

ANALISIS REGRESI *ROBUST* DENGAN PENDUGA *METHOD OF MOMENT* (MM) UNTUK MENGATASI DATA YANG TERIDENTIFIKASI PENCILAN BERDASARKAN DATA PRODUKSI KEDELAI DI INDONESIA

Muliyani¹, Noeryanti²

^{1,2}Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta

Email : ¹MULIYANI212330@gmail.com, ²snoeryanti@yahoo.com

Abstract. *Robust regression analysis using Method of Moment estimation to resolve the data that is outlier identify. The purposes of this study is to show how to apply Method of Moment estimation in solving outlier that occur the case of data of soybean production in Indonesia 2015. Robust regression analysis using MM estimation with Tukey Bisquare weighting is used to solve the outlier data in the data of soybean production in Indonesia 2015. Robust regression analysis is a regression method that is used when there is a outlier data. A method of MM estimation is the combination of method S estimation and M estimation. The result of this research with the case of soybean production in Indonesia 2015 found four datas that are identified as outliers, they are the 8, 12, 15 and 18 of data. Firstly this case is identified by the least square method, but one of classical regression assumption, that is the normality regression assumption is not fulfilled, then the case of soybean production is estimated by MM estimation with Tukey Bisquare weighting in robust regression. Afterwards, both of method are compared based on MSE value. The result of the comparison shows that MM estimation is the better method with the smaller MSE value and the model. It is obtained from MM estimation with $\hat{Y} = 13009 + 1,56 X_1 + 1,11 X_2 + 35,7 X_3 - 235 X_4 + 0,0224 X_5$.*

Keywords: *Least Square Method, Outlier, MM Estimation, Robust Regression.*

Abstrak. Analisis regresi *robust* menggunakan penduga *Method of Moment* (MM) untuk mengatasi data yang teridentifikasi pencilan. Tujuan dari penelitian ini adalah penerapan penduga-MM dalam mengatasi adanya pencilan pada kasus data produksi kedelai di Indonesia tahun 2015. Untuk mengatasi data pencilan pada data produksi kedelai di Indonesia tahun 2015, digunakan analisis regresi *robust* yaitu penduga *Method of Moment* (MM) dengan pembobot *Tukey Bisquare*. Regresi *robust* merupakan metode regresi yang digunakan ketika terdapat data pencilan. Metode penduga-MM merupakan gabungan dari metode penduga-S dan penduga-M. Hasil dari penelitian dengan kasus produksi kedelai di Indonesia tahun 2015 ini, terdapat empat amatan yang teridentifikasi sebagai pencilan yaitu amatan ke-8, 12, 15 dan 18. Kasus tersebut awalnya dianalisis dengan metode kuadrat terkecil tetapi karena salah satu asumsi klasik regresi tidak terpenuhi yaitu asumsi kenormalan maka kasus produksi kedelai di Indonesia tahun 2015 diestimasi menggunakan penduga-MM dengan pembobot *Tukey Bisquare* dalam regresi *robust*. Kemudian kedua metode tersebut dibandingkan berdasarkan nilai MSE yang diperoleh. Hasil dari perbandingan tersebut menunjukkan bahwa penduga-MM merupakan metode yang paling baik dengan nilai MSE yang lebih kecil dan model yang dihasilkan dari penduga-MM yaitu $\hat{Y} = 13009 + 1,56 X_1 + 1,11 X_2 + 35,7 X_3 - 235 X_4 + 0,0224 X_5$.

Kata kunci: *Metode Kuadrat Terkecil, Pencilan, Penduga-MM, Regresi Robust.*

1. Pendahuluan

Produksi kedelai pada tahun 2013 (780 juta ton) sudah berada di bawah produksi kedelai pada tahun 2012 (844 juta ton) atau turun sebesar 64 juta ton. Penurunan produksi kedelai ini terjadi karena beberapa faktor, yaitu luas panen, curah hujan, intensitas cahaya, kelembaban dan kebutuhan pupuk. Namun untuk tahun 2014, produksi kedelai kembali naik menjadi 955 juta ton, karena berbagai terobosan yang dilakukan pemerintah, seperti perluasan areal tanam, pemberian bantuan benih maupun sarana produksi pertanian serta insentif bagi petani agar mereka bergairah menanam kedelai. Menurut BPS (Aram II) pada tahun 2015 produksi kedelai kembali melonjak naik yaitu 964 juta ton. Kenaikan produksi ini disebabkan oleh perluasan luas panen sebesar 61 ribu ha dengan produktivitas mengalami kenaikan sebesar 15,68 kuintal per ha [13]. Dalam kasus produksi kedelai di Indonesia tahun 2015, banyaknya produksi kedelai masing-masing provinsi sangat berbeda, sehingga diasumsikan terdapat pencilan. Data pencilan tersebut tidak boleh

dibuang begitu saja karena akan mempengaruhi model prediksi serta menghasilkan estimasi parameter yang kurang tepat.

Untuk itu perlu dilakukan pengidentifikasian pencilan untuk mengetahui data produksi kedelai di Indonesia tahun 2015 teridentifikasi pencilan atau tidak, jika teridentifikasi pencilan maka untuk mengatasi pencilan pada data produksi kedelai digunakan analisis regresi *robust* dengan penduga-MM menggunakan pembobot *Tukey Bisquare*. Penelitian ini bertujuan untuk mengetahui gambaran produksi kedelai di Indonesia tahun 2015 dan penerapan analisis regresi *robust* dalam mengatasi data produksi kedelai di Indonesia tahun 2015 yang teridentifikasi pencilan. Hasil dari metode penduga-MM akan dibandingkan dengan metode kuadrat terkecil berdasarkan nilai MSE yang diperoleh.

Penerapan analisis regresi *robust* pada penelitian sebelumnya telah dilakukan oleh Candraningtyas S (*Jurnal Gaussian*, 2013) untuk menentukan persamaan regresi linier berganda yang mengandung pencilan pada data bangkitan *software* Minitab. Irfagitami N (*E-Jurnal Matematika*, 2014) untuk menghasilkan penduga yang kekar terhadap pencilan pada data bangkitan program R versi 2.15.3. Suyanti (*UNNES Journal of Mathematics*, 2014) untuk melihat tingkat efektifitas metode-metode regresi *robust* pada data ketahanan pangan di Jawa Tengah tahun 2007 dan data *survival time* yang teridentifikasi pencilan.

Pada penelitian ini akan dibahas regresi *robust* penduga-MM menggunakan pembobot *Tukey Bisquare* meliputi penerapannya dalam mengatasi pencilan yang mempengaruhi pendugaan model regresi.

Regresi Linear Berganda

Regresi linear berganda merupakan analisis yang menyatakan hubungan antara satu peubah terikat dengan beberapa peubah bebas dalam bentuk regresi [4]. Model regresi linier berganda [9] secara umum dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + \varepsilon_i \quad \text{untuk } i = 1, 2, \dots, n \quad (1)$$

Apabila dinyatakan dalam bentuk persamaan matriks model regresi linear berganda didefinisikan dengan matriks-matriks sebagai berikut:

$$Y = X \beta + \varepsilon \quad (2)$$

dengan:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Pendugaan Kuadrat Terkecil

Metode kuadrat terkecil (penduga kuadrat terkecil) merupakan penduga parameter regresi dengan meminimumkan jumlah kuadrat semua simpangan dari garis yang sebenarnya [9].

Model estimasi dari persamaan (1) sebagai berikut:

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \dots + b_k X_{ik}, i = 1, 2, \dots, n. \quad (3)$$

Persamaan estimasi di atas bila dinyatakan dalam bentuk matriks sebagai berikut

$$\hat{Y} = Xb \quad (4)$$

Untuk memperoleh nilai galat dan *JKG* digunakan persamaan sebagai berikut:

$$e_i = Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - b_3 X_{i3} - \dots - b_k X_{ik} \quad (5)$$

$$JKG = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - b_3 X_{i3} - \dots - b_k X_{ik})^2 \quad (6)$$

Untuk memperoleh b_0, b_1, \dots, b_k akan diminimumkan jumlah kuadrat galat. Caranya dengan melakukan penurunan parsial pada *JKG* terhadap setiap komponen vektor b dan menyamakannya dengan nol [10], sehingga diperoleh persamaan berikut:

$$b = (X'X)^{-1}X'Y \quad (7)$$

maka b sebagai penaksir bagi β sehingga memenuhi sifat *BLUE* (*Best Linear Unbiased Estimator*) artinya estimator bersifat terbaik, tidak bias (*unbias*) dan memiliki variansi minimum.

Asumsi Klasik

a. Uji Linearitas

Dasar pengambilan keputusan uji linearitas dapat dilakukan dengan cara melihat nilai signifikansi atau dengan melakukan uji F . Apabila nilai signifikansi lebih dari α atau jika nilai F_{hitung} kurang dari nilai F_{tabel} maka dapat dikatakan bahwa terdapat hubungan yang linear antara peubah bebas dengan peubah terikat atau pengujian linearitas terpenuhi [9].

b. Uji Multikolinearitas

Hubungan antara peubah bebas dapat dilakukan dengan cara memeriksa besaran *Variance Inflation Factor (VIF)*. Apabila nilai VIF yang diperoleh lebih besar dari 10 menandakan bahwa data mengalami masalah multikolinearitas [9]. Besaran nilai VIF untuk X_j dirumuskan sebagai berikut:

$$VIF_j = \frac{1}{1-R_j^2}, \quad j = 1, 2, \dots, k \quad (8)$$

c. Uji Autokorelasi

Statistik uji yang sering digunakan dalam pengujian asumsi ini adalah statistik d dari *Durbin-Watson* [9]. Statistik uji d *Durbin-Watson* dihitung dengan rumus sebagai berikut:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (9)$$

d. Uji Heteroskedastisitas

Heteroskedastisitas dapat dideteksi dengan besar r_s , yaitu koefisien korelasi peringkat dari Spearman [9] yang dirumuskan sebagai berikut:

$$r_{sd} = 1 - 6 \left[\frac{\sum d_i^2}{n(n^2-1)} \right] \quad (10)$$

e. Uji Normalitas

Pengujian asumsi normalitas yang paling sering digunakan adalah uji normalitas Kolmogorov-Smirnov [7]. Metode Kolmogorov-Smirnov didasarkan pada nilai D yang didefinisikan sebagai berikut:

$$D = \max_x [|F_T - F_S|] \quad (11)$$

Mengidentifikasi Pencilan dan Amatan Berpengaruh

Pencilan adalah amatan yang terpisah jauh dari kumpulan data lainnya, yang mana pencilan ini mengasilkan sisaan yang besar dan mempunyai pengaruh yang sangat besar terhadap model taksirannya [6].

1. Mengidentifikasi Pencilan

Untuk mengidentifikasi pencilan dapat dideteksi dengan menggunakan metode sebagai berikut:

a. Metode *Scatter Plot*

Metode ini didapatkan dari model regresi yang dapat dilakukan dengan cara memplot antara sisaan (ε) dengan nilai prediksi Y (\hat{Y}). Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya pencilan [1].

b. Nilai Leverage

Nilai leverage untuk mengidentifikasi amatan pencilan ditinjau dari nilai-nilai X dapat dirumuskan sebagai berikut:

$$h_{ii} = X_i(X'X)^{-1}X_i' \quad (12)$$

Apabila nilai $h_{ii} > 2p/n$ maka dianggap pencilan, dengan p adalah banyaknya parameter, k adalah banyaknya peubah bebas dan n adalah jumlah data [6].

c. Sisaan Dibuang Terstudentkan

Pengaruh pencilan amatan Y berdasarkan pada pemeriksaan sisaan dibuang terstudentkan d_i^* yang dirumuskan sebagai berikut:

$$d_i^* = \left[\frac{n-p-1}{JKG(1-h_{ii})-e_i^2} \right]^{1/2} \quad (13)$$

Amatan dengan $|d_i^*| > t_{(n-p-1)}$ dapat dianggap sebagai pencilan [6].

2. Mengidentifikasi Amatan Berpengaruh

Untuk mengetahui suatu amatan berpengaruh atau tidak, dapat ditentukan berdasarkan metode *DFFITs*.

a. *DFFITs*

DFFITs dapat dihitung dengan menggunakan persamaan berikut ini:

$$(DFFITs)_i = d_i^* \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} \tag{14}$$

Apabila nilai $|DFFITs| > 2\sqrt{p/n}$, mengindikasikan bahwa kasus ke-*i* yang tidak disertakan merupakan amatan berpengaruh [8].

Analisis Regresi *Robust*

Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil yang resisten terhadap adanya pencilan [3].

1. Estimasi *Method of Moment* (MM)

Estimasi *Method of Moment* (MM) menggabungkan estimasi *high breakdown point* dan efisiensi statistik. Fungsi pembobot yang digunakan dalam penduga-MM yaitu fungsi *Tukey Bisquare* dengan persamaan sebagai berikut :

$$w_i(u_i) = \begin{cases} \frac{u_i(1-(\frac{u_i}{c})^2)^2}{u_i}, & |u_i| < c \\ 0, & |u_i| \geq c \end{cases} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| < c \\ 0, & |u_i| \geq c \end{cases} \tag{15}$$

dengan $u_i = \frac{e_i}{\hat{\sigma}_s}$ dan *c* merupakan konstanta pembobot *Tukey Bisquare* sebesar 4,685 [1].

2. Pendugaan β dengan Metode *Weight Least Square*

Menggunakan metode *WLS* maka persamaan regresi menjadi

$$Y_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{i1}^* + \dots + \hat{\beta}_k^* \quad i = 1, 2, \dots, n \tag{16}$$

Persamaan (16) akan menjadi

$$\frac{Y_i}{\sigma_i^2} = \hat{\beta}_0 \left(\frac{1}{\sigma_i^2} \right) + \hat{\beta}_1 \left(\frac{X_{i1}}{\sigma_i^2} \right) + \dots + \hat{\beta}_k \left(\frac{X_{ik}}{\sigma_i^2} \right)$$

$$W_i Y_i = W_i (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik}) \quad \text{dengan } W_i = \frac{1}{\sigma_i^2} \tag{17}$$

Dinyatakan dalam bentuk matriks, maka persamaan (17) menjadi

$$WY = WX\hat{\beta} \tag{18}$$

dengan

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix}, W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & \vdots \\ \vdots & \dots & \ddots & 0 \\ 0 & \dots & 0 & w_n \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

$$\sum_{i=1}^n W_i e_i^2 = \sum_{i=1}^n W_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_k X_{ik})^2 \tag{19}$$

Untuk memperoleh $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ akan diminimumkan jumlah kuadrat galat terboboti. Caranya dengan melakukan penurunan parsial pada $\sum_{i=1}^n W_i e_i^2$ terhadap setiap komponen vektor $\hat{\beta}$ dan menyamakannya dengan nol [1], sehingga diperoleh persamaan berikut:

$$\hat{\beta} = (X'WX)^{-1} X'WY \tag{20}$$

Pengujian Parameter

a. Uji Simultan

Untuk menguji atau mengukur hubungan antara peubah bebas dengan peubah terikat yang menjelaskan dari persamaan regresi secara menyeluruh [9] disebut uji statistik *F*, dengan hipotesis sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ Artinya tidak ada pengaruh peubah bebas terhadap model yang terbentuk secara bersama-sama

H_1 : Minimal ada satu $\beta_k \neq 0$ Artinya ada peubah bebas yang berpengaruh terhadap model yang terbentuk secara bersama-sama.

Uji statistik F dinyatakan dalam bentuk rumus sebagai berikut :

$$F_{hitung} = \frac{MSR}{MSE} = \frac{SSR/(p-1)}{SSE/(n-k)} \quad (21)$$

b. Uji Parsial

Untuk menguji pengaruh masing-masing peubah bebas terhadap peubah terikat [9] disebut statistik uji t , dengan hipotesis sebagai berikut:

H_0 : $\beta_i = 0$ (Peubah bebas ke- i tidak berpengaruh terhadap peubah terikat)

H_1 : $\beta_i \neq 0$ (Minimal ada satu peubah bebas ke- i berpengaruh terhadap peubah terikat) untuk $i = 1, 2, \dots, p-1$ dengan p adalah banyaknya parameter.

Uji statistik t dinyatakan dalam bentuk rumus sebagai berikut :

$$t_{hitung} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \quad (22)$$

2. Metode

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data produksi kedelai di Indonesia tahun 2015 yang diperoleh langsung dari situs Badan Pusat Statistik. Variabel yang digunakan dalam penelitian ini adalah data produksi kedelai di Indonesia tahun 2015 dengan peubah yang diamati yaitu peubah terikat (Y) adalah produksi kedelai (ton) dan peubah bebas (X) yaitu luas panen (Ha), curah hujan (mm), intensitas cahaya (%), kelembaban (%) dan pupuk (ton). Adapun prosedur dalam penelitian ini, yaitu sebagai berikut:

a. Penggambaran data

Penggambaran data dilakukan dengan menggunakan analisis deksriptif data mengenai data produksi kedelai di Indonesia tahun 2015.

b. Pembentukan pendugaan model regresi linear berganda

Untuk pendugaan model regresi linear berganda digunakan metode kuadrat terkecil dengan menggunakan persamaan (3).

c. Pengujian asumsi-asumsi klasik

Model regresi akan dapat dijadikan alat estimasi yang tidak bias jika telah memenuhi persyaratan BLUE (*Best Linear Unbiased Estimator*). Untuk itu perlu dilakukan pengujian asumsi klasik, yaitu uji linearitas, uji multikolinearitas, uji autokorelasi, uji heteroskedastisitas dan uji normalitas.

d. Pengidentifikasi pencilan dan amatan berpengaruh

Mengidentifikasi pencilan dilakukan menggunakan metode *scatter plot*, nilai leverasi dan s issan dibuang terstudentukan. Apabila data teridentifikasi pencilan maka akan dilakukan pengidentifikasi amatan berpengaruh dengan metode *DFFITs* untuk mengetahui data yang teridentifikasi pencilan akan mempengaruhi model atau tidak.

e. Prosedur regresi *robust* dengan penduga-MM menggunakan pembobot *Tukey Bisquare*

- Menghitung nilai residual (e_i) dan nilai $\hat{\sigma}_s$
- Menghitung nilai (residual yang distadarisasi ($|u_i|$))
- Menghitung nilai pembobot (w_i) dengan menggunakan fungsi pembobot *Tukey Bisquare*
- Menghitung nilai parameter-parameter penduga-MM berdasarkan fungsi pembobot.

f. Pengujian parameter penduga-MM menggunakan uji simultan dan uji parsial.

g. Melakukan perbandingan hasil metode kuadrat terkecil dengan penduga-MM berdasarkan nilai R^2 dan MSE .

3. Hasil Dan Pembahasan

2.1 Deskriptif Produksi Kedelai dan Faktor-Faktor yang Mempengaruhinya

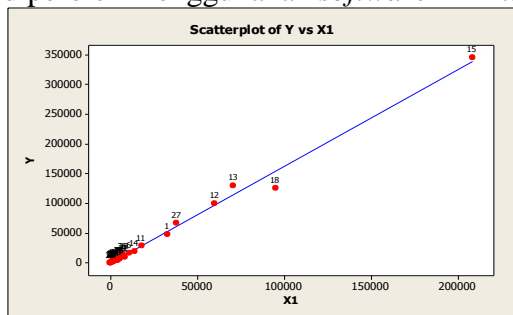
Analisis deksriptif produksi kedelai di Indonesia tahun 2015 diperoleh dengan menggunakan *software* SPSS yang hasil analisisnya disajikan seperti pada Tabel 1.

Tabel 1. Statistik Deskriptif Peubah Bebas dan Peubah Terikat

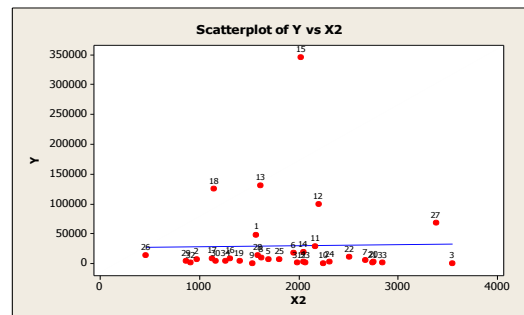
Peubah	Mean	Variansi	Minimum	Maksimum
Produksi Kedelai	29162,0882	4288484385	1	344998
Luas Panen	18592,8235	1599407492	1	208067
Curah Hujan	1871,0324	512878,945	460,9	3548
Intensitas Cahaya	67,0465	125	46,97	85,05
Kelembaban	79,5265	15,997	70	86,9
Pupuk	96377,7353	38499830784	40	914333

2.2 Plot Produksi Kedelai Terhadap Faktor-Faktor yang Mempengaruhinya

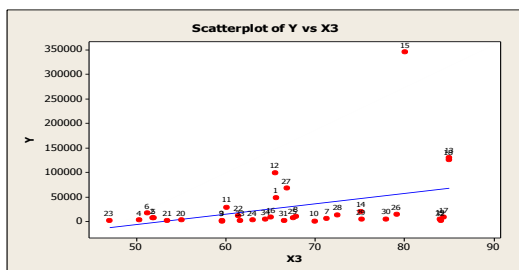
Cara melihat bentuk pola data produksi kedelai di Indonesia tahun 2015 terhadap faktor-faktor yang mempengaruhinya yaitu dengan menggunakan *scatter* plot yang diperoleh menggunakan *software* Minitab 16.



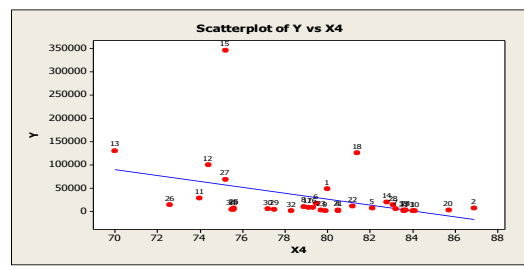
Gambar 1. Plot Data Produksi Kedelai dengan Luas Panen



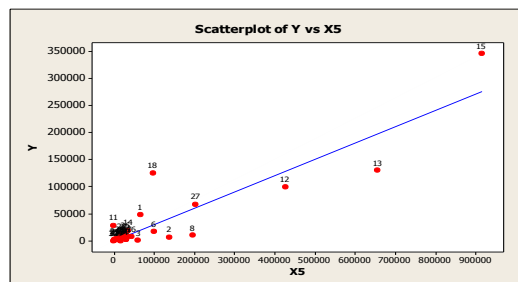
Gambar 2. Plot Data Produksi Kedelai dengan Curah Hujan



Gambar 3. Plot Data Produksi Kedelai dengan Intensitas Cahaya



Gambar 4. Plot Data Produksi Kedelai dengan Kelembaban



Gambar 5. Plot Data Produksi Kedelai dengan Pupuk

2.3 Pendugaan Model Regresi dengan Metode Kuadrat Terkecil

Pendugaan model regresi dimaksudkan untuk mengekspresikan hubungan linear antara peubah terikat (produksi kedelai) dan peubah bebas (luas panen, curah hujan, intensitas cahaya, kelembaban dan pupuk) yang diperoleh dari *software* Minitab 16. Berdasarkan *output* diperoleh nilai parameter-parameter regresi, sehingga diperoleh model regresi yang terbentuk yaitu:

$$\hat{Y} = 24930 + 1,43X_1 + 1,75X_2 - 9,0X_3 - 368X_4 + 0,0444X_5$$

2.4 Pengujian Asumsi Klasik

a. Uji Linearitas

Berdasarkan *output* diperoleh nilai F_{hitung} sebesar 2031,03 dan nilai kritis F sebesar 2,56. Jadi, nilai F_{hitung} lebih besar dari nilai kritis F maka disimpulkan bahwa peubah terikat (produksi kedelai) dan peubah bebas (luas panen, curah hujan, intensitas cahaya, kelembaban dan pupuk) diasumsikan linear dalam parameter-parameter regresinya atau asumsi linearitas terpenuhi.

b. Uji Multikolinearitas

Berdasarkan *output* diperoleh nilai *Variance Inflation Factor* (*VIF*) masing-masing peubah bebas kurang dari 10, hal ini menandakan bahwa data tidak mengalami masalah multikolinearitas atau asumsi multikolinearitas terpenuhi.

c. Uji Autokorelasi

Berdasarkan *output* diperoleh nilai d sebesar 1,958, $d_L = 1,1439$ dan $d_U = 1,8076$ sedangkan $4 - d_U = 2,1924$, sehingga nilai d berada dalam batas $d_L < d < 4 - d_U$ sebagai kriteria penerimaan hipotesis nol sehingga dapat disimpulkan bahwa tidak ada korelasi antar galat atau dengan kata lain asumsi uji autokorelasi terpenuhi.

d. Uji Heteroskedastisitas

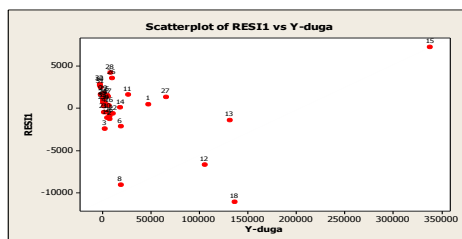
Berdasarkan *output* diketahui bahwa nilai P masing-masing peubah bebas lebih dari tingkat signifikansi α yang ditetapkan. Hal ini menunjukkan bahwa hipotesis nol diterima yang artinya tidak terdapat heteroskedastisitas sehingga asumsi heteroskedastisitas terpenuhi maka tidak terjadi masalah heteroskedastisitas pada model regresi.

e. Uji Normalitas

Berdasarkan analisis diperoleh nilai P -value sebesar 0,010 dengan α sebesar 0,05. Hal ini berarti nilai P -value ($0,010 < \alpha$ ($0,05$)) maka hipotesis nol ditolak, sehingga dapat dikatakan bahwa nilai-nilai galat dari model regresi tidak berdistribusi normal atau asumsi normalitas tidak terpenuhi.

2.5 Mengidentifikasi Pencilan Poduksi Kedelai di Indonesia Tahun 2015

1. Metode *Scatter* plot



Gambar 6. Plot Data antara Nilai Sisaan dan Nilai Dugaan

2. Nilai Leverasi

Berdasarkan *output* diperoleh nilai-nilai leverasi dengan 5 peubah bebas dan banyaknya amatan adalah 34 yakni $h_2 = 0,453275$, $h_{13} = 0,53031$, $h_{15} = 0,730512$, $h_{18} = 0,524562$. Nilai ini tersebut melebihi kriteria dua kali rata-rata nilai leverasi sebesar 0,352941, sehingga amatan ke-2, 13, 15, dan 18 dianggap sebagai pencilan ditinjau dari nilai-nilai X .

3. Sisaan Dibuang Terstudentkan

Berdasarkan *output* diperoleh nilai $|d_8^*| = 2,98465$, $|d_{12}^*| = 2,04245$, $|d_{15}^*| = 5,058694$, dan $|d_{18}^*| = 7,22896$. Amatan-amatan tersebut melebihi $t_{(0,95,31)} = 1,70$, sehingga amatan ke-8, 12, 15, dan ke-18 terdeteksi sebagai pencilan ditinjau dari nilai-nilai Y .

Untuk mengetahui apakah amatan ke-2, 8, 12, 13, 15, dan ke-18 yang terdeteksi sebagai pencilan juga merupakan amatan berpengaruh maka dideteksi dengan nilai *DFFITs*.

2.6 Pendeteksian Amatan Berpengaruh pada Produksi Kedelai di Indonesia Tahun 2015 yang Teridentifikasi Pencilan

1. *DFFITs*

Berdasarkan *output* diperoleh nilai $|DFFITs_{(8)}| = 1,03861$, $|DFFITs_{(12)}| = 0,92093$, $|DFFITs_{(15)}| = 8,442172$, dan $|DFFITs_{(18)}| = 7,64144$. Nilai-nilai *DFFITs* tersebut yang lebih besar dari nilai dua kali akar $p/n = 0,840186$ yakni amatan ke-8, 12, 15, dan 18, sehingga pengamatan tersebut teridentifikasi sebagai amatan berpengaruh.

2.7 Prosedur Regresi *Robust* dengan Penduga *Method of Moment*

1. Menghitung nilai e_i

$$e_1 = Y_1 - \hat{Y}_1 = 47910 - 47515,24 = 394,756$$

Demikian seterusnya sampai pada pengamatan ke-34 seperti pada. Residual $e_1^{(1)}$ pada langkah pertama digunakan untuk menghitung pembobot awal $w_i^{(1)}$ dengan menggunakan pembobot *Tukey Bisquare*.

2. Menghitung nilai $\hat{\sigma}_s$

$$\hat{\sigma}_s = \sqrt{\frac{n \sum_{i=1}^n (e_i^2) - (\sum_{i=1}^n e_i)^2}{n(n-1)}} = \sqrt{\frac{34(389380584) - 1846648}{34(34-1)}} = 3434,788$$

3. Menentukan nilai $|u_i|$ dengan rumus $|u_i| = \frac{e_i}{\hat{\sigma}_s}$

$$|u_1| = \frac{394,756}{3434,788} = 0,114929$$

Demikian seterusnya untuk pengamatan ke-2 sampai pengamatan ke-34.

4. Menentukan nilai $w_i^{(1)}$ dengan menggunakan pembobot *Tukey Bisquare*.

$$w_1^{(1)} = \left[1 - \left(\frac{|u_i|}{c} \right)^2 \right]^2 = \left[1 - \left(\frac{0,114929}{4,685} \right)^2 \right]^2 = 0,998797$$

Demikian seterusnya untuk amatan ke-2 sampai amatan ke-34.

Berikut ini merupakan nilai $\hat{\beta}_j$ pada iterasi demi iterasi yang didapat dari *software* Minitab 16.

Tabel 2. Nilai $\hat{\beta}_j$ Iterasi Demi Iterasi

Iterasi	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
1	19490	1,48	1,40	13,8	-306	0,0359
2	13667	1,54	1,12	33,7	-241	0,0250
3	13014	1,55	1,10	35,7	-235	0,0227
4	12905	1,56	1,10	35,6	-233	0,0223
5	12999	1,56	1,11	35,7	-235	0,0224
6	13008	1,56	1,11	35,7	-235	0,0224
7	13009	1,56	1,11	35,7	-235	0,0224
8	13009	1,56	1,11	35,7	-235	0,0224

Berdasarkan Tabel 2 terlihat bahwa koefisien regresi konvergen pada iterasi ke-8 dengan pendugaan model yang terbentuk :

$$\hat{Y} = 13009 + 1,56 X_1 + 1,11 X_2 + 35,7 X_3 - 235 X_4 + 0,0224 X_5$$

2.8 Pengujian Parameter Model Penduga-MM

1. Pengujian Parameter Secara Simultan

Berdasarkan *output* diperoleh nilai F_{hitung} sebesar 5851,74 dan nilai kritis F sebesar 2,57, karena nilai F_{hitung} melebihi nilai kritis F maka hipotesis nol ditolak, sehingga dapat disimpulkan peubah bebas secara simultan signifikan terhadap model yakni peubah luas panen, curah hujan, intensitas cahaya, kelembaban dan pupuk.

2. Pengujian Parameter Secara Parsial

Berdasarkan *output* diperoleh nilai $|t_{hitung}|$ masing-masing peubah X_1, X_2, X_3, X_4 , dan X_5 yaitu 57,90; 1,82; 0,85; 2,00; dan 4,25 dan nilai kritis t sebesar 1,70, sehingga dapat disimpulkan bahwa peubah luas panen, curah hujan, kelembaban dan pupuk berpengaruh terhadap produksi kedelai. Sedangkan parameter yang tidak signifikan yakni peubah X_3 dengan nilai $|t_{hitung}|$ lebih kecil dari 1,70, yang artinya bahwa peubah intensitas cahaya tidak berpengaruh terhadap produksi kedelai.

2.9 Perbandingan Hasil Metode Kuadrat Terkecil dengan Penduga-MM pada Model yang Konvergen

Tabel 3. Perbandingan Hasil Metode Kuadrat Terkecil dengan Penduga-MM

Metode	Koefisien Determinasi	MSE	Standar Error
MKT	99,7%	13897488	3727,933
Penduga-MM	99,9%	4460298	2111,942

Berdasarkan Tabel 3 diperoleh bahwa penduga-MM lebih baik dari metode kuadrat terkecil. Hal ini dapat dilihat dari nilai MSE dan standar error penduga-MM yang lebih kecil dari metode kuadrat terkecil serta nilai koefisien determinasi yang lebih tinggi dari metode kuadrat terkecil, sehingga regresi *robust* dengan metode penduga-MM dapat menjadi solusi permasalahan Metode kuadrat terkecil terhadap data yang terdapat pencilan.

4. Kesimpulan

Penggambaran produksi kedelai di Indonesia tahun 2015 menunjukkan bahwa peubah bebas pupuk memiliki nilai rata-rata yang tinggi, sedangkan peubah bebas intensitas cahaya dan kelembaban memiliki nilai rata-rata yang rendah. Dari penggambaran produksi kedelai di Indonesia tahun 2015 juga menunjukkan bahwa terdapat beberapa amatan yang jauh dari amatan lainnya yang ditunjukkan pada hasil plot data produksi kedelai di Indonesia tahun 2015 terhadap faktor-faktor yang mempengaruhinya, sehingga analisis yang sesuai digunakan adalah analisis regresi *robust*.

Pendeteksian pencilan pada data produksi kedelai di Indonesia tahun 2015 diperoleh amatan yang teridentifikasi sebagai pencilan yaitu amatan ke-2, 8, 12, 13, 15 dan 18. Sedangkan pendeteksian amatan berpengaruh dilakukan dengan menggunakan metode *DFFITs*, sehingga diperoleh amatan ke-8, 12, 15 dan 18 sebagai amatan berpengaruh.

Penerapan analisis regresi *robust* dengan penduga-MM menggunakan pembobot *Tukey Bisquare* pada data produksi kedelai di Indonesia tahun 2015 diperoleh model persamaannya yaitu $\hat{Y} = 13009 + 1,56 X_1 + 1,11 X_2 + 35,7 X_3 - 235 X_4 + 0,0224 X_5$ dengan nilai R^2 sebesar 99,9% dengan nilai MSE sebesar 4460298. Adapun peubah bebas yang berpengaruh

pada produksi kedelai di Indonesia tahun 2015 yaitu luas panen (X_1), curah hujan (X_2), kelembaban (X_4) dan pupuk (X_5).

Ucapan Terimakasih

Dalam penyusunan tulisan ini, banyak pihak yang telah memberikan dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada seluruh dosen dan pimpinan Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta.

Daftar Pustaka

- [1] Ardiyanti H, 2011, *Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan estimasi-MM karena Pengaruh Outlier dalam Analisis Regresi Linear*, *UNNES Journal of Mathematics*, Universitas Negeri Semarang, Semarang
- [2] Candraningtyas S, Safitri D dan Ispriyanti D, 2013, *Regresi Robust MM-Estimator untuk Penanganan Pencilan pada Regresi Linier Berganda*, *Jurnal Gaussian*, Volume 2, Nomor 4, 394-404, UNDIP, Semarang.
- [3] Chen C, 2003, *Robust Regression and Outlier Detection with the ROBUSTREG Procedure*, *SUGI Paper*, 265-27.
- [4] Gujarati D N, 1978, *Ekonometrika Dasar*, terjemahan Sumarno Zain, Penerbit Erlangga, Jakarta.
- [5] Irfagutami Nia N P, Srinadi I G A M dan Sumarjaya I W, 2014, *Perbandingan Regresi Robust Penduga MM dengan Metode Random Sample Consensus dalam Menangani Pencilan*, *E-Jurnal Matematika*, Vol.3, No.2, 45-52, Universitas Udayana, Bali.
- [6] Neter J & Wasserman W, 1974, *Applied Linear Statistical Models (Regression, Analysis of Variance, and Experimental Designs)*, Richard D. Irwin Inc, Ontario.
- [7] Noeryanti, 2012, *Metode Statistika II*, AKPRIND PRESS, Yogyakarta.
- [8] Rowlings O J, Pantula G S, dan Dickey A D, 1998, *Applied Regresi Analysis A Research Tool, Second Edition*, Library of congress cataloging-in-publication data, Springer-Verlag New York Inc, New York.
- [9] Supranto J, 2001, *Statistik: Teori dan Aplikasi, Edisi Keenam*, Penerbit Erlangga, Jakarta.
- [10] Suryowati K, 2016, *Analisis Pseudoinvers dan Aplikasinya pada Regresi Linear Berganda, Prosiding Seminar Aplikasi Sains & Teknologi (SNAST) PeriodeIV*, Yogyakarta.
- [11] Suyanti, 2014, *Deteksi Outlier Menggunakan Diagnosa Regresi Berbasis Estimator Parameter Robust*, *UNNES Journal of Mathematics*, Universitas Negeri Semarang, Semarang.
- [12] Walpole RE & Myers RH, 1995, *Ilmu Peluang dan Statistikan untuk Insinyur dan Ilmuwan Edisi keempat*, terjemahan RK Sembiring, Penerbit ITB, Bandung.
- [13] ——— (2016), Statistik Publikasi Pertanian, diakses tanggal 22 September 2016 pada situs <http://www.bps.go.id/>