

Analisis Regresi TELBS Untuk Menentukan Pengaruh Lahan Kopi Terhadap Produksi Kopi di Indonesia Tahun 2023 Menggunakan Bahasa Pemrograman Python

Muhammad Dhafin Qinthar Ramdhani, Nurul Gusriani, Firdaniza

Program Studi S-1 Matematika, Fakultas MIPA, Universitas Padjadjaran

Email: muhammad20216@mail.unpad.ac.id, nurul.gusriani@unpad.ac.id, firdaniza@unpad.ac.id

Abstrak

Indonesia, sebagai salah satu produsen kopi terbesar dunia, terkenal dengan ragam kopi berkualitas tinggi seperti Arabika, Robusta, dan Liberika. Produksi kopi dipengaruhi oleh banyak faktor, termasuk luas lahan perkebunan. Data produksi kopi dapat mengandung pencilan yang disebabkan oleh perubahan cuaca, serangan hama, praktik pertanian yang tidak konsisten, atau kesalahan pencatatan. Hal tersebut dapat diatasi menggunakan regresi robust, dimana salah satu estimasinya adalah estimasi Tabatabai Eby Li Bae Singh (TELBS). Metode estimasi TELBS melakukan estimasi parameter model dengan cara meminimumkan fungsi objektif. Pada paper ini, ditentukan model estimasi TELBS pada data produksi kopi di Indonesia pada tahun 2023 dengan variabel dependennya adalah jumlah produksi kopi dan variabel independennya adalah luas lahan perkebunan kopi. Pengujian parameter menggunakan uji t menunjukkan bahwa luas lahan perkebunan kopi signifikan dalam mempengaruhi produksi kopi pada tahun tersebut dengan taraf signifikansi 0,05. Hasil dari model estimasi TELBS menunjukkan koefisien determinasi sebesar 96,51%, menunjukkan bahwa model regresi tersebut mampu menjelaskan sebagian besar variasi dalam data.

Kata Kunci: TELBS, pencilan, regresi, koefisien determinasi, produksi kopi

Abstract

Indonesia, as one of the world's largest coffee producers, is renowned for its diverse range of high-quality coffees such as Arabica, Robusta, and Liberica. Coffee production is influenced by various factors, including the extent of plantation land. Coffee production data may contain outliers due to factors like weather changes, pest attacks, inconsistent farming practices, or recording errors. These challenges can be addressed using robust regression methods, with one such estimation being Tabatabai Eby Li Bae Singh (TELBS) estimation. TELBS estimates model parameters by minimizing an objective function. In this study, a TELBS estimation model was applied to Indonesian coffee production data in 2023, with the dependent variable being coffee production quantity and the independent variable being plantation land area. Parameter testing using t-tests indicated that plantation land area significantly influences coffee production in that year at a significance level of 0.05. The TELBS estimation model yielded a coefficient of determination of 96.51%, demonstrating its capability to explain a substantial portion of the data's variance.

Keywords: TELBS, outliers, regression, coefficient of determination, coffee production

1 PENDAHULUAN

Indonesia adalah salah satu produsen kopi terbesar di dunia, dikenal dengan berbagai jenis kopi berkualitas tinggi seperti kopi Arabika, Robusta, dan Liberika. Kopi merupakan komoditas penting bagi perekonomian Indonesia, tidak hanya sebagai sumber devisa negara melalui ekspor, tetapi juga sebagai sumber mata pencaharian bagi jutaan petani kopi di seluruh nusantara. Produksi kopi di Indonesia dipengaruhi oleh berbagai faktor, salah satunya adalah luas lahan yang digunakan

untuk menanam kopi. Oleh karena itu, diperlukan penelitian untuk melihat apakah luas lahan kopi secara langsung mempengaruhi produksi kopi.

Data mengenai produksi kopi berpotensi mengandung pencilan, yang dapat disebabkan oleh berbagai faktor seperti perubahan cuaca ekstrem, serangan hama, praktik pertanian yang tidak konsisten, atau kesalahan dalam pencatatan data. Pencilan dalam data dapat diatasi dengan menggunakan regresi robust, salah satunya adalah metode regresi robust estimasi Tabatabai Eby Li Bae Singh (TELBS). Menurut Tabatabai *et al.*, (2012), estimasi TELBS menghasilkan bias dan Mean Squared Error (MSE) yang lebih baik dibandingkan *ordinary least square* (OLS), estimasi jenis maximum likelihood (estimasi M), dan metode estimasi momen (estimasi MM). Metode estimasi TELBS juga mampu mendeteksi pencilan pada variabel terikat dan variabel bebas.

Gusriani & Firdaniza (2021) menggunakan model regresi TELBS untuk studi tentang Demam Berdarah Dengue (DBD) dan mencatatkan nilai koefisien determinasi yang mencapai 94,71%. Li & Eby (2018) melakukan perbandingan antara metode TELBS dengan beberapa metode robust lainnya seperti estimasi M, Least Trimmed Squares (LTS), scale estimator (estimasi S), dan estimasi MM. Hasilnya menunjukkan bahwa TELBS secara konsisten mengungguli metode-metode tersebut dalam sebagian besar kasus. Secara keseluruhan, TELBS menunjukkan kinerja yang stabil dalam menangani data pencilan di berbagai kasus.

Penelitian ini mengkaji estimasi parameter model regresi TELBS serta melakukan pengujian parameter untuk mengevaluasi apakah ada hubungan signifikan antara variabel independen dan variabel dependen secara parsial. Data yang dianalisis berasal dari produksi kopi di Indonesia pada tahun 2023, yang diperoleh dari Badan Pusat Statistik (BPS). Variabel yang diteliti meliputi jumlah produksi kopi (Y) dan luas lahan perkebunan kopi (X).

2 KAJIAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi terdiri dari dua variabel yang dihubungkan, yaitu variabel terikat (Y) dan variabel bebas (X). Persamaan umum dari regresi linear adalah

$$y = X\beta + e. \quad (1)$$

Metode yang sering digunakan dalam analisis regresi adalah *Ordinary Least Squares* (OLS). OLS bertujuan untuk meminimumkan jumlah kuadrat residual (Sembiring, 2003). Berikut merupakan taksiran untuk regresi OLS:

$$\beta = (X^tX)^{-1}(X^ty). \quad (2)$$

Salah satu kelemahan dari metode OLS adalah keberadaan pencilan dalam data. Pencilan dalam data akan menyebabkan hasil yang bias.

2.2 Metode Regresi Estimasi TELBS

Salah satu metode yang tahan terhadap keberadaan pencilan adalah regresi *robust*. Salah satu estimasi dalam regresi *robust* adalah estimasi Tabatabai Eby Li Bae Singh (TELBS). Metode TELBS dilakukan dengan meminimumkan fungsi objektif

$$\min_{\beta} \sum_{i=1}^n \frac{\rho_{\omega}(t_i)}{L_i} \quad (3)$$

di mana

$$\rho_{\omega}(t_i) = 1 - \text{Sech}(\omega t_i) \tag{4}$$

$$t_i = \frac{(y_i - x_i^t \beta)(1 - h_{ii})}{\hat{\sigma}} \tag{5}$$

Variabel h_{ii} merupakan elemen diagonal *hat matrix* yang diperoleh menggunakan persamaan

$$H = X(X^t X)^{-1} X^t \tag{6}$$

Menurut Tabatabai *et al.*, (2012) fungsi ρ merupakan fungsi diferensiabel dan ω merupakan konstanta kesesuaian yang bernilai 0,628. Nilai $\hat{\sigma}$ dapat diperoleh menggunakan salah satu penduga $\hat{\sigma}$ yang ditemukan oleh Rousseeuw dan Croux (1993), yaitu:

$$\hat{\sigma} = 1,1926 \times \text{Med}\{\text{Med}|e_i - e_j|\} \tag{7}$$

Nilai 1,1926 ditujukan agar $\hat{\sigma}$ menjadi estimasi yang mendekati tak bias (Rousseeuw & Croux, 1993). Nilai L_i diperoleh menggunakan persamaan

$$L_i = \sum_{j=1}^k \text{Max}\{M_j, |x_{ij}|\} \tag{8}$$

di mana

$$M_j = \text{Med}\{|x_{1j}|, |x_{2j}|, \dots, |x_{nj}|\} \tag{9}$$

Untuk meminimumkan persamaan (3), diperlukan turunan fungsi objektif terhadap β_0 dan β_j lalu disamakan dengan nol sehingga menghasilkan persamaan

$$\sum_{i=1}^n \frac{\psi_{\omega}(t_i)}{L_i} \frac{\partial t_i}{\partial \beta_j} = 0 \tag{10}$$

di mana $\psi_{\omega}(t_i)$ merupakan turunan dari ρ_{ω} , yaitu

$$\psi_{\omega}(x) = \frac{\partial \rho_{\omega}(t_i)}{\partial t_i} = \omega \text{Sech}(\omega t_i) \text{Tanh}(\omega t_i) \tag{11}$$

Didefinisikan fungsi pembobot w_{ii} (Tabatabai *et al.*, 2012) yaitu

$$w_{ii} = \frac{\psi_{\omega}(t_i)(1 - h_{ii})}{\hat{\sigma} e_i L_i} \tag{12}$$

di mana w_{ii} merupakan elemen diagonal dari matriks diagonal \mathbf{W} berukuran $n \times n$. Kemudian persamaan (9) dapat ditulis sebagai

$$\sum_{i=1}^n w_{ii} e_i \frac{\partial (y_i - \beta_0 - x_{i1}\beta_1 - x_{i2}\beta_2 - \dots - x_{ik}\beta_k)}{\partial \beta_j} = 0 \quad (13)$$

Persamaan (13) diturunkan terhadap $\beta_0, \beta_1, \dots, \beta_j$ sehingga didapatkan taksiran estimasi TELBS

$$\hat{\beta}_{\text{TELBS}} = (X^t W X)^{-1} X^t W y. \quad (14)$$

2.3 Koefisien Determinasi

Koefisien determinasi (R^2) merupakan nilai yang digunakan untuk mengukur besarnya pengaruh variabel bebas (X) terhadap variasi variabel terikat (Y). Koefisien determinasi (R^2) pada analisis regresi *robust* dengan metode estimasi TELBS didefinisikan sebagai berikut (Tabatabai *et al.*, 2012):

$$R^2 = 1 - \left(\frac{\text{Med}|e_i|}{\text{Med}|y_i - \text{Med } \hat{y}_i|} \right)^2 \quad (15)$$

di mana e_i merupakan nilai residual yang didapatkan dari iterasi terakhir. R^2 bernilai antara 0 sampai 1. Jika nilai R^2 mendekati 1 menunjukkan tingkat ketepatan model yang semakin baik dalam menerangkan variasi data.

2.4 Uji t

Uji t adalah metode statistik yang digunakan untuk mengukur apakah ada pengaruh signifikan antara variabel bebas dan variabel terikat secara parsial. Uji t dinyatakan dalam persamaan berikut:

$$t_1 = \frac{\hat{\beta}_1}{S_1} \quad (16)$$

Hasil dari perhitungan t_1 dibandingkan dengan nilai t tabel. Kriteria uji pengambilan keputusannya adalah jika t tabel $< t_1$, maka terdapat pengaruh dari variabel bebas yang signifikan secara parsial terhadap variabel terikat.

3 METODE PENELITIAN

Metode penelitian yang digunakan adalah kuantitatif dengan data sekunder yang diperoleh dari BPS. Terdapat satu variabel terikat dan satu variabel bebas, yaitu jumlah produksi kopi (Y) dan lahan kopi (X). Metode yang digunakan untuk menganalisis serta memodelkan adalah regresi *robust* estimasi TELBS dengan bantuan bahasa pemrograman Python. Hasil dari model yang diperoleh akan dicari koefisien determinasi untuk menilai seberapa baik model tersebut dalam menjelaskan variasi dalam data. Selain itu, model yang diperoleh diuji menggunakan uji t untuk melihat signifikansi dari variabel bebas terhadap variabel terikat.

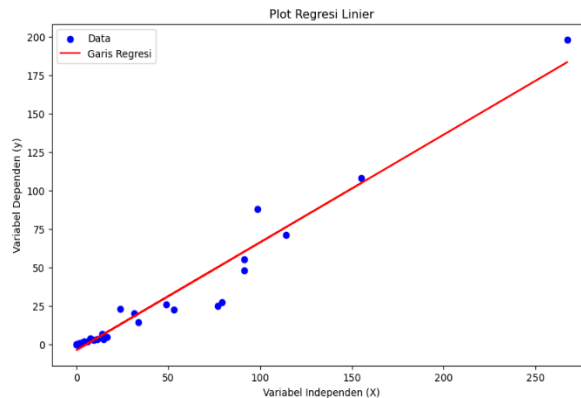
4 HASIL DAN PEMBAHASAN

4.1 Hasil Regresi OLS

Tahap pertama dalam penelitian ini adalah analisis data menggunakan estimasi OLS. Berikut merupakan hasil analisisnya:

Tabel 1. Koefisien Regresi dan Determinasi dengan Metode Kuadrat Terkecil

	Estimasi
<i>Intercept</i>	-3,7675
<i>X</i>	0,6999
R^2	0,954



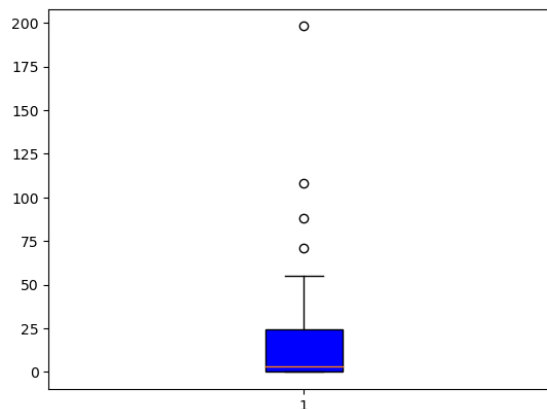
Gambar 1. Plot Regresi OLS

Dari Tabel 1 dapat diperoleh model regresi OLS sebagai berikut:

$$\hat{Y}_{OLS} = -3,7675 + 0,6999X \tag{17}$$

Nilai koefisien determinasi yang diperoleh 0,954 atau 95,4% yang artinya sekitar 95,4% produksi padi disebabkan oleh luas lahan kopi (*X*) sedangkan sisanya yaitu 4,6% ditentukan oleh faktor lain yang tidak masuk dalam model. Dari plot regresi OLS yang ditampilkan pada Gambar 1, terlihat bahwa garis regresi terpengaruh oleh data yang diduga sebagai pencilan.

Diagram *boxplot* dari data produksi kopi di Indonesia ditunjukkan oleh Gambar 2.



Gambar 2. Diagram *boxplot* Produksi Kopi di Indonesia 2023

Hasil *boxplot* di atas memperlihatkan adanya 4 titik di luar batas, yang menunjukkan bahwa terdapat 4 wilayah dengan jumlah produksi kopi yang lebih tinggi dibandingkan wilayah lainnya. Ini mengindikasikan adanya kemungkinan pencilan dalam data produksi kopi.

4.2 Uji t

Model regresi estimasi TELBS yang diperoleh diuji menggunakan uji parameter t pada penelitian ini tingkat signifikansi yang ditetapkan adalah $\alpha = 0,05$ dan derajat kebebasannya adalah 29. Dari tingkat signifikansi dan derajat kebebasan yang telah ditetapkan, diperoleh nilai t tabel sebesar 2,045. Dengan bantuan pemrograman Python, diperoleh nilai t untuk masing-masing variabel yang ditunjukkan oleh Tabel 2.

Tabel 2. Hasil perhitungan nilai t untuk data Produksi Kopi di Indonesia tahun 2023

	Nilai t
X	25,677

Berdasarkan Tabel 2, nilai t untuk variabel X diketahui lebih besar daripada nilai t tabel, yang menunjukkan adanya pengaruh luas lahan kopi terhadap jumlah produksi kopi di Indonesia pada tahun 2023.

4.3 Hasil Regresi TELBS

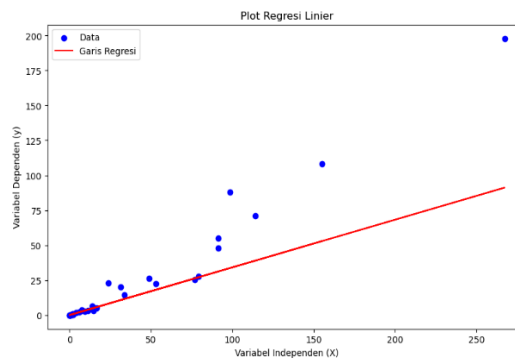
Analisis berikutnya dilakukan menggunakan metode TELBS. Nilai koefisien regresi yang diperoleh dari OLS digunakan sebagai nilai awal dalam iterasi pertama. Dengan menggunakan bahasa pemrograman Python, dibuat program untuk menghitung estimasi TELBS pada data produksi kopi di Indonesia tahun 2023. Dalam penelitian ini, dibutuhkan 32 iterasi untuk mencapai konvergensi. Berikut merupakan *syntax* Python untuk metode TELBS:

```
y_pred = beta[0] + beta[1] * x1
ei = y - y_pred
ti = (ei * (1 - hii))/gamma
abs_diff_matrix = np.abs(np.subtract.outer(ei, ei))
median_j = np.median(abs_diff_matrix, axis=0)
median_i = np.median(median_j)
gamma = 1.1926 * median_i
for i in range(34):
    w[i] = (psi(ti[i])*(1-hii[i]))/(gamma*ei[i]*L[i])
X_transpose = np.transpose(X)
weighted_X = X_transpose.dot(np.diag(w)).dot(X)
weighted_X_inv = np.linalg.pinv(weighted_X)
weighted_X_y = X_transpose.dot(np.diag(w)).dot(y)
beta = weighted_X_inv.dot(weighted_X_y)
NORM = np.linalg.norm(beta - prev_beta)
```

Hasil estimasi TELBS ditampilkan pada Tabel 3.

Tabel 3. Koefisien Regresi dan Determinasi dengan Metode TELBS Berdasarkan Uji t

	Estimasi
<i>Intercept</i>	-0,0954
X	0,3415
R^2	0,9651



Gambar 3. Plot Regresi TELBS

Dari Tabel 3 dapat diperoleh model regresi TELBS sebagai berikut:

$$\hat{Y}_{TELBS} = -0,0954 + 0,3415X \quad (18)$$

Nilai koefisien determinasi yang diperoleh lebih besar dibandingkan dengan nilai koefisien determinasi metode OLS. Nilai yang diperoleh 0,9651 atau 96,51% yang artinya sekitar 96,51% produksi padi disebabkan oleh luas lahan kopi (X) sedangkan sisanya yaitu 3,49% ditentukan oleh faktor lain yang tidak masuk dalam model. Dari plot regresi TELBS yang ditunjukkan gambar 3, dapat dilihat bahwa garis regresi tidak terpengaruh oleh data yang diduga pencilan.

5 SIMPULAN

Berdasarkan hasil dan pembahasan diatas, dapat disimpulkan bahwa luas lahan kopi (X) berpengaruh terhadap jumlah produksi kopi. Nilai koefisien determinasi yang diperoleh dari metode TELBS cukup baik, yaitu sebesar 0,9651 atau 96,51% yang artinya sekitar 96,51% produksi padi disebabkan oleh luas lahan kopi (X) sedangkan sisanya yaitu 3,49% ditentukan oleh faktor lain yang tidak masuk dalam model.

DAFTAR PUSTAKA

- Gusriani, N., & Firdaniza. (2021). *Model Kasus Demam Berdarah Dengue (DBD) di Kabupaten Majalengka Tahun 2016 Berdasarkan Regresi TELBS*. 17(1), 5–13.
<https://doi.org/10.24198/jmi.v17.n1.2021.5-13>
- Li, H., & Eby, W. M. (2018). A Comparative Study of TELBS Robust Linear Regression. *American Review of Mathematics and Statistics*, 6(1), 1–16.
<https://doi.org/10.15640/arms.v6n1a1>
- Rousseeuw, P., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, 88, 1273–1283.
- Sembiring. (2003). *Analisis Regresi*. ITB.
- Tabatabai, Eby, W., Li, Bae, sejong, & Singh, K. (2012). TELBS robust linear regression method. *Open Access Medical Statistics*, 65. <https://doi.org/10.2147/oams.s37395>