

Pemodelan Jumlah Penduduk Miskin di Jawa Timur dengan Generalized Linier Model

Noviana Pratiwi¹, Maria Jefin Paput²

^{1,2}Jurusan Statistika, IST AKPRIND Yogyakarta

Email: ¹novianapratiwi@akprind.ac.id, ²mariapaput05@gmail.com

Abstract

Poverty as a development issue is an important thing to pay attention. Therefore, it is necessary to analyze to prevent an increase in the number of poor people in an area, one of which is by considering the things that affect poverty estimates and knowing the model of the number of poor people that can later be used as predictions. East Java is the region in Indonesia with the highest population in Indonesia in 2019. One way to find out the influencing factors and at the same time to make a model of the number of poor people in East Java is by modeling with the Generalized Linear Model (GLM) method. General Linear Model (GLM). GLMs aim to determine the causal relationship and the influence of the independent variable on the dependent variable. The GLMs model in this study uses the Gaussian family because the research response variable (Y) is an integer. The results show that of the seven estimated models, only one model that meets the significance is the model with the dependent variable population density (X3). So it can be concluded that between Area (X1), Open Unemployment Rate (X2), and Population Density (X3), population density has a significant effect on the number of poor people with a 95% confidence level.

Keywords: GLMs, Number of Poor Population, Area, Open Unemployment Rate, Population Density

Abstrak

Kemiskinan sebagai isu pembangunan merupakan hal yang penting untuk diperhatikan. Oleh karena itu perlu dilakukan analisis untuk mencegah peningkatan jumlah penduduk miskin di suatu wilayah, salah satunya adalah dengan mempertimbangkan hal-hal yang mempengaruhi jumlah kemiskinan serta mengetahui model jumlah penduduk miskin yang nantinya bisa digunakan sebagai prediksi kedepan. Jawa Timur merupakan wilayah di Indonesia yang menempati jumlah penduduk tertinggi di Indonesia pada tahun 2019. Salah satu cara untuk mengetahui faktor yang mempengaruhi dan sekaligus untuk membuat model mengenai jumlah penduduk miskin di Jawa Timur adalah dengan pemodelan dengan metode Generalised Linier Model (GLMs). Generalized Linear Model (GLMs). GLMs bertujuan untuk mengetahui hubungan sebab-akibat, pengaruh dari variable independent terhadap variable dependent. Model GLMs dalam penelitian ini menggunakan family Gaussian karena variabel respon penelitian (Y) merupakan integer. Hasil menunjukkan bahwa dari tujuh model yang diestimasi hanya satu model yang memenuhi signifikansi yaitu model dengan variable dependent kepadatan penduduk (X3). Jadi dapat disimpulkan bahwa diantara Luas Wilayah (X1), Tingkat Pengangguran Terbuka (X2), dan Kepadatan Penduduk (X3), kepadatan penduduk berpengaruh signifikan terhadap jumlah penduduk miskin dengan tingkat kepercayaan 95%.

Kata Kunci: GLMs, Jumlah Penduduk Miskin, Luas Wilayah, Tingkat Pengangguran Terbuka, Kepadatan Penduduk

1. PENDAHULUAN

Pembangunan dalam bidang ekonomi dilaksanakan untuk dapat mewujudkan kemakmuran dan kesejahteraan masyarakat. Berbagai strategi pembangunan ekonomi dilakukan oleh pemerintah untuk berorientasi pada peningkatan pertumbuhan ekonomi dan pemerataan ekonomi yang berkeadilan yaitu dengan menurunkan jumlah penduduk miskin, namun dalam realitasnya masih terdapat diskrepansi antara harapan dan kenyataan di lapangan berupa kemiskinan (Soejoto dan Karisma, 2011). Terdapat empat alasan pentingnya mengangkat kemiskinan sebagai isu pembangunan menurut world bank. Pertama, penurunan kemiskinan merupakan prioritas dalam pembangunan. Sehingga, kemiskinan harus diukur secara akurat, agar keberadaan penduduk miskin dapat dijangkau dengan lebih mudah. Kedua, agar sebaran penduduk miskin dapat diidentifikasi dengan mudah. Ketiga, sebagai bahan evaluasi terhadap program pemerintah terkait penanganan kemiskinan. Keempat, mengevaluasi efektivitas lembaga pemerintah dalam penanganan kemiskinan. Oleh karena itu, perlu dilakukan analisis untuk mencegah peningkatan jumlah penduduk miskin di Indonesia, salah satunya adalah dengan mempertimbangkan hal-hal yang mempengaruhi jumlah kemiskinan tersebut. Provinsi Jawa Timur dipilih sebagai daerah kajian dalam penelitian ini dikarenakan pada tahun 2019 Jawa Timur menempati peringkat pertama dengan jumlah penduduk miskin terbesar di Indonesia.

Secara statistic, faktor-faktor yang mempengaruhi jumlah kemiskinan dapat dilihat dengan melakukan pengolahan data dengan metode Analisa regresi. Analisis regresi adalah salah satu metode untuk menentukan hubungan sebab-akibat antara satu variabel dengan variabel yang lain. Variabel penyebab disebut dengan variabel independen, variabel penjelas atau variabel X. Sementara variabel yang terkena akibat dikenal sebagai variabel yang dipengaruhi, variabel dependen, variabel terikat, variabel respon atau variabel Y. Pendugaan kurva regresi digunakan untuk menjelaskan hubungan antara peubah penjelas dengan peubah respon. Pendekatan pendugaan yang paling sering digunakan adalah pendekatan parametrik. Asumsi yang mendasari pendekatan ini adalah kurva regresi dapat diwakili oleh suatu model parametrik (Hardle, 1990)

Generalized linier model (GLMs) merupakan perluasan dari model regresi linier dengan asumsi prediktor memiliki efek linier akan tetapi tidak mengasumsikan distribusi tertentu dari variabel respon dan digunakan ketika variabel respon merupakan anggota dari keluarga eksponensial (Nelder dan Weddeburn, 1972). Santi (2021) menggunakan GLMs untuk memodelkan jumlah kasus malaria dengan variabel respon berdistribusi Poisson. Penelitian mengenai factor yang mempengaruhi jumlah penduduk miskin sebelumnya diteliti oleh Indah (2018), hasilnya menunjukkan bahwa tingkat pengangguran dan tingkat pendidikan berpengaruh positif dan signifikan terhadap kemiskinan sedangkan jumlah penduduk tidak berpengaruh signifikan terhadap kemiskinan.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan jenis penelitian asosiatif yang bertujuan untuk menguji pengaruh tiga variabel yaitu pengaruh luas wilayah, tingkat pengangguran terbuka, dan kepadatan penduduk terhadap jumlah penduduk miskin. Desain penelitian yang digunakan adalah hipotesis, yaitu luas wilayah berpengaruh terhadap jumlah penduduk miskin di Jawa Timur, tingkat pengangguran terbuka berpengaruh terhadap jumlah penduduk miskin di Jawa

Timur, dan kepadatan penduduk berpengaruh terhadap jumlah penduduk miskin di Jawa Timur. Lokasi penelitian dilakukan di Kabupaten/Kota di Jawa Timur. Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang dikeluarkan dan dipublikasikan oleh Badan Pusat Statistik (BPS) yang berkaitan dengan objek penelitian. Variabel-variabel yang digunakan yaitu jumlah penduduk miskin (Y) sebagai variabel respon atau dependen dan luas wilayah (X1), tingkat pengangguran terbuka (X2), dan kepadatan penduduk (X3) sebagai variabel penjelas atau independen.

Teknik analisis data yang digunakan untuk memecahkan permasalahan dalam penelitian ini adalah teknik analisis Generalized Linear Model (GLMs). Perbedaan GLMs dengan GLM sebelumnya terletak pada bahwa linieritas dapat ditampilkan dalam satu wadah yang memuat berbagai Teknik analisis statistik. GLMs mengakomodasi berbagai model linier baik itu yang memerlukan asumsi normal maupun yang non normal. Generalized Linier Models (GLMs) bertujuan untuk mengetahui hubungan sebab-akibat, pengaruh dari variabel independen terhadap variabel dependen. Keunggulan GLMs dibandingkan dengan regresi linier biasa terletak pada distribusi (bentuk kurva) variabel dependen. Variabel dependen pada GLMs tidak diisyaratkan berdistribusi normal (kurva lonceng simetris), akan tetapi distribusi-distribusi yang termasuk keluarga eksponensial, yaitu; Binomial, Poisson, Binomial Negative, Normal, Gamma, Invers Gaussian. Linieritas antar variabel dependen dan independen dalam GLMs tidak harus linier, namun transformasi Y melalui link function diasumsikan berhubungan secara linier. Homoskedastisitas varian dalam GLMs juga tidak harus terpenuhi, misalnya pada kasus overdispersi pada regresi logistik.

Model linier tergeneralisir ada tiga komponen yang penting yaitu:

- komponen distribusi, yaitu y berdistribusi keluarga eksponensial,
- komponen prediktor linier, yaitu $\eta = x^T \beta_i$,
- fungsi link yaitu fungsi monoton dan diferensiabel g sehingga $\eta_i = g(\mu_i)$ atau dalam bentuk yang lebih lengkap.

$$g(\mu_i) = g[E(Y_i)] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} \dots \dots \dots (1)$$

Sebaran anggota keluarga eksponensial yang digunakan penulis adalah sebaran Gaussian, dengan karakteristik (1) Sekala kontinu dengan rentang $-\infty < y < \infty$, (2) Simetris, dan (3) Varians tidak bergantung mean (varians konstan).

Kombinasi linier dari prediktor, yang biasa disebut sebagai prediktor linear, $\eta = \sum_{ij=0}^p x_{ij} \beta_j$ terbuka untuk mengambil nilai sembarang bilangan real. Untuk itu diperlukan sebuah fungsi yang menghubungkan dan sekaligus mensinkronkan antara respon dengan prediktor linier. Fungsi ini disebut fungsi penghubung (link function). Dengan demikian fungsi link sekaligus berfungsi untuk menjaga linieritas agar prediktor tetap linier dan normalitas agar ruang rentang antar prediktor linier dan y atau μy tetap sinkron. Diantara fungsi- fungsi link yang dapat digunakan, ada yang disebut fungsi link kanonik yaitu fungsi hubungan yang terjadi pada saat $b(\theta) = \eta = \sum_{j=0}^p \beta_j x_j$. Untuk sebaran Gaussian fungsi link kanonik adalah identitas $\mu_i = \eta_i$.

Estimasi parameter dalam GLMS menggunakan maximum Likelihood Estimation (MLE). MLE merupakan Teknik yang digunakan untuk mencari titik tertentu untuk memaksimumkan sebuah fungsi. Langkah yang perlu dilakukan dengan metode MLE adalah dengan membuat fungsi distribusi likelihood, membuat transformasi fungsi dalam bentuk $\ln$, menurunkan secara parsial terhadap parameter dan menyamadengkan dengan nol.

Pemilihan model terbaik dari beberapa model yang disajikan, dipilih berdasarkan kriteria yang digunakan. Adapun pilihan kriteria sebagai perbandingan model untuk memperoleh model terbaik, antara lain Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) dan Root Mean Square Error (RMSE). AIC adalah suatu kriteria yang menyeimbangkan goodness of fit model berdasarkan nilai likelihood dengan banyaknya parameter dari model. Sehingga, model yang terbaik adalah model dengan nilai Akaike's Information Criterion (AIC) terkecil. Nilai AIC diperoleh dari persamaan sebagai berikut:

$$AIC = -2l(\hat{\theta}) + 2q \dots\dots\dots(2)$$

Dengan $l(\hat{\theta})$ adalah nilai likelihood dari model yang dihadapi dan q adalah banyaknya parameter dalam model.

Pengujian analisis dalam GLMs diperlukan untuk melihat bagaimana model yang dianalisis bisa digunakan atau tidak. Uji yang dilakukan adalah uji inferensia diantaranya adalah uji parameter dan uji kecocokan model. Uji parameter dilakukan untuk mengetahui apakah parameter dalam model GLMs signifikan atau tidak, sedangkan uji kecocokan model dilakukan dengan membandingkan model dengan model yang paling mungkin cocok yaitu model yang didalamnya banyak parameter.

Uji parameter dilakukan dengan mengambil hipotesis berupa koefisien tidak layak masuk model sebagai H_0 dan koefisien layak untuk model sebagai hipotesis alternatifnya. Statistik Uji yang dipakai adalah t dengan rumus

$$t_{hitung} = \frac{r_{x_i y} \sqrt{n-2}}{\sqrt{1-r_{x_i y}^2}} \dots\dots\dots(4)$$

Dengan $r_{x_i y}$ merupakan korelasi y dan x_i untuk parameter β ke- i dan n merupakan jumlah observasi. H_0 akan ditolak jika statistic uji t_{hitung} lebih kecil daripada nilai $t_{tabel(\alpha, n)}$.

Uji kecocokan model mengambil Hipotesis Model tidak layak digunakan untuk hipotesis nol dan model layak digunakan untuk hipotesis alternatifnya. Statistik uji yang digunakan adalah X_p^2 dengan rumus

$$X_p^2 = \sum_{ij} \frac{(f_{ij} - E_{ij})^2}{E_{ij}} \dots\dots\dots(5)$$

Hipotesis nol akan ditolak jika statistic uji $X_{p_{hitung}}^2$ lebih kecil daripada nilai $X_{p_{tabel}}^2$.

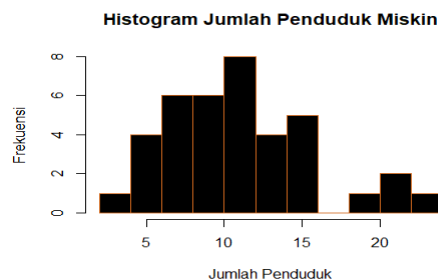
Langkah-langkah dalam analisis GLMs adalah sebagai berikut :

- Mengumpulkan data, kesalahan dalam proses pengumpulan data akan membuat proses analisis menjadi sulit dan hasil yang diberikan akan menjadi rancu.
- Menentukan variable independent dan variable dependen, kedua variable ini merupakan variable yang sering digunakan dalam penelitian karena mempunyai kemampuan analisis yang luas. Penentuan variable ini dari data sebelumnya berdasarkan pada teori keilmuan yang terkait.
- Melakukan statistika diskriptif, ukuran numerik dalam statistic deskriptif menggambarkan beberapa karakteristik data.

- Pengujian linieritas, Pengujian linieritas merupakan syarat sebelum dilakukan uji GLMs, secara umum uji ini bertujuan untuk mengetahui apakah dua variable memiliki hubungan yang linier secara signifikan atau tidak. Data yang baik biasanya memiliki hubungan linier antara variable independent dan data dependent.
- Identifikasi Distribusi variable dependent dan pemilihan fungsi link, Langkah ini dilakukan dengan mencocokkan data pada kurva distribusi. Setelah diketahui distribusi variable dependent kemudian ditentukan fungsi link yang sesuai dengan distribusi tersebut.
- Pemodelan GLMs, melakukan estimasi parameter untuk membuat beberapa model yang mungkin terjadi.
- Uji parameter untuk menyeleksi parameter yang masuk model
- Uji signifikan untuk menyeleksi model yang signifikan
- Pemilihan model terbaik, metode yang digunakan dalam pemilihan model terbaik adalah dengan metode AIC.

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini adalah Luas Wilayah (X_1), Tingkat Pengangguran Terbuka (X_2), dan Kepadatan Penduduk (X_3) sebagai variabel independent, dan variable dependen adalah Jumlah Penduduk Miskin (Y) di Jawa Timur, Tahun 2020. Variabel independe dan variable dependen pada data telah berbentuk numerik. Histogram data dapat dilihat pada gambar berikut



Gambar 1. Histogram jumlah penduduk miskin

Pada Gambar 1 menunjukkan bahwa data mendekati berdistribusi gaussian, sehingga fungsi link yang digunakan adalah dari distribusi gaussian. Selanjutnya, Karena variabel independen yang digunakan ada tiga, maka dilakukan kombinasi dalam pembentukan model. Diperoleh 7 model yang akan dianalisis. Rangkuman Uji Signifikan koefisien dan konstanta untuk 7 model, dapat dilihat pada Tabel 1

Tabel 1. Table rangkuman uji signifikan koefisien dan konstanta untuk 7 model

Parameter	Estimate	Std. Error	t Value	Pr > t
Model 1				
(Intercept)	15.9814789	2.2719115	7.034	4.04e-08
X1	-0.0005016	0.0006837	-0.734	0.46821
X2	-0.3360940	0.4051122	-0.830	0.41253
X3	-0.0012685	0.0004485	-2.829	0.00779
Model 2				
(Intercept)	16.1538238	2.4878796	6.493	1.74e-07

X1	0.0004965	0.0006415	0.774	0.44418
X2	-1.0236714	0.3550111	-2.883	0.00668
Model 3				
(Intercept)	14.5717894	1.5013765	9.706	1.84e-11
X1	-0.0005423	0.0006789	-0.799	0.429795
X3	-0.0014918	0.0003572	-4.177	0.000187
Model 4				
(Intercept)	15.1440216	1.9512948	7.761	4.09e-09
X2	-0.3574310	0.4013923	-0.890	0.37929
X3	-0.0010987	0.0003816	-2.879	0.00676
Model 5				
(Intercept)	9.5888674	1.1001853	8.716	2.14e-10
X1	0.0011387	0.0006599	1.725	0.093
Model 6				
(Intercept)	17.3148	1.9736	8.773	1.82e-10
X2	-1.1191	0.3311	-3.380	0.00176
Model 7				
(Intercept)	13.5643864	0.8105335	16.735	< 2e-16
X3	-0.0013226	0.0002862	-4.622	4.75e-05

Diketahui bahwa H_0 adalah variabel independen mempunyai pengaruh bermakna terhadap variabel dependen dengan tingkat kepercayaan 95%. H_0 ditolak jika $P - value < \alpha$.

Selanjutnya kita melakukan pemilihan model dengan menggunakan AIC dan Hasilnya menunjukkan seperti table dibawah ini

Tabel 1. Table nilai AIC dari model

Model	AIC
Model 1	213.1232
Model 2	219.1536
Model 3	211.8848
Model 4	211.7200
Model 5	225.2529
Model 6	217.7984
Model 7	210.5713

Model yang baik adalah model yang memiliki nilai AIC terkecil. Berdasarkan output di atas, diperoleh bahwa model ke-7 adalah model terbaik, dan merupakan model yang akan digunakan. Maka dengan menggunakan Persamaan (1) diperoleh Model 7;

$$g(\mu_i) = g[E(Y)] = 13,5643864 - 0,0013226X_3$$

Persamaan di atas dapat diketahui bahwa Apabila variabel lain bernilai konstan maka Jumlah Penduduk Miskin (Y) akan berubah dengan sendirinya sebesar nilai konstanta yaitu 13,5643864. Apabila variabel lain bernilai konstan maka Jumlah Penduduk akan berkurang sebesar 0,0013226 setiap satu satuan Kepadatan Penduduk (X_3).

Nilai Pseudo- R^2 sebesar 7,963679% dimana variasi dari variabel dependen dapat dijelaskan oleh model yang terbentuk. Artinya masih banyak variable diluar Kepadatan Penduduk yang bisa dilakukan untuk mengetahui factor-faktor yang mempengaruhi jumlah penduduk miskin di Indonesia.

4. KESIMPULAN

Hasil pemilihan model dengan menggunakan AIC terkecil adalah model dengan persamaan $g(\mu_i) = g[E(Y)] = 13,5643864 - 0,0013226X_3$. Persamaan di disamping dapat disimpulkan bahwa Apabila variabel lain bernilai konstan maka Jumlah Penduduk Miskin (Y) akan berubah dengan sendirinya sebesar nilai konstanta yaitu 13,5643864 dan Apabila variabel lain bernilai konstan maka Jumlah Penduduk Miskin akan berkurang sebesar 0,0013226 setiap satu satuan Kepadatan Penduduk (X3).

Berdasarkan uji signifikan model, variable Kepadatan Penduduk (X3) berpengaruh signifikan terhadap variable Jumlah Penduduk Miskin (Y) dengan tingkat kepercayaan 95%. Nilai Pseudo-R² sebesar 7,963679% menunjukkan variasi dari variabel dependen dapat dijelaskan oleh model yang terbentuk. Analisis ini dapat dijadikan bahan evaluasi bagi Pemerintah Jawa Timur dalam mengambil keputusan untuk kebijakan dalam mengendalikan potensi peningkatan jumlah penduduk miskin di Jawa Timur.

REFERENCES

- Adi P., Sudarsana A., 2018, Analisis Pengaruh Tingkat Pengangguran Terbuka, Kesempatan Kerja, Dan Tingkat Pendidikan Terhadap Tingkat Kemiskinan Pada Kabupaten/Kota Di Provinsi Bali, *Jurnal Ekonomi Pembangunan Universitas Udayana*, No. 3, Vol. 7: <https://ojs.unud.ac.id/index.php/eep/article/view/37696>.
- Agustina, Syechalad, Hamzah, 2020, Pengaruh Jumlah Penduduk, Tingkat Pengangguran dan Tingkat Pendidikan Terhadap Kemiskinan di Provinsi Aceh, <https://doi.org/10.21157/j.ked.hewan.v%25vi%25i.13022>.
- Astutik, Santoso, 2020, Analisis Faktor yang Mempengaruhi Kemiskinan di Jawa Timur Pendekatan Spasial, <https://jimfeb.ub.ac.id/index.php/jimfeb/article/viewFile/7051/6076>.
- BPS Jatim. (2019). Konsep Kemiskinan. Badan Pusat Statistik: Jawa Timur.
- Jamilatuzzahr, Eko C., Herliansyah, 2018, Aplikasi Generalized Linear Model pada R, Ed.1, Innosain, Yogyakarta.
- Indah Agustina, N. (2018). PENGARUH JUMLAH PENDUDUK MISKIN DAN KONSUMSI RUMAH TANGGA TERHADAP INDEKS PEMBANGUNAN MANUSIA DI INDONESIA (Doctoral dissertation, UNIVERSITAS SULTAN AGENG TIRTAYASA).
- Maya S., Abi W., Sudarwanto, 2021, Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model, *Jurnal Statistika dan Aplikasinya*, No. 1, Vol. 5: <https://doi.org/10.21009/JSA.05111>.
- Rosadi, D., 2010, Analisis Ekonometrika & Runtun Waktu Terapan dengan R, Penerbit Andi, Yogyakarta.
- Santi, V. M., & Wiyono, A. (2021). Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model. *Jurnal Statistika dan Aplikasinya*, 5(1), 112-120.
- Soejoto, A, Karisma, A, 2011, Pertumbuhan Ekonomi Dan Pengangguran Terhadap Kemiskinan Di Jawa Timur, *Jurnal Mahasiswa Universitas Negri Surabaya*.