# BAB III
# METODA *LEAST SQUARE*

Metoda *least square* merupakan suatu teknik penyelesaian permasalahan yang penting dan dimanfaatkan dalam banyak bidang aplikasi. Metoda ini banyak digunakan untuk mencari / mengetahui adanya hubungan atau korelasi di antara dua variabel. Kondisi demikian akan sangat berguna jika dapat diketahui dan dipahami dengan baik. Sebagai contoh, metode *least square* dapat digunakan untuk mengetahui apakah ada kerelasian antara pengeluaran sebagai upaya promosi dengan hasil penjualan. Contoh lain tentu masih banyak, mulai skala nilai yang relatif kecil hingga pada tingkatan nilai yang sangat besar.

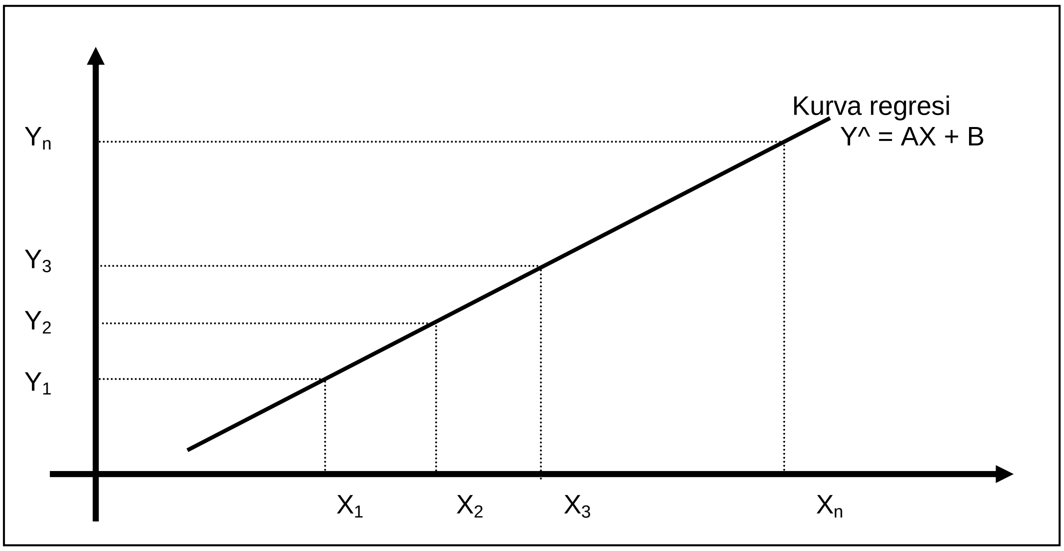
Metoda *least square* diperlukan untuk melakukan analisa apakah terdapat hubungan di antara dua variabel yang ditinjau, seberapa kuat hubungan yang terjadi, dan bahkan dapat dipergunakan untuk peramalan kondisi mendatang dengan mamanfaatkan persamaan *regresi* atau *kurva regresi* yang dihasilkan.

Secara umum terdapat sejumlah variasi bentuk kurva regresi yang dihasilkan oleh persamaan regresi di antara variabel-variabel. Salah satu bentuk kurva regresi yang paling sederhana adalah berupa garis lurus yang dapat diselesaikan dengan metoda regresi linier sederhana. Pembahasan pada bagian ini akan meninjau dua metoda biasa diterapkan untuk anallisa dalam berbagai bidang, yakni *regersi linier sederhana* dan *regresi polinomial* (regresi suku banyak). Kedua teknik tersebut sama-sama memanfaatkan metoda *least square* dalam proses penyelesaiannya. Kedua metoda tersebut sering dijumpai dan dimanfaatkan untuk melakukan analisa hubungan di antara variabel-variabel terhadap penyelesaian permasalahan dalam kehidupan sehari-hari.

## 3.1. Regresi Linier Sederhana

Metoda regresi linier sederhana digunakan untuk mencari nilai-nilai koefisien A dan B dalam persamaan pendekatan kurva penduga **A^ = A +BX**. Kurva tersebut dimanfaatkan untuk menunjukkan hubungan atau korelasi antara sejumlah pasangan data **X** dan **Y** yang mempunyai kecenderungan hubungan linier. Fungsi *least square* dalam metoda ini adalah meminimalkan jumlah kuadrat kesalahan antara titik-titik koordinat data yang sebenarnya dan titik koordinat yang dihasilkan oleh persamaan regresi. Metoda yang dilakukan adalah dengan meminimalkan jumlah kuadrat selisih jarak tegak siku-siku antara titik koordinat plot data asli dan titik koordinat pada kurva regresi untuk setiap pasangan data X dan Y.

Dengan cara tersebut maka diharapkan bahwa pendekatan garis regresi yang diperoleh dapat mendekati sempurna, yaitu bahwa titik-titik koordinat hasil perhitungan hampir sama dengan titik-titik koordinat pada pasangan data aslinya. **Gambar 3.1** menunjukkan pendekatan metoda *least square* pada regresi linier sederhana.




**Gambar 3.1** : Fungsi *least square* pada regresi linier sederhana

Proses penyelesaian permasalahan dengan menggunakan metoda regresi linier sederhana dimulai dengan membaca cacah pasangan data yang ada, di sini dinyatakan dengan variabel N. Dengan suatu proses berulang sebanyak N kali, masing-masing pasangan data dibaca sebagai X[I] dan Y[I]. Indeks I digunakan untuk mengidentifikasi setiap pasangan data X dan Y. Proses kemudian dilanjutkan untuk menghitung harga-harga berikut :

$\sum X[I]$ : jumlah semua data X

$\sum Y[I]$ : jumlah semua data Y

$\sum X[I] \times Y[I]$ : jumlah semua data X x Y

$\sum (X[I])^2$ : jumlah semua data X dikuadratkan

Pada akhirnya nilai-nilai koefisien persamaan regresi A dan B dapat dihitung dengan formula sebagai berikut :

$$B = \frac{\sum (X\,I] \times Y[I]) - \{\sum Y[I]/N\}}{\sum X[I]^2 - \{(\sum X[I])^2 / 2\}}$$

$$A = \overline{Y} - B\,\overline{X}$$

Keterangan,

$\overline{Y} =$ harga rata-rata Y, dihitung denganformula $\sum Y[I]/N$

$\overline{X} =$ harga rata-rata X, dihitung denganformula $\sum X[I]/N$

Jika diselesaikan dengan program komputer dengan memanfaatkan operasi perulangan, maka perhitungan tersebut dapat dikerjakan dengan mudah, cepat dan relatif sederhana.

*Flowchart* prosedur untuk mencari koefisien-koefisien persamaan dengan metoda regresi linier sederhana ditunjukkan pada **Gambar 3.2**, dimana didalamnya digunakan beberapa variabel yaitu :

N : menyatakan cacah pasangan X dan Y

I : menyatakan variabel pencacah (*counter*)

X[I] : menyatakan harga data X ke-I

Y[I] : menyatakan harga data Y ke-I

JUMX : menyatakan jumlah semua harga data X

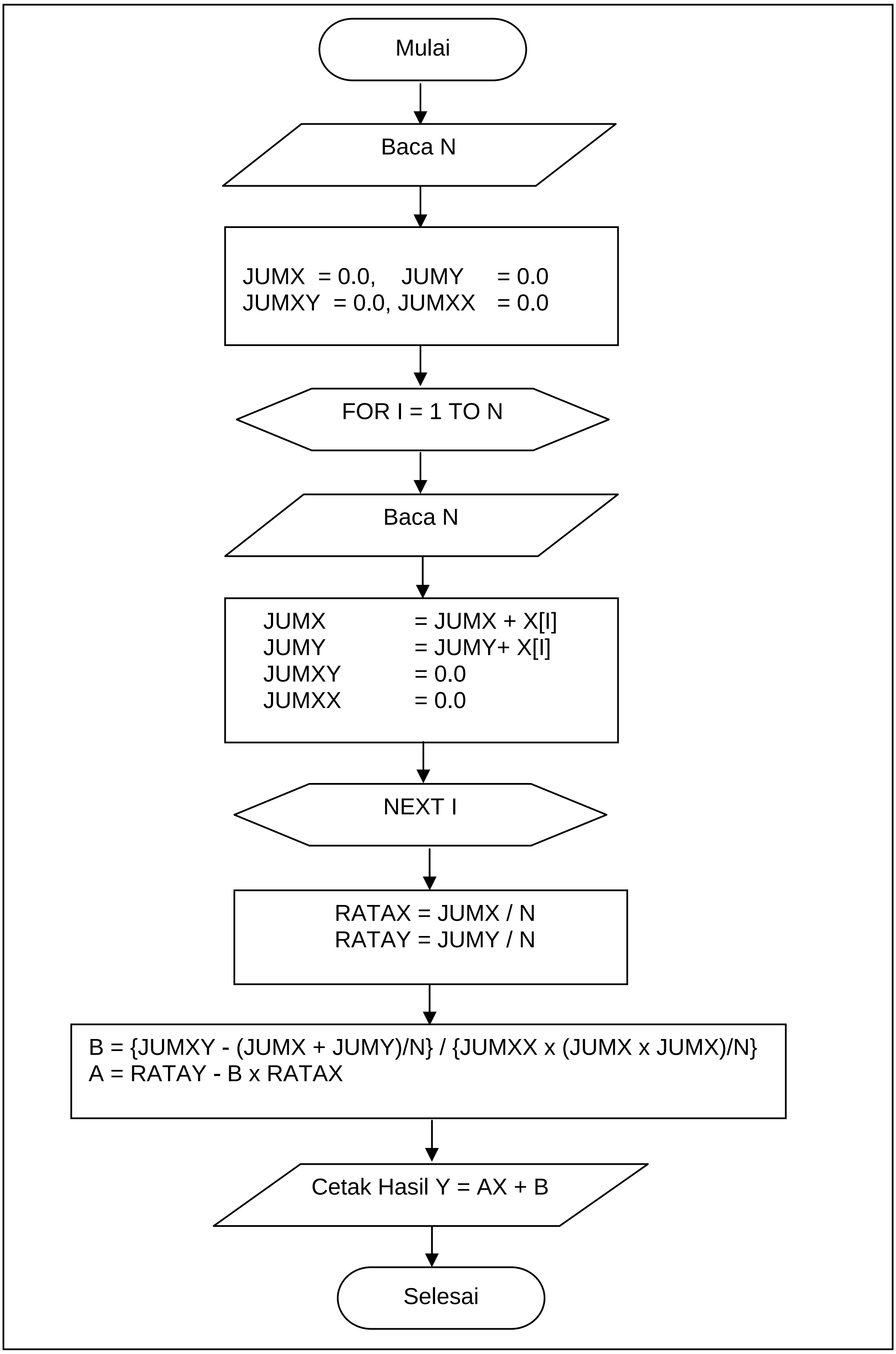
JUMY : menyatakan jumlah semua harga data Y

JUMXY : menyatakan jumlah semua harga perkalian data X dan Y

JUMXX : menyatakan jumlah semua harga data X dikuadratkan

RATAX : menyatakan harga rata-rata data X

RATAY : menyatakan harga rata-rata data Y




**Gambar 3.2** : *Flowchart* prosedur mencari koefisien persamaan regresi dengan metoda regresi linier

Selanjutnya solusi dalam bentuk algoritma untuk prosedur pencarian harga-harga koefisien persamaan regresi dengan metoda regresi linier sederhana dapat dituliskan sebagai berikut :

1. Mulai
2. Tentukan harga-harga awal

   JUMX = 0.0

   JUMY = 0.0

   JUMXY = 0.0

   JUMXX = 0.0
3. Proses berulang untuk menghitung jumlah semua data X,Y, XxY dan $X^2$

   FOR I = 1 TO N

      JUMX = JUMX + X[I]

      JUMY = JUMY + Y[I]

      JUMXY = JUMXY + X[I] x Y[I]

      JUMXX = JUMXX + X[I] x X[I]
4. Hitung rata-rata harga data X dan Y

   RATAX = JUMX/N

   RATAY = JUMY/N
5. Hitung koefisien-koefisien persamaan regresi

   B = {JUMXY - (JUMX x JUMY) / N} / {JUMXX - (JUMX x JUMX) /N}

   A = RATAY - B x RATAX
6. Cetak hasil

   Y = A + BX
7. Selesai

### 3.2. Regresi Polinomial

Regresi polinomial merupakan suatu metoda yang digunakan untuk mencari nilai-nilai koefisien $B_0$, $B_1$, $B_2$, $B_3$, $B_4$ ,…. $B_M$, pada persamaan pendekatan kurva regresi dalam regresi polinomial. Sebagaimana dalam metoda regresi linier sederhana, kurva tersebut digunakan untuk menggambarkan hubungan / korelasi antara sejumlah pasangan data X dan Y.

Sebagai ilustrasi, jika N menyatakan cacah pasangan data yang akan dihitung harga koefisien-koefisien regresinya, yaitu sebagai berikut :

| Data | I | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | … | N |
| $X_1$ | $X_1$ | $X_2$ | $X_3$ | $X_5$ | … | $X_N$ |
| $Y_1$ | Y1 | $Y_3$ | $Y_4$ | $Y_5$ | … | $Y_N$ |

Persamaan regresi polinomial dinotasikan sebagai berikut ini :

$Y\hat{} = B_0X^0 + B_1X^1 + B_2X^2 + B_3X^3 + \ldots + B_MX^M$

Keterangan:

M : menunjukkan orde persamaan regresi polinomial pada kurva regresi

$X^0$ : 1

Sebagaimana dalam metoda regresi linier sederhana, penggunaan metoda *least square* di sini juga dimaksudkan untuk mencari pendekatan kurva garis penduga yang paling sempurna, yaitu dengan meminimalkan jumlah kuadrat selisih jarak tegak siku-siku antara titik koordinat plot data asli dan titik koordinat pada kurva regresi untuk setiap pasangan data X dan Y. Untuk menyelesaikan perhitungan dalam metoda regresi polinomial, diperlukan prosedur untuk perhitungan perkalian dan invers matrik. Perhitungan perkalian matrik dan invers matrik ini akan dibahas secara khusus pada **BAB IX**.

Secara garis besar prosedur penyelesaian untuk mencari koefisien-koefisien regresi dalam regresi polinomial dapat dituliskan dalam bentuk algoritma berikut :

Masukkan cacah pasangan data (=N), dan setiap pasangan data X[I] dan Y[I].

1. Mulai
2. Tentukan orde persamaan regresi polinomial (=M)
3. Tentukan persamaan-persamaan regresinya

   $Y_0 = B_0X_0^0 + B_1X_0^1 + B_2X_0^2 + B_3X_0^3 + \ldots + B_MX_0^M$

   $Y_1 = B_0X_1^0 + B_1X_1^1 + B_2X_1^2 + B_3X_1^3 + \ldots + B_MX_1^M$

$Y_2 = B_0X_2^0 + B_1X_2^1 + B_2X_2^2 + B_3X_2^3 + \ldots + B_MX_2^M$

$\vdots$

$Y_N = B_0X_N^0 + B_1X_N^1 + B_2X_N^2 + B_3X_N^3 + \ldots + B_MX_N^M$

4. Ubahlah persamaan-persamaan tersebut ke dalam notasi perkalian matrik sebagai berikut ini :

$$\begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \\ \bullet \\ \bullet \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_0^1 & X_0^2 & X_0^M \\ 1 & X_1^1 & X_1^2 & X_1^M \\ 1 & X_2^1 & X_2^2 & X_2^M \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ 1 & X_N^1 & X_N^2 & X_N^M \end{pmatrix} \bullet \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ \bullet \\ \bullet \\ B_n \end{pmatrix}$$

5. Perkalian matrik pada langkah-4 dapat dituliskan dalam notasi sebagai berikut :

   $Y = X \,.\, B$

   Atau :

   $X \,.\, B = Y$

   Untuk menyelesaikan persamaan tersebut, masing-masing ruas dikalikan dengan $X^{-1}$ (kebalikan atau invers matrik X), sehingga diperoleh hasil sebagai berikut :

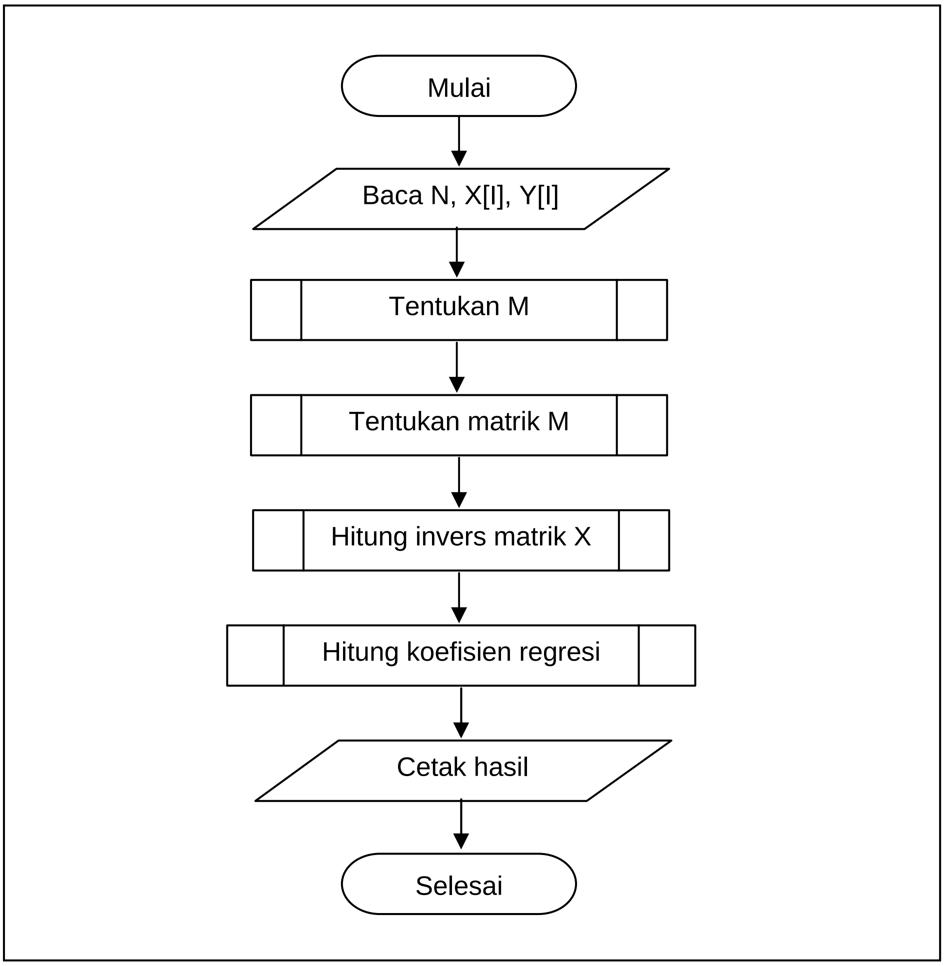
   $(X^{-1} \,.\, X) \,.\, B = X^{-1} \,.\, Y$

   $B = X^{-1} \,.\, Y,$
6. Hitung invers matrik X
7. Setelah $X^{-1}$ ditemukan, hasilnya dikalikan dengan matrik Y

   $B = X^{-1} \,.\, Y$
8. Hasil yang diperolah merupakan nilai-nilai koefisien persaman regresi yang dicari, yaitu : $B_0$, $B_1$, $B_2$, $B_3$, …,$B_M$
9. Cetak hasil

   $Y\hat{} = B_0X^0 + B_1X^1 + B_2X^2 + B_3X^3 + \ldots + B_MX^M$
10. Selesai

Langkah umum penyelesaian untuk mencari harga-harga koefisien $B_0$, $B_1$, $B_2$, $B_3$, … $B_M$ dalam regresi polinomial yang dapat digambarkan ke dalam flowchart sistem seperti ditunjukkan pada gambar **Gambar 3.3**. Tentu saja *flowchart* sistem tersebut masih dapat dibuat ke dalam *flowchart* prosedur yang lebih rinci.




**Gambar 3.3** : *Flowchart* sistem mencari koefisien-koefisien persamaan pada regresi polinomial

Sebagai pelengkap analisa terhadap pasangan data X dan Y, maka perlu disertakan perhitungan-perhitungan untuk beberapa parameter berikut :

1. Korelasi sederhana, dihitung dengan formula sebagai berikut :

$$\frac{\sum (X I] - \overline{X} - \{\sum Y[I] - \overline{Y}\}}{\sum (X[I] - \overline{X})^2 x (\sum Y[I] - \overline{Y})^2}$$

Keterangan:

$\overline{X}$ : harga rata-rata data X

$\overline{Y}$ : harga rata-rata data Y

2. Tabel Analisa Varians

| **Sumber Variasi** | **Derajat Bebas** | **Jumlah Kuadrat** | **Rata-rata Kuadrat** | **Nilai F** |
|---|---|---|---|---|
| Regresi | M | SSR | MSR | $F_{hit}$ |
| Deviasi | N – M – 1 | SSD | MSD | - |
| Total | N - 1 | SST | - | - |

Keterangan :

M : orde regresi polinomial

N : cacah pasangan data

SSR : $(\sum Y[I] - \overline{Y})^2$

SST : $(\sum Y[I] - \overline{Y})^2$

SSD : SST - SST

MSR : SSR / M

MSD : SSD / (N –M -1)

$F_{hit}$ : MSR / MSD

3. Koefisien determinasi, dihitung dengan formula sebagai berikut :

   $R^2$ = SSR / SST

4. Koefisien korelasi ganda, dihitung dengan formula sebagai berikut :

   R = SQRT($R^2$)

5. Ralat baku dari koefisien regresi, dihitung dengan formula sebagai berikut :

   SD = SQRT {(SSD x X[I]) / (N – M -1)}

   Keterangan:

   I = 0, 1, 2, 3, …, M

Sebagai keterangan tambahan, metoda *least square* yang menghasilkan persamaan garis regresi dan kurva regresi, dapat dipergunakan sebagai dasar analisa yang lebih lanjut untuk peramalan (*forecasting*) kondisi di masa yang akan datang berdasarkan data lampau yang diketahui. Caranya, adalah dengan mensubstitusikan nilai-nilai variabel (=X) ke dalam persamaan regresi. Sehingga akan diperoleh harga-harga Y untuk setiap harga X yang disubstitusikan tersebut. Perhitungannya sangat mudah dilakukan, dan hasilnya relatif cukup akurat.