ESTIMASI *METHOD OF MOMENT* DENGAN FUNGSI PEMBOBOT HUBER PADA PEMODELAN DATA TERIDENTIFIKASI PENCILAN

Herlina[1), Ruslan[2) ,
Universitas Halu Oleo
Jl. HEA Mokodompit Kampus Baru UHO Kendari
email: herlinastik@gmail.com
email: rushlan_@yahoo.com

Abstract

The purpose of this study is to obtain a robust estimator using the moment method using Huber weighting in multiple linear regression modeling on the identified data and compare the results of the analysis between weighted Bisquare Tukey and Huber weights. The usual methods used to estimate regression parameters include the Least Square Method (MKT). Robust regression is a method that can generate a robust parameter estimator of the outliers. A robust estimator is relatively unaffected to large changes in small pieces of data or small changes in large parts of the data. The MM estimation is a good method to cope with the results of robust estimators (robust) with high breakdown points and high efficiency and involves multiple weights, the weighted Bisquare Tukey and Huber weights. The result of the analysis shows that the data is identified by the method of moment (MM) using weighted Bisquare Tukey more robust than using Huber weights it is aimed with the coefficient of determinant value (R^2) and the middle quadratic error value (KTG) gives result Which is better and more efficient.

Keywords: *Least square Method, Evalengism, Weighing Bisquare Tukey, Weighted Huber, Robust Regression with MM Estimation*

1. PENDAHULUAN

Regresi linear adalah metode statistika yang digunakan untuk membentuk model hubungan antara peubah terikat atau dependen (Y) dengan satu atau lebih peubah bebas atau independen (X). Apabila banyaknya peubah bebas hanya ada satu, maka disebut sebagai regresi linear sederhana, sedangkan apabila terdapat lebih dari satu peubah bebas, maka disebut sebagai regresi linear berganda (Kurniawan, 2008).

Dalam sejumlah data hubungan sebenarnya jarang dapat diketahui akan tetapi hubungan tersebut dapat diestimasi berdasarkan data pengamatan. Dalam menentukan estimator terbaik sangat dipengaruhi oleh penggunaan metode. Metode yang biasa digunakan untuk mengestimasi parameter regresi antara lain adalah Metode Kuadrat Terkecil (MKT). Metode ini memberikan hasil yang terbaik dalam menaksir nilai- nilai koefisien parameter regresinya serta memiliki sifat tak berbias, efisien dan konsisten (Neter *et al.*, 1997).

MKT ini memerlukan pengujian asumsi klasik yang harus dipenuhi. Beberapa asumsinya antara lain adalah komponen sisaan memenuhi asumsi kenormalan, kehomogenan ragam, multikolinearitas dan tidak terjadi autokorelasi. Dalam berbagai kasus tidak jarang ditemukan berbagai hal yang menyebabkan tidak terpenuhinya asumsi klasik tersebut. Hal ini akan mempengaruhi model dengan interpretasi, oleh karena itu saat asumsi klasik tidak terpenuhi, metode ini harus dihindari.

Salah satu penyebab tidak terpenuhinya asumsi klasik adalah adanya pencilan yang menyebabkan pelanggaran terhadap asumsi kenormalan. Pencilan adalah pengamatan yang jauh dari pusat data yang dapat mengakibatkan estimasi koefisien regresi yang diperoleh tidak tepat (Walpole & Myers, 1995). Hal ini dapat ditunjukan dengan nilai standar error yang besar apabila menggunakan MKT.

Pencilan dapat dilihat sebagai pengamatan dengan sisaan yang cukup besar (Soemartini, 2007). Pencilan seringkali mempunyai pengaruh besar terhadap model taksirannya. Oleh karena itu suatu pencilan akan diperiksa secara seksama dan kemudian diputuskan apakah data itu harus dihilangkan atau masih dapat dipertahankan dan jika dihilangkan, apakah data tersebut mempunyai pengaruh terhadap proses pendugaan model (Neter *et al.*, 1997).

Salah satu metode regresi yang digunakan ketika terdapat data pencilan adalah metode regresi *robust*. Regresi *robust* merupakan suatu metode yang dapat menghasilkan penduga parameter yang kekar terhadap pencilan.

Metode estimasi dalam regresi *robust* diantaranya estimasi M (*Maximum Likelihood Type*), LTS (*Least Trimmed Square*), estimasi LMS (*Least Median of Square*), estimasi S (*Scale*) dan estimasi MM (*Method of Moment*) (Cholin Chen, 202:1). Estimasi M adalah metode paling sederhana yang banyak digunakan dan mempunyai nilai efisiensi yang tinggi, sedangkan estimasi S, LST, LMS adalah estimasi dengan nilai *breakdown point* tinggi, akan tetapi estimasi S mempunyai nilai *breakdown point* yang tinggi diantara ketiganya. Dan estimasi yang menghasilkan estimator yang *robust* serta menghasilkan *breakdown point* yang tinggi dengan efesiensi tinggi adalah estimasi MM, sehingga metode estimasi ini baik untuk menanggulangi pencilan.

Dalam estimasi MM terdapat beberapa pembobot, yaitu pembobot *Bisquare Tukey* dan pembobot *Huber*. Pembobot *Bisquare Tukey* telah digunakan pada penelitian sebelumnya oleh Syawaluddin (2013).

Berdasarkan uraian di atas penulis mengangkat judul “Estimasi *Method of Moment* (MM) dengan fungsi pembobot *Huber* pada data teridentifikasi pencilan.

2. KAJIAN LITERATUR DAN PENGEMBANGAN HIPOTESIS

2.1. Regresi Linear Berganda

Model regresi linear berganda secara umum dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{i,k} + \varepsilon_i \tag{2.1}$$

Dalam notasi matriks, model regresi linear berganda dapat ditulis sebagai berikut:

$$Y = X\beta + \varepsilon \tag{2.2}$$

Y merupakan peubah terikat yang merupakan matriks yang berukuran $n \times 1$, X adalah matriks yang berukuran $n \times (k+1)$, β adalah matriks koefisien regresi yang berukuran $(k+1) \times 1$ dan $\boldsymbol{\varepsilon}$ adalah sisaanyang berukuran $n \times 1$, dalam notasi matriks dapat dituliskan sebagai berikut (Neter *et al.*, 1997):

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.3}$$

dimana: Y adalah vektor respon, β adalah vektor parameter, X adalah matriks rancangan, dan ε adalah vektor peubah acak normal bebas dengan nilai harapan $E\{\boldsymbol{\varepsilon}\} = 0$ dan matriks ragam peragam $\sigma^2\{\varepsilon\} = \sigma^2 I$.

2.2. Penduga Kuadrat Terkecil

Metode yang sering digunakan dalam parameter regresi adalah metode kuadrat terkecil. Metode ini bertujuan untuk meminimumkan jumlah kuadrat semua simpangan dari garis yang sebenarnya. Vektor model taksiran untuk $b_0, b_1, \ldots, b_{k+1}$ dapat dirumuskan sebagai berikut:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2.4}$$

 ISBN : 2685-5852

Nilai $\boldsymbol{b}$ adalah nilai dugaan bagi unsur-unsur β, $(\boldsymbol{X'X})^{-1}$ adalah matriks peubah bebas yang berukuran $(k+1)\times(k+1)$, $\boldsymbol{X'}$ adalah matriks peubah bebas yang berukuran $(k+1)\times n$ dan $\boldsymbol{Y}$ adalah matriks peubah terikat yang berukuran $n\times 1$. Untuk mendapatkan nilai $\hat{Y}$ dengan menggunakan rumus berikut:

$$\hat{Y} = Xb \tag{2.5}$$

dengan $\hat{Y}$ adalah nilai duga, X adalah data pada kasus ke-i yang merupakan matriks dengan ukuran $n\times k$ dan b adalah nilai yang diperoleh dari persamaan 4, sedangkan untuk memperoleh nilai sisaanya adalah sebagai berikut:

$$e_i = y_i - \hat{y}_i \tag{2.6}$$

nilai e_i adalah nilai sisaan untuk data ke-i, $\boldsymbol{y_i}$ adalah data pada kasus ke-i yang merupakan matriks dengan ukuran $n\times 1$ dan $\hat{y}_i$ adalah nilai dugaan data pada kasus ke-i.

2.3 Asumsi-asumsi Regresi Linear

asumsi utama dalam pemodelan regresi, yaitu:

1. Multikolinearitas yaitu analisis peubah bebas X_1, X_2, ..., X_k, saling bebas atau tidak berkorelasi secara kuat dan signifikan.
2. Kebebasan galat, asumsi ini menyatakankan bahwa galat dari model regresi, ε_1, ε_2,...,ε_n saling bebas. Jika asumsi ini tidak terpenuhi berarti terjadi masalah korelasi antara galat.
3. Asumsi homoskedastisitas berarti galat dari model regresi menpunyai ragam σ^2 yang sama. Apabila asumsi ini tidak terpenuhi maka terjadi masalah heteroskedastisitas.
4. Normalitas yaitu galat atau residual dari model regresi ε_1, ε_2, ..., ε_n berdistribusi normal [ε-N(0,σ2)].

2.4. Analisis Ragam

Ragam setiap data Y_i adalah σ^2 yang yang berasal dari sisaan ε_i dalam model regresi, sehingga nilainya harus diduga untuk memperoleh gambaran tentang keragaman Y_i. Untuk memperoleh jumlah kuadrat regresi (JKR) dapat dirumuskan sebagai berikut:

$$\boldsymbol{JKR = b'X'Y - \left(\frac{1}{n}\right)Y'JY} \tag{2.12}$$

dimana $\boldsymbol{b'}$ adalah matriks yang diperoleh dari persamaan 4, $\boldsymbol{X'}$ adalah matriks peubah bebas yang berukuran $n\times k$, $\boldsymbol{Y}$ adalah matriks peubah terikat yang berukuran $n\times 1$, n adalah banyaknya data dan $\boldsymbol{J}$ adalah matriks yang semua unsurnya adalah 1 dan ukuran $n\times 1$.

Untuk memperoleh jumlah kuadrat galat (JKG) dapat menggunakan rumus sebagai berikut:

$$\boldsymbol{JKG = e'e} \tag{2.13}$$

dimana $\boldsymbol{e}$ adalah nilai yang diperoleh dari persamaan 6 dan $\boldsymbol{e'} = [e_1, e_2, \dots, e_n]$.

Untuk memperoleh jumlah kuadrat total (JKT) dapat diproleh dengan rumus sebagai berikut:

$$\boldsymbol{JKT = Y'Y - \left(\frac{1}{n}\right)Y'JY} \tag{2.14}$$

Untuk menentukan nilai kuadrat tengah regresi (KTR) dengan menggunakan persamaan berikut ini:

$$KTR = \frac{JKR}{1} \tag{2.15}$$

Untuk menentukan nilai kuadrat tengah galat (KTG) dengan mengunakan persamaan berikut ini:

$$KTG = \frac{JKG}{n-1} \tag{2.16}$$

Untuk mendapatkan nilai koefisien determinansi (R^2) digunakan persamaan berikut:

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT} \tag{2.17}$$

2.5 Mengidentifikasi Pencilan

Untuk mengedintifikasi pencilan dapat dideteksi dengan menggunakan metode grafik untuk melihat pencilan secara *visual* dan menggunakan metode perhitungan dengan menggunakan nilai *leverasi* untuk melihat pencilan X dan sisaan dibuang ter-*student*-kan untuk melihat pencilan Y (Neter *et al.*, 1997).

 ISBN : 2685-5852

1. Metode *Scatter plot*

Salah satu metode grafik yang digunakan yaitu metode *Scatter Plot*, metode ini didapatkan dari model regresi yang dapat dilakukan dengan cara memplot sisaan (ε) dengan nilai prediksi Y ($\hat{Y}$). Jika terdapat satu atau beberapa yang terletak jauh dari pada kumpulan data keseluruhan maka hal ini mengindentifikasi adanya pencilan (Soemartini, 2007).

2. Nilai h_{ii} (*Leverasi*)

Untuk mengidentifikasi data pencilan terhadap nilai X digunakan nilai *leverasi* (h_{ii}). Nilai *leverasi* yang besar mengidetifikasi bahwa data ke-*i* berada jauh dari pusat semua data X. Suatu nilai *leverasi* dianggap besar jika lebih dari dua kali nilai rataan semua *leverasi*. Rataan nilai *leverasi* dapat dirumuskan oleh:

$$\bar{h} = \frac{1}{n}\sum_{i=1}^{n} h_{ii} = \frac{p}{n} \tag{2.18}$$

dengan demikian, nilai *leverasi* lebih besar dari 2*p/n* yang dianggap mengidentifikasi data pencilan ditinjau dari nilai-nilai X dapat dirumuskan sebagai berikut:

$$\boldsymbol{h_{ii} = X_i(X'X)^{-1}X_i'} \tag{2.19}$$

nilai $\boldsymbol{h}_{\text{ii}}$ adalah suatu ukuran jarak antara nilai-nilai X, $\boldsymbol{X_i}$ adalah vektor baris ke-*i* yang berukuran $1 \times (k+1)$, $\boldsymbol{(X'X)^{-1}}$ adalah matriks yang berukuran peubah bebas $(k+1) \times (k+1)$ dan matriks $\boldsymbol{X_i'}$ adalah vektor kolom ke-*i* berukuran $(k+1) \times 1$. Apabila nilai $h_{ii} > 2p/n$, maka dianggap pencilan, dimana p adalah banyaknya parameter dan n adalah jumlah data (Neter, 1997).

3. Sisaan Dibuang Ter-*student*-kan

Untuk mengidentifikasi pencilan Y, akan dicari sisaan dibuang ter-*student*-kan yang bernilai mutlak terbesar, kemudian menggunakan sebaran *t*, untuk melihat data yang dianggap sebagai pencilan. Pengaruh pencilan data Y berdasarkan pada pemeriksaan sisaan dibuang ter-*student*-kan d_i^* yang dirumuskan sebagai berikut:

$$d_i^* = e_i\left[\frac{n-p-1}{JKG(1-h_{ii}-e_i{}^2)}\right]1/2 \tag{2.20}$$

dimana d_i^* adalah sisaan dibuang ter-*student*-kan, e_i adalah nilai yang diperoleh dari persamaan 6, h_{ii} adalah nilai leverasi yang diperoleh dari persamaan 19 dan JKG adalah jumlah kuadrat galat yang dihasilkan dari fungsi regresi dengan menyetarakan kasus ke-*i* yang diperoleh dari persamaan 13.

Diketahui bahwa $d_i^* \sim t_{(n-p-1,\alpha)}$, setiap sisaan dibuang ter-*student*-kan menyebar mengikuti sebaran *t* dengan derajat bebas $(n-p-1)$. data dengan $|d_i^*| > t_{(n-p-1)}$ dapat dianggap sebagai pencilan (Cook & Weisberg, 1982).

2.6 Mengidentifikasi Data Berpengaruh

Suatu data dikatakan berpengaruh apabila ketidaksertaan data ini akan menyebabkan perubahan besar dari pada proses pendugaan model. Untuk mengetahui suatu data berpengaruh atau tidak dapat ditentukan berdasarkan metode DFFITS.

DFFITS merupakan nilai perubahan dalam harga yang diprediksi ($\hat{Y}$) jika data tertentu dikeluarkan. DFFITS dapat dihitung dengan persamaan berikut ini:

$$(\text{DFFITS})_i = d_i^*\left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} \tag{2.21}$$

apabila nilai $|\text{DFFITS}| > 2\sqrt{p/n}$, maka kasus ke-*i* yang tidak disertakan merupakan data berpengaruh (Rowlings *et al.*,1998).

2.7 Regresi Robust

1. *Estimasi Method of Moment* (MM)

MM dikenalkan oleh Yohai pada tahun 1987 yang menghubungkan suatu *high breakdown point* dengan efisiensi statistik (Wilcox, 2005). Estimasi S menjamin nilai

 ISBN : 2685-5852

breakdown point yang tinggi dan estimasi M yang membuatnya mempunyai efisiensi yang tinggi. Estimasi MM juga mengunakan iterasi untuk mencari estimasi parameter regresi.

Fungsi pembobot untuk Huber dapat dituliskan sebagai berikut:

$$w_i = \begin{cases} 1, & |e_i{}^*| \le c \\ \dfrac{c}{|e_i{}^*|}, & |e_i{}^*| > c \end{cases} \tag{2.22}$$

dengan w_i merupakan fungsi pembobot iterasi dimana $e_i{}^* = \frac{e_i}{\hat{\sigma}_s}$ dan c merupakan konstanta pembobot *Huber* sebesar 1,345 serta $e_i{}^*$ adalah nilai residual yang distandarisasi (Yohai, 1987).

Untuk nilai $\hat{\sigma}_s$ dapat diperoleh dengan rumus sebagai berikut:

$$\hat{\sigma}_s = \frac{MAR}{0{,}6745} \tag{2.23}$$

dimana MAR adalah *median absolute residual* yang dapat dicari dengan rumus $\frac{1}{n}\sum_i^n \left|Y_i - \hat{Y}_i\right|$

2. *Breakdown point*

Breakdown point adalah salah satu cara yang digunakan untuk mengukur kekekaran suatu estimator. *Breakdown point* merupakan proporsi minimum dari banyaknya pencilan yang dapat ditangani sebelum pengamatan tersebut mempengaruhi model dibandingkan seluruh data pengamatan. Regresi robust yang mempunyai *Breakdown point* tinggi adalah regresi robust dengan metode estimasi *Scale* (S), *Least Trimmed Square* (LTS) dan *Method of moment* (MM).

3. Penyelesaian untuk $\tilde{\beta}$

Untuk meminimumkan ρ (fungsi obyektif) dari residualnya, dicari turunan parsial pertama ρ terhadap β_j, dimana $\beta_j = 0, 1, 2, \ldots, k$ dan disama dengan 0. Ini memberikan $p = k + 1$ dengan sistem persamaan:

$$\sum_{i-1}^{n} x_{ij}\psi\left[y_i \sum_{j=0}^{k} x_{ij}\beta_j\right] = 0 \tag{2.24}$$

dengan $\psi = p'$ dan ψ merupakan fungsi *influence* yang digunakan dalam memperoleh bobot, x_{ij} adalah observasi ke-*i* pada regressor ke-*j* dan $x_{i0} = I$.
didefenisikan suatu fungsi pembobot.

$$w(e_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s}\right)}{\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\hat{\sigma}_s}} \tag{2.25}$$

dimisalkan $w_i = w(e_i)$, maka persamaan (25) dapat ditulis:

$$\sum_{i=1}^{n} x_{ij}w_i\left[y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right] = 0, j = 1,2, \ldots k \tag{2.26}$$

persamaan 25 diselesaikan dengan IRLS (*Iterasi Reweighted Least*), estimasi awal koefisien $\tilde{\beta}^{(1)}$ dan residual $e_i{}^{(1)}$ diambil dari regresi robust dengan *high breakdown point* (estimasi S), untuk bobot permulaan $w_i{}^{(1)} = w\left(e_i{}^{(1)}\right)$, maka $p = k$ persamaan 21 dapat ditulis:

$$\sum_{i=0}^{n} x_{ij}w_i{}^{(1)}\left[y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right] = 0 \tag{2.27}$$

untuk regresi berganda, persamaan 21 menjadi:

$$\sum_{i=0}^{n} x_{ij}w_i{}^{(1)}\left[y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right] = 0 \tag{2.28}$$

$$\sum_{i=0}^{n} w_i^{(1)} x_i y_i - \sum_{i=0}^{n}\sum_{j=0}^{k} x_{ij}^2 w_i^{(1)} \widehat{\beta_J} = 0$$
$$\sum_{i=0}^{n}\sum_{j=0}^{k} x_{ij}^2 w_i^{(1)} \widehat{\beta_J} = \sum_{i=0}^{n} w_i^{(1)} x_i y_i$$

dalam bentuk matriks dapat ditulis:

$$\boldsymbol{X'W^{(1)}X\widetilde{\beta_J}^{(2)} = X'W^{(1)}y}$$
$$\boldsymbol{(X'W^{(1)}X)^{-1}X'W^{(1)}X\widetilde{\beta_J}^{(2)} = (X'W^{(1)}X)^{-1}\, X'W^{(1)}y}$$
$$\boldsymbol{\widetilde{\beta}_{jpx1}}^{(2)} = (\boldsymbol{X'W^{(1)}X})_{\boldsymbol{pxp}}{}^{-1}(\boldsymbol{X'W^{(1)}Y_{px1}}) \qquad (2.29)$$

pada langkah selanjutnya mengitung kembali bobot $w_i^{(2)}$ menggunakan $\widetilde{\beta_J}^{(2)}$dan skala parameter $\hat{\sigma}_s$, untuk $w_i^{(m)}$ bobot yang diberikan, dapat diperoleh estimator $\widetilde{\beta_J}^{(m+1)} = (X'^{W^m}X)^{-1}X'^{W^m}Y$ sampai $\sum_{i=0}^{n}\left|e_i^{(m)}\right|$ konvergen (selisih nilai $\tilde{\beta}^{(m+1)}$ dan $\tilde{\beta}^{(m)}$ mendekati 0 atau sama dengan), dengan m banyaknya iterasi (Kurniawan, 2008).

2.8 Pengujian Parameter Regresi Berganda

1. Pengujian Simultan

Untuk menguji atau mengukur signifikansi model hubungan antara peubah bebas dengan peubah terikat yang menjelaskan dari persamaan regresi secara menyeluruh disebut uji *F*.

Hipotesis untuk uji signifikansi model secara simultan:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_j = 0$ (Tidak ada pengaruh peubah bebas terhadap peubah terikat)

H_1 : paling sedikit ada satu $\beta_j \neq 0$ dengan $j = 1, 2, \ldots, k$ (minimal ada satu peubah bebas yang berpengaruh terhadap peubah terikat)

Statistik Uji *F* dinyatakan dalam bentuk rumus sebagai berikut:

$$F_{hitung} = \frac{MSR}{MSE} \qquad (2.29)$$

dimana,

MSR = Rata-rata kuadrat regresi

MSE = Rata-rata kuadrat sisaan

Kriteria Uji:

Tolak H_0 jika $F_{hitung} > F_{tabel}$ atau nilai $P \leq \alpha$ dan sebaliknya. Artinya minimal ada satu peubah bebas yang mempengaruhi peubah terikat (Sembiring, 2003).

2. Pengujian Parsial

Untuk menguji signifikansi parameter model secara masing-masing antara peubah bebas terhadap peubah terikat dapat digunakan statistik uji *t*.

Hipotesis untuk uji signifikansi model secara parsial:

H_0 : $\beta_i = 0$ (Tidak ada pengaruh peubah bebas ke-*i* terhadap peubah terikat)

H_1 : $\beta_i \neq 0$ (Ada pengaruh peubah bebas ke-*i* terhadap peubah terikat untuk $i = 1, 2, \ldots, p-1$)

Statistik uji *t* dinyatakan dalam rumus sebagai berikut:

$$t_{hitung} = \frac{\beta_i}{se(\beta_i)} \qquad (2.30)$$

Kriteria uji:

Tolak H_0 jika $\left|t_{hitung}\right| > t_{tabel}$ atau nilai $P \leq \alpha$ dan sebaliknya. Dalam hal ini dapat dikatakan bahwa ada pengaruh peubah bebas ke-*i* terhadap peubah terikat secara statistik (Sembiring, 2003).

3. METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder tentang pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana dengan peubah yang diamati sebagai berikut (Syawaluddin, 2013).

Y adalah pendapatan bersih nelayan (Rp.juta/trip)
X_1 adalah jenis alat tangkap yang digunakan oleh nelayan (unit)
X_2 adalah jumlah armada penangapan (kendaraan) (unit)
X_3 adalah armada tenaga kerja yang digunakan dalam proses penangkapan
X_4 adalah pengalaman kerja (pertahun)
X_5 adalah biaya yang dikeluarkan dalam membiayai seluruh proses penangkapan (Rp.ribu/trip).

3.2 Prosedur Penelitian

Adapun langkah–langkah analisis data dalam penelitian ini adalah sebagai berikut:

1. Membetuk penaksiran parameter regresi dengan menggunakan persamaan 4
2. Pengujian asumsi model regresi linear
3. Pengujian parameter model secara simultan dan parsial menggunakan persamaan 29 dan 30
4. Mendeteksi pencilan dengan metode *scatter plot*, nilai *leverasi* dan sisaan dibuang ter-*student*-kan serta mendeteksi data berpengaruh dengan *DFFITS* menggunakan bantuan *software* Minitab
5. Penanganan data teridentifikasi pencilan dengan metode estimasi MM pada analisis *robust* menggunakan fungsi pembobot Huber dengan prosedur berikut:
   a. Menghitung parameter model regresi menggunakan estimasi MKT, sehingga didapatkan awal $\widetilde{\beta_J}^{(1)}$ dan residual $e_i{}^{(1)}$ yang diperlukan sebagai nilai awal
   b. Residual e_i pada langkah pertama digunakan untuk menghitung skala estimasi $\hat{\sigma}$ dan hitung pula pembobot awal $w_{i,0}$
   c. Residual $e_i{}^{(1)}$ dengan skala estimasi $\hat{\sigma}_s$ pada langkah kedua digunakan dalam iterasi awal sebagai penaksiran WLS untuk menghitung koefisien regresi.
   d. Menghitung bobot baru $w_i{}^{(2)}$ dengan skala estimasi dari iterasi awal WLS
   e. Mengulang langkah b,c,d (dengan skala estimasi tetap konstan) sampai mendapatkan $\sum_{i=0}^{n}\left|e_i{}^{(m)}\right|$ konvergen (selisih nilai $\tilde{\beta}^{(m+1)}$ dan $\tilde{\beta}^{(m)}$ mendekati 0, dengan m banyaknya iterasi).
6. Pengujian parameter model secara simultan dan parsial mengunakan persamaan 29 dan 30 pada model yang telah konvergen
7. Interpretasi hasil
8. Menarik kesimpulan.

2 HASIL PENELITIAN

4.1 Model Regresi Linear Berganda

Metode kuadrat terkecil bertujuan meminimumkan jumlah kuadrat galat. Berdasarkan hasil analisis, diperoleh nilai-nilai parameter regresi, sehingga model regresi yang terbentuk adalah sebagai berikut:

$$\hat{Y} = 0{,}0308 + 0{,}0130\,X_1 - 0{,}0006\,X_2 - 0{,}0849\,X_3 + 0{,}00028\,X_4 + 0{,}000901\,X_5$$

4.2 Pengujian Asumsi Regresi Linear

1. Uji Multikolinearitas

Berdasarkan hasil analisis, diperoleh nilai VIF secara berturut-turut yaitu 85,600; 106,917; 22,777; 1,159 dan 75,595. Hal ini menunjukkan bahwa peubah X_1, X_2, X_3 dan X_5 memliki nilai VIF lebih besar dari 10, artinya data mengalami masalah multikolinearitas.

Sedangkan pada peubah X_4 diperoleh nilai VIF lebih kecil dari 10, hal ini menunjukkan data tidak mengalami masalah multikolinearitas.

2. Uji Kebebasan Galat

pengujian kebebasan galat menggunakan statistik *d* dari *Durbin-Wiston*.

Hipotesis yang diuji adalah sebagai berikut:

$H_0 : \rho = 0$ (tidak ada korelasi antara galat)

$H_1 : \rho \neq 0$ (ada korelasi antara galat)

Berdasarkan tabel batas-batas statistik *d* dari *Durbin-Wiston*, untuk $n = 40$ dan lima peubah bebas ($k = 5$), dengan nilai kritis pada $\alpha = 0{,}05$ diketahui nilai $d_L = 1{,}2305$ dan $d_U = 1{,}7859$ sedangkan $4 - d_L = 2{,}7695$. Nilai *d* yang dihasilakan adalah 1,704 berada dalam batasan $d_L \leq d \leq d_U$, sehingga tidak dapat disimpulkan apakah nilai-nilai saling berkorelasi atau tidak.

3. Uji Homoskedastisitas

Pengujian homoskedastisitas menggunakan koefisien korelasi peringkat spearman seperti pada persamaan 9. Hipotesisi yang diuji adalah sebgai berkut:

$H_0 : \sigma^2 =$ tidak terjadi heteroskedastisitas

$H_1 : \sigma^2 \neq$ terjadi heteroskedastisitas

Model regresi terdiri dari lima peubah bebas, dengan demikian berdasarkan analisis diperoleh lima koefisien korelasi peringkat spearman. Pengujian tingkat penting koefisien korelasi ini menggunkan perbandingan nilai *P* dengan tingkat signifikansi $\alpha = 0{,}05$.

Tabel 1. Koefisien korelasi peringkat spearman (r_s) antara peubah bebas dan galat berdasarkan keluaran *SPSS* seperti pada lampiran (5).

			X_1	X_2	X_3	X_4	X_5
Spearman's rho	Galat	Koefisien korelsi	0,092	0,175	0,286	0,204	0,108
		Nilai *P*	0,572	0,335	0,074	0,206	0,508
		N	40	40	40	40	40

Berdasarkan Tabel 1, diketahui bahwa nilai *P* untuk masing-masing peubah bebas lebih dari tingkat signifikan yang ditetapkan. Hal ini menunjukkan bahwa tidak terjadi heteroskedastisitas.

4. Uji Normalitas

Uji asumsi normalitas menggunakan uji *kolmogorov-smirnov*. Pengujian ini membandikan distribusi dari nilai-nilai galat distribusi normal baku.

Hipotesis yang diuji adalah:

H_0 : galat menyebar normal $[\varepsilon \sim N(0, \sigma^2)]$

H_1 : galat tidak menyebar normal $[\varepsilon \nsim N(0, \sigma^2)]$

Berdasarkan hasil analisis, diperoleh nilai *P* sebesar $0{,}024 < \alpha\ (0{,}05)$, maka hipotesis nol ditolak, sehingga dapat disimpulkan bahwa nilai-nilai galat dari model regresi tidak berdistribusi normal.

Berdasarkan hasil pengujian asumsi di atas, diketahui ada dua uji yang tidak terpenuhi yaitu uji multikolinearitas dan uji normalitas, hal ini disebabkan oleh kemungkinan adanya pencilan pada data.

4.3 Pengujian Parameter Model

1. Pengujian secara Simultan

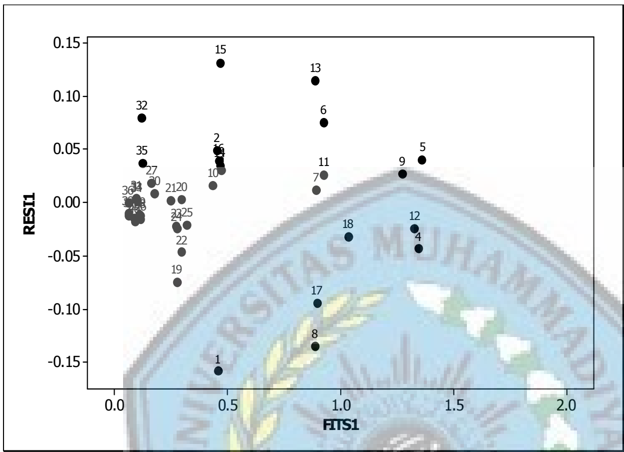
Berdasarkan hasil analisis, diperoleh nilai statistik uji *F* sebesar 374,91 dan nilai *P* sebesar 0,000 dengan DF_1 sebesar 5, DF_2 sebesar 34 dan selang kepercayaan yang digunakan sebesar 0,95 diperoleh nilai kritis *F* sebesar 2,49 seperti pada Lampiran 13. Karena nilai F_{hitung} (374,91) > F_{tabel} (2,49) atau nilai $P\ (0{,}000) \leq \alpha\ (0{,}05)$, maka hipotesis nol ditolak, sehingga dapat disimpulkan bahwa minimal ada satu variable X terhadap pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana atau dapat dikatakan peubah bebas secara simultan signifikan terhadap model yang terbentuk.

2. Pengujian secara Parsial

Berdasarkan hasil analisis diperoleh nilai $|t_{hitung}|$ dari kelima peubah bebas X_1, X_2, X_3, X_4 dan X_5 secra berturut-turut yaitu sebesar (2,98; 0,04; 2,29; 0,25; dan 2,75) dan nilai *P* sebesar 0,002 dengan DF sebesar 34 dan selang kepercayaan yang digunakan 0,95 diperoleh nilai kritis *t* sebesar 1,69. Hal ini berarti hasil uji signifikansi parameter model secara parsial dari kelima peubah bebas yaitu X_1, X_2, X_3, X_4 dan X_5 hanya dua peubah bebas yang tidak berpengaruh terhadap pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana yaitu peubah bebas X_2 dan peubah bebas X_4.

4.4 Mengidentifikasi Pencil

1. *Scatter Plot*



Gambar 1. Plot data antara nilai dugaan dan nilai sisaan

Berdasarkan Gambar 1, dapat dilihat bahwa amatan ke-1, 8, 15 dan 17 merupakan amatan yang dipredikisi sebagai pencilan terhadap nilai sisaan ($\boldsymbol{\varepsilon}$) dan nilai prediksi ($\widehat{\boldsymbol{Y}}$), karena data-data tersebut terpisah jauh dari kumpulan data lainnya.

2. Menentukan Nilai *Leverasi*

Berdasarkan hasil analisis untuk nilai-nilai h_{ii} dengan 5 peubah bebas diperoleh baahwa $h_9 = 0{,}376671$, $h_{17} = 0{,}475196$, $h_{18} = 0{,}638354$, dimana nilai tersebut melebihi kriteria dua kali rataan nilai *leverasi* sebesar 0,30, sehingga data ke-9, 17 dan 18 dianggap sebagai pencilan yang ditinjau dari nilai-nilai X.

3. Menentukan Sisaan Dibuang Ter-*student*-kan

Berdasarkan hasil analisis diperoleh nilai $|d_1^*| = 3{,}13666$, $|d_8^*| = 2{,}62026$, $|d_{13}^*| = 2{,}15586$, $|d_{15}^*| = 2{,}60621$ dan $|d_{17}^*| = 2{,}32111$. Data-data tersebut melebihi $t_{(0,95,33)} = 1{,}69$, sehingga data ke-1, 8, 13, 15 dan ke-17 terdeteksi sebagai pencilan ditinjau dari nilai-nilai Y.

4. Mengidentifikasi Data Berpengaruh

Untuk mengetahui apakah data ke-1, 8, 9, 13, 15, 17 dan ke-18 yang terdeteksi sebagai pencilan juga merupakan data berpengaruh maka dideteksi dengan nilai DFFITS mengunakan persamaan 21. Berdasarkan hasil analisis dipeoleh nilai DFFITS secara berturut-turut yakni (-1,03167; -0,9993; -0,4554; -0,7560; -1,08872; -2,208; dan -1,1660) sehingga dapat disimpulkan bahwa nilai-nilai yang lebih besar dari nilai dua kali akar $p/n = 0{,}7745$ yakni data ke-1, 8, 15, 17 dan 18, jadi pengamatan tersebut teridintifikasi sebgai data berpengaruh.

4.5 Prosedur Regresi *Robust* dengan Estimasi *Method of Moment*

1. Menghitung nilai e_i dengan rumus persamaan 6. Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung pembobot awal $w_i^{(1)}$ dengan menggunkan pembobot *Huber* dengan koefisien *r* yang digunakan adalah 1,345.
2. Menghitung nilai $\hat{\sigma}_s$ berdasarkan nilai e_i nilai masing-masing observasi $\hat{\sigma}_s = \frac{MAR}{0,6745} = \frac{\frac{1}{n}\sum_1^n |Y_i - \hat{Y}_i|}{0,6745}$
3. menentukan nilai $e_i{}^*$ dengan rumus: $e_i{}^* = \frac{e_i}{\hat{\sigma}_s}$
4. Menentukan nilai $w_{i,0}$ dengan menggunkan pembobot *Huber* menggunakan rumus pada persamaan 22. Berikut nilai $\widehat{\beta_J}$ dari setiap iterasi yang diperoleh dari *software minitab*.

Tabel 2. Nilai $\widehat{\beta_J}$ Hasil Iterasi

Iterasi	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
0	(0,03084;	0,013038;	0,00056;	0,08491;	0,000277;	0,0009013)
1	(0,04659;	0,013450;	0,00213;	0,08285;	0,0004821;	0,0009167)
2	(0,04990;	0,013492;	0,00251;	0,08307;	0,0006324;	0,0009272)
3	(0,05085;	0,013462;	0,00269;	0,08307;	0,0006690;	0,0009348)
4	(0,05119;	0,013433;	0,00278;	0,08304;	0,0006788;	0,0009396)
5	(0,05133;	0,013418;	0,00283;	0,00283;	0,0006827;	0,0009419)
6	(0,05140;	0,013406;	0,00285;	0,08300;	0,0006837;	0,0009435)
7	(0,05144;	0,013404;	0,00286;	0,08300;	0,0006850;	0,0009438)
8	(0,05145;	0,013401;	0,00287;	0,08299;	0,0006850;	0,0009443)
9	(0,05139;	0,013415;	0,00284;	0,08301;	0,0006851;	0,0009424)
10	(0,05137;	0,013415;	0,00283;	0,08301;	0,0006843;	0,0009423)
11	(0,05138;	0,013416;	0,00283;	0,08302;	0,0006846;	0,0009423)
12	(0,01538;	0,013416;	0,00283;	0,08302;	0,0006848;	0,0009423)
13	(0,05138;	0,013416;	0,00283;	0,08302;	0,0006848;	0,0009423)

Berdasarkan Tabel 2, di atas terlihat bahwa koefisien regresi konvergen pada iterasi ke-13 dengan model yang terbentuk:

$$\hat{Y}_B = 0,0514 + 0,0134\, X_1 - 0,0028\, X_2 - 0,0830\, X_3 + 0,000685\, X_4 + 0,00094\, X_5$$

dimana $\hat{Y}_B$ adalah nilai duga terboboti yang telah konvergen.

4.6 Pengujian Parameter Model setelah Konvergen

1. Pengujian secara Simultan

Berdasarkan Lampiran 11, diperoleh nilai statistik uji *F* sebesar 530,28 dan nilai *P* sebesar 0,000 dengan DF_1 sebesar 5, DF_2 sebesar 34 dan selang kepercayaan yang digunakan sebesar 0,95 diperoleh nilai kritis *F* sebesar 2,49 seperti pada Lampiran 13. Karena nilai F_{hitung} (530,28) > F_{tabel} (2,49) atau nilai *P* (0,000) $\leq \alpha$ (0,05), maka hipotesis nol ditolak, sehingga dapat disimpulkan bahwa minimal ada satu variable X yang berpengaruh terhadap variable pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana.

2. Pengujian secara Parsial

Berdasarkan hasil analisis diperoleh nilai $|t_{hitung}|$ dari kelima peubah bebas X_1, X_2, X_3, X_4 dan X_5 secra berturut-turut yaitu sebesar (3,66; 0,26; 2,67; 0,72; dan 3,42) dan nilai *P* sebesar 0,002 dengan DF sebesar 34 dan selang kepercayaan yang digunakan 0,95 diperoleh nilai kritis *t* sebesar 1,69. Hal ini berarti hasil uji signifikansi parameter model secara parsial dari kelima peubah bebas yaitu X_1, X_2, X_3, X_4 dan X_5 hanya dua peubah bebas yang tidak berpengaruh terhadap pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana yaitu peubah bebas X_2 dan peubah bebas X_4.

 ISBN : 2685-5852

4.7 Perbandingan Hasil MKT dengan *Metode of Moment* menggunakan pembobot *Tukey's Bisquare* dan *Huber*

Tabel 3. Hasil Perbandingan MKT dengan *Metode of Moment* menggunakan pembobot *Tukey's Bisquare* dan *Huber*

Metode	Parameter yang Signifikan	R^2	KTG	Konvergen Pada Itersi
MKT dengan melibatkan pencilan	X_1, X_3 dan X_5	0,982	0,0035	-
MKT tanpa pencilan dan tidak berpengaruh	X_1, X_3 dan X_5	0,982	0,0032	-
HUBER	X_1, X_3 dan X_5	0,987	0,0025	13
TUKEY	X_1, X_2, X_3 dan X_5	0,995	0,0010	12

5. SIMPULAN

Berdasarkan hasil analisis Regresi *robust* dengan estimasi *Method of Moment* menggunakan pembobot *Huber* dalam pemodelan data teridentifikasi pencilan diperoleh estimator yang kekar dengan model sebagai berikut:

$$\hat{Y} = 0{,}05138 + 0{,}013416\,X_1 - 0{,}00283\,X_2 - 0{,}08302\,X_3 + 0{,}000685\,X_4 + \ 0{,}0009423\,X_5$$

Dari hasil analisis Regresi *robust* dengan estimasi *Method of Moment* menggunakan pembobot *Tukey's Bisquare* dalam pemodelan data teridentifikasi pencilan diperoleh model sebagai berikut:

$$\hat{Y} = 0{,}07521 + 0{,}014299\,X_1 - 0{,}005715\,X_2 - 0{,}08312\,X_3 + 0{,}00185\,X_4 \ + \ 0{,}0009698\,X_5$$

Model lebih kekar (*robust*) dibandingkan dengan menggunakan pembobot *Huber* dalam pemodelan data teridentifikasi pencilan. Hal ini ditunjukan dengan:

a. Hasil iterasi pada pembobot *Tukey's Bisquare* lebih efisien dibandingkan dengan pembobot *Huber*.
b. Hasil analisis pada pembobot *Tukey's Bisquare* mengenai faktor-faktor yang berpengaruh terhadap pendapatan bersih nelayan di Kecamatan Rumbia Kabupaten Bombana, disimpulkan bahwa pendapatan bersih nelayan dipengaruhi oleh empat peubah bebas yaitu jenis alat tangkap (X_1), jumlah armada (X_2), armada tenaga kerja (X_3) dan biaya yang dikeluarkan (X_5), sedangkan hasil analisis pada pembobot *Huber* dipengaruhi oleh tiga peubah bebas yaitu jenis alat tangkap (X_1), armada tenaga kerja (X_3) dan biaya yang dikeluarkan (X_5),
c. Nilai koefisien determinansi (R^2) pada pembobot *Tukey's Bisquare* sebesar 0,995 lebih besar dibandingkan dengan nilai koefisien determinansi (R^2) pada pembobot *Huber* yaitu sebesar 0,987
d. Nilai Kuadrat Tengah Galat (KTG) pada pembobot *Tukey's Bisquare* sebesar 0,0010 lebih kecil dibandingkan dengan nilai KTG pada pembobot *Huber* yaitu sebesar 0,0025.

6. REFERENSI

Agresti, A 1996. *Introduction to Categorical Data*. New York: Jhon Willey & Sons.

Chatterjee, S. & Ali, S.H. 2006. *Regression Analysis by Example – 4th ed*. New jersey: Jhon Willey & Sons, Inc., Publication.

Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUST REG Procedure*. North Carolina: SAS Institute. www.sas.com.

Cook, R.D. & Weisberg, S.1982. *Residual and Influence In Regression*. New York: Jhuon wiley.

Gujarati, D.1978. Ekonometrika Dasar. Jakarta: Penerbit Erlangga.

Kurniawan, D. 2008. *Regresi Linear*. http://neddeni.wordpress.com.

Neter, J., Wasserman, W., & Kutner, M.H. 1997. *Model Linear Terapan.* Buku II. Sumantri, B., penerjemah. Bogor: IPB.

Rowlings, O., John, P., & Dickey, A.D. 1998. *Applied Regresi Analysis*. *A research tool*. *Second edition*. *Library of congresr cataloging-inpublication* data.

Sembiring, R.K. 2003. *Analisis Regresi*. Edisi kedua. Bandung: Institut Teknologi Bandung.

Soemartini. 2007. *Pencilan (outlier)* . Jatinangor: Penerbit Universitas Padjajaran.

Steel, R.G.S & Torrie, J.H. 1993. *Prinsip Dasar Prosedur Statistika Suatu Pendekatan Biometrika*. Jakarta: Gramedia Ustaka Utama.

Syawaluddin. 2013. *Estimasi Method of Moment pada Pemodelan data Teridentifikasi pencilan*. Kendari: Universitas Halu Oleo.

Walpole, R.E. & Myres, R.H. 1995. *Ilmu Peluang dan Statistikan untuk Insinyur dan Ilmuan*. RK Sembiring, penerjemah. Bandung: ITB.

Wilcox, R.R. 2005. *Introduction to Robust Estimation and Hypothesis*. San Diego: Academic Press.

Yohai, V.J. 1987. *High Breakdown Point and High Efficiency Robust Estimates For Regression Annalys of Statistic.* Vol. 15, No.20, 642-656.