Universitas Gadjah Mada
Fakultas Teknik
Departemen Teknik Sipil dan Lingkungan

REGRESI

Statistika dan Probabilitas

Kurva Regresi

- Mencari garis/kurva yang mewakili serangkaian titik data
- Ada dua cara untuk melakukannya, yaitu
  - Regresi
  - Interpolasi
- Aplikasi di bidang enjiniring
  - Pola perilaku data (*trend analysis*)
  - Uji hipotesis (*hypothesis testing*)

Kurva Regresi

- Pemakaian regresi
  - Apabila data menunjukkan tingkat kesalahan yang cukup signifikan atau menunjukkan adanya *noise*
  - Untuk mencari satu kurva tunggal yang mewakili pola umum perilaku data
  - Kurva yang dicari tidak perlu melewati setiap titik data

Kurva Regresi

- Interpolasi
  - Diketahui bahwa data sangat akurat
  - Untuk mencari satu atau serangkaian kurva yang melewati setiap titik data
  - Untuk memperkirakan nilai-nilai di antara titik-titik data
- Ekstrapolasi
  - Mirip dengan interpolasi, tetapi untuk memperkirakan nilai-nilai di luar kisaran titik-titik data

Kurva Regresi terhadap Data Pengukuran

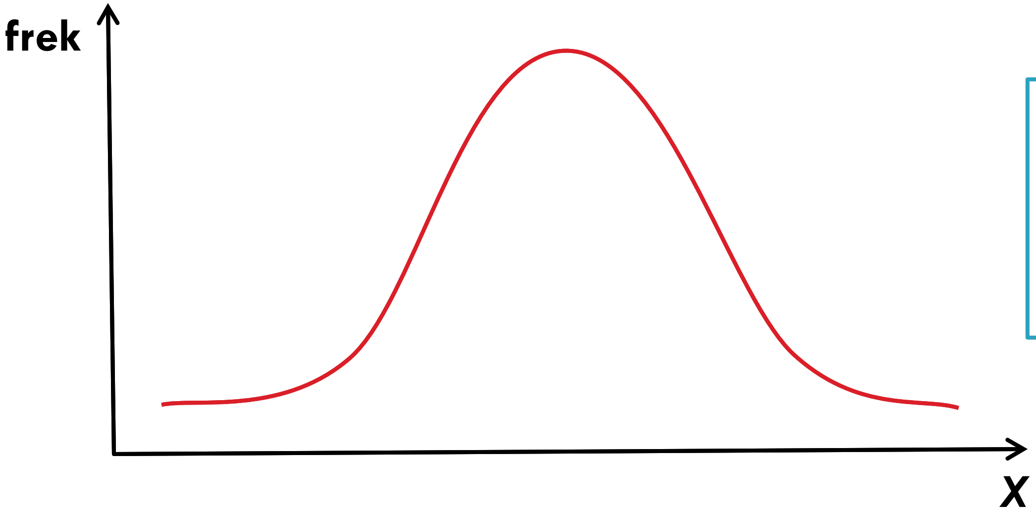
- Analisis pola perilaku data
  - Pemanfaatan pola data (pengukuran, eksperimen) untuk melakukan perkiraan
  - Apabila data persis (akurat): interpolasi
  - Apabila data tak persis (tak akurat): regresi
- Uji hipotesis
  - Pembandingan antara hasil teori atau hasil hitungan dengan hasil pengukuran

Beberapa Besaran Statistik

merepresentasikan sebaran data

- Rata-rata aritmatik, *mean* ➡ $\bar{y} = \frac{1}{n}\sum y_i$
- Deviasi standar, simpangan baku, *standard deviation* ➡ $s_y = \sqrt{\frac{S_t}{n-1}}$ $\quad S_t = \sum (y_i - \bar{y})^2$
- Varian ('ragam'), *variance* ➡ ${s_y}^2 = \frac{S_t}{n-1}$
- *Coefficient of variation* ➡ $c_v = \frac{s_y}{\bar{y}} 100\%$

Distribusi Probabilitas



Distribusi Normal

salah satu distribusi/sebaran data yang sering dijumpai adalah distribusi normal

Regresi

Regresi linear: metode kuadrat terkecil

Regresi hubungan tak-linear yang dilinearkan

Regresi: Metode Kuadrat Terkecil

- Mencari satu kurva atau satu fungsi (pendekatan) yang sesuai dengan pola umum yang ditunjukkan oleh data
  - Datanya menunjukkan kesalahan yang cukup signifikan
  - Kurva tidak perlu memotong setiap titik data
- Regresi linear
- Regresi persamaan-persamaan tak-linear yang dilinearkan
- Regresi tak-linear

Regresi: Metode Kuadrat Terkecil

- Bagaimana caranya?
  - Program komputer
  - *Spreadsheet* (Microsoft Excel)
  - Program aplikasi: Matlab, Octave, Scilab

Regresi Linear

- Kesalahan atau residu (e) adalah perbedaan antara nilai *y* sesungguhnya (data *y*) dan *y* nilai pendekatan (y_{reg}) menurut persamaan linear $a_0 + a_1x$.

$$e = y - y_{reg} = y - (a_0 + a_1 x)$$

- Meminimumkan jumlah kuadrat residu tersebut

$$\min[S_r] = \min\left[\sum {e_i}^2\right] = \min\left[\sum (y - a_0 - a_1 x)^2\right]$$

Regresi Linear

- Bagaimana cara mencari koefisien a_0 dan a_1?
  - Diferensialkan persamaan tersebut dua kali, masing-masing terhadap a_0 dan a_1.
  - Samakan kedua persamaan hasil diferensiasi tersebut dengan nol.

$$\frac{\partial S_r}{\partial a_0} = -2\sum(y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum(y_i - a_0 - a_1 x_i)x_i$$

Regresi Linear

- Selesaikan persamaan yang didapat untuk mencari a_0 dan a_1

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum {x_i}^2 - (\sum x_i)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

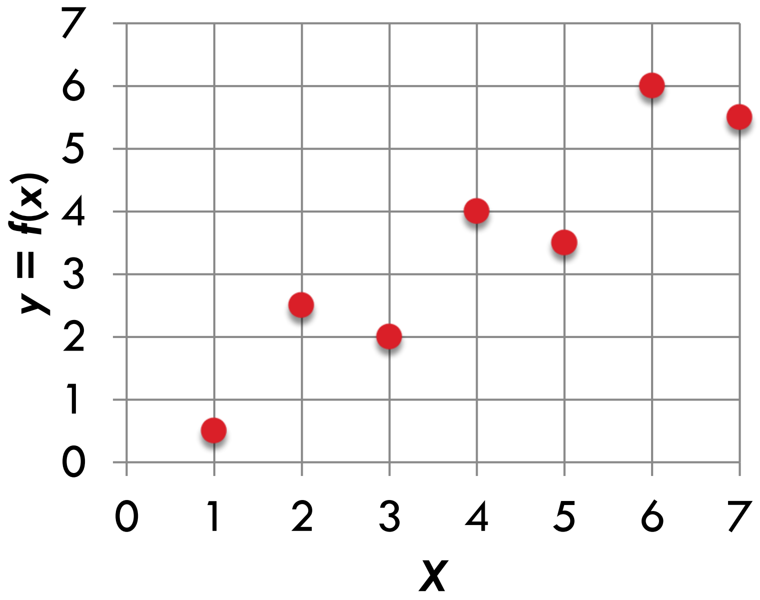
- dalam hal ini, $\bar{y}$ dan $\bar{x}$ masing-masing adalah nilai y rata-rata x rata-rata

Contoh Regresi Linear

Tabel data

i	x_i	$y_i = f(x_i)$
0	1	0.5
1	2	2.5
2	3	2
3	4	4
4	5	3.5
5	6	6
6	7	5.5

Grafik/kurva data



Hitungan Regresi Linear

i	x_i	y_i	$x_i\ y_i$	x_i^2	y_{reg}	$(y_i - y_{reg})^2$	$(y_i - y_{mean})^2$
0	1	0.5	0.5	1	0.910714	0.168686	8.576531
1	2	2.5	5	4	1.75	0.5625	0.862245
2	3	2.0	6	9	2.589286	0.347258	2.040816
3	4	4.0	16	16	3.428571	0.326531	0.326531
4	5	3.5	17.5	25	4.267857	0.589605	0.005102
5	6	6.0	36	36	5.107143	0.797194	6.612245
6	7	5.5	38.5	49	5.946429	0.199298	4.290816
$\sum =$	**28**	**24.0**	**119.5**	**140**	$\sum =$	**2.991071**	**22.71429**

Hitungan Regresi Linear

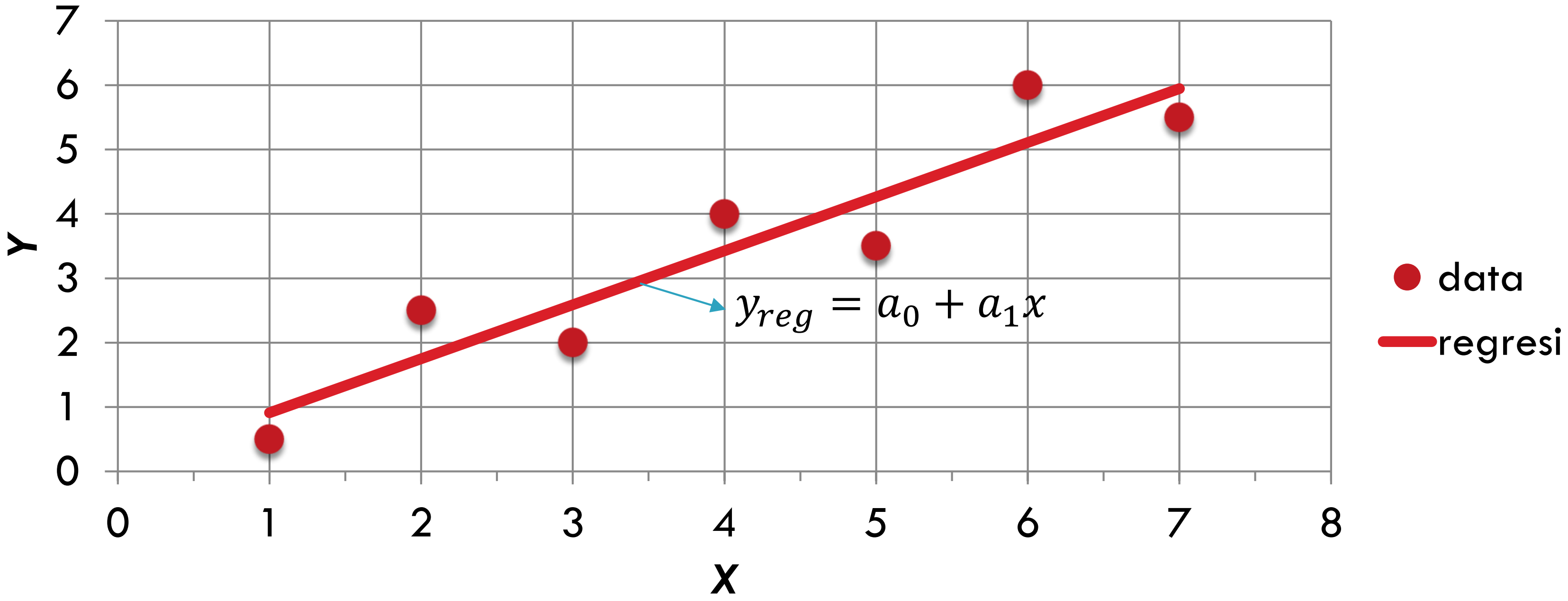
$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum {x_i}^2 - (\sum x_i)^2} = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.839286$$

$$\bar{y} = \frac{24}{7} = 3.4$$

$$\bar{x} = \frac{28}{7} = 4$$

$$a_0 = \bar{y} - a_1\bar{x} = 3.4 - 0.839286(4) = 0.071429$$

Hitungan Regresi Linear



Regresi Linear

- Kuantifikasi kesalahan
  - Kesalahan standar

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} \qquad S_r = \sum (y_i - a_0 - a_1 x_i)^2$$

  - Perhatikan kemiripannya dengan simpangan baku

$$s_y = \sqrt{\frac{S_t}{n-1}} \qquad S_t = \sum (y_i - \bar{y})^2$$

Regresi Linear

- Beda antara kedua kesalahan tersebut menunjukkan perbaikan atau pengurangan kesalahan

$$r^2 = \frac{S_t - S_r}{S_t}$$ → koefisien determinasi (*coefficient of determination*)

$$r = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum {x_i}^2 - \left(\sum x_i\right)^2}\sqrt{n\sum {y_i}^2 - \left(\sum y_i\right)^2}}$$ → koefisien korelasi (*correlation coefficient*)

$-1 \le r \le +1$

Hitungan regresi linear

$$S_r = \sum (y_i - a_0 - a_1 x_i)^2 = 2.991071$$

$$S_t = \sum (y_i - \bar{y})^2 = 22.71429$$

$$r^2 = \frac{S_t - S_r}{S_t} = \frac{22.71429 - 2.991071}{22.71429} = 0.868318$$

$$r = 0.931836$$

$$\boxed{-1 \leq r \leq +1}$$

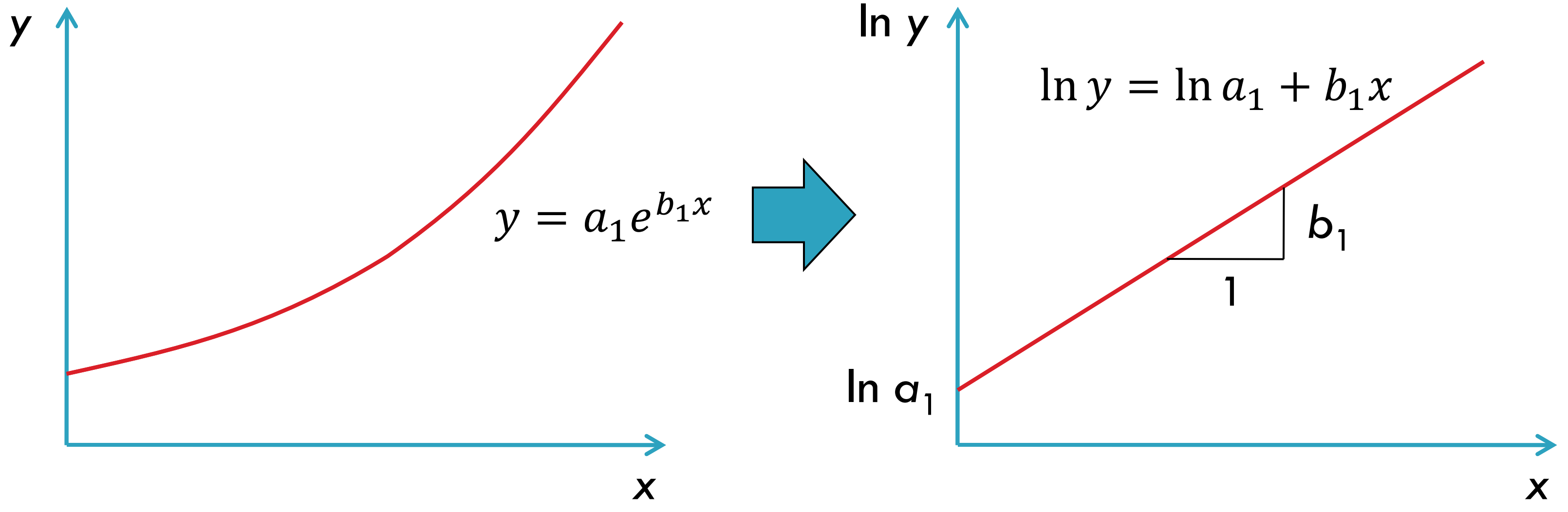
Regresi

Linearisasi persamaan tak-linear

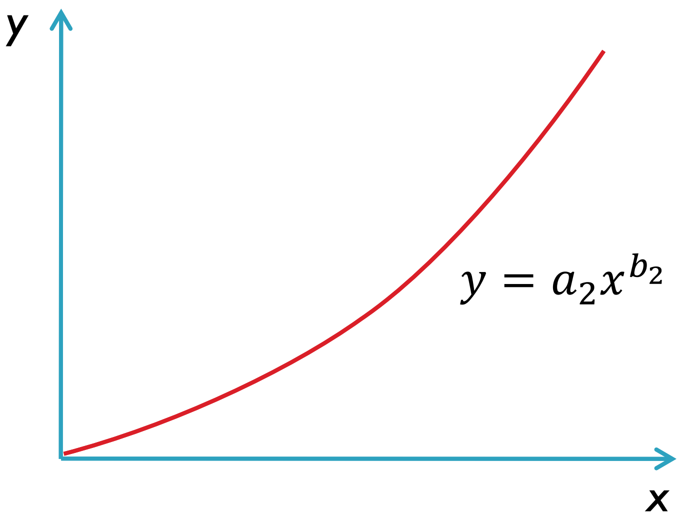
Regresi Linear

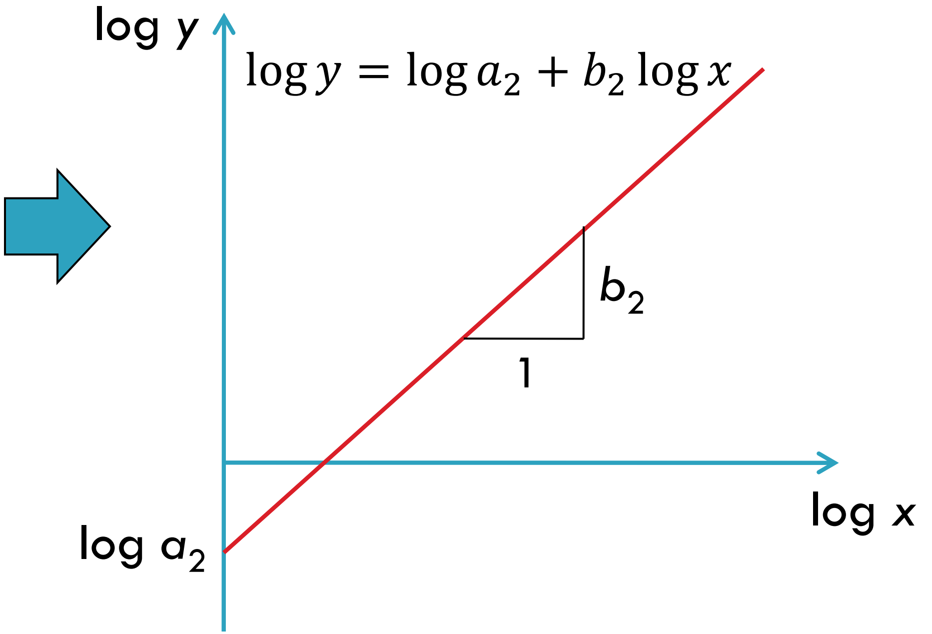
- Linearisasi persamaan-persamaan tak-linear
  - Logaritmik menjadi linear
  - Eksponensial menjadi linear
  - Pangkat (polinomial tingkat $n > 1$) menjadi linear (polinomial tingkat 1)
  - Dll.

Linearisasi Persamaan Tak-Linear

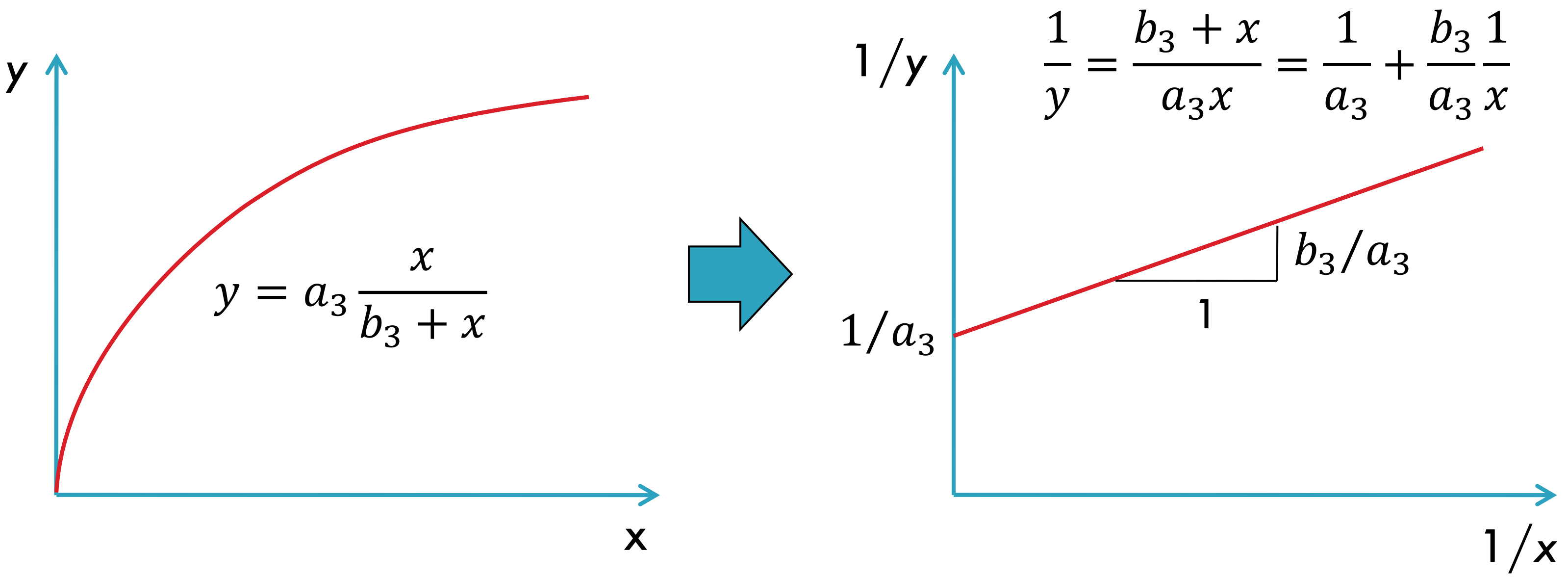


Linearisasi Persamaan Tak-Linear





Linearisasi Persamaan Tak-Linear



Regresi

Regresi Polinomial

Regresi Polinomial

- Sebagian data teknik, walau menunjukkan pola yang tampak jelas, namun pola itu tidak dapat direpresentasikan sebagai sebuah garis lurus
- Regresi linear menjadi tidak cocok untuk menggambarkan pola hubungan antara *dependent variable* dan *independent variables*
- Opsi regresi
  - Cara 1: transformasi koordinat (persamaan nonlinear menjadi persamaan linear)
  - Cara 2: regresi polinomial

Regresi Polinomial

- Polinomial derajat *m*

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

- Kuadrat error

$$S_r = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 {x_i}^2 - \cdots - a_m {x_i}^m\right)^2$$

Regresi Polinomial

- Metode kuadrat-terkecil, diperluas untuk keperluan "pencocokan" data ke sebuah persamaan polinomial derajat *m*
- Persamaan di kanan disamakan dengan nol dan kemudian diselesaikan untuk mendapatkan $a_0, a_1, \ldots$

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m\right)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i\left(y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m\right)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^2\left(y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m\right)$$

$$\vdots$$

$$\frac{\partial S_r}{\partial a_m} = -2\sum x_i^m\left(y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m\right)$$

Regresi Polinomial

$$a_0 n + a_1 \sum_{i-1}^{n} x_i + a_2 \sum_{i=1}^{n} {x_i}^2 + \cdots + a_m \sum_{i=1}^{n} {x_i}^m = \sum_{i=1}^{n} y_i$$

$$a_0 \sum_{i-1}^{n} x_i + a_1 \sum_{i=1}^{n} {x_i}^2 + a_2 \sum_{i=1}^{n} {x_i}^3 + \cdots + a_m \sum_{i=1}^{n} {x_i}^{m+1} = \sum_{i=1}^{n} x_i y_i$$

$$a_0 \sum_{i=1}^{n} {x_i}^2 + a_1 \sum_{i=1}^{n} {x_i}^3 + a_2 \sum_{i=1}^{n} {x_i}^4 + \cdots + a_m \sum_{i=1}^{n} {x_i}^{m+2} = \sum_{i=1}^{n} {x_i}^2 y_i$$

.

.

.

$$a_0 \sum_{i=1}^{n} {x_i}^m + a_1 \sum_{i=1}^{n} {x_i}^{m+1} + a_2 \sum_{i=1}^{n} {x_i}^{m+2} + \cdots + a_m \sum_{i=1}^{n} {x_i}^{2m} = \sum_{i=1}^{n} {x_i}^m y_i$$

Regresi Polinomial

- Ada $m+1$ persamaan linear equations yang mengandung $m+1$ *unknowns*, yaitu $a_0, a_1, a_2, \dots, a_m$
- Metode penyelesaian persamaan di atas
  - Eliminasi Gauss
  - Gauss-Jordan
  - Iterasi Jacobi
  - Inversi matriks

Regresi Polinomial

- Contoh
  - Lakukan regresi untuk mendapatkan polinomial kudratik terhadap data dalam tabel

    $y = a_0 + a_1 x + a_2 x^2$

x_i	y_i
0	2.1
1	7.7
2	13.6
3	27.2
4	40.9
5	61.1

- Jawab

$$y = 2.47857 + 2.35929x + 1.86071x^2$$

$$r^2 = 1 - \frac{S_r}{S_t} = 1 - \frac{3.74657}{2513.39} = 0.9985$$

$$r = 0.9993$$

Regresi

Regresi linear ganda (*multiple linear regression*)

Regresi Linear Ganda

- Misal variabel y adalah fungsi linear dua variabel bebas x_1 dan x_2

$$y = a_0 + a_1 x_1 + a_2 x_2$$

  - Koefisien a_0, a_1, a_2 pada persamaan di atas dapat ditemukan dengan metode kuadrat terkecil kesalahan (*error*)

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})^2$$

Regresi Linear Ganda

- Diferensial parsial persamaan tersebut terhadap masing-masing koefisien adalah sbb.

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n} x_{1i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum_{i=1}^{n} x_{2i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$

- Samakan persamaan diferensial tsb dengan nol dan atur suku-suku dalam persamaan

$$n a_0 + \sum_{i=1}^{n} x_{1i} a_1 + \sum_{i=1}^{n} x_{2i} a_2 = \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} x_{1i} a_0 + \sum_{i=1}^{n} {x_{1i}}^2 a_1 + \sum_{i=1}^{n} x_{1i} x_{2i} a_2 = \sum_{i=1}^{n} x_{1i} y_i$$

$$\sum_{i=1}^{n} x_{2i} a_0 + \sum_{i=1}^{n} x_{1i} x_{2i} a_1 + \sum_{i=1}^{n} {x_{2i}}^2 a_2 = \sum_{i=1}^{n} x_{2i} y_i$$

Regresi Linear Ganda

- Persamaan-persamaan linear tersebut dapat dituliskan dalam bentuk persamaan matriks sbb.

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}x_{1i} & \sum_{i=1}^{n} x_{2i}x_{1i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}x_{1i} & \sum_{i=1}^{n} x_{2i}x_{2i} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \end{Bmatrix}$$

Regresi Linear Ganda

- Inversi matriks

$$\begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}x_{1i} & \sum_{i=1}^{n} x_{2i}x_{1i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}x_{1i} & \sum_{i=1}^{n} x_{2i}x_{2i} \end{bmatrix}^{-1} \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \end{Bmatrix}$$

Contoh

- Temukanlah persamaan linear yang mewakili pola sebaran data dalam tabel di samping ini.
- Jawab

$$y = 5 + 4x_1 - 3x_2$$

$$r^2 = 1$$

x_1	x_2	y
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

Regresi Linear Ganda

- Regresi linear ganda dapat dipakai untuk "*best fitting*" persamaan berpangkat (*power equations*)

$$y = a_0 {x_1}^{a_1} {x_2}^{a_2} \dots {x_m}^{a_m}$$

- Persamaan di atas sangat bermanfaat dalam kasus *fitting* data eksperimen
- Persamaan di atas diubah menjadi

$$\log y = \log a_0 + a_1 \log x_1 + a_2 \log x_2 + \cdots + a_m \log x_m$$

- Persamaan di atas adalah bentuk persamaan regresi linear ganda dalam skala logaritma

Regresi

General linear least-squares

Metode Kuadrat-terkecil Linear

- Ketiga metode kuadrat terkecil untuk mendapatkan persamaan linear yang telah dibahas dalam bab ini, yaitu linear sederhana, polinomial, dan linear ganda dapat dirangkum sebagai berikut

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m$$

  - $z_0, z_1, \dots, z_m$ adalah $m+1$ persamaan
  - $m+1$ adalah jumlah variabel bebas
  - $n+1$ adalah jumlah titik data
- Bentuk matriks persamaan di atas adalah

$$\{Y\} = [Z]\{A\}$$

Metode Kuadrat-terkecil

$$\{Y\} = [Z]\{A\} \Rightarrow [Z]^T[Z]\{A\} = [Z]^T\{Y\}$$

$$[Z] = \begin{bmatrix} a_{01} & a_{11} & \cdot & \cdot & \cdot & a_{m1} \\ a_{02} & a_{12} & \cdot & \cdot & \cdot & a_{m2} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ a_{0n} & a_{1n} & \cdot & \cdot & \cdot & a_{mn} \end{bmatrix}$$

- $\{Y\}$ data variabel tak-bebas
- $[Z]$ adalah matriks data variabel bebas
- $\{A\}$ koefisien

$$S_r = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} a_j z_{ji}\right)^2$$

Metode Kuadrat-terkecil

$$[Z]^T[Z]\{A\} = [Z]^T\{Y\}$$

- Cara penyelesaian
  - dekomposisi matriks LU
  - metode Cholesky
  - inversi matriks $\longrightarrow$ $\{A\} = \left[[Z]^T[Z]\right]^{-1}[Z]^T\{Y\}$

Terima kasih