II. TINJAUAN PUSTAKA

2.1 Pencilan

2.1.1 Definisi Pencilan

Menurut Ferguson (1961), pencilan didefinisikan sebagai suatu data yang menyimpang dari sekumpulan data yang lain. Menurut Barnett (1981), pencilan adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Menurut R.K Sembiring (1950) Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefesien regresi (Soemartini, 2007).

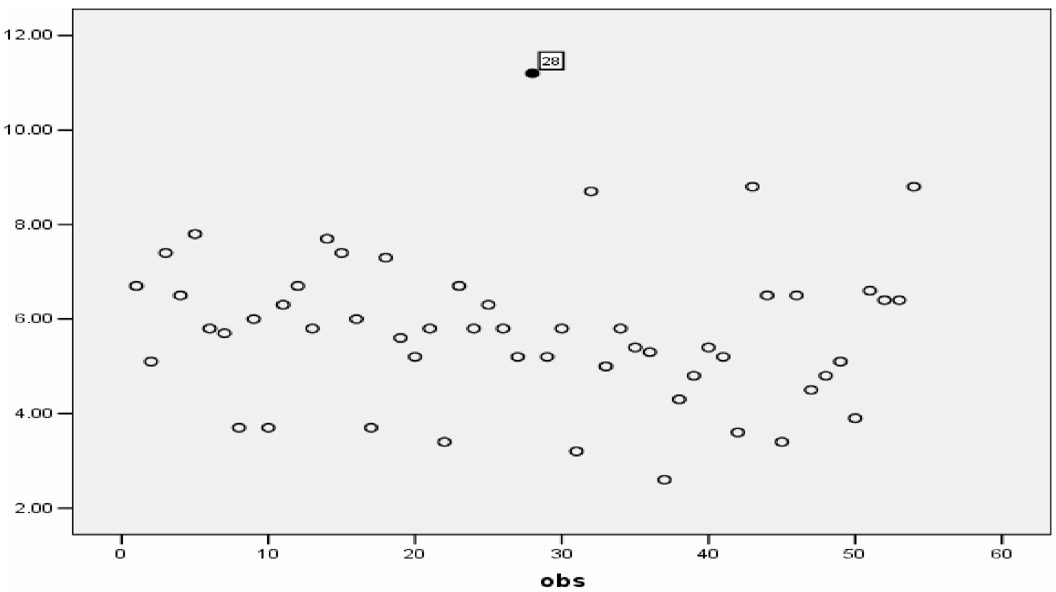
Menurut Hair, dkk. (1995), Pencilan adalah data yang muncul yang memiliki karakteristik unik yang terlihat sangat jauh berbeda dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi
(Cogito Ergo Sum, 2010).

2.1.2 Pendeteksian Pencilan

Terdapat banyak cara untuk mengidentifikasi adanya pencilan atau tidak pada sekumpulan data. Di sini, akan dijelaskan dua cara untuk mengidentifikasi pencilan, diantaranya adalah diagram pencar, boxplot.

1. Diagram Pencar

Untuk melihat apakah terdapat pencilan atau tidak pada sekumpulan data dapat dilakukan dengan memplot data dengan observasi ke-i (i=1, 2, 3, ..., n)seperti pada gambar dibawah ini :



Dari contoh di atas, dapat dilihat bahwa terdapat salah satu data yaitu observasi ke-28 yang mengindikasikan pencilan.

Kelemahan data dari metode ini adalah keputusan bahwa data adalah suatu pencilan sangat tergantung pada *judgement* peneliti. Oleh karena itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut.

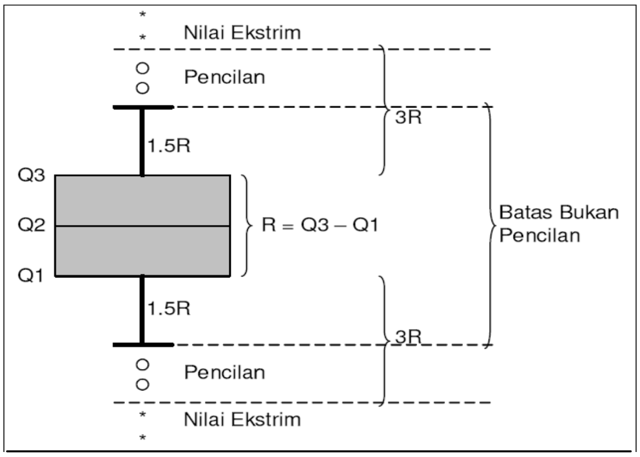
2. Boxplot

Pencilan dapat dideteksi dengan menggunakan boxplot. Metode ini sangat terkenal dalam mendeteksi pencilan. Metode ini menggunakan nilai quartil. Quartil 1,2,dan 3 akan membagi sebuah urutan data menjadi empat bagian.

Jangkauan (*Interquartile (IQR))* didefinisikan sebagai selisih antara quartil 1 dan quartil 3, atau IQR = Q3-Q1.

Menurut Soemartini (2007), Data-data pencilan dapat ditentukan, yaitu nilai yang kurang dari 1.5*IQR terhadap quartil 1 dan nilai yang lebih dari 1.5*IQR terhadap quartil 3.

Gambar 2. Skema identifikasi pencilan menggunakan IQR atau *boxplot*



2.1.3 Pengaruh Pencilan *(Outlier)*

Pencilan *(outlier)* berpengaruh terhadap proses analisa data, salah satunya terhadap nilai mean dan standar deviasi. Oleh karena itu, keberadaan pencilan *(outlier)* dalam suatu pola data harus dihindari. Pencilan *(outlier)* dapat menyebabkan hal-hal berikut :

1. Variance data menjadi besar
2. Interval data dan range menjadi lebar
3. Mean tidak dapat menunjukkan nilai yang sebenarnya (bias), dan

4. Pada beberapa analisa data, outlier dapat menyebabkan kesalahan dalam pengambilan keputusan dan kesimpulan.

2.2 Analisis Regresi

Analisis Regresi adalah salah satu metode statistika yang dapat dipergunakan untuk menyelidiki atau membangun model hubungan antara beberapa variabel (Usman, 2001).

Model regresi linier, biasa dituliskan sebagai berikut:

$$\boldsymbol{y = x\beta + e} \qquad \text{dengan } \boldsymbol{\varepsilon \sim N(0, \sigma^2 I)}$$

Dimana: $\boldsymbol{y}$ = vektor n x 1 variabel tak bebas

$\mathbf{x}$ = matriks n x k variabel bebas

$\boldsymbol{\beta}$ = vektor k x 1 koefisien variabel bebas

$\boldsymbol{\varepsilon}$ = vektor n x 1 variabel acak galat dengan E(ε) = 0 dan matriks ragam peragam $\sigma^2(\varepsilon) = \sigma^2 I$

Menurut Myers (1990), asumsi-asumsi pada analisis regresi adalah sebagai berikut :

1. Galat menyebar normal.

   $\varepsilon_I \sim N(0,\sigma^2)$

2. Ragam galat homogen.

   $Var(\varepsilon_i) = \sigma^2 ; i = 1,2,\ldots\ldots,n$

3. Nilai ε_i adalah bebas satu dengan yang lainnya.

$E(\varepsilon_i) = 0$ dan $E(\varepsilon_i^2) = \sigma^2$

4. X dan Y terkait secara linier. Untuk setiap nilai X dihubungkan maka akan membentuk garis lurus.

Dalam analisis regresi, terdapat dua model regresi, yaitu :

1. Model Regresi Linier Sederhana

   $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

2. Model Regresi Linier Berganda

   $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots\ldots. + \beta_k X_{ki} + \varepsilon_i$

Dimana :

$\beta_0, \beta_1, \ldots\ldots\ldots\ldots, \beta_k =$ Koefisien Regresi

X_i = Variabel bebas (*Regressor*)

Y_i = Variabel tak bebas (*Regressand*)

ε_i = Galat atau *Error*

(Myers, 1990).

2.3 Metode Kuadrat Terkecil (MKT)

Metode Kuadrat Terkecil (MKT) atau sering juga disebut dengan metode OLS (*Ordinari Least Square*) diperkenalkan oleh Carl Friedrick Gauss seorang matematikawan Jerman. Metode Kuadrat Terkecil (MKT) merupakan salah satu metode penduga parameter (b_o,b_1) yang terbaik karena bersifat tak bias dan konsisten. Metode kuadrat terkecil akan menghasilkan ragam(varian) minimum bagi parameter regresi Prinsip dasar metode kuadrat terkecil adalah dengan meminimumkan jumlah kuadrat galat (JKG minimum), sehingga menghasilkan penduga yang mempunyai kesalahan terkecil.

Dengan menggunakan Persamaan linier untuk pendugaan garis regresi linier, metode kuadrat terkecil dapat diuraikan dengan notasi matematika yaitu sebagai berikut:

$$\hat{y}_i = b_o + b_1 x_i$$

Jarak vertikal antara titik observasi (x_i, y_i) dan titik $(x_i, \hat{y}_i)$ pada garis dugaan dapat ditulis :

$$|y_i - \hat{y}_i| \; atau \; |y_i - \hat{b}_0 - \hat{b}_i x_i|$$

Jumlah kuadrat dari semua jarak ini ditulis:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_o - b_1 x_i)^2$$

Solusi dari metode kuadrat terkecil dapat dilakukan sebagai berikut:

$$S(b_{o,} b_1) = \sum_{i=1}^{n} (y_i - b_o - b_1 x_i)^2$$

$$\frac{dS_{(b_0, b_1)}}{db_o} = -2\sum_{i=1}^{n} (y_i - b_o - b_1 x_i) = 0$$

$$\frac{dS_{(b_0,b_1)}}{db_1} = -2\sum_{i=1}^{n}(y_i - b_o - b_1 x_i)x_i = 0$$

Dengan menyederhanakan kedua persamaan ini maka diperoleh:

$$nb_o + b_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$b_0\sum_{i=1}^{n} x_i + b_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$ ...... (Persamaan normal kuadrat terkecil)

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)/n}{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}$$ dan $$b_o = \overline{Y} - b_1\overline{X}$$

Persamaan garis kuadrat terkecil yang didapat adalah:

$\hat{y} = b_o + b_1 X$ atau $\hat{y} = \overline{Y} + b_1(X - \overline{X})$

Persamaan garis diatas dapat digunakan untuk memprediksi Y oleh nilai X yang berpadanan.

Selama asumsi-asumsi regresi dipenuhi oleh data, maka dugaan metode kuadrat terkecil bersifat tak bias dengan varians minimum. Karena memenuhi kedua sifat ini maka MKT dikenal sebagai penduga yang BLUE (*Best Linear Unbiased Estimator*).

2.4 *Mean Square Error* (MSE)

Jika $\hat{\beta}$ penduga tak bias dari β, maka $E((\hat{\beta})-\beta)^2$ sama dengan ragam penduga $\hat{\beta}$. Tetapi, jika suatu $\hat{\beta}$ penduga yang bias dari β, maka $E((\hat{\beta})-\beta)^2$ disebut *Mean Square Error (MSE)* atau kuadrat tengah galat dari penduga $\hat{\beta}$.

MSE ($\hat{\beta}$) = $E((\hat{\beta})-\beta)^2$

Bukti :

MSE ($\hat{\beta}$) = $E(\hat{\beta}^2-2\hat{\beta}\beta+\beta^2)$

$= E(\hat{\beta}^2)-2E(\hat{\beta})\beta+\beta^2)$

$= \{E(\hat{\beta}^2)-2E(\hat{\beta})\beta+\beta^2\}+\{(E(\hat{\beta}))^2-(E(\hat{\beta}))^2\}$

$= \{E(\hat{\beta}^2)-(E(\hat{\beta}))^2\} + \{(E(\hat{\beta}))^2-2E(\hat{\beta})\beta+\beta^2\}$

$= \{E(\hat{\beta}^2)-(E(\hat{\beta}))^2\} + \{(E(\hat{\beta})-\beta)^2\}$

= Var ($\hat{\beta}$) + (Bias ($\hat{\beta}$))2

2.5 *Robust*

Regresi *robust* diperkenalkan oleh Andrews (1972) dan merupakan metode regresi yang digunakan ketika distribusi dari *error* tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model (Ryan, 1997). Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh pencilan , sehingga dihasilkan model yang *robust* atau *resistance* terhadap pencilan.

Prosedur *robust* ditunjukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data pencilan, dan juga bersifat otomatis dalam menanggulangi data pencilan. Beberapa metode penduga dalam regresi *robust* diantaranya Penduga M, *Least Trimmed Square* (LTS), Penduga MM, Penduga S, dan *Least Mean Square* (LMS).

2.6 Penduga-MM

Penduga MM (*MM-estimator*) diperkenalkan oleh Yohai (1987), yaitu sebuah metode yang secara simultan mempunyai dua sifat, yaitu penduga yang bersifat *breakdown point* tinggi dan efisiensi tinggi, atau dengan kata lain Penduga MM (*MM-estimator*) bertujuan menghasilkan sebuah penduga yang *breakdown point* tinggi serta mempertahankan efisiensi baik, dimana *breakdown point* dan efisiensi merupakan sifat terpenting dalam penduga *robust.*

Breakdown point adalah jumlah maksimum data terkontaminasi (pencilan) yang dapat ditoleransi oleh suatu metode. Adapun yang termasuk *breakdown point* tinggi diantaranya, *Least Median Square* (LMS), *Least Trimmed Square* (LTS), Penduga S, dan Penduga MM. Dikatakan *breakdown point* terkecil jika nilai *breakdown point*nya 1/n. Adapun yang termasuk *breakdown point* terkecil diantaranya, Metode Kuadrat Terkecil (MKT), dan penduga M (Montgomeri *et.al.,*1992).

Efisiensi sampel terbatas dari penduga *robust* didefinisikan sebagai perbandingan nilai Kuadrat Tengah Galat (KTG) atau *Mean Square Error* (MSE) antara Metode Kuadrat Terkecil (MKT) dengan penduga *robust*

(Montgomeri *et.al.*,1992).

Yohai (1987) memperkenalkan penduga MM dalam tiga tahap, diantaranya :

1. Menghitung nilai penduga awal. Penduga awal yang digunakan adalah penduga LTS (*Least Median Square*).

   Perhitungan penduga LTS, yaitu meminimumkan jumlah kuadrat galat terhadap subhimpunan data berukuran h yang dapat dirumuskan sebagai berikut.:

$$\hat{\beta}_{LTS} = \min_{\beta} \sum_{i=1}^{h} e_i^2$$

$$= \min_{\beta} \sum_{i=1}^{h} (y_i - \hat{y}_i)^2$$

$$= \min_{\beta} \sum_{i=1}^{h} (y_i - x\beta)^2$$

, dengan h memenuhi $\frac{(3n+p+1)}{4} \leq h \leq n$

   Solusi $\hat{\beta}$ pada persamaan di atas dapat diperoleh dengan menggunakan turunan atau differensial seperti pada penyelesaian penduga MKT. Hanya pada LTS persamaan tersebut dihitung pada subhimpunan data terbaik yang berukuran h.

2. Menghitung parameter skala $(\hat{\sigma})$ dari penduga M, menggunakan galat berdasarkan penduga awal.

   Persamaan yang digunakan untuk solusi penyelesaiannya, yaitu sebagai berikut:

$$\hat{\sigma}_{(m+1)} = \text{median}\left\{\frac{|e|}{0.6745}\right\}$$

$$= \text{median}\left\{\frac{|y_i - x\hat{\beta}_{LTS}{}^{m}|}{0.6745}\right\} \quad ; i = 1, 2, ..., n \; ; m = \text{iterasi } 0, 1,2,..,m$$

3. Menghitung penduga akhir berdasarkan residual penduga awal dengan menggunakan rumus penduga M.

   Prinsip dasar penduga-M adalah meminimumkan fungsi objektif:

$$\sum_{i=1}^{n}\rho\left(\frac{e_i}{\hat{\sigma}}\right)$$

$$e_{(i)}^2 = (y_i - \hat{y}_i)^2 = (y_i - x_i\beta)^2$$

   Jika ψ turunan dari ρ, maka fungsi objektif diatas akan menjadi bentuk persamaan :

$$\sum_{i=1}^{n}\psi\left(\frac{e_i}{\hat{\sigma}}\right)x_i = 0$$

   dalam perhitungan fungsi psi (ψ) juga digunakan fungsi pembobot (*weight*) yang dinotasikan :

$$w_i = \frac{\psi(e_i)}{(e_i)}$$

Dengan fungsi pembobot w_i , maka persamaannya menjadi:

$$\sum_{i=1}^{n} w_i \left(\frac{e_i}{\hat{\sigma}} \right) x_i = 0$$

Jika dibuat dalam bentuk matriks maka Persamaan penduga M adalah sebagai berikut.

$$\hat{\beta}_{(m+1)} = (x' w_{(m)} x)^{-1} x' w_{(m)})y$$

; dengan $w_m = \begin{cases} \left(1 - \left(\frac{e_i / \hat{\sigma}}{c} \right)^2 \right)^2 & ; |e_i / \hat{\sigma}| < c \\ o & ;\ lainnya \end{cases}$

Setelah menyelesaikan fungsi diatas, maka akan didapat parameter penduga M, dan akan menjadi solusi nilai penduga MM.