

# Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Jawa Tengah dengan Analisis Regresi *Robust* Estimasi *Generalized Scale* (GS)

Andini Wati<sup>1</sup>, Yuliana Susanti<sup>2</sup>, Hasih Pratiwi<sup>3</sup>

<sup>1,2,3</sup>Program Studi Statistika, Fakultas MIPA, Universitas Sebelas Maret

**Abstract.** Pembangunan suatu negara memiliki tujuan utama untuk menyejahterakan rakyatnya sehingga keberhasilan pembangunan tidak hanya dinilai dari pertumbuhan ekonominya namun juga dari kualitas manusianya. Indeks Pembangunan Manusia (IPM) adalah salah satu indikator yang berguna untuk mengukur capaian pembangunan manusia di suatu wilayah tertentu. Berdasarkan data BPS, IPM di Jawa Tengah selalu mengalami kenaikan sejak tahun 2017 dan mencapai kategori tinggi dalam capaian status pembangunan. Penelitian ini menggunakan analisis regresi *robust* estimasi GS karena data IPM di Jawa Tengah mengandung pencilan. Tujuan dari penelitian ini adalah untuk mengetahui variabel yang signifikan mempengaruhi IPM di Jawa Tengah. Penelitian ini menggunakan data yang bersumber dari BPS dengan variabel dependen IPM di Jawa Tengah tahun 2021 dan keempat variabel independen yaitu harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita disesuaikan, dan umur harapan hidup. Hasil yang diperoleh pada penelitian ini yaitu keempat variabel independen berpengaruh signifikan terhadap IPM dengan nilai  $R^2$  *adjusted* sebesar 99.62% yang berarti bahwa keempat variabel independen tersebut merupakan faktor-faktor yang mempengaruhi variabel IPM di Jawa Tengah 2021.

**Keyword.** Analisis Regresi, Regresi *Robust*, Estimasi GS, Indeks Pembangunan Manusia

## 1. Pendahuluan

Manusia merupakan pembentuk sebuah bangsa dan menjadi kekayaan bangsa yang sebenarnya. Sudah sewajarnya apabila manusia diharapkan dapat menjadi objek sekaligus subjek dalam pembangunan sehingga dapat memberikan perubahan yang positif dalam bidang ekonomi, politik, sosial, budaya, dan lainnya. Tujuan utama dari suatu pembangunan adalah kesejahteraan rakyat sehingga keberhasilan pembangunan dinilai bukan hanya berdasarkan pertumbuhan ekonominya saja, namun juga dari peningkatan kualitas manusianya.

Indeks Pembangunan Manusia (IPM) adalah salah satu indikator yang berguna untuk mengukur capaian pembangunan manusia di suatu wilayah tertentu [1]. Berdasarkan data yang bersumber dari Badan Pusat Statistik (BPS), sejak tahun 2017, Jawa Tengah telah mencapai kategori tinggi dalam capaian status pembangunan.

Penelitian terkait IPM sebelumnya pernah dilakukan di Jawa Timur tahun 2019 dengan menggunakan model regresi *robust* estimasi-M [2]. Selanjutnya, penelitian juga dilakukan pada IPM di Jawa Tengah tahun 2019-2020 dengan pendekatan regresi *robust* fungsi pembobot *Bisquare Tukey* pada estimasi-M dan estimasi-S [3].

Berdasarkan data BPS, tingkat IPM di Jawa Tengah terus mengalami kenaikan setiap tahunnya sehingga dilakukan penelitian untuk menguji faktor yang signifikan mempengaruhi IPM di Jawa Tengah dengan keempat variabel independen yaitu harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita disesuaikan, dan umur harapan hidup. Data IPM yang digunakan pada penelitian ini mengandung pencilan sehingga penelitian ini menggunakan metode regresi *robust* estimasi *Generalized Scale* (GS). Regresi *robust* adalah metode analisis yang digunakan untuk menganalisis data yang memiliki pencilan. Regresi *robust* berguna untuk mendeteksi pencilan serta memberikan hasil resisten terhadap adanya pencilan tersebut [4]. Regresi *robust* estimasi GS merupakan solusi minimisasi estimasi *Maximum likelihood* (M) dengan residu skala berpasangan yang merupakan perluasan dari estimasi *Scale* (S).

## 2. Metode Penelitian

Penelitian dilakukan melalui pemodelan pada data IPM di Provinsi Jawa Tengah menurut kabupaten/kota tahun 2021 dengan regresi *robust* estimasi GS. Jumlah data yang digunakan pada penelitian ini yaitu sebanyak 35 data dan dengan menggunakan 4 variabel independen yaitu harapan lama sekolah ( $X_1$ ), rata-rata lama sekolah ( $X_2$ ), pengeluaran per kapita disesuaikan ( $X_3$ ), dan umur harapan hidup ( $X_4$ ). Data tersebut merupakan data sekunder yang bersumber dari *website* BPS.

### 2.1. Regresi linear berganda

Regresi linear berganda yaitu regresi yang memiliki variabel independen lebih dari satu [5]. Dalam regresi linear berganda  $\varepsilon_i$  berdistribusi normal dengan mean nol dan varians konstan dan residu tidak memiliki korelasi satu sama lain. Estimator dari regresi ini menggunakan Metode Kuadrat Terkecil (MKT) [6]. Bentuk matriks dari regresi linear berganda sebagai berikut:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I) \quad (1)$$

dengan

$Y$  : vektor variabel dependen berukuran  $n \times 1$

$X$  : matriks variabel independen berukuran  $n \times (k + 1)$

$\beta$  : vektor parameter regresi berukuran  $(k + 1) \times 1$

$\varepsilon$  : vektor sesatan berukuran  $n \times 1$

Sedangkan nilai estimasi untuk  $Y$  dan  $\varepsilon$  adalah

$$\hat{Y} = X\hat{\beta} \text{ dan } \hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} \quad (2)$$

Pada model regresi dilakukan uji asumsi klasik untuk melihat apakah model memenuhi asumsi klasik atau tidak. Uji asumsi klasik yang dilakukan yaitu uji normalitas, homoskedastisitas, non autokorelasi, dan multikolinieritas. Regresi linear diasumsikan setiap  $\varepsilon_i$  berdistribusi normal dengan lambang  $\varepsilon_i \sim N(0, \sigma^2)$  [7]. Salah satu uji normalitas pada residu yang dapat digunakan yaitu uji Shapiro Wilk. Asumsi klasik selanjutnya yang harus terpenuhi adalah tidak adanya gejala heteroskedastisitas [8]. Uji homoskedastisitas yang digunakan yaitu uji Breusch Pagan. Uji ini dirumuskan dengan  $\phi = \frac{1}{2}$  (jumlah kuadrat regresi). Variansi residu homogen dinyatakan dengan  $H_0$  dan  $H_1$  merupakan sebaliknya.  $H_0$  ditolak jika nilai  $p - \text{value} < \alpha$ , dengan  $k$  yaitu banyaknya variabel independen. Uji non autokorelasi digunakan untuk melihat ada atau tidaknya korelasi antara residu dalam model. Salah satu uji statistik yang digunakan untuk uji non autokorelasi yaitu uji Durbin Watson. Pada uji ini  $H_0$  menyatakan tidak terdapat autokorelasi antar residu dan  $H_1$  sebaliknya.  $H_0$  ditolak jika nilai  $p - \text{value} < \alpha$ . Pendeteksian suatu model regresi terhadap multikolinieritas dapat dilakukan dengan melihat nilai *Variance Inflation Factors* (VIF) [9]. Apabila VIF bernilai lebih dari 10 maka menunjukkan multikolinieritas yang kuat.

### 2.2. Pencilan

Salah satu metode untuk menunjukkan suatu pencilan yaitu dengan nilai *Difference in Fitted Value* (DFFITS). DFFITS merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke- $i$  terhadap nilai estimasi  $\hat{Y}_i$  [6]. Suatu data dikatakan pencilan jika nilai dari  $|(DFFITS)_i| > 2\sqrt{(k + 1)/n}$  dengan  $k$  adalah banyaknya variabel independen.

### 2.3. Regresi Robust Estimasi GS

Estimasi *Generalized Scale* (GS) adalah solusi minimasi estimasi M dengan skala selisih residu yang digunakan adalah residu skala berpasangan. Estimasi GS dengan  $\hat{\sigma}_{GS}$  diperoleh dari estimasi M skala selisih residu berpasangan yang merupakan solusi :

$$\frac{1}{\binom{n}{2}} \sum_{i < i'} \rho \left( \frac{\Delta e_{ii'}(\beta)}{\hat{\sigma}_{GS}} \right) = \delta \quad [10]$$

dengan  $\rho$  adalah fungsi pembobot *Tukey's Biweight*,  $\delta$  adalah nilai *breakdown point*, dan

$$\hat{\sigma}_{GS} = \frac{\text{median } |e_i - \text{median}(e_i)|}{0.6745}.$$

Estimasi GS  $\hat{\beta}$  dapat dinyatakan dalam bentuk lain dengan mencari penurunan terhadap  $\beta$  diperoleh :

$$\sum_{i < i'}^n \rho' \left( \frac{\Delta e_{ii'}(\beta)}{\hat{\sigma}_{GS}} \right) = \sum_{i < i'}^n \psi \left( \frac{\Delta e_{ii'}(\beta)}{\hat{\sigma}_{GS}} \right) = 0 \quad (3)$$

Turunan dari  $\rho$  yaitu  $\psi$  yang disebut sebagai fungsi pengaruh. Bentuk ini dapat ditulis dengan  $w_{iGS}(u_i) = \frac{\psi(u_i)}{u_i}$  merupakan fungsi pembobot dimana nilai  $u_i = \frac{\Delta e_{ii'}}{\hat{\sigma}_{GS}}$  dan

$$w_{iGS} = \begin{cases} \left(1 - \left(\frac{u_i}{0.9958}\right)^2\right)^2 & |u_i| \leq 0.9958, \\ 0 & |u_i| > 0.9958 \end{cases} \quad \begin{matrix} \text{iterasi} = 1 \\ \text{iterasi} > 1 \end{matrix}$$

Nilai residu awal yang digunakan pada estimasi GS adalah residu berpasangan yang diperoleh dari residu estimasi M yang sudah konvergen. Proses dilanjutkan dengan MKT terboboti secara iterasi yang disebut *Iteratively Reweighted Least Square (IRLS)* hingga mencapai konvergen [11]. Berikut algoritma untuk mengestimasi parameter model regresi *robust* estimasi GS:

1. Menentukan estimasi awal dengan MKT sehingga diperoleh nilai awal  $\hat{\beta}_0$ .
2. Menghitung nilai residu berpasangan  $\Delta e_i(\beta) = \Delta e_{12}(\beta), \dots, \Delta e_{(n-1)n}(\beta)$ .
3. Menghitung nilai  $\hat{\sigma}_{GS}$ .
4. Menghitung nilai  $u_i = \frac{e_i}{\hat{\sigma}_{GS}}$ .
5. Menghitung pembobot  $w_{iGS}$ .
6. Menghitung estimator  $\hat{\beta}$  dengan metode *iteratively reweighted least squares* dengan pembobot  $w_{iGS}$ .
7. Mengulangi langkah (2) hingga (6) sehingga diperoleh nilai  $\hat{\beta}$  yang konvergen

### 2.4. Teknik analisis data

Berikut langkah-langkah analisis yang dilakukan pada penelitian ini:

1. Mengestimasi parameter regresi linear dengan Metode Kuadrat Terkecil (MKT).
2. Menguji asumsi klasik dari model MKT.
3. Mendeteksi pencilan.
4. Mengestimasi parameter regresi *robust* dengan menggunakan estimasi GS.
5. Menguji signifikansi parameter model regresi *robust* estimasi GS.
6. Membuat kesimpulan.

## 3. Hasil Penelitian

Pada subbab ini menunjukkan hasil analisis yang dilakukan pada penelitian ini. Hasil-hasil tersebut meliputi metode kuadrat terkecil, identifikasi pencilan, estimasi model regresi *robust* estimasi GS, dan uji signifikansi parameter.

### 3.1. Metode kuadrat terkecil

Pemodelan regresi dengan menggunakan metode MKT pada data IPM di Jawa Tengah menurut kabupaten/kota tahun 2021 yaitu

$$\hat{Y} = 5.814 + 0.9345X_1 + 1.304X_2 + 0.0008746X_3 + 0.4534X_4 \quad (1)$$

Nilai  $R^2$  *adjusted* yang didapatkan sebesar 99.85% yang berarti bahwa variabel independen yang digunakan yaitu harapan lama sekolah ( $X_1$ ), rata-rata lama sekolah ( $X_2$ ), pengeluaran per kapita disesuaikan ( $X_3$ ), dan umur harapan hidup ( $X_4$ ) mampu menjelaskan variabel dependen yaitu IPM sebesar 99.85% sedangkan sisanya yaitu sebesar 0.15% dipengaruhi dan dijelaskan variabel lain yang tidak ada dalam model.

Hasil uji asumsi klasik dari model didapatkan bahwa dari empat asumsi klasik, model tidak memenuhi asumsi normalitas. Pada uji normalitas menggunakan Shapiro Wilk diperoleh  $p\text{-value} = 0.0148 < \alpha = 0.05$  maka  $H_0$  gagal ditolak yang berarti bahwa residu berdistribusi tidak normal. Pada uji homoskedastisitas dengan menggunakan uji Breusch-Pagan didapatkan  $p\text{-value} = 0.5816 > \alpha = 0.05$  sehingga  $H_0$  gagal ditolak yang berarti bahwa variansi sisaan homogen. Uji asumsi selanjutnya yaitu uji non autokorelasi yang dilakukan dengan menggunakan uji Durbin-Watson. Berdasarkan uji yang dilakukan diperoleh  $p\text{-value} = 0.3815 > 0.05$  maka  $H_0$  gagal ditolak yang berarti bahwa tidak terdapat autokorelasi. Uji multikolinieritas dilakukan dengan menggunakan nilai *Variance Inflation Factors* (VIF). Diperoleh nilai VIF untuk  $X_1 = 5.853443$ ,  $X_2 = 8.767351$ ,  $X_3 = 3.322934$ ,  $X_4 = 2.121159$  yang berarti nilai VIF dari keempat variabel  $X < 10$  maka  $H_0$  gagal ditolak sehingga dapat disimpulkan tidak terdapat multikolinearitas antar variabel independen.

### 3.2. Identifikasi pencilan

Pendeteksian pencilan dilakukan dengan menggunakan metode DFFITS. Melalui perhitungan nilai DFFITS data dikatakan merupakan pencilan jika  $|DFFITS| > 2\sqrt{\frac{k+1}{n}} = 2\sqrt{\frac{5}{35}} = 0.75593$ .

Hasil perhitungan menunjukkan nilai DFFITS pada pengamatan ke-27 = -0.893733174, pengamatan ke-32 = -1.580238844, pengamatan ke-33 = -1.108864339. Nilai DFFITS pada pengamatan ke-27, 32, dan 33 > nilai pembanding yaitu 0.75593 sehingga disimpulkan bahwa data tersebut merupakan pencilan.

### 3.3. Estimasi model regresi robust estimasi GS

Awal proses perhitungan dimulai dengan menentukan estimasi awal koefisien regresi yang diperoleh dari MKT yaitu  $\hat{\beta}^0 = (5.814; 0.9345; 1.304; 0.0008746; 0.4534)$ . Kemudian dilakukan perhitungan regresi *robust* estimasi GS sesuai algoritma dan menghitung estimator  $\hat{\beta}$  dengan metode *iteratively reweighted least squares* dengan pembobot  $w_{iGS}$  hingga mendapatkan hasil  $\hat{\beta}$  yang konvergen. Berikut merupakan nilai  $\hat{\beta}$  pada setiap iterasi dengan menggunakan estimasi GS.

**Tabel 1.** Regresi *robust* estimasi GS

| Iterasi | Estimasi GS     |                 |                 |                 |                 |
|---------|-----------------|-----------------|-----------------|-----------------|-----------------|
|         | $\hat{\beta}^0$ | $\hat{\beta}^1$ | $\hat{\beta}^2$ | $\hat{\beta}^3$ | $\hat{\beta}^4$ |
| 1       | 4.68968         | 1.06299         | 1.26023         | 0.00090         | 0.45754         |
| 2       | 4.36378         | 1.24180         | 1.24180         | 0.00091         | 0.45515         |
| 3       | 4.29362         | 1.12716         | 1.23529         | 0.00092         | 0.45230         |
| ...     | ...             | ...             | ...             | ...             | ...             |
| 26      | 4.30453         | 1.15140         | 1.22981         | 0.00093         | 0.44626         |
| 27      | 4.30454         | 1.15139         | 1.22981         | 0.00093         | 0.44626         |
| 28      | 4.30454         | 1.15139         | 1.22981         | 0.00093         | 0.44626         |

Berdasarkan Tabel 1 didapatkan bahwa pada iterasi 28 nilai  $\hat{\beta}$  yang baru memiliki nilai yang sama dengan nilai  $\hat{\beta}$  awal sehingga proses berhenti pada iterasi tersebut. Jadi, model regresi yang diperoleh sebagai berikut

$$\hat{Y} = 4.30454 + 1.15139X_1 + 1.22981X_2 + 0.00093X_3 + 0.44626X_4 \quad (5)$$

Model persamaan regresi (5) menunjukkan bahwa apabila keempat variabel independen bernilai nol atau konstan maka nilai IPM adalah sebesar 4.30454, setiap kenaikan satu satuan harapan lama sekolah akan menaikkan IPM sebesar 1.15139, setiap kenaikan satu satuan rata-rata lama sekolah akan menaikkan IPM sebesar 1.22981, setiap kenaikan satu satuan pengeluaran per kapita disesuaikan akan menaikkan IPM sebesar 0.00093, dan setiap kenaikan satu satuan umur harapan hidup saat bayi akan menaikkan IPM sebesar 0.44626.

Nilai  $R^2$  *adjusted* yang diperoleh sebesar 99.62% berarti bahwa variabel independen yang digunakan yaitu harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita disesuaikan, dan umur harapan hidup saat bayi mampu menjelaskan variabel dependen yaitu IPM di Jawa Tengah tahun 2021 sebesar 99.62% sedangkan sisanya yaitu sebesar 0.38% dipengaruhi dan dijelaskan variabel lain yang tidak ada dalam model.

### 3.4. Uji signifikansi parameter

Pada model regresi *robust* terdapat dua uji signifikansi parameter yaitu uji serentak dan parsial. Uji serentak digunakan untuk mengetahui bahwa variabel independen secara bersamaan berpengaruh terhadap model. Uji parsial berguna untuk menguji pengaruh dari setiap variabel independen terhadap variabel dependen [12]. Berdasarkan uji serentak yang dilakukan dengan uji simultan  $F$  didapatkan nilai  $F_{hitung} = 2281.623 > F_{tabel(0.05;4;30)} = 2.689$  sehingga  $H_0$  ditolak yang berarti bahwa variabel independen  $X_1, X_2, X_3$ , dan  $X_4$  secara serentak berpengaruh terhadap variabel dependen. Selanjutnya dilakukan uji parsial  $t$  untuk menguji pengaruh dari setiap variabel.

**Tabel 2.** Pengujian parsial  $t$

| Variabel | $t_{hitung}$ | $t_{tabel}$ |
|----------|--------------|-------------|
| $X_1$    | 33.482       | 2.042       |
| $X_2$    | 45.629       | 2.042       |
| $X_3$    | 69.745       | 2.042       |
| $X_4$    | 49.891       | 2.042       |

Berdasarkan Tabel 2 diperoleh nilai  $t_{hitung}$  dari keempat variabel independen lebih besar dari nilai  $t_{(0.025;30)} = 2.042$  maka  $H_0$  ditolak yang berarti variabel independen  $X_1, X_2, X_3$ , dan  $X_4$  memiliki pengaruh signifikan terhadap IPM di Jawa Tengah.

## 4. Kesimpulan dan Saran

Berdasarkan hasil analisis yang dilakukan dapat diambil kesimpulan bahwa faktor-faktor yang mempengaruhi IPM di Jawa Tengah yaitu variabel harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita disesuaikan, dan umur harapan hidup dengan nilai  $R^2$  *adjusted* sebesar 99.62%. Model regresi *robust* estimasi GS pada IPM di Jawa Tengah 2021 sebagai berikut

$$\hat{Y} = 4.30454 + 1.15139X_1 + 1.22981X_2 + 0.00093X_3 + 0.44626X_4 \quad (6)$$

Interpretasi dari model tersebut yaitu apabila keempat variabel independen bernilai nol atau konstan maka nilai IPM adalah sebesar 4.30454, setiap kenaikan satu satuan harapan lama sekolah akan menaikkan IPM sebesar 1.15139, setiap kenaikan satu satuan rata-rata lama sekolah akan menaikkan IPM sebesar 1.22981, setiap kenaikan satu satuan pengeluaran per kapita disesuaikan akan menaikkan IPM sebesar 0.00093, dan setiap kenaikan satu satuan umur harapan hidup saat bayi akan menaikkan IPM sebesar 0.44626.

Saran yang dapat disampaikan pada penelitian ini yaitu pada penelitian selanjutnya dapat melakukan analisis IPM dengan menambah jumlah variabel dan dapat menggunakan metode analisis regresi *robust* dengan estimasi yang berbeda.

## 5. Daftar Pustaka

- [1] Badan Pusat Statistik. (2016). Indeks Pembangunan Manusia 2015. *Jakarta. Badan Pusat Statistik Republik Indonesia.*
- [2] Puspitasari, M.T.S., Susanti, Y. and Handajani, S.S. (2021). Model Regresi Robust untuk Indeks Pembangunan Manusia di Jawa Timur dengan Estimasi M. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 659-665)
- [3] Nurbaroqah, A., Pratikno, B. and Supriyanto, S. (2022). Pendekatan Regresi Robust dengan Fungsi Pembobot Bisquare Tukey pada Estimasi-M dan Estimasi-S. *Jurnal Ilmiah Matematika dan Pendidikan Matematika*, 14(1), pp.19-30
- [4] 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
- [5] Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons
- [6] Susanti, Y., Pratiwi, H., & Sulistijowati, S. (2021). *Regresi Robust Teori dan Terapannya*. UNS Press, Surakarta.
- [7] Gujarati, D. (2007). *Dasar-dasar Ekonometrika, Edisi ketiga, jilid 1*. Julius A. Mulyadi, S.E, penerjemah. Jakarta: Erlangga. Terjemahan dari: Essentials of Econometrics.
- [8] Gujarati, D. N. (2003). *Basic Econometrics*. New York: The McGrawHill Companies, Inc.
- [9] Montgomery, D. C., Peck, E. A. (1992). *Introduction to Linear Regression Analysis, 2 nd edition*. New York: John Wiley & Sons, Inc.
- [10] Croux, C., Rousseeuw, P., & Hossjer, O. (1994). *Generalized S-Estimators*, *Journal of the American Statistical Association*, 89:428, 1271-1281.
- [11] Fox, J. & Weisberg, S. (2010). *Robust Regression in R*. Appendix to An R and S-Plus Companion to Applied Regression, Second Edition.
- [12] Handajani, S.W., Susanti, Y., & Ratnasari, A. (2021). *Metode Statistika dengan R*. Surakarta: UNS Press.