

Pemodelan Kepadatan Penduduk Jawa Tengah dengan Regresi *Robust* Estimasi *Generalized Scale* (GS)

Adibatus Salma Hety¹, Yuliana Susanti², Isnandar Slamet³

^{1,2,3}Program Studi Statistika, Universitas Sebelas Maret

Abstract. Provinsi Jawa Tengah berada pada posisi ke-5 sebagai provinsi paling padat penduduk di Indonesia pada tahun 2021. Tingkat kepadatan penduduk berguna untuk melihat dinamika penduduk guna mengambil berbagai kebijakan. Oleh karena itu, kepadatan penduduk penting untuk dilihat perkembangannya. Kepadatan penduduk dipengaruhi oleh banyak faktor. Pemodelan faktor yang mempengaruhi kepadatan penduduk dapat diselesaikan dengan analisis regresi, namun karena asumsi kenormalan tidak terpenuhi dan terdapat pencilan, maka digunakan regresi yang tahan terhadap pencilan yaitu regresi *robust*. Estimasi parameter regresi *robust* dapat dilakukan dengan beberapa metode. Salah satunya yaitu estimasi *Generalized Scale* (GS). Tujuan penelitian ini yaitu menentukan model regresi untuk kepadatan penduduk di Jawa Tengah tahun 2021 menggunakan regresi *robust* estimasi GS dengan variabel dependen yaitu kepadatan penduduk Jawa Tengah tahun 2021 (Y) dan empat variabel independen yaitu upah minimum kabupaten (X_1), laju pertumbuhan penduduk (X_2), luas wilayah (X_3), dan tingkat pengangguran terbuka (X_4). Berdasarkan hasil analisis diperoleh model regresi untuk estimasi kepadatan penduduk Jawa Tengah tahun 2021 yaitu $\hat{y} = 1562,02 + 0,00019X_1 + 1103,67X_2 - 430,88X_3 - 304,34X_4$. Hasil uji signifikansi parameter menyatakan bahwa keempat variabel independen berpengaruh signifikan terhadap variabel dependen. Nilai R^2 *adjusted* sebesar 79% yang berarti bahwa variabel independen menyumbang sebesar 79% terhadap variabel dependen, sedangkan 21% lainnya dipengaruhi oleh faktor lain diluar model.

Keyword. estimasi GS, kepadatan penduduk, regresi *robust*

1. Pendahuluan

Kepadatan penduduk merupakan perbandingan jumlah masyarakat dengan luas daerah yang dihuninya (Audey dan Ariusni, 2017). Menurut Christiani dkk. (2014) tingginya kepadatan penduduk dapat menyebabkan berbagai persoalan kependudukan, misalnya tingkat kemiskinan, kekurangan lapangan kerja, dan kriminalitas. Selain itu kepadatan penduduk dapat berdampak pada usaha peningkatan kualitas penduduk yang terhambat. Jawa Tengah berada pada posisi ke-5 sebagai provinsi paling padat penduduk di Indonesia (Dukcapil, 2021). Oleh karena itu, kepadatan penduduk di Jawa Tengah perlu diperhatikan perkembangannya. Kepadatan penduduk dipengaruhi oleh banyak faktor. Faktor-faktor yang diduga mempengaruhi kepadatan penduduk antara lain Upah Minimum Kabupaten, laju pertumbuhan penduduk, luas wilayah, dan Tingkat Pengangguran Terbuka. Diperlukan adanya analisis faktor-faktor yang mempengaruhi kepadatan penduduk di Jawa Tengah. Pemodelan faktor tersebut dapat dilakukan dengan analisis regresi, namun karena data berdistribusi tidak normal serta terdapat pencilan maka estimasi regresi dengan Metode Kuadrat Terkecil (MKT) kurang tepat digunakan. Salah satu cara analisis regresi dimana estimasinya tetap menggunakan pencilan yang sudah diolah yaitu dengan menggunakan model regresi *robust*. Terdapat beberapa metode estimasi parameter Regresi

robust, salah satunya yaitu estimasi *Generalized Scale* (GS). Estimasi GS merupakan pengembangan dari estimasi *Scale* (S). Estimasi ini menggunakan standar deviasi residu berpasangan sebagai skalanya. Selain itu pencilan yang terdapat pada variabel dependen dan variabel independen dapat diatasi oleh estimasi GS.

2. Landasan Teori

2.1. Regresi Linear Berganda

Regresi linear berganda adalah regresi yang memiliki variabel independen lebih dari satu. Teknik regresi linear berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel independen terhadap variabel dependen. Estimasi regresi ini menggunakan Metode Kuadrat Terkecil (MKT). Model umum regresi linear berganda adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i, i = 1, 2, \dots, n \quad (1)$$

dengan e_i berdistribusi normal dengan *mean* nol dan variansi konstan dan residu tidak saling berkorelasi.

2.2. Uji Asumsi Klasik.

Regresi linear berganda memiliki beberapa syarat yang dikenal dengan uji asumsi klasik. Terdapat empat uji asumsi klasik, yang pertama yaitu normalitas. Salah satu metode untuk menguji normalitas pada residu adalah uji Saphiro Wilks. Uji ini valid untuk menguji kenormalan dengan ukuran sampel antara 3 sampai 50 (Peng dkk, 2004). Pada taraf signifikansi α , asumsi kenormalan tidak terpenuhi jika nilai $p\text{-value} < \alpha$. Uji asumsi kedua yaitu uji homoskedastisitas. Uji ini digunakan untuk menguji ada tidaknya ketidaksamaan variansi dari residu suatu pengamatan ke pengamatan lain (Gujarati, 2003). Uji homoskedastisitas dari variansi residu dapat dilakukan dengan uji Breusch Pagan. Pada taraf signifikansi α , variansi residu dinyatakan tidak homogen jika nilai $p\text{-value} < \alpha$. Uji ketiga yaitu uji non autokorelasi yang bertujuan untuk menguji ada tidaknya korelasi antara residu dalam model regresi linear. Uji non autokorelasi dilakukan dengan menggunakan uji Durbin Watson (Gujarati, 2003). Asumsi non autokorelasi terpenuhi jika $p\text{-value} > \alpha$. Uji asumsi keempat yaitu uji non multikolinearitas. Uji ini dilakukan untuk mengetahui ada tidaknya korelasi antara variabel independen. Uji ini dilakukan dengan menentukan nilai *Variance Inflation Factor* (VIF). Apabila nilai dari $VIF < 10$ maka asumsi non multikolinearitas terpenuhi (Susanti dkk, 2021).

2.3. Pencilan.

Pencilan adalah pengamatan yang jauh dari pusat data. Salah satu metode untuk menunjukkan suatu pencilan yaitu dengan nilai *Difference in Fitted Value* (DFFITS) (Susanti, dkk. 2021). Rumus DFFITS adalah sebagai berikut:

$$(DFFITS) = t_i \left(\frac{h_{ii}}{1-h_{ii}} \right)^{1/2}, i = 1, 2, \dots, n \quad (2)$$

Suatu data dikatakan pencilan jika nilai dari $|(DFFITS)_i| > 2\sqrt{(k+1)/n}$ dengan k adalah banyaknya variabel independen.

2.4. Regresi Robust Estimasi Generalized Scale (GS)

Estimasi GS merupakan solusi untuk meminimasi estimasi M dengan skala selisih residu berpasangan. Misalkan $\hat{\beta}$ hasil estimasi dari β , $e(\hat{\beta}) = (e_1(\hat{\beta}), \dots, e_n(\hat{\beta}))'$ merupakan vektor residu dengan vektor selisih residu berpasangan sebagai berikut.

$$\begin{aligned} ei(\beta) &= yi - \sum_{j=0}^k x_{ij}\beta_j, 1 \leq i \leq n \\ \Delta ei(\beta) &= (\Delta e_{12}(\beta), \dots, \Delta e_{(n-1)n}(\beta))' \\ \Delta e_{ii'}(\beta) &= ei(\beta) - ei'(\beta), 1 \leq i \leq i' \leq n \end{aligned} \quad (4)$$

dengan $\hat{\sigma}_{GS} = 1,483$ Mr dimana Mr adalah median dari selisih residu berpasangan, yang merupakan solusi:

$$\frac{1}{\binom{n}{2}} \sum_{i < i'} \rho \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right) = \delta \quad (5)$$

ρ adalah fungsi pembobot Tukey's Biweight dan δ adalah nilai *breakdown point*. Estimasi $\hat{\beta}$ dapat dinyatakan dalam bentuk lain dengan mencari penurunan terhadap β diperoleh:

$$\begin{aligned} \sum_{i < i'}^n \rho' \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right) &= 0 \\ \sum_{i < i'}^n \psi \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right) &= 0 \end{aligned} \quad (6)$$

ψ disebut fungsi pengaruh yang merupakan turunan dari ρ . Bentuk ini dapat ditulis sebagai berikut:

$$w_i(u_i) = \frac{\psi(u_i)}{u_i} \quad (7)$$

Persamaan (7) merupakan fungsi pembobot dimana nilai $u_i = \frac{\Delta e_{ii'}}{\hat{\sigma}_{GS}}$ dan nilai *tuning constant* yang digunakan adalah 0,9958 (Chen, C. 2002). Nilai residu awal yang digunakan pada estimasi GS adalah residu berpasangan yang diperoleh dari residu estimasi M yang sudah konvergen. Proses dilanjutkan dengan MKT terbobot secara iterasi atau yang disebut *Iteratively Reweighted Least Square* (IRLS) hingga mencapai konvergen.

2.5. Uji Signifikansi Parameter

Uji signifikansi parameter meliputi uji simultan dan parsial. Uji simultan digunakan untuk mengetahui bahwa variabel independen secara bersamaan berpengaruh terhadap variabel dependen. Uji ini dilakukan menggunakan uji F. Pada taraf signifikansi α , paling tidak ada satu variabel independen yang berpengaruh terhadap variabel dependen jika nilai $p - value > \alpha$ atau $F_{hitung} > F_{(\alpha, k, n-k-1)}$. Uji parameter parsial digunakan untuk mengetahui variabel independen mana saja dalam model regresi linear yang berpengaruh pada variabel dependennya. Uji ini dilakukan menggunakan uji t. Pada taraf signifikansi α , variabel independen ke-k berpengaruh terhadap variabel dependen jika nilai $p - value < \alpha$ atau $|t_{hitung}| > t_{\frac{\alpha}{2}; (n-k-1)}$.

2.6. Koefisien Determinasi (R^2)

R^2 mengukur proporsi variasi dalam variabel dependen (Y) yang dijelaskan oleh variabel independen (X). R^2 *adjusted* juga mengukur proporsi variasi, tetapi menyesuaikan jumlah variabel independen dalam model. Rumus R^2 *adjusted* menurut Kuzmič, P. dkk. (2004) adalah sebagai berikut:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right] \quad (8)$$

3. Metode Penelitian

Penelitian ini menggunakan data data sekunder tahun 2021 yang diambil dari *website* Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Obyek penelitian yang akan dijadikan sebagai variabel dependen yaitu kepadatan penduduk di Jawa Tengah berdasarkan 35 kabupaten/kota sebagai Y. Sedangkan variabel independent yang digunakan ada 5, yaitu Upah Minimum Kabupaten/Kota (UMK) sebagai X_1 , laju pertumbuhan penduduk sebagai X_2 , persentase luas wilayah sebagai X_3 , dan Tingkat Pengangguran Terbuka (TPT) sebagai X_4 .

Tahapan analisis pada penelitian ini adalah; menentukan prediksi model menggunakan metode MKT, menguji asumsi regresi linear berganda, mendeteksi pencilan dengan metode DFITS, menghitung nilai residu e_i , menghitung selisih residu e_i berpasangan, menghitung $\hat{\sigma}_{GS}$, menghitung nilai pembobot $w_i(u_i)$,

menggunakan MKT terbobot untuk mendapatkan penduga kuadrat terkecil terbobot, menjadikan selisih residu berpasangan langkah (e) sebagai selisih residu berpasangan awal langkah (d) sehingga diperoleh nilai $\hat{\sigma}_{GS}$ dan pembobot $w_i(u_i)$ yang baru, melakukan pengulangan iterasi sampai konvergen, melakukan uji signifikansi model dengan uji F simultan dan t parsial, dan menginterpretasikan modelnya.

4. Hasil Penelitian

4.1. Regresi Linear Berganda.

Model regresi dari kepadatan penduduk kabupaten/kota di Jawa Tengah tahun 2021 dengan MKT yaitu:

$$\hat{y} = 8444 + 0,00239X_1 + 3223X_2 - 751,4X_3 - 544,6X_4$$

Nilai R^2 *adjusted* sebesar 77% yang berarti bahwa variabel upah minimum kabupaten (X_1), laju pertumbuhan penduduk (X_2), luas wilayah (X_3), dan tingkat pengangguran terbuka (X_4) menyumbang sebesar 77% pada variabel kepadatan penduduk sedangkan 23% lainnya dipengaruhi oleh variabel lain diluar model. Hasil uji simultan menyatakan bahwa paling tidak ada satu variabel dependen yang berpengaruh signifikan terhadap variabel dependen. Selain itu, hasil uji parsial menyatakan bahwa keempat variabel independen masing-masing berpengaruh pada variabel dependen.

4.2. Uji Asumsi Klasik.

Berdasarkan uji Shapiro Wilks didapatkan nilai $p\text{-value} = 0,026 < \alpha = 0,05$ yang berarti bahwa residu tidak berdistribusi normal. Uji Homoskedastisitas dengan Breusch-Pagan menghasilkan nilai $p\text{-value} = 0,07755 > \alpha = 0,05$ yang berarti bahwa variansi residu homogen. Uji Non Autokorelasi dengan Durbin Watson menghasilkan nilai $p\text{-value} = 0,426 > \alpha = 0,05$ yang berarti tidak terdapat autokorelasi antar residu. Nilai VIF untuk keempat variabel independen yaitu $X_1 = 1,176$, $X_2 = 1,189$, $X_3 = 1,218$, dan $X_4 = 1,16$. Keempat nilai VIF tersebut kurang dari 10, maka dapat disimpulkan bahwa asumsi non multikolinearitas terpenuhi.

4.3. Pencilan.

Pendeteksian pencilan dilakukan dengan menggunakan nilai DFFITS. Terdapat empat pengamatan pencilan yaitu data ke-1, 28, 31, 33, dan 35 pada variabel independen dimana nilai $|DFFITS| > 2\sqrt{\frac{(k+1)}{n}}$ = 0,75592.

4.4. Regresi Robust Estimasi Generalized Scale (GS)

Penentuan model regresi *robust* estimasi GS dimulai dengan menghitung estimasi awal koefisien regresi atau $\hat{\beta}^0$. Berdasarkan perhitungan MKT, didapatkan $\hat{\beta}^0 = (8444; 0,00239; 3223; -751,4; -544,6)$. Selanjutnya menghitung nilai e_i , selisih e_i berpasangan, $\hat{\sigma}_{GS}$, dan $w_i(u_i)$. Setelah itu dilakukan proses iterasi menggunakan MKT terbobot dengan pembobot $w_i(u_i)$ hingga didapatkan nilai $\hat{\beta}$ yang konvergen. Iterasi dari estimasi GS disajikan pada Tabel 1.

Tabel 1. Hasil iterasi

Iterasi ke-	$\hat{\beta}$
1	(4309,89145; 0,00103; 1829,33909; -538,45021; -432,61015)
2	(2486,88086; 0,00024; 1370,26445; -458,81371; -356,07293)
3	(1840,92404; 0,00005; 1191,54937; -35,24964; -322,78491)
...	...
28	(1562,02117; 0,00019; 1103,67193; -430,88432; -304,33810)
29	(1562,02116; 0,00019; 1103,67193; -430,88432; -304,33810)
30	(1562,02116; 0,00019; 1103,67193; -430,88432; -304,33810)

Hasil iterasi menunjukkan bahwa parameter koefisien regresi konvergen pada iterasi ke-30. Sehingga

model yang didapatkan yaitu sebagai berikut:

$$\hat{y} = 1562,02 + 0,00019X_1 + 1103,67X_2 - 430,88X_3 - 304,34X_4$$

Nilai R^2 *adjusted* yang dihasilkan dari model tersebut yaitu sebesar 79% yang berarti bahwa upah minimum kabupaten (X_1), laju pertumbuhan penduduk (X_2), luas wilayah (X_3), dan tingkat pengangguran terbuka (X_4) menyumbang sebesar 79% pada kepadatan penduduk, sedangkan 21% lainnya dipengaruhi oleh faktor lain diluar model.

4.5. Uji Signifikansi Parameter

Hasil uji simultan menghasilkan nilai $F_{hitung} = 4,45 > F_{tabel(0,05;3;30)} = 2,69$ dan $p-value = 0,000 < \alpha = 0,05$ yang berarti paling tidak terdapat satu dari variabel upah minimum kabupaten, laju pertumbuhan penduduk, luas wilayah, atau tingkat pengangguran terbuka yang berpengaruh signifikan terhadap kepadatan penduduk. Selanjutnya dilakukan uji t dan didapatkan nilai seperti pada tabel 2 berikut:

Tabel 2. Hasil uji t			
Variabel	t_{hitung}	$p-value$	$T_{tabel(0,025;30)}$
X_1	2,41	0,037	2,04
X_2	-3,56	0,001	
X_3	-7,48	0,000	
X_4	6,60	0,000	

Berdasarkan tabel 2, diperoleh keempat nilai $|t_{hitung}| > t_{tabel}$ dan $p-value < \alpha = 0,05$, maka dapat disimpulkan bahwa keempat variabel yaitu upah minimum kabupaten, laju pertumbuhan penduduk, luas wilayah, dan tingkat pengangguran terbuka masing-masing berpengaruh signifikan terhadap kepadatan penduduk.

5. Kesimpulan

Model regresi *robust* estimasi *Generalized Scale* (GS) untuk kepadatan penduduk kabupaten/kota di Jawa Tengah yaitu:

$$\hat{y} = 1562,02 + 0,00019X_1 + 1103,67X_2 - 430,88X_3 - 304,34X_4$$

Nilai R^2 *adjusted* yang dihasilkan dari model tersebut sebesar 79%. Berdasarkan model yang diperoleh, maka setiap kenaikan upah sebesar Rp100.000 akan menaikkan kepadatan penduduk sebesar 19 jiwa/km², kenaikan laju pertumbuhan penduduk sebesar 1% akan menaikkan kepadatan penduduk sebesar 1103 jiwa/km², kenaikan presentase luas wilayah sebesar 1% akan menurunkan kepadatan penduduk sebesar 430 jiwa/km², dan kenaikan TPT sebesar 1 satuan akan menurunkan kepadatan penduduk sebesar 304 jiwa/km².

6. Daftar Pustaka

- [1] Audey, R. P., dan Ariusni. 2017. Pengaruh Kualitas Sumber Daya Manusia terhadap Tingkat Kriminalitas di Indonesia. *Jurnal Kajian Ekonomi Dan Pembangunan*. Vol. 1, No. 2, pp. 653-666.
- [2] Bayutama, I., 2017. Perbandingan keakuratan model regresi menggunakan estimasi M dan estimasi LTS pada produksi padi di Indonesia. Prosiding Seminar Nasional, Universitas Sebelas Maret.
- [3] Bkti, D. R. 2011. Regresi Robust dengan M-Estimation. statisticsanalyst.files.wordpress.com. Diunduh Tanggal 29 Juni 2022.
- [4] Cahya, A. 2019. Analisis Faktor-Faktor Penyebab Migrasi Penduduk Jawa Akibat Pertumbuhan Penduduk yang Tinggi. Makalah Tugas Akhir. Banjarmasin: Universitas Lambung Mangkurat.
- [5] Cahyawati, S., Tanuji, dan Abdiati. 2009. Efektivitas Metode Regresi Robust Penduga Welsch dalam Mengatasi Pencilan pada Pemodelan Regresi Linear Berganda. *Jurnal Penelitian Sains*. Vol. 12, No.1(A).

- [6] Cahyono W.E. 2011. Kajian Tingkat Pencemaran Sulfur Dioksida dari Industri di Beberapa Daerah di Indonesia. *Berita Dirgantara*. Vol. 12, No. 4, pp. 132.
- [7] Chen, C. 2002. Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Paper 265-27. North Carolina: SAS Institute Inc. An R Companion to Applied Regression, Second Edition. New York.
- [8] Christiani, C., Tedjo, P., dan Martono, B. 2014. Analisis Dampak Kepadatan Penduduk terhadap Kualitas Hidup Masyarakat Provinsi Jawa Tengah. *Jurnal Ilmiah UNTAG*. pp. 102– 114.
- [9] Edriani, T. S., Rahmadani, A., dan Noor, D. 2021. Analisis Hubungan Kepadatan Penduduk dengan Pola Penyebaran COVID-19 Provinsi DKI Jakarta menggunakan Regresi Robust. *Indonesian Journal of Applied Mathematics*. Vol. 1, No. 2, pp. 51 - 60.
- [10] Faeni, Y., A. 2021. Pemodelan Hubungan Kepadatan Penduduk dan Indeks Kualitas Udara (IKU) di Indonesia Menggunakan Regresi Kuantil Smoothing Splines. *Jurnal Sistem Cerdas*. Vol. 4, No. 1, pp. 56-66.
- [11] Gujarati, D. N. 2003. *Basic Econometrics*. The McGrawHill Companies Inc, New York.
- [12] Kuzmič, P., Hill, C., dan Janc, W. 2004. Practical Robust Fit of Enzyme Inhibition Data. *Methods in Enzymology*, Academic Press. Vol. 383, pp. 366-381.
- [13] Peng, G., Lilly, E. and Indiana. 2004. *Testing Normality of Data Using SAS*. United States.
- [14] Susanti, Y., Pratiwi, H., dan Qona'ah, N. 2021. *Regresi Robust Teori dan Terapannya*. UNS Press, Surakarta.
- [15] Susanti, Y., Pratiwi, H., H., S. S., and Liana, T. 2014. M Estimation, S Estimation, and MM Estimation. *International Journal of Pure and Applied Mathematics*. 90: 349-360.
- [16] Widiyatmoko, H. 2018. Analisis Faktor Pengaruh Komposisi Penduduk, Aksesibilitas dan Sosio-Ekonomi Terhadap Kepadatan Penduduk di Kabupaten Klaten. Makalah Tugas Akhir. Surakarta: Universitas Muhammadiyah Surakarta.