

Pertemuan IV

PENCOCOKAN KURVA: ANALISA REGRESI

Materi Minggu Ini

- Pencocokan Kurva ▶
- Regresi Kuadrat Terkecil (RKT) ▶
 - RKT untuk Kurva Linier ▶
 - RKT untuk Kurva Non-Linier ▶
- Regresi Polynomial ▶
- Tugas IV ▶

Pencocokan Kurva (1)

Seringkali data tersajikan dalam bentuk rangkaian nilai diskrit (deretan angka² dalam urutan yang kontinu), tanpa disertai bentuk fungsi yang menghasilkan data tsb.

Dalam kasus di atas, kita dapat men-“generate” fungsi sederhana untuk mengaproksimasi bentuk fungsi sebenarnya dengan memanfaatkan rangkaian data yang ada.

Mengapa bentuk fungsi begitu penting bagi kita?...

Selain untuk memenuhi kebutuhan proses numeris, seperti integrasi atau mendapatkan solusi pendekatan dari persamaan differensial, seringkali kita harus menganalisa tren atau melakukan pengujian hipotesa terhadap nilai² diskrit yang dihasilkan oleh fungsi tsb.

Pencocokan Kurva (2)

Terhadap keberadaan rangkaian/pasangan data yang tidak diketahui fungsi asalnya, terdapat 2 hal yang dapat dilakukan :

Pertama, berusaha mencari bentuk kurva (fungsi) yang mewakili rangkaian data diskrit tersebut.

Kedua, berusaha meng-estimasi/memperkirakan nilai fungsi pada titik² tertentu yang belum diketahui.

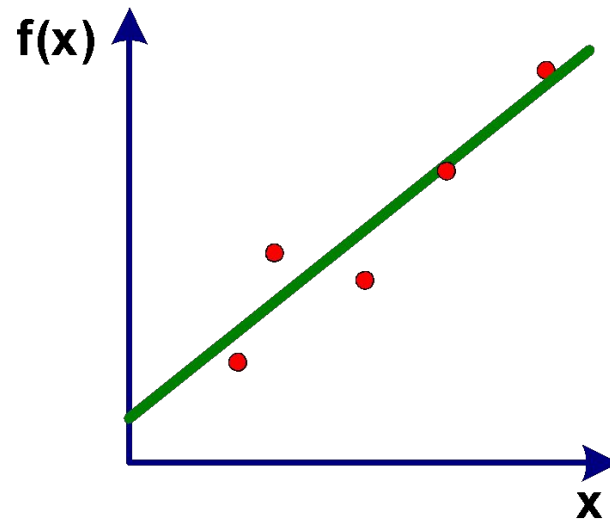
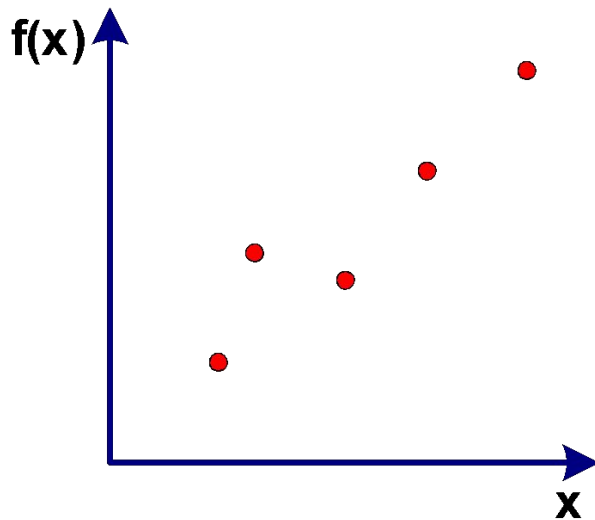
Keduanya dikenal sebagai teknik Pencocokan Kurva (curve fitting).

Pencocokan Kurva (3)

Pendekatan² yang lazim digunakan untuk melakukan Pencocokan Kurva antara lain adalah :

Regresi Kuadrat Terkecil (*least-square regresion*)

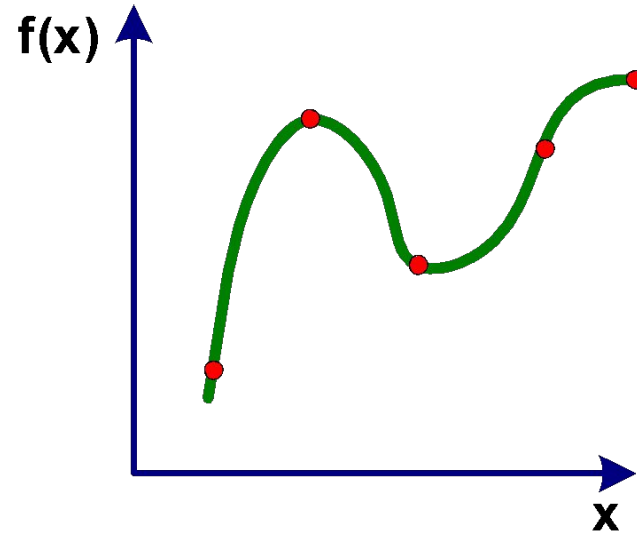
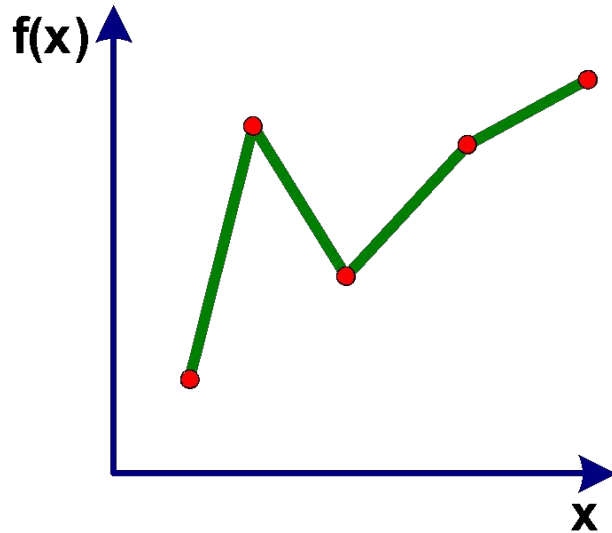
Metode ini digunakan apabila data yang tersaji memiliki tingkat kesalahan berarti (akurasi) rendah. Anda hanya perlu membuat sebuah garis lurus yang merepresentasikan tren dari data² tsb secara umum.



Pencocokan Kurva (3)

Interpolasi

Jika tingkat ketelitian data yang kita miliki lebih baik, maka metode *Interpolasi* dapat dipakai. Kita gunakan segmen² garis lurus untuk menghubungkan titik² data (interpolasi linier). Atau, dengan menggunakan kurva (interpolasi polynomial) untuk memperoleh hasil yang lebih baik.



Analisis regresi menggunakan sedikit notasi dan perhitungan statistik.
Ini artinya, ada sedikit yang perlu anda ingat kembali... ▶

Sedikit Notasi Statistik



Tahun (x_i)	Debit Air = y_i (m^3/det)
1991	8,52
1992	3,33
1993	7,85
1994	7,65
1995	10,91
1996	4,17
1997	3,40
1998	8,00
1999	13,4
2000	5,40
2001	8,87
2002	4,73
2003	7,40
2004	6,88
2005	5,00

Rerata/rata-rata data (\bar{y}) adalah jumlah nilai data (Σy) dibagi jumlah data (n) :

$$\bar{y} = \frac{\Sigma y_i}{n}$$

Deviasi Standar (σ) atau penyebaran nilai data adalah :

$$\sigma = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n - 1}}$$

Atau dapat pula dinyatakan dalam bentuk Varians (σ^2) :

$$\sigma^2 = \frac{\Sigma (y_i - \bar{y})^2}{n - 1}$$

Regresi Kuadrat Terkecil (1)

Jika data yang tersaji memiliki tingkat kesalahan yang cukup signifikan, maka penggunaan interpolasi bukanlah pilihan yang bijaksana. Karena (kemungkinan besar) hasil pendekatannya akan kurang memuaskan.

Ada cara yang lebih mudah, yaitu dengan membuat kurva yang dapat mewakili titik² data tersebut.

Katakan kurva ini adalah kurva dari fungsi $g(x)$ yang 'mirip' dengan fungsi sebenarnya.

Tetapi jika penyebaran titik datanya sangat besar? ...

Rasanya koq sulit cara di atas bisa berhasil dengan baik 😞

Kecuali jika, ... kita dapat membuat kurva buatan, $g(x)$, yang mampu meminimalkan perbedaan (selisih) dengan kurva aslinya, $f(x)$.

Teknik ini yang disebut dengan **metode Regresi Kuadrat Terkecil**.

Regresi Kuadrat Terkecil (2)

Secara umum prosedur untuk mengaplikasikan metode Regresi Kuadrat Terkecil (RKT) ini adalah sbb :

1. Titik² data diplot ke dalam koordinat cartesian. Dari pola datanya bisa dilihat, apakah kurva pendekatan yang akan kita buat berbentuk garis lurus atau garis lengkung;
2. Tentukanlah sebuah fungsi $g(x)$ yang dpt mewakili fungsi titik² data $f(x)$:
$$g(x) = a_0 + a_1x + a_2x^2 + \dots + a_rx^r$$
3. Jika a_0, a_1, \dots, a_r adalah parameter fungsi $g(x)$, maka tentukan nilai parameter² tsb sdmk hingga $g(x)$ dpt mendekati titik² data;
4. Jika koordinat titik² data adalah $M(x_i, y_i)$, maka selisih dengan fungsi $g(x)$ adalah :

$$\begin{aligned} {}_iE &= M_i - G_i \\ &= y_i - g(x_i; a_0, a_1, \dots, a_r) \\ &= y_i - (a_0 + a_1x_i + a_2x_i^2 + a_3x_i^3 + \dots + a_rx_i^r) \end{aligned}$$

Regresi Kuadrat Terkecil (3)



5. Tingkat kesalahan fungsi $g(x)$ diukur melalui rumus berikut :

$$D^2 = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n \{ y_i - g(x_i) \}^2$$

6. Cari nilai parameter a_0, a_1, \dots, a_r sdmk hingga D^2 dapat semimum mungkin. D^2 bernilai minimum jika turunan pertamanya terhadap $a_0, a_1, \dots, a_r = 0$.

$$\frac{\partial D^2}{\partial a_0} = 0 \quad \frac{\partial D^2}{\partial a_1} = 0 \quad \dots \quad \frac{\partial D^2}{\partial a_r} = 0$$

7. Persamaan no 6 di atas akan memberikan nilai parameter a_0, a_1, \dots, a_r . Sehingga persamaan kurva yang mewakili titik² data dapat diperoleh.

RKT untuk Kurva Linier (1)

Permasalahan kita akan menjadi sederhana apabila kurva (yang diwakili oleh titik² data) berbentuk garis lurus.

Seperti anda ketahui, bahwa persamaan umum sebuah garis lurus adalah :

$$g(x) = a + bx$$

Dan jumlah-kuadrat-kesalahan dapat dihitung melalui persamaan :

$$D^2 = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n \{ y_i - a - bx_i \}^2$$

dimana

$$a = y - b x$$

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

RKT untuk Kurva Linier (5)

Persamaan (4) dan (6) dapat dimanfaatkan untuk mendapatkan koefisien a dan b, sehingga fungsi $g(x)$ dapat diperoleh.

Sedangkan untuk mengetahui derajat kesesuaian dari persamaan yang dicari, dapat dihitung melalui koefisien korelasi yang berbentuk :

$$r^2 = \frac{D_+^2 - D^2}{D_+^2} \quad \dots (7)$$

dengan,

r = koefisien korelasi

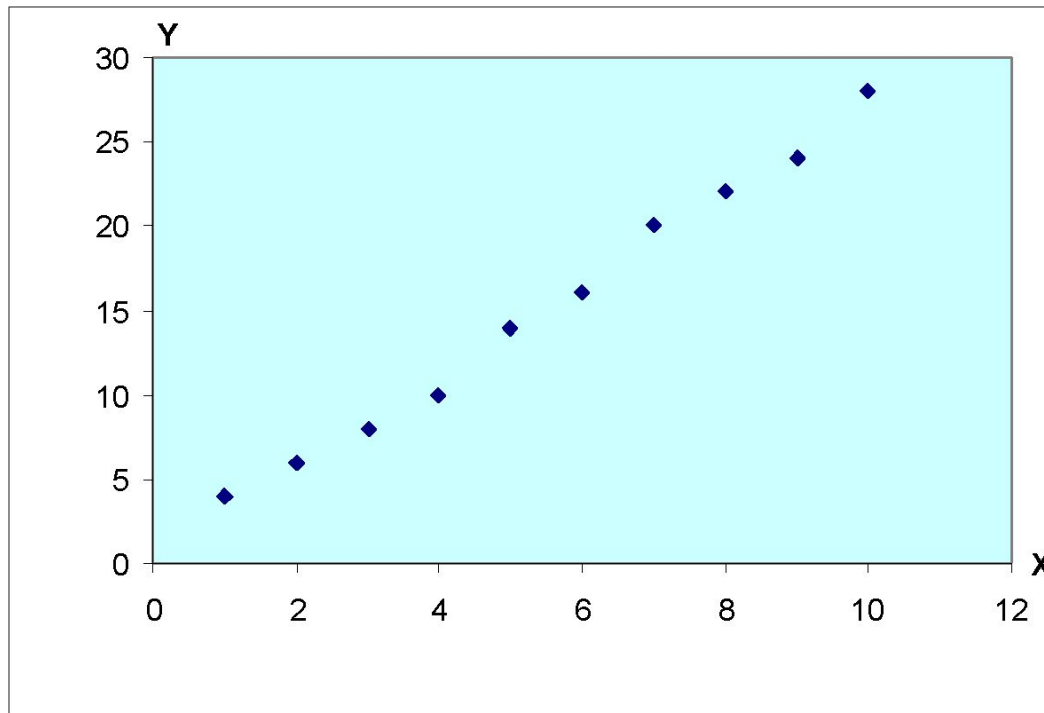
$$D_+^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \quad D^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Jika $r = 1$, maka fungsi $g(x)$ memiliki tingkat kesesuaian yang tinggi dengan persamaan aslinya ($f(x)$). Atau $r = 0$ jika sebaliknya.

RKT untuk Kurva Linier (6)

contoh : tentukan persamaan garis yang mewakili data berikut :

x	1	2	3	4	5	6	7	8	9	10
y	4	6	8	10	14	16	20	22	24	28



No	x_i	y_i	$x_i \cdot y_i$	x_i^2
1	1	4	4	1
2	2	6	12	4
3	3	8	24	9
4	4	10	40	16
5	5	14	70	25
6	6	16	96	36
7	7	20	140	49
8	8	22	176	64
9	9	24	216	81
10	10	28	280	100
Σ	55	152	1058	385

RKT untuk Kurva Linier (7)



Nilai rerata untuk x dan y adalah :

$$\bar{x} = \sum x / n = 55/10 = 5,5$$

$$\bar{y} = \sum y / n = 152/10 = 15,2$$

Jika persamaan umum garis dinyatakan sebagai : $y = a + bx$,
dan,

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{10 \cdot 1058 - 55 \cdot 152}{10 \cdot 385 - (55)^2} = \frac{2220}{825} = 2,690909$$

$$a = \bar{y} - b\bar{x} = 15,2 + 2,690909 \cdot 5,5 = 30$$

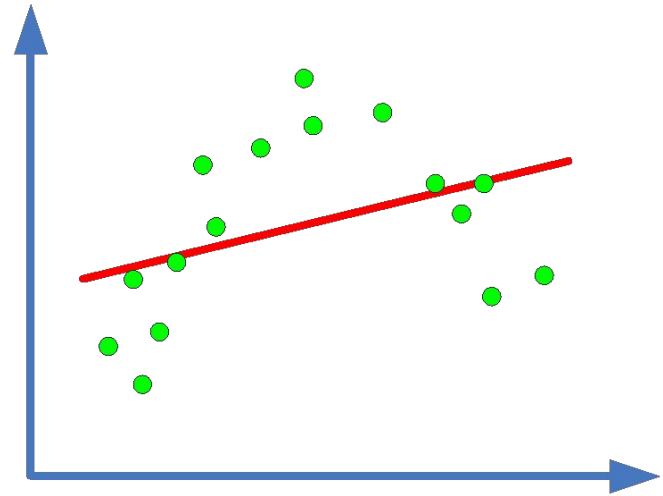
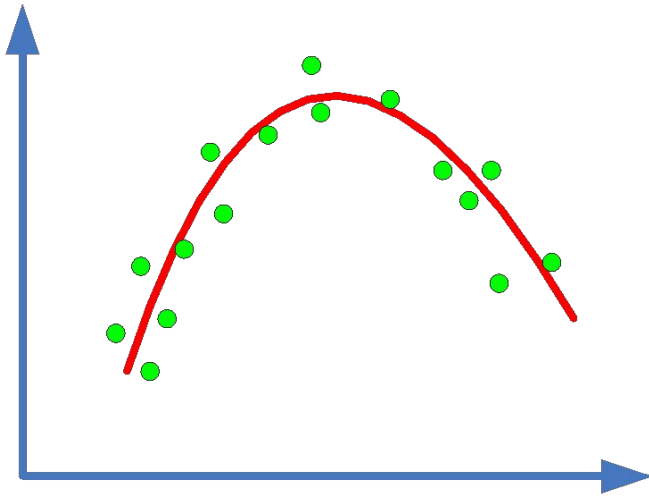
Jadi persamaan garis yang mendekati rangkaian data tersebut adalah :

$$y = 30 + 2,69x$$

RKT untuk Kurva Non-Linier (1)

Di dalam praktek akan sering kita jumpai kasus dimana plotting titik² data memiliki tren berupa kurva lengkung.

Sehingga pendekatan melalui RKT utk kurva linier menjadi kurang optimal/sesuai untuk digunakan.



Karena kurva lengkung yang didekati dengan sebuah garis lurus tentu akan menimbulkan kesalahan yang cukup berarti.

RKT untuk Kurva Non-Linier (2)

Kecuali untuk beberapa bentuk fungsi yang memang dapat didekati dengan metode Linierisasi Kurva Non-Linier. Fungsi² tersebut antara lain :

1. Fungsi Eksponensial

$y = a e^{bx}$ dengan a dan b adalah konstanta

persamaan di atas dapat dilinierkan dengan logaritma-natural spt berikut :

$$\ln y = \ln a + b x \ln e$$

jika $\ln e = 1$, maka $\ln y = \ln a + b x$

persamaan di atas berbentuk garis lurus dengan kemiringan b , dan memotong sumbu $\ln y$ di $\ln a$.

2. Fungsi Berpangkat

fungsi berpangkat adalah contoh lain fungsi dengan kurvanya yang non-linier.

$$y = a x^b \quad \text{dengan } a \text{ dan } b \text{ adalah konstanta}$$

me-linier-kan fungsi di atas juga dapat dilakukan menggunakan persamaan logaritmik spt berikut :

$$\log y = b \log x + \log a$$

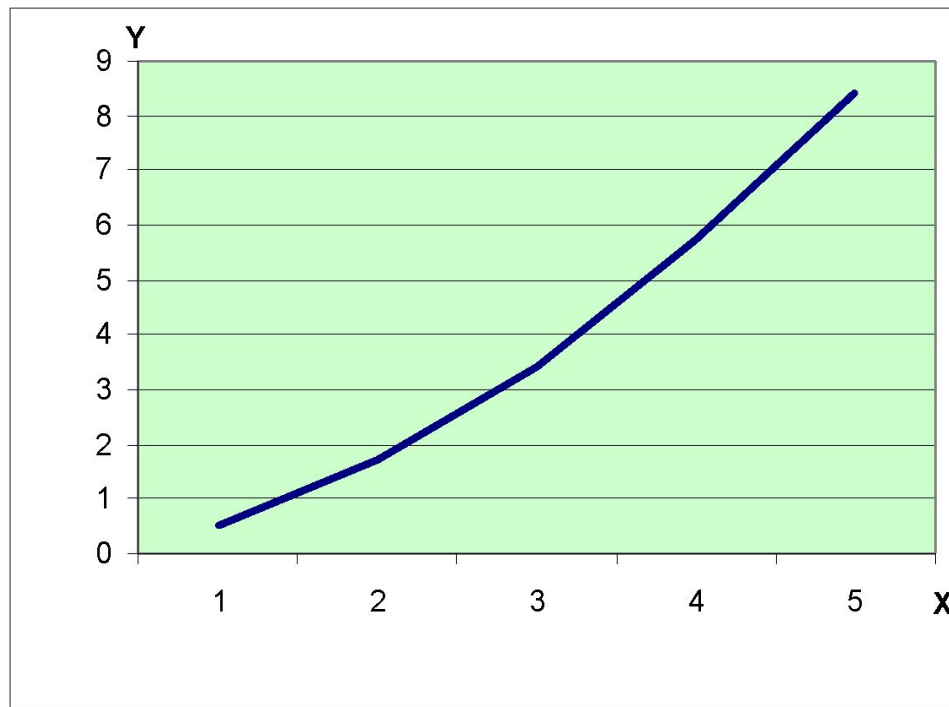
persamaan di atas berbentuk garis lurus dengan kemiringan b dan memotong sumbu $\log y$ di $\log a$.

RKT untuk Kurva Non-Linier (4)

Contoh : Tentukan persamaan kurva lengkung yang diwakili serangkaian data berikut :

x	1	2	3	4	5
y	0,5	1,7	3,4	5,7	8,4

Penyelesaian masalah di atas dilakukan melalui 2 *fashion*, transformasi log dan ln.



RKT untuk Kurva Non-Linier (5)

Transformasi log (fungsi asli diamsusikan sebagai fungsi berpangkat)

Misal persamaan yang dicari adalah : $y = ax^b$

Persamaan tersebut dapat dilinierisasi melalui fungsi logaritmik sbb : $\log y = \log ax^b$

atau dapat pula dinyatakan sebagai : $\log y = b \log x + \log a$

Sehingga jika dimisalkan $q = \log x$, $p = \log y$, $A = \log a$, $B = b$, akan diperoleh tabel :

No	x_i	y_i	q_i	P_i	$q_i \cdot p_i$	$(q_i)^2$
1	1	0,5	0	- 0,301	0	0
2	2	1,7	0,3010	0,2304	0,0693	0,0906
3	3	3,4	0,4771	0,5315	0,2536	0,2276
4	4	5,7	0,6020	0,7559	0,4550	0,3624
5	5	8,4	0,6990	0,9243	0,6461	0,4886
Σ		19,7	2,0791	2,1411	1,4240	1,1692

RKT untuk Kurva Non-Linier (6)

Dari tabel tersebut dapat diperoleh beberapa parameter penting, seperti :

$$\bar{q} = \sum \log x_i / n = 2,0791 / 5 = 0,4158$$

$$\bar{p} = \sum \log y_i / n = 2,1411 / 5 = 0,42822$$

Sedangkan koefisien A dan B dihitung melalui persamaan (4) dan (6) :

$$B = \frac{n \sum q_i p_i - \sum q_i \sum p_i}{n \sum q_i^2 - (\sum q_i)^2} = \frac{5 (1,4240) - 2,0791 (2,1411)}{5 (1,1692) - 2,0791 (2,0791)} = \frac{2,6684}{1,5233} = 1,7572$$

$$A = \bar{p} - B \bar{q} = 0,42822 - 1,7572 \cdot 0,4158 = - 0,3024$$

Karena $A = \log a$ maka $a = 0,4984$

Karena $B = b$ maka $b = 1,7572$

Dengan demikian fungsi yang dicari adalah : $y = 0,4984 x^{1,7572}$

RKT untuk Kurva Non-Linier (7)

Transformasi ln (fungsi asli diasumsikan sebagai fungsi eksponensial)

Misal persamaan yang dicari adalah : $y = a e^{bx}$

atau dapat pula dinyatakan sebagai : $\ln y = \ln a + bx$

Sehingga jika dimisalkan $q = x$, $p = \ln y$, $A = \ln a$, $B = b$, maka dpt diperoleh tabel :

No	x_i	y_i	q_i^2	p_i	$q_i \cdot p_i$
1	1	0,5	1	- 0,6931	- 0,6931
2	2	1,7	4	0,5306	1,0612
3	3	3,4	9	1,2238	3,6714
4	4	5,7	16	1,7405	6,962
5	5	8,4	25	2,1282	10,641
Σ	15	19,7	55	4,93	21,6425

RKT untuk Kurva Non-Linier (8)

Dari tabel tersebut dapat diperoleh beberapa parameter penting, seperti :

$$\bar{q} = \sum x_i / n = 15 / 5 = 3$$

$$\bar{p} = \sum \ln y_i / n = 2,0791 / 5 = 0,4158$$

Sedangkan koefisien A dan B dihitung melalui persamaan (4) dan (6) :

$$B = \frac{n \sum q_i p_i - \sum q_i \sum p_i}{n \sum q_i^2 - (\sum q_i)^2} = \frac{5 (21,6425) - 15 (4,93)}{5 (55) - (15)^2} = \frac{34,2625}{50} = 0,6852$$

$$A = \bar{p} - B \bar{q} = 0,4158 - 0,68525 \cdot 3 = - 1,63995$$

Karena $A = \ln a$ maka $a = 0,34447$

Karena $B = b$ maka $b = 0,68525$

Dengan demikian fungsi yang dicari adalah : $y = 0,34447 e^{0,68525x}$

RKT untuk Kurva Non-Linier (9)

Sekarang waktunya memilih. Mana di antara 2 pendekatan yang memberikan akurasi lebih bagus. Caranya adalah dengan menghitung koefisien korelasi (7) :

$$r^2 = \frac{D_+^2 - D^2}{D_+^2}$$

dengan

$$D_+^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\sum_{i=1}^n$$

$$D^2 = \sum_{i=1}^n (y_i - ax^b)^2$$

$$D^2 = \sum (y_i - ae^{bx})^2$$

No	x_i	y_i	Transformasi log			Transformasi ln		
			$g(x_i)$	D^2	D_t^2	$g(x_i)$	D^2	D_t^2
1	1	0,5	0,4984	0,000003	11,8336	0,6835	0,03367	11,8336
2	2	1,7	1,6848	0,000231	5,0176	1,3563	0,11813	5,0176
3	3	3,4	3,4354	0,00125	0,2916	2,6912	0,50240	0,2916
4	4	5,7	5,6953	0,000022	3,0976	5,3401	0,12953	3,0976
5	5	8,4	8,4296	0,000876	19,8916	10,5963	4,82373	19,8916
Σ	15	19,7		0,00238	40,132		5,60746	40,132

RKT untuk Kurva Non-Linier (9)

Dari tabel tersebut dapat dicari nilai r untuk transformasi log :

$$r = \left[\frac{D_+^2 - D^2}{D_+^2} \right]^{\frac{1}{2}} = \left[\frac{40,132 - 0,00238}{40,132} \right]^{\frac{1}{2}} = 0,99997$$

Sedangkan untuk transformasi ln :

$$r = \left[\frac{D_+^2 - D^2}{D_+^2} \right]^{\frac{1}{2}} = \left[\frac{40,132 - 5,60746}{40,132} \right]^{\frac{1}{2}} = 0,92751$$

Dapat dilihat bahwa koefisien korelasi r untuk transformasi log lebih mendekati nilai 1 dibanding transformasi ln. Sehingga bisa disimpulkan bahwa **transformasi log memberikan pendekatan yang lebih baik.**

Regresi Polynomial (1)

Persamaan garis lurus dapat didekati dg **Metode Kuadrat Terkecil**.

Sementara untuk kurva lengkung, pendekatan yang cukup logis adalah melalui transformasi log atau ln.

Atau, untuk kurva lengkung dapat didekati dengan **Regresi Polynomial**.

Persamaan polynomial orde-r mempunyai bentuk :

$$y = a_0 + a_1x + a_2x^2 + \dots + a_rx^r$$

Jumlah kuadrat kesalahan dr pendekatan polynom di atas adalah :

$$D^2 = \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_rx_i^r)^2$$

Regresi Polynomial (2)

Jika didiferensiasi terhadap setiap koefisiennya, maka persamaan (8) akan menjadi :

$$\frac{\partial D^2}{\partial a_0} = -2 \sum_{i=1}^n (a_0 - a_1 x - a_2 x^2 - \dots - a_r x^r)$$

$$\frac{\partial D^2}{\partial a_1} = -2 \sum_{i=1}^n x_i (a_0 - a_1 x - a_2 x^2 - \dots - a_r x^r)$$

$$\frac{\partial D^2}{\partial a_2} = -2 \sum_{i=1}^n x_i^2 (a_0 - a_1 x - a_2 x^2 - \dots - a_r x^r)$$

⋮

$$\frac{\partial D^2}{\partial a_r} = -2 \sum_{i=1}^n x_i^r (a_0 - a_1 x - a_2 x^2 - \dots - a_r x^r)$$

Bentuk persamaan²nya menyerupai sistem persamaan

Regresi Polynomial (3)

$$\begin{pmatrix} n & \sum x & \sum x_i^2 & \dots & \sum x_i^r \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{r+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \dots & \sum x_i^{r+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^r & \sum x_i^{r+1} & \sum x_i^{r+2} & \dots & \sum x_i^{r+r} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^r y_i \end{pmatrix}$$

Berarti kita bisa menyatakan himpunan persamaan turunan tersebut menjadi persamaan matriks $AX = B$. Dan selanjutnya kita dapat menggunakan metode eliminasi Gauss, Gauss-Jordan, dll untuk mencari nilai a_0, a_1, \dots, a_r .

Regresi Polynomial (4)

contoh : carilah persamaan kurva polynomial orde-2 yang mewakili data berikut :

x	0	1	2	3	4	5
y	2,1	7,7	13,6	27,2	40,9	61,1

Persamaan polynomial orde-2 memiliki bentuk :

$$g(x) = a_0 + a_1x + a_2x^2$$

Diferensialisasi D^2 terhadap setiap koefisiennya akan menghasilkan bentuk :

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix}$$

Regresi Polynomial (5)



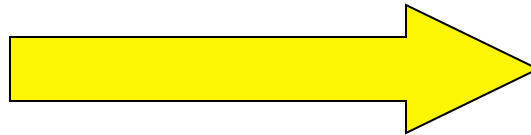
n	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1	0	2,1	0	0	0	0	0
2	1	7,7	1	1	1	7,7	7,7
3	2	13,6	4	8	16	27,2	54,4
4	3	27,2	9	27	81	81,6	244,8
5	4	40,9	16	64	256	163,6	654,4
6	5	61,1	25	75	625	305,5	1527,5
Σ	15	397,4	55	175	979	585,6	2488,8

Dengan menggunakan data tabel, maka diperoleh :

$$6a_0 + 15a_1 + 55a_2 = 397,4$$

$$15a_0 + 55a_1 + 175a_2 = 585,6$$

$$55a_0 + 175a_1 + 979a_2 = 2488,8$$



$$a_2 = 0,1922732$$

$$a_1 = -33,20882$$

$$a_0 = 147,49288$$

Jadi persamaan kurva yang dicari adalah :

$$y = 147,49288 - 33,20882x + 0,1922732x^2$$

Latihan (1)

1. Tentukan: (a) rerata; (b) deviasi standar; dan (c) varian; dari data-data berikut

0,95 1,42 1,54 1,55 1,63
1,32 1,15 1,47 1,95 1,25
1,46 1,47 1,92 1,35 1,05
1,85 1,74 1,65 1,78 1,71
2,39 1,82 2,06 2,14 2,27

2. Gunakan regresi kuadrat terkecil untuk menaksir fungsi garis lurus dari data berikut :

x	1	3	5	7	10	12	13	16	18	20
y	3	2	6	5	8	7	10	9	12	10

3. Gunakan regresi kuadrat terkecil untuk menaksir fungsi garis lurus dari data berikut :

x	4	6	8	10	14	16	20	22	24	28	28	34	36	38
y	30	18	22	28	14	22	16	8	20	8	14	14	0	8

4. Gunakan regresi kuadrat terkecil untuk menaksir fungsi kurva dari data berikut :

x	1	2	2,5	4	6	8	8,5
y	0,4	0,7	0,8	1,0	1,2	1,3	1,4

Latihan (2)

5. Gunakan regresi kuadrat terkecil untuk menaksir fungsi kurva dari data berikut :

x	2,5	3,5	5	6	7,5	10	12,5	15	17,5	20
y	5	3,4	2	1,6	1,2	0,8	0,6	0,4	0,3	0,3

6. Gunakan regresi polynomial untuk menaksir fungsi kurva dari data berikut :

x	0,05	0,4	0,8	1,2	1,6	2,0	2,4
y	550	750	1.000	1.400	2.000	2.700	3.750

7. Gunakan regresi polynomial untuk menaksir fungsi kurva dari data berikut :

x	0	2	4	6	9	11	13	15	17	19	23	25	28
y	1,2	0,6	0,4	-0,2	0	-0,6	-0,4	-0,2	-0,4	0,2	0,4	1,2	1,8