

PERBANDINGAN METODE ORDINARY LEAST SQUARE (OLS) DAN REGRESI ROBUST

Ibnu Dharma Syahputra¹, Abdul Karim²

^{1,2} Program Studi Statistika FMIPA Universitas Muhammadiyah Semarang

Alamat e-mail dharmasyahputra77@gmail.com

ABSTRACT

Human Development Index (HDI) is a comparative measurement of life expectancy, literacy, education. A country that is said to be advanced can be reflected if one of its reference is the Human Development Index (HDI), which of course explains how much human development in a country. Indonesia has human resources that can be explored and explored to show a significant index of human development. The Human Development Index (IPM) is a composite index that is influenced by health indicators represented by life expectancy (AHH), education indicators represented by the average school length (RLS) and economic indicators represented by public purchasing power (PPP) . Decentralization of Central Kalimantan Province is expected to make arrangements that can realize the life of the people of Central Kalimantan to be more prosperous and healthy. One of the elements in human development is to make the people to live longer and healthier lives. Data source used in this research is secondary data of HDI in Central Kalimantan obtained from Central Bureau of Statistics of Central Kalimantan Province 2014. Variable used in this research is endogenous variable (Y) and exogenous variable (X). Variable endogenous (Y) in the research that is HDI (percent) while the variable (X) in the study there are 3 that is Life Expectancy Numbers, Perkapita Revenue, Average Learning Length. Using OLS method and ROBUST method of both methods the best model that is on robotic method

Keywords: HDI, OLS, ROBUST

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan. Suatu negara yang dikatakan maju dapat tercermin jika yang dijadikan acuan salah satunya adalah Indeks Pembangunan Manusia (IPM), yang tentu saja menjelaskan seberapa besar perkembangan manusia disuatu negara. Indonesia memiliki sumber daya manusia yang bisa dieksplorasi dan digali sehingga menunjukkan indeks pembangunan manusia yang signifikan. Indeks Pembangunan Manusia (IPM) merupakan indeks komposit yang dipengaruhi oleh indikator kesehatan yang diwakili oleh angka harapan hidup (AHH), indikator pendidikan yang diwakili oleh rata-rata lama sekolah (RLS) dan indikator ekonomi yang diwakili oleh daya beli masyarakat (PPP).

Desentralisasi Provinsi Kalimantan Tengah diharapkan mampu melakukan penataan yang dapat mewujudkan kehidupan masyarakat Kalimantan Tengah agar semakin sejahtera dan sehat. Salah satu unsur dalam pembangunan manusia adalah lebih mengupayakan agar penduduk dapat mencapai “usia hidup” yang panjang dan sehat. Berdasarkan Visi dan Misi Dinas Kesehatan Kalimantan Tengah (Profil Kesehatan Provinsi Kalimantan Tengah 2016). Maka pembangunan kesehatan masyarakat Kalimantan Tengah diarahkan pada fokus pencapaian program yaitu “ Masyarakat Kalimantan Tengah Sehat dan Berkualitas”.

Berdasarkan Teori H.L. Blum, derajat keehatan masyarakat dengan indikatornya

angka kematian(mortalitas) dan angka kesakitan (morbiditas) sangat dipengaruhi oleh empat factor yaitu factor lingkungan, factor perilaku, factor pelayanan kesehatan, dan factor keturunan. Indikator utama dari pencapaian visi tersebut adalah angka tingkat kesehatan. Melihat keadaan tersebut peneliti ingin mengetahui berapa Indeks Pembangunan Manusia di Kalimantan Tengah.

2. KAJIAN LITERATUR

OLS (*Ordinary Least Square*) merupakan metode regresi yang meminimumkan jumlah kesalahan (error) kuadrat. Metode estimasi parameter yang digunakan adalah metode OLS (*Ordinary Least Square*), yaitu menduga koefisien regresi (β) dengan meminimumkan kesalahan (error) . Adapun penaksir parameternya adalah sebagai berikut :

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Dimana $\hat{\beta}$ adalah vektor dari parameter yang diestimasi berukuran $(p + 1) \times 1$, X adalah matriks variabel prediktor berukuran $n \times (p + 1)$ dan y vektor observasi dari variabel respon berukuran $n \times 1$

Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal atau adanya pencilan berpengaruh pada model (Ryan,1997). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh pencilan. Robust artinya parameter model tidak banyak berubah ketika sampel baru diambil dari populasi. Penggunaan umum dari regresi robust adalah pada data yang terdapat outlier. Deteksi outlier mencakup determinasi dimana residu(error = prediksi – hasil aktual) adalah nilai positif atau negatif ekstrim. Outlier dapat benar-benar mengacau pada sample mean akan tetapi memiliki efek relatif kecil pada sample median.

Metode regresi *robust* menurut Huber (1981: 43) mempunyai tiga estimasi, yaitu estimasi L (kombinasi linear dari statistik *order*/terurut), estimasi M (estimasi dengan maksimum likelihood) dan estimasi R (estimasi yang berasal dari uji rank). Estimasi M lebih fleksibel dan dapat digunakan untuk menyelesaikan masalah estimasi multiparameter. Dalam menentukan estimasi parameter, pada aplikasinya estimasi M lebih mudah digunakan dibandingkan dengan estimasi maupun estimasi L.

3. METODE PENELITIAN

a. Sumber data dan variable penelitian

Sumber data yang digunakan pada penelitian ini adalah data sekunder IPM di Kalimantan Tengah yang diperoleh dari Badan Pusat Statistik Provinsi Kalimantan Tengah tahun 2014.

Variabel yang digunakan dalam penelitian ini yaitu variabel endogenous(Y) dan variabel eksogenous (X). Variabel endogenous (Y) dalam penelitian yaitu IPM (persen) sedangkan variabel (X) dalam penelitian ada 3 yaitu Angka Harapan Hidup, Pendapatan Perkapita, Rata-rata Lama Belajar.

b. Metode analisis

Regresi Linear Berganda

Regresi linear berganda adalah teknik statistika yang digunakan untuk menentukan model hubungan satu variabel respon (Y) dengan melibatkan lebih dari satu variabel prediktor hingga p variabel prediktor dimana banyaknya p kurang dari jumlah observasi (n). Adapun model regresi berganda sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$$

dimana Y_i merupakan nilai variabel dependen dalam pengamatan ke- i , $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah parameter yang tidak diketahui nilainya, $X_{i1}, X_{i2}, \dots, X_{ip}$ adalah nilai dari variabel independen dari pengamatan ke- i , dan ε_i adalah error random dan berdistribusi normal dengan rata-rata nol dan varians σ^2 .

OLS (*Ordinary Least Square*) merupakan metode regresi yang meminimumkan jumlah kesalahan (error) kuadrat. Metode estimasi parameter yang digunakan adalah metode OLS (*Ordinary Least Square*), yaitu menduga koefisien regresi (β) dengan meminimumkan kesalahan (error). Adapun penaksir parameternya adalah sebagai berikut :

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Dimana $\hat{\beta}$ adalah vektor dari parameter yang diestimasi berukuran $(p + 1) \times 1$, X adalah matriks variabel prediktor berukuran $n \times (p + 1)$ dan y vektor observasi dari variabel respon berukuran $n \times 1$

Regresi robust merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal atau adanya pencilan berpengaruh pada model (Ryan,1997). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh pencilan. Robust artinya parameter model tidak banyak berubah ketika sampel baru diambil dari populasi. Penggunaan umum dari regresi robust adalah pada data yang terdapat outlier. Deteksi outlier mencakup determinasi dimana residu(error = prediksi – hasil aktual) adalah nilai positif atau negatif ekstrim. Outlier dapat benar-benar mengacau pada sample mean akan tetapi memiliki efek relatif kecil pada sample median.

Metode regresi *robust* menurut Huber (1981: 43) mempunyai tiga estimasi, yaitu estimasi L (kombinasi linear dari statistik *order*/terurut), estimasi M (estimasi dengan maksimum likelihood) dan estimasi R (estimasi yang berasal dari uji rank). Estimasi M lebih fleksibel dan dapat digunakan untuk menyelesaikan masalah estimasi multiparameter. Dalam menentukan estimasi parameter, pada aplikasinya estimasi M lebih mudah digunakan dibandingkan dengan estimasi maupun estimasi L.

Dalam menilai hasil kedua metode dengan membandingkan standar error kedua metode dengan OLS yang terdapat *outlier*. Apabila standar error yang dihasilkan metode regresi robust lebih kecil dari OLS, maka regresi robust dapat menganalisis data tanpa membuang *outlier* dan menghasilkan estimasi yang resisten terhadap *outlier*. Sehingga dapat dikatakan regresi robust dapat mengatasi kelemahan OLS terhadap pengaruh *outlier*.

4. HASIL DAN PEMBAHASAN

Tabel 1. Uji Normalitas

| Model | <i>p-value</i> |
|---------------|----------------|
| <i>OLS</i> | 0.1171 |
| <i>Robust</i> | 0.04722 |

Berdasarkan tabel normalitas diatas diperoleh bahwa metode OLS berdistribusi normal sedangkan ROBUST tidak berdistribusi normal pada taraf signifikansi 5%

Tabel 2. Uji Signifikansi Model

| | | |
|--|--------------|-----------|
| | Signifikansi | parameter |
|--|--------------|-----------|

| Parameter | OLS | Robust | OLS | Robust |
|-----------|----------|----------|---------|---------|
| INTERCEPT | 0.576 | 0.60574 | 2.27141 | -6.9198 |
| AHH | 1.09e-06 | 0.00803 | 0.56368 | 0.7324 |
| PP | 2.04e-07 | 0.00114 | 1.31002 | 0.9859 |
| RLS | 5.18e-07 | 4.67e-09 | 1.71093 | 1.8282 |

Berdasarkan uji signifikansi diatas diperoleh hasil untuk metode OLS terdapat 3 variabel yang signifikan yaitu Angka Harapan Hidup, Pendapatan Perkapita, Rata-rata Lama Sekolah. Sedangkan pada metode robust juga terdapat 1 variabel yang signifikan yaitu Rata-rata Lama Sekolah pada taraf signifikan 5%. Sedangkan Dari tabel parameter diatas maka di peroleh model sebagai berikut :

$$\text{OLS } IPM = 2.27141 + 0.56368 \text{ AHH} - 1.31002 \text{ PP} + 1.71093 \text{ RLS}$$

$$\text{Robust } IPM = -6.9198 + 0.7324 \text{ AHH} - 0.9859 \text{ PP} + 1.8282 \text{ RLS}$$

Tabel 3. Kriteria Keباikan Model

| Model | R-squared |
|--------|-----------|
| OLS | 0.9883 |
| Robust | 0.9959 |

Berdasarkan tabel kriteria kebaikan model diatas diperoleh nilai R-squared pada metode Robust lebih besar dibandingkan metode OLS sehingga dapat dikatakan bahwa model robust lebih baik dibandingkan metode OLS. Untuk nilai R-squared dari robust sendiri yaitu 0.9959 yang berarti sebesar 99.06% model dipengaruhi oleh faktor-faktor tersebut sedangkan sisanya dipengaruhi oleh faktor lain.

5. KESIMPULAN

Berdasarkan hasil analisis diatas diperoleh kesimpulan bahwa Pada metode OLS variabel yang signifikan adalah Angka Harapan Hidup, Pendapatan Perkapita, Rata-rata Lama Sekolah, sedangkan pada metode robust yaitu hanya Rata-rata Lama Sekolah, dari kedua metode tersebut model yang baik yaitu pada metode robus dengan nilai r-square sebesar 0.9959. akan tetapi dilihat dari uji normalitasnya diperoleh bahwa metode OLS berdistribusi normal sedangkan ROBUST tidak berdistribusi normal pada taraf signifikansi 5%, jadi metode yang terbaik yaitu menggunakan metode OLS.

6. REFERENSI

Badan pusat statistik provinsi kalimantan tengah,2014.data ipm provinsi kalimantan tengah tahun 2014

Musafirah.,Raupong., dan Nasrah Sirajang. "Perbandingan Metode Robust Least Trimmed Square Dengan Metode Scale Dalam Mengestimasi Parameter Regresi Linear Berganda Untuk Data Yang Mengandung Pencilan".Universitas Hasanuddin

Mashitah., Arief Wibowo. dan Diah Indriani. "Metode *Robust Regression on Ordered Statistics* (ROS) pada Data Tersensor Kiri dengan *Outlie*" [hal 148-157]. Universitas Airlangga

Prasetyo, Teguh., Ni Wayan Surya Wardhani dan Waego Hadi Nugroho. "Perbandingan Metode *Robust Generalized-M Schweppe One-Step Estimator* (Gm-S1s) dan Metode *Robust M-Estimator* Untuk Menangani Pencilan Pada Regresi Linier Berganda". Universitas Brawijaya.