Analisis Numerik
Regresi dan Interpolasi

Ahmad Rio Adriansyah

STT Terpadu - Nurul Fikri

ahmad.rio.adriansyah@gmail.com
arasy@nurulfikri.ac.id

April 20, 2017

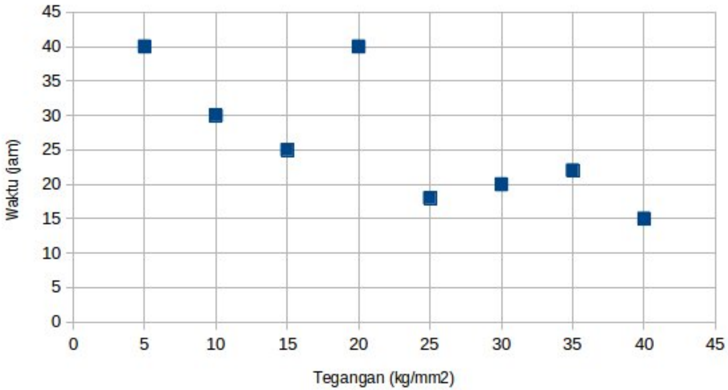
Pengukuran (dari percobaan) biasanya data diskrit

Contoh :
Berikut adalah data pengujian kekuatan baja

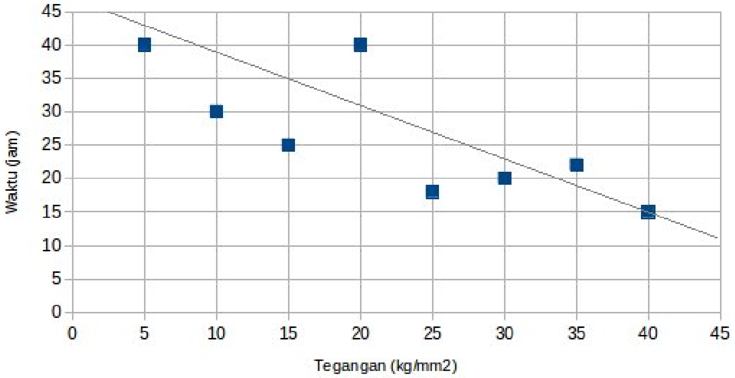
Tegangan (kg/mm^2)	5	10	15	20	25	30	35	40
Waktu patah (*jam*)	40	30	25	40	18	20	22	15

Bagaimana jika ingin mengetahui waktu patah baja jika diberi tegangan sebesar 18 kg/mm^2 tanpa harus melakukan pengukuran lagi?

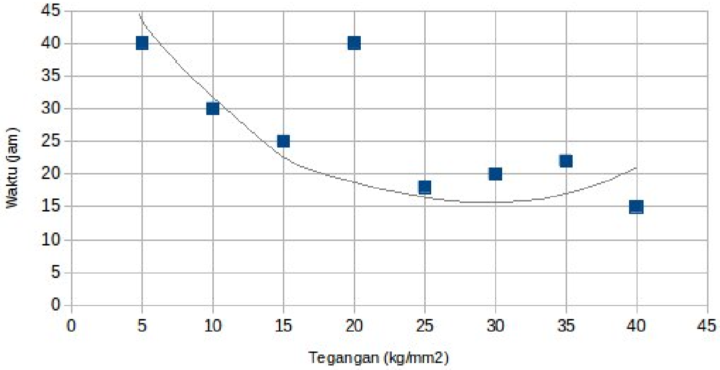
Bisa didekati dengan sebuah fungsi



Bisa didekati dengan sebuah fungsi linier



Bisa didekati dengan sebuah fungsi kuadratik



Pencocokan Kurva

Dalam numerik, disebut sebagai pencocokan kurva(*curve fitting*).

- Regresi
  Untuk data yang kurang teliti, kurva tidak harus melalui semua titik. Kurva dibuat sedemikian rupa sehingga kuadrat dari selisih antara titik data dan kurva (hampiran) sekecil mungkin. Metode regresi dengan cara ini disebut sebagai regresi kuadrat terkecil (*least square regression*).
- Interpolasi
  Untuk data dengan ketelitian tinggi, lebih cocok jika dibuat kurva yang melalui semua titik. Jika kurva yang digunakan adalah sebuah polinom, maka pekerjaan ini disebut interpolasi polinom. Polinom yang digunakan disebut polinom interpolasi.

Regresi Linier

Titik-titik data (x_i, y_i) ingin kita hampiri dengan sebuah garis lurus $f(x) = mx + c$ sehingga selisihnya minimum. Garis ini disebut sebagai garis regresi.

Selisih (penyimpangan/deviasi) antara nilai titik data dan nilai hampiran dapat dituliskan sebagai berikut

$$\begin{aligned} r_i &= y_i - f(x_i) \\ &= y_i - (mx_i + c) \end{aligned}$$

dan total kuadrat deviasinya adalah

$$R = \sum r_i^2 = \sum (y_i - mx_i - c)^2$$

Regresi Linier

Untuk mendapatkan R yang minimum, maka turunannya harus bernilai 0 (dari mata kuliah kalkulus/matematika dasar)

$$\frac{\partial R}{\partial c} = -2 \sum (y_i - c - mx_i) = 0$$
$$\frac{\partial R}{\partial m} = -2 \sum x_i (y_i - c - mx_i) = 0$$

berarti

$$\sum (y_i - c - mx_i) = 0 \quad \rightarrow \quad \sum y_i - \sum c - \sum mx_i = 0$$
$$\sum x_i (y_i - c - mx_i) = 0 \quad \rightarrow \quad \sum x_i y_i - \sum cx_i - \sum mx_i^2 = 0$$

Regresi Linier

$$\sum y_i = \sum c + \sum mx_i$$
$$\sum x_i y_i = \sum cx_i + \sum mx_i^2$$

karena m dan c konstan, persamaannya dapat diubah menjadi

$$\sum y_i = nc + m \sum x_i$$
$$\sum x_i y_i = c \sum x_i + m \sum x_i^2$$

Regresi Linier

Persamaan tersebut dapat diubah menjadi persamaan matriks

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} c \\ m \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

dan diselesaikan dengan metode eliminasi Gauss, menghasilkan

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

$$c = \frac{\sum y_i - m \sum x_i}{n}$$

Regresi Linier

Contoh:

Berikut adalah data pengujian kekuatan baja

Tegangan (kg/mm^2)	5	10	15	20	25	30	35	40
Waktu patah (jam)	40	30	25	40	18	20	22	15

Untuk mempermudah perhitungan digunakan tabel

i	x_i	y_i	x_i^2	$x_i y_i$
1	5	40	25	200
2	10	30	100	300
3	15	25	225	375
4	20	40	400	800
5	25	18	625	450
6	30	20	900	600
7	35	22	1225	770
8	40	15	1600	600
$n = 8$	$\sum x_i = 180$	$\sum y_i = 210$	$\sum x_i^2 = 5100$	$\sum x_i y_i = 4095$

Regresi Linier

Dari hasil tersebut didapat persamaan matriks

$$\begin{bmatrix} 8 & 180 \\ 180 & 5100 \end{bmatrix} \begin{bmatrix} c \\ m \end{bmatrix} = \begin{bmatrix} 210 \\ 4095 \end{bmatrix}$$

dan diselesaikan dengan metode eliminasi Gauss, menghasilkan

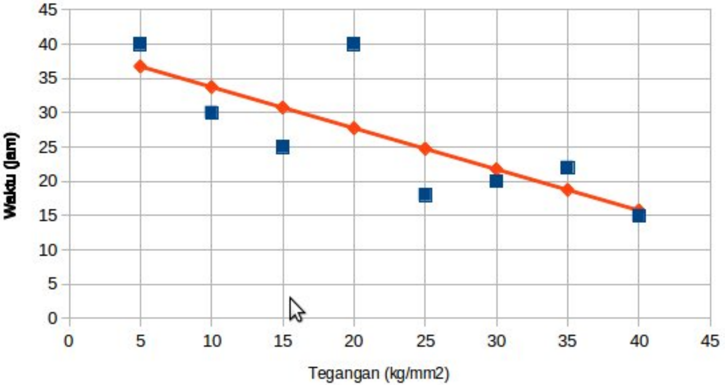
$$\begin{aligned} m &= \frac{(8 * 4095) - (180 * 210)}{(8 * 5100) - (180)^2} \\ &= \frac{-5040}{8400} = \boxed{-0,6} \end{aligned}$$

$$\begin{aligned} c &= \frac{210 - (-0,6 * 180)}{8} \\ &= \frac{318}{8} = \boxed{39,75} \end{aligned}$$

Regresi Linier

Berarti garis yang dapat meregresi titik-titik data tersebut dengan deviasi yang minimum (garis regresi) adalah

$$f(x) = -0,6x + 39,75$$



Galat

Untuk mengetahui seberapa bagus fungsi garis regresi tersebut, kita dapat mengukurnya dengan galat RMS (*Root-mean-square error*)

$$\epsilon_{RMS} = \sqrt{\left(\frac{\sum |f(x_i) - y_i|^2}{n}\right)}$$

Semakin kecil ϵ_{RMS} nya, maka semakin bagus fungsi regresi untuk mencocokkan titik-titik data tersebut.

ϵ_{RMS} dari data di atas adalah 5,7825

Latihan

Tentukan fungsi linier untuk mencocokkan titik-titik data berikut dengan metode regresi. Hitung pula galat RMS nya

- 1

x	1,0	1,5	2,0	2,5	3,0
y	2,0	3,2	4,1	4,9	5,9

- 2

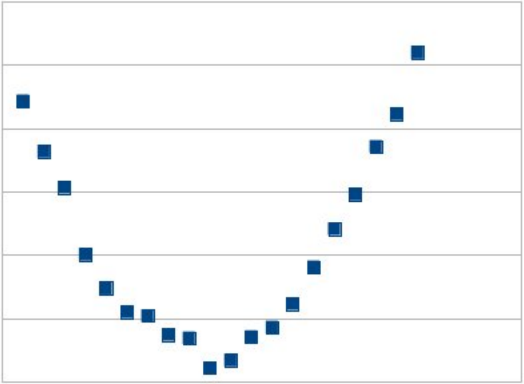
x	0,1	0,4	0,5	0,7	0,7	0,9
y	0,61	0,92	0,99	1,52	1,47	2,03

Linierisasi

Tidak semua data cocok diregresikan dengan garis lurus. Beberapa data persebarannya mengikuti fungsi yang tidak linier.

Linierisasi

Contoh :



Linierisasi

Tidak semua data cocok diregresikan dengan garis lurus. Beberapa data persebarannya mengikuti fungsi yang tidak linier.

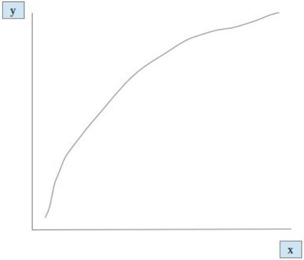
Tahap awal dalam menentukan regresi apa yang cocok digunakan seharusnya adalah dugaan secara visual. Titik-titik data digambarkan ke dalam diagram kartesius, lalu kita memperkirakan fungsi apa yang cocok untuk titik-titik data tersebut.

Beberapa fungsi yang non linier, dapat diubah menjadi fungsi yang linier dengan langkah-langkah tertentu. Proses ini disebut sebagai linierisasi/pelinieran.

Persamaan Pangkat Sederhana

Jika sebuah data memiliki kecenderungan lengkung (tak linier), dan di saat x mendekati tak hingga y menuju tak hingga, maka kita bisa menduga data tersebut cocok jika diregresikan dengan fungsi

$$y = ax^b$$



Persamaan Pangkat Sederhana

Persamaan tersebut dilinerisasi dengan cara memberikan operator *ln* ke kedua sisi persamaan.

$$
\begin{aligned}
y &= ax^b \\
ln(y) &= ln(ax^b) \\
ln(y) &= ln(a) + ln(x^b)
\end{aligned}
$$

$$
\begin{aligned}
ln(y) &= ln(a) + b * ln(x) \\
&\downarrow \\
Y &= c + m * X
\end{aligned}
$$

Persamaan Pangkat Sederhana

Dengan mengambil nilai-nilai

$$
\begin{aligned}
Y &= ln(y) \\
X &= ln(x) \\
c &= ln(a) \\
m &= b
\end{aligned}
$$

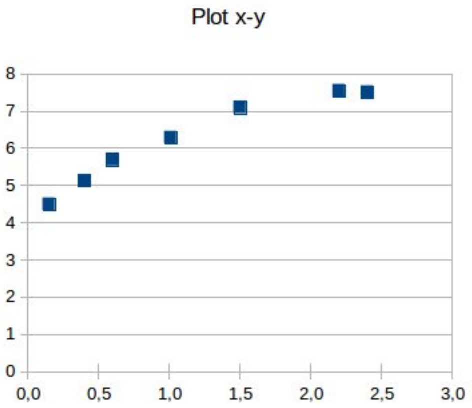
kita bisa mendapatkan persamaan linier $Y = mX + c$. Berarti titik-titik data yang kita punya juga harus disesuaikan.

Persamaan Pangkat Sederhana

Contoh:

Cocokkan data berikut dengan fungsi $y = ax^b$

x_i	y_i
0,1500	4,4964
0,4000	5,1284
0,6000	5,6931
1,0100	6.2884
1,5000	7.0989
2,2000	7,5507
2,4000	7,5106



Persamaan Pangkat Sederhana

Kita akan dekati titik-titik data tersebut dengan fungsi $y = ax^b$, jadi kita perlu peubah baru X dan Y.

i	x_i	y_i	$X_i = ln(x_i)$	$Y_i = ln(y_i)$	X_i^2	$X_i Y_i$
1	0,1500	4,4964	-1,8971	1,5033	3,5990	-2,8519
2	0,4000	5,1284	-0,9163	1,6348	0,8396	-1,4980
3	0,6000	5,6931	-0.5108	1,7393	0,2609	-0,8884
4	1,0100	6.2884	0.0100	1,8387	0,0001	0,0184
5	1,5000	7.0989	0,4055	1,9599	0,1644	0,7947
6	2,2000	7,5507	0,7885	2,0216	0,6217	1,5940
7	2,4000	7,5106	0,8755	2,0163	0,7665	1,7653
		$\sum$	**-1,2447**	**12,7139**	**6,2522**	**-1,0659**

Persamaan Pangkat Sederhana

Dari tabel tersebut, kita bisa menghitung nilai m dan c dengan cara

$$
\begin{aligned}
m &= \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \\
&= \frac{7 * (-1,0659) - (-1,2447) * 12,7139}{7(6,2522) - (-1,2447)^2} \\
&= 0,1981
\end{aligned}
$$

$$
\begin{aligned}
c &= \frac{\sum Y_i - m\sum X_i}{n} \\
&= \frac{12,7139 - 0,1981 * (-1,2447)}{7} \\
&= 1,8515
\end{aligned}
$$

*catatan : nilai X dan Y yang digunakan adalah nilai yang sudah diubah, bukan nilai asalnya.

Persamaan Pangkat Sederhana

Kita mendapatkan persamaan $Y = 0,1981X + 1,8515$ untuk meregresi titik-titik data tersebut. Tapi ingat, fungsi tersebut meregresi titik-titik yang sudah ditransformasi ke X dan Y.

Untuk mendapatkan fungsi yang benar, harus kita ubah fungsi tersebut menjadi fungsi $y = ax^b$ dengan

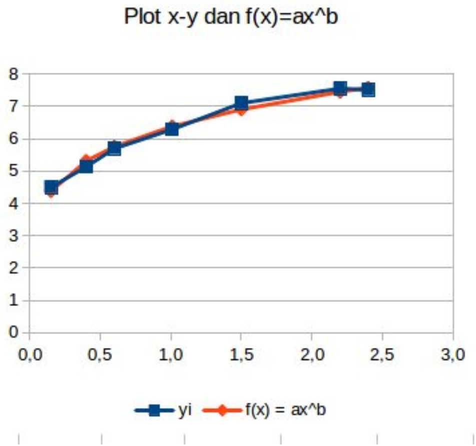
$a = e^c = e^{1,8515} = \boxed{6,3694}$ (dari persamaan $c = ln(a)$)

$b = m = \boxed{0,1981}$ (dari persamaan $m = b$)

Kita dapat fungsi regresinya adalah

$$y = 6,3694\ x^{0,1981}$$

Persamaan Pangkat Sederhana



Persamaan Pangkat Sederhana

Jika kita hitung galat dari fungsi tersebut dengan galat RMS

$$
\begin{aligned}
\epsilon_{RMS} &= \sqrt{\left(\frac{\sum |f(x_i) - y_i|^2}{n}\right)} \\
&= \sqrt{\frac{0,1153}{7}} \\
&= 0,1284
\end{aligned}
$$

kita mendapatkan nilai galat yang mendekati nol.

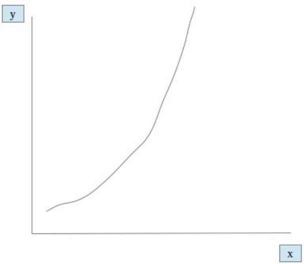
Persamaan Eksponensial

Serupa dengan persamaan pangkat sederhana, kita bisa regresikan dengan fungsi eksponensial jika data yang kita miliki

- cenderung lengkung (tak linier) dan
- di saat x mendekati tak hingga y menuju tak hingga.

Selain itu, kedivergenan data lebih cepat dibanding persamaan pangkat. Fungsi yang dimaksud adalah

$$y = ae^{bx}$$



Persamaan Eksponensial

Persamaan tersebut dilinerisasi dengan cara memberikan operator *ln* ke kedua sisi persamaan.

$$\begin{aligned} y &= ae^{bx} \\ ln(y) &= ln(ae^{bx}) \\ ln(y) &= ln(a) + ln(e^{bx}) \\ ln(y) &= ln(a) + bx * ln(e) \end{aligned}$$

$$\begin{aligned} ln(y) &= ln(a) + b * x \\ &\downarrow \\ Y &= c + m * X \end{aligned}$$

Persamaan Eksponensial

Dengan mengambil nilai-nilai

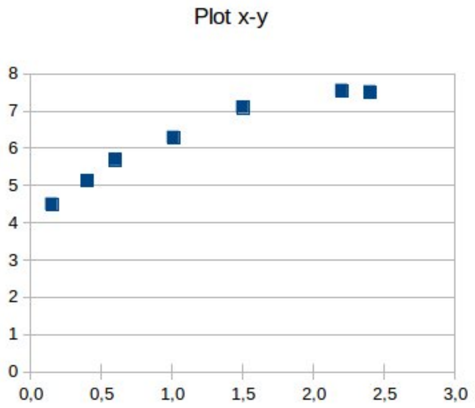
$$
\begin{aligned}
Y &= ln(y) \\
X &= x \\
c &= ln(a) \\
m &= b
\end{aligned}
$$

kita bisa mendapatkan persamaan linier $Y = mX + c$. Berarti titik-titik data yang kita punya juga harus disesuaikan.

Contoh:

Cocokkan data berikut dengan fungsi $y = ae^{bx}$

x_i	y_i
0,1500	4,4964
0,4000	5,1284
0,6000	5,6931
1,0100	6.2884
1,5000	7.0989
2,2000	7,5507
2,4000	7,5106



Persamaan Eksponensial

Kita akan dekati titik-titik data tersebut dengan fungsi $y = ae^{bx}$. Berbeda dengan persamaan pangkat sederhana, di sini kita hanya perlu peubah baru Y karena $X = x$.

i	x_i	y_i	$Y_i = ln(y_i)$	x_i^2	$x_i Y_i$
1	0,1500	4,4964	1,5033	0,0225	0,2255
2	0,4000	5,1284	1,6348	0,1600	0,6539
3	0,6000	5,6931	1,7393	0,3600	1,0436
4	1,0100	6.2884	1,8387	1,0201	1,8571
5	1,5000	7.0989	1,9599	2,2500	2,9399
6	2,2000	7,5507	2,0216	4,8400	4,4475
7	2,4000	7,5106	2,0163	5,7600	4,8391
$\sum$	**8,2600**		**12,7139**	**14,4126**	**16,0066**

Persamaan Eksponensial

Dari tabel tersebut, kita bisa menghitung nilai m dan c dengan cara yang sama

$$
\begin{aligned}
m &= \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \\
&= \frac{7 * (16,0066) - 8,2600 * 12,7139}{7(14,4126) - (8,2600)^2} \\
&= 0,2152
\end{aligned}
$$

$$
\begin{aligned}
c &= \frac{\sum Y_i - m \sum X_i}{n} \\
&= \frac{12,7139 - 0,2152 * 8,2600}{7} \\
&= 1,5623
\end{aligned}
$$

Persamaan Eksponensial

Kita mendapatkan persamaan $Y = 0,2152X + 1,5623$ untuk meregresi titik-titik data tersebut.

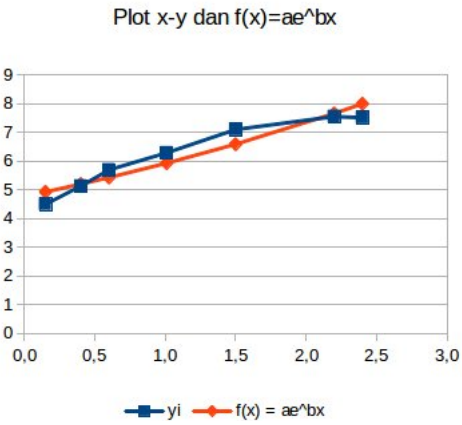
Untuk mendapatkan fungsi yang benar, harus kita ubah fungsi tersebut menjadi fungsi $y = ae^{bx}$ dengan

$$a = e^c = e^{1,5623} = \boxed{4,7698} \qquad \text{(dari persamaan } c = ln(a))$$
$$b = m = \boxed{0,2152} \qquad \text{(dari persamaan } m = b)$$

Kita dapat fungsi regresinya adalah

$$y = 4,7698\ e^{0,2152x}$$

Persamaan Eksponensial



Persamaan Eksponensial

Jika kita hitung galat dari fungsi tersebut dengan galat RMS

$$
\begin{aligned}
\epsilon_{RMS} &= \sqrt{\left(\frac{\sum |f(x_i) - y_i|^2}{n}\right)} \\
&= \sqrt{\frac{0,8985}{7}} \\
&= 0,3583
\end{aligned}
$$

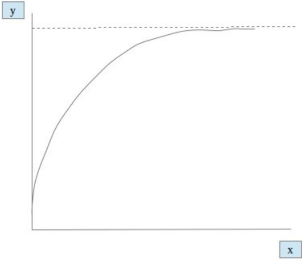
Model Laju Pertumbuhan Jenuh

Jika sifat dari data yang kita miliki adalah

- cenderung lengkung (tak linier) dan
- di saat x mendekati tak hingga y menuju ke suatu titik (asimtot)

kita bisa coba regresikan data tersebut dengan fungsi

$$y = \frac{ax}{b+x}$$



Model Laju Pertumbuhan Jenuh

Persamaan tersebut dilinerisasi dengan cara

$$
\begin{aligned}
y &= \frac{ax}{b+x} \\
\frac{1}{y} &= \frac{b+x}{ax} \\
\frac{1}{y} &= \frac{b}{ax} + \frac{x}{ax}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{y} &= \frac{b}{a}\frac{1}{x} + \frac{1}{a} \\
&\downarrow \\
Y &= m * X + c
\end{aligned}
$$

Model Laju Pertumbuhan Jenuh

Dengan mengambil nilai-nilai

$$Y = \frac{1}{y} \qquad c = \frac{1}{a}$$

$$X = \frac{1}{x} \qquad m = \frac{b}{a}$$

kita bisa mendapatkan persamaan linier $Y = mX + c$. Berarti titik-titik data yang kita punya juga harus disesuaikan.

Model Laju Pertumbuhan Jenuh

i	x_i	y_i	$X_i = \frac{1}{x}$	$Y_i = \frac{1}{y}$	X_i^2	$X_i Y_i$
1	0,1500	4,4964	6,6667	0,2224	44,4444	1,4827
2	0,4000	5,1284	2,5000	0,1950	6,2500	0,4875
3	0,6000	5,6931	1,6667	0,1757	2,7778	0,2928
4	1,0100	6.2884	0,9901	0,1590	0,9803	0,1574
5	1,5000	7.0989	0,6667	0,1409	0,4444	0,0939
6	2,2000	7,5507	0,4545	0,1324	0,2066	0,0602
7	2,4000	7,5106	0,4167	0,1331	0,1736	0,0555
		$\sum$	**13,3613**	**1,1585**	**55,2772**	**2,6299**

Model Laju Pertumbuhan Jenuh

$$
\begin{aligned}
m &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \\
&= \frac{7 * (2,6299) - 13,3613 * 1,1585}{7(55,2772) - (13,3613)^2} \\
&= 0,0141
\end{aligned}
$$

$$
\begin{aligned}
c &= \frac{\sum Y_i - m\sum X_i}{n} \\
&= \frac{1,1585 - 0,0141 * 13,3613}{7} \\
&= 0,1387
\end{aligned}
$$

Model Laju Pertumbuhan Jenuh

Kita mendapatkan persamaan $Y = 0,0141X + 0,1387$ untuk meregresi titik-titik data tersebut.

Untuk mendapatkan fungsi yang benar, harus kita ubah fungsi tersebut menjadi fungsi $y = \dfrac{ax}{b+x}$ dengan

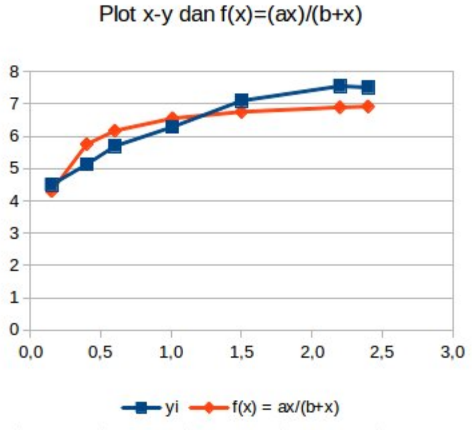
$a = \dfrac{1}{c} = \dfrac{1}{0,1387} = \boxed{7,2116}$ (dari persamaan $c = \dfrac{1}{a}$)

$b = ma = 0,0141 * 7,2116 = \boxed{0,1014}$ (dari persamaan $m = \dfrac{b}{a}$)

Kita dapat fungsi regresinya adalah

$$y = \frac{7,2116x}{0,1014 + x}$$

Model Laju Pertumbuhan Jenuh



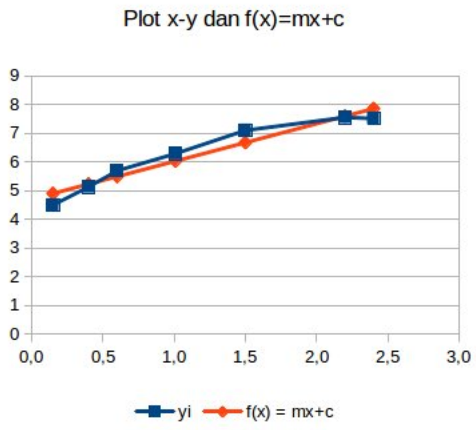
Model Laju Pertumbuhan Jenuh

Jika kita hitung galat dari fungsi tersebut dengan galat RMS

$$\begin{aligned}\epsilon_{RMS} &= \sqrt{\left(\frac{\sum |f(x_i) - y_i|^2}{n}\right)} \\ &= \sqrt{\frac{1,6242}{7}} \\ &= 0,4817\end{aligned}$$

Perbandingan

Sebagai perbandingan, jika menggunakan regresi linier langsung, maka didapatkan fungsi garis $y = 1,3146x + 4,7011$. Galat RMSnya sebesar 0,2885.



Perbandingan

No.	Metode	Fungsi	Galat RMS
1	Linier	$y = mx + c$	0,2885
2	Pangkat Sederhana	$y = ax^b$	0,1284
3	Eksponensial	$y = ae^{bx}$	0,3583
4	Laju Pertumbuhan Jenuh	$y = \dfrac{ax}{b + x}$	0,4817

Mana yang terbaik untuk titik-titik data tersebut?

Fungsi Lain

No.	Fungsi	Bentuk Linier	Perubahan Peubah dan Konstanta
1	$y = mx + c$	$y = mx + c$	-
2	$y = ax^b$	$ln(y) = b * ln(x) + ln(a)$	$Y = ln(y), X = ln(x)$ $a = e^c$
3	$y = ae^{bx}$	$ln(y) = b * x + ln(a)$	$Y = ln(y), X = x$ $a = e^c$
4	$y = \frac{ax}{b+x}$	$\frac{1}{y} = \frac{b}{a}\frac{1}{x} + \frac{1}{a}$	$Y = \frac{1}{y}, X = \frac{1}{x}$ $a = \frac{1}{c}, b = ma$
5	$y = a + \frac{b}{x}$	$y = b\frac{1}{x} + a$	$Y = y, X = \frac{1}{x}$

Fungsi Lain

No.	Fungsi	Bentuk Linier	Perubahan Peubah dan Konstanta
6	$y = \dfrac{a}{b+x}$	$y = -\dfrac{1}{b}(xy) + \dfrac{a}{b}$	$Y = y, X = xy$ $b = \dfrac{-1}{m}, a = \dfrac{-c}{m}$
7	$y = \dfrac{1}{a+bx}$	$\dfrac{1}{y} = bx + a$	$Y = \dfrac{1}{y}, X = x$
8	$y = (a+bx)^2$	$\sqrt{y} = bx + a$	$Y = \sqrt{y}, X = x$
9	$y = ax\ e^{-bx}$	$ln(\dfrac{y}{x}) = -bx + ln(a)$	$Y = ln(\dfrac{y}{x}), X = x$ $a = e^c, b = -m$

Latihan

Diberikan titik-titik data sebagai berikut

x	1	2	3	4	5
y	0,6	0,9	4,3	7,6	12,6

- Cocokkan titik-titik data tersebut dengan fungsi $f(x) = ae^{bx}$ dan fungsi $f(x) = ax^b$
- Hitung deviasi $y_i - f(x_i)$, lalu hitung galat RMS masing-masing. Yang mana fungsi hampiran yang lebih baik?