang telah diperoleh adalah sebagai berikut:

1. Berdasarkan hasil dari pengujian dengan menggunakan pendekatan baru dari metode transformasi Box-Cox dengan data kemiskinan diperoleh bahwa ketika λ bernilai -2 sampai 0 dan λ bernilai 1 sampai 2 memberikan hasil p-value yang kecil. Sebaliknya ketika λ bernilai 0.1 sampai 0.9 memberikan hasil p-value yang tinggi. Pada rentang λ bernilai 0.1 sampai 0.9 didapat bahwa nilai p-value yang tertinggi pada saat λ sebesar 0.6. Sehingga pendugaan nilai λ yang optimal yaitu 0.6.
2. Berdasarkan hasil pengujian dengan menggunakan pendekatan baru dari metode transformasi Box-Cox didapat model regresi sebagai berikut:

$$y^{3/5} = (0.0168^{5/3}) + (1.2410^{5/3}X_1) - (2.3160^{5/3}X_2) + (0.000066^{5/3}X_3)$$

Berdasarkan hasil pemodelan regresi diperoleh 2 variabel prediktor yang berpengaruh terhadap Jumlah kemiskinan (Y) di Provinsi Jawa Barat tahun 2022 yaitu Indeks Pembangunan Manusia (X_2) dan Upah Minimum Kerja (X_3)

Acknowledge

Penulis sangat bersyukur kepada Allah SWT atas nikmat, rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan penelitian ini dengan lancar. Serta mengucapkan terimakasih kepada Bapak Nusar selaku dosen pembimbing yang telah memberikan ilmu serta bimbingan kepada penulis.

Daftar Pustaka

- [1] Draper & Smith., 1992. Analisis Regresi Terapan. Kedua ed. Jakarta: Gramedia Pustaka Utama
- [2] Velez, J. I., Correa, J. C. & Ramos, F. M., 2015. A new approach to the Box-Cox transformation. *Frontiers in Applied Mathematics and Statistics*, p. 1
- [3] Kurniawan, R. & Yuniarto, B., 2016. *Analisis Regresi Dasar dan Penerapannya dengan R*. Jakarta: K E N C A N A
- [4] Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W., 2005. Applied Linier Statistical Models. In: New York: McGraw-Hill Irwin
- [5] Yendra, R. & Noviadi, E. T., 2015. Perbandingan Estimasi Parameter pada Distribusi Eksponensial dengan Menggunakan Metode Maksimum Likelihood dan Metode Bayesian. *Sains, Matematika dan Statistika*, Volume 1, p. 2.

- [6] Fransiska, W., Nugroho, S. & Faisal, F., n.d. Transformasi Box-Cox dalam Analisis Regresi Linier Sederhana.
- [7] Hajarisman, N., 2014. Beberapa Bentuk Transformasi dalam Analisis Regresi
- [8] Hernandez & Johnson, 1980. The large-sample behavior of transformations to normality. Volume 75, pp. 855-61.
- [9] Alam, M. R. N., & Mutaqin, A. K. (2023). Pemodelan Distribusi Poisson-Sujatha pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia. *Jurnal Riset Statistika*, 71–78. <https://doi.org/10.29313/jrs.v3i1.1944>