Koefisien Korelasi Karl Pearson

Analisis korelasi → untuk mengetahui kekuatan relasi linier yang diperoleh dari model regresi

Kekuatan linier ini diukur dengan koefisien korelasi r :

$$r = \frac{N\left(\sum_{i=1}^{N} X_i Y_i\right) - \left(\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i\right)}{\sqrt{\left[N\sum_{i=1}^{N} X_i^{2} - \left(\sum_{i=1}^{N} X_i\right)^2\right]\left[N\sum_{i=1}^{N} Y_i^2 - \left(\sum_{i=1}^{N} Y_i\right)^2\right]}}$$

$$-1 \leq r \leq 1$$

contoh

X_1	X_2	Y	X_1X_2	X_1Y	X_2Y	X_1^2	X_2^2	Y^2
5	5	6	25	30	30	25	25	36
4	5	5	20	20	25	16	25	25
7	8	8	56	56	64	49	64	64
6	6	6	36	36	36	36	36	36
4	5	5	20	20	25	16	25	25
6	5	6	30	36	30	36	25	36
7	5	6	35	42	30	49	25	36
5	4	5	20	25	20	25	16	25
6	7	7	42	42	49	36	49	49
8	7	8	56	64	56	64	49	64
6	5	6	30	36	30	36	25	36
4	7	5	28	20	35	16	49	25
68	69	73	398	427	430	404	413	457

$$r_{x_1x_2} = \frac{n\left(\sum x_1x_2\right) - \left(\sum x_1\right)\left(\sum x_2\right)}{\sqrt{\left(n\sum x_1^2 - \left(\sum x_1\right)^2\right)\left(n\sum x_2^2 - \left(\sum x_2\right)^2\right)}}$$

$$= \frac{12(398) - (68.69)}{\sqrt{\left(12.404 - (68)^2\right)\left(12.413 - (69)^2\right)}}$$

$$= 0.401918$$

Uji sig. Koef. Korelasi
misal : reg.Sederhana ant X1 dg y

$$i.\quad H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

$$ii.\quad \alpha = 0.05$$

$$r_{x_1y} = \frac{n\left(\sum x_1 y\right) - \left(\sum x_1\right)\left(\sum y\right)}{\sqrt{\left(n\sum x_1^2 - \left(\sum x_1\right)^2\left(n\sum y^2 - \left(\sum y\right)^2\right)\right)}}$$

$$= \frac{12(427) - (68.73)}{\sqrt{\left(12.404 - (68)^2\right)\left(12.457 - (73)^2\right)}} = 0.858678$$

X_1	X_2	Y	X_1X_2	X_1Y	X_2Y	X_1^2	X_2^2	Y^2
5	5	6	25	30	30	25	25	36
4	5	5	20	20	25	16	25	25
7	8	8	56	56	64	49	64	64
6	6	6	36	36	36	36	36	36
4	5	5	20	20	25	16	25	25
6	5	6	30	36	30	36	25	36
7	5	6	35	42	30	49	25	36
5	4	5	20	25	20	25	16	25
6	7	7	42	42	49	36	49	49
8	7	8	56	64	56	64	49	64
6	5	6	30	36	30	36	25	36
4	7	5	28	20	35	16	49	25
68	69	73	398	427	430	404	413	457

$$t = \frac{r_{x_1y}\sqrt{n-2}}{\sqrt{1-r_{x_1y}^2}}$$

$$= \frac{0.858678\sqrt{12 - 2}}{\sqrt{1-(0.858678)^2}}$$

$$= 5.298129$$

Bandingkan dengan t tabel $\mathrm{t}(\alpha, \mathrm{n} - 2) = \mathrm{t}(0.05, 10) = 1.812$

Karena t>1.812 maka Ho ditolak, jadi ada korelasi positif antara X1 dengan y

Contoh X1, X2 dan y

Akan dibentuk model Regresi

→ Secara komputerisasi :

		B
1	(Constant)	.639
	X1	.570
	X2	.385

a. Dependent Variable: Y

1. Bentuk model

$$\hat{y} = 0.639 + 0.570X_1 + 0.2385X_2$$

2. misal dengan uji, diketahui X1 dan X2 linier terhadap Y

3. Uji F

Diperuntukkan guna melakukan uji hipotesis koefisien (slop) regresi secara bersamaan.

i. H_0 : $\beta_2 = \beta_3 = \beta_4 = \ldots\ldots\ldots\ldots = \beta_k = 0$

H_1 : Tidak demikian (paling tidak ada satu slop yang $\neq 0$)

Dimana: k adalah banyaknya variabel bebas.

ii. Tingkat signifikansi 5%

iii. Tabel ANOVA

Tabel ANOVA

Sumber	JK	df	RK	F Hitung
Regresi	JKR	k	RKR = JKR/k	$F = \frac{RKR}{RKS}$
Sesatan	JKS	n-k-1	RKS= JKS/(n-k-1)	
Total	JKT	n-1		

Bandingkan F_{Hit} dengan $F_{\alpha(k,n-k-1)}$

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11.547	2	5.773	37.929	.000[a]
	Residual	1.370	9	.152		
	Total	12.917	11			

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Ho ditolak karena F=37.929>F(2,9,0.05=4.26)

Artinya H1 diterima, dkl hubungan antara X1, X2 dengan Y berarti

- **Standard Error (Kesalahan Baku)**

Prinsip OLS: meminimalkan error. Oleh karena itu, ketepatan dari nilai dugaan sangat ditentukan oleh *standard error* dari masing-masing penduga. Adapun standard error dirumuskan sebagai berikut:

$$Se = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-2}} = \sqrt{\frac{SST-SSR}{n-2}} = \sqrt{\frac{\sum Y^2 - b\sum XY}{n-2}} = \sqrt{RK_S}$$

$$\text{cth. } S_e = \sqrt{RK_S} = \sqrt{0.152} = 0.390$$

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945[a]	.894	.870	.390

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Standard error coef

$$2\,\text{var}, s_{bi} = \sqrt{\frac{Se}{\left(\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right)\left(1 - r_{12}^2\right)}}$$

$$s_{b1} = \sqrt{\frac{0.390}{\left(404 - \frac{(68)^2}{12}\right)\left(1 - 0.402^2\right)}} = 0.099$$

dengan cara yang sama, diperoleh

$$s_{b2} = 0.106$$

Model		Unstandardized Coefficients	
		B	Std. Error
1	(Constant)	.639	.649
	X1	.570	.099
	X2	.385	.106

a. Dependent Variable: Y

4. Uji t

- Pengujian koefisien regresi secara individu.

  i. H_{obj} : $\beta_j = 0$

  H_{1bj} : $\beta_j \neq 0$; j = 0, 1, 2........, k k adalah koefisien slop.
- Ii. α=5%
- Iii.

$$t_1 = \frac{0.570}{0.099} = 5.758 \Rightarrow 5.758 > t_{\frac{0.05}{2},9} = 2.262 \Rightarrow H_{0b1}\text{ditolak}$$

$$t_2 = \frac{0.385}{0.106} = 3.632 \Rightarrow 3.632 > t_{\frac{0.05}{2},9} = 2.262 \Rightarrow H_{0b2}\text{ditolak}$$

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.639	.649		.984	.351
	X1	.570	.099	.685	5.778	.000
	X2	.385	.106	.432	3.646	.005

Asumsi-asumsi dasar OLS

Pendugaan OLS akan bersifat BLUE (Best Linier Unbiased Estimate) jika memenuhi 3 asumsi utama, yaitu:

- Tidak ada multikolinieritas
- Tidak mengandung Heteroskedastisitas
- Bebas dari autokorelasi

Multikolinieritas

- Multikolinieritas: adanya hubungan linier antara regressor.

  Misalkan terdapat dua buah regressor, X_1 dan X_2. Jika X_1 dapat dinyatakan sebagai fungsi linier dari X_2, misal : $X_1 = \gamma X_2$, maka ada kolinieritas antara X_1 dan X_2. Akan tetapi, bila hubungan antara X_1 dan X_2 tidak linier, misalnya $X_1 = X2^2$ atau $X_1 = \log X_2$, maka X_1 dan X_2 tidak kolinier.

Contoh Data Perfect Multikolinieritas

X_1	X_2	X_3
12	48	51
16	64	65
19	76	82
23	92	96
29	116	118

Nilai-nilai yang tertera dalam tabel menunjukan bahwa Antara $X1$ dan $X2$ mempunyai hubungan: $X2 = 4X1$. Hubungan seperti inilah yang disebut dengan *perfect multicollinearity*.

Akibat Multikolinieritas

- Varians besar (dari taksiran OLS)
- Interval kepercayaan lebar (variansi besar $\Rightarrow$ Standar Error besar $\Rightarrow$ Interval kepercayaan lebar)
- R^2 tinggi tetapi tidak banyak variabel yang signifikan dari uji t.
- Terkadang taksiran koefisien yang didapat akan mempunyai nilai yang tidak sesuai dengan substansi, sehingga dapat menyesatkan interpretasi.

Contoh Ilustrasi

Konsumsi (Y)	Pendapatan (X_1)	Kekayaan (X_2)
40	50	500
50	65	659
65	80	856
90	110	1136
85	100	1023
100	120	1234
110	140	1456
135	190	1954
140	210	2129
160	220	2267

Ilustrasi

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.800	4.696		2.726	.030
	Pendapatan_X1	-1.414	1.199	-2.152	-1.179	.277
	Kekayaan_X2	.202	.117	3.141	1.721	.129

a. Dependent Variable: Konsumsi_Y

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13864.440	2	6932.220	195.620	.000[a]
	Residual	248.060	7	35.437		
	Total	14112.500	9			

a. Predictors: (Constant), Kekayaan_X2, Pendapatan_X1

b. Dependent Variable: Konsumsi_Y

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.991[a]	.982	.977	5.953	1.028

a. Predictors: (Constant), Kekayaan_X2, Pendapatan_X1

b. Dependent Variable: Konsumsi_Y

- Model:

$$Y = 12{,}8 - 1{,}414X_1 + 0{,}202\ X_2$$

SE (4,696) (1,199) (0,117)

t (2,726) (-1,179) (1,721)

$$R^2 = 0{,}982$$

- R^2 relatif tinggi, yaitu 98,2%. Artinya?
- Uji t tidak signifikan. Artinya?
- Koefisien X1 bertanda negatif. Artinya?

Ilustrasi: misal model dipecah

- **Dampak Pendapatan pada Konsumsi**

$$Y = 14{,}148 + 0{,}649X_1$$
$$\text{SE} \quad (5{,}166) \quad (0{,}037)$$
$$t \quad (2{,}739) \quad (17{,}659)$$
$$R^2 = 0{,}975$$

  R^2 tinggi, Uji t signifikan, dan tanda X_1 positif.

- **Dampak Kekayaan pada Konsumsi**

$$Y = 13{,}587 + 0{,}0635X_2$$
$$\text{SE} \quad (4{,}760) \quad (0{,}003)$$
$$t \quad (2{,}854) \quad (19{,}280)$$
$$R^2 = 0{,}979$$

- R^2 tinggi, Uji t signifikan, dan tanda X_2 positif.

- X_1 **dan** X_2 **menerangkan variasi yang sama. Bila 1 variabel saja cukup, kenapa harus dua?**

Mendeteksi Multikolinieritas dengan Uji Formal

1. Eigenvalues dan Conditional Index

- Aturan yang digunakan adalah: **Multikolinieritas ditengarai ada didalam persamaan regresi bila nilai Eigenvalues mendekati 0**.
- Hubungan antara Eigenvalues dan Conditional Index (CI) adalah sebagai berikut:

$$CI = \sqrt{\frac{\max \ eigenvalues}{\min \ eigenvalues}}$$

Jika CI berada antara nilai 10 sampai 30: kolinieritas moderat.

Bila CI mempunyai nilai diatas 30: kolinieritas yang kuat.

2. VIF dan Tolerance

Nilai tolerance di atas 0.1 dan VIF di bawah 10 mengindikasikan tidak ada multikolinearitas antar variabel bebas

$$\mathrm{VIF}_j = \frac{1}{(1 - R_j^2)} \qquad ; j = 1,2,\ldots\ldots,k$$

R_j^2 adalah koefisien determinasi antara variabel bebas ke-j dengan variabel bebas lainnya.

Jika $R_j^2 = 0$ atau antar variabel bebas tidak berkorelasi, maka nilai VIF = 1.

Jika $R_j^2 \neq 0$ atau ada korelasi antar variabel bebas, maka nilai VIF > 1.

Oleh karena itu, dapat disimpulkan bahwa kolinieritas tidak ada jika nilai VIF mendekati angka 1

Cth sebelumnya

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	12.800	4.696		2.726	.030		
	Pendapatan_X1	-1.414	1.199	-2.152	-1.179	.277	.001	1326.670
	Kekayaan_X2	.202	.117	3.141	1.721	.129	.001	1326.670

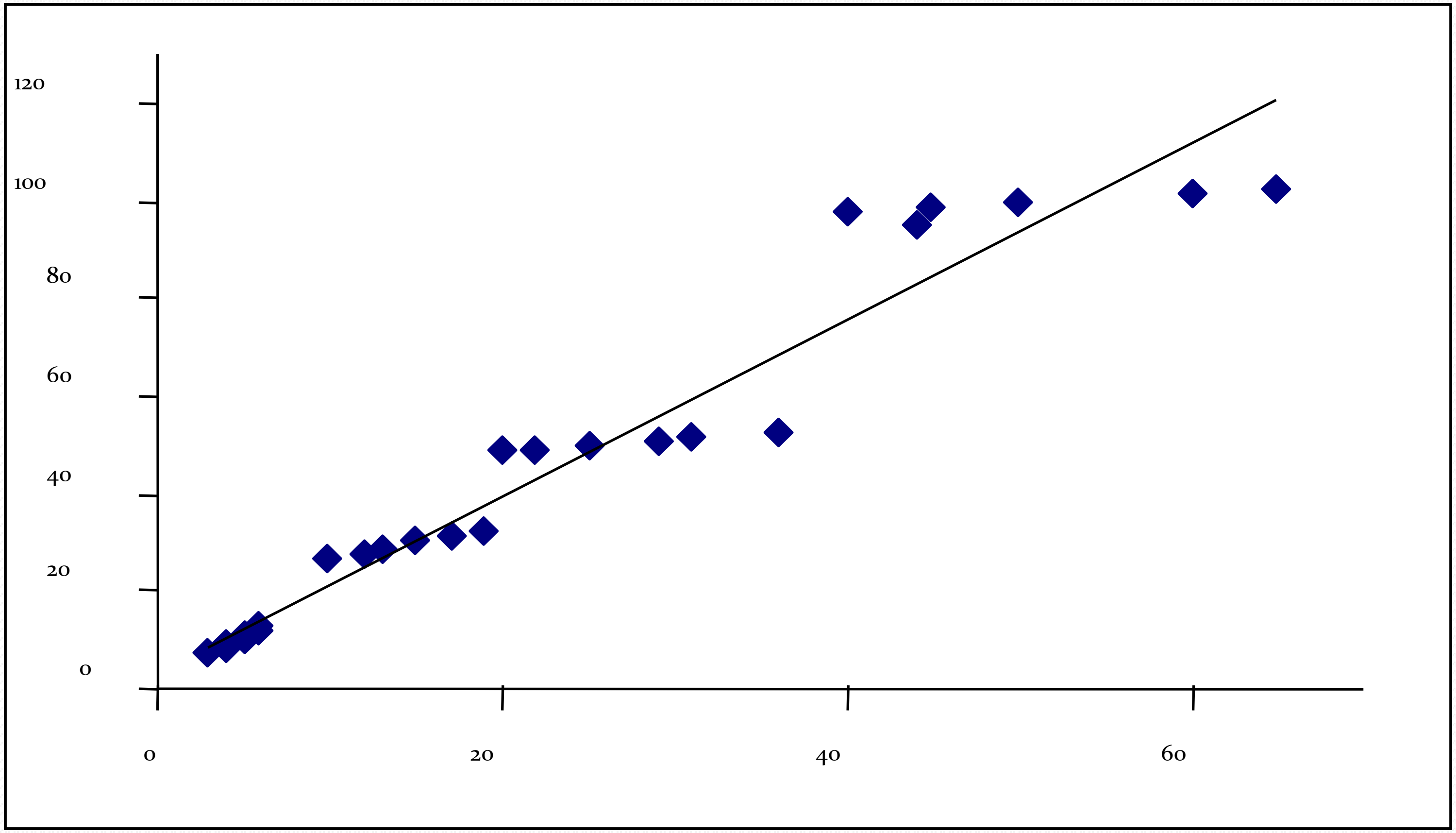
a. Dependent Variable: Konsumsi_Y

Mengatasi multikolinieritas

- Melihat informasi sejenis yang ada
- Tidak mengikutsertakan salah satu variabel yang kolinier
  - Banyak dilakukan.
  - Hati-hati, karena dapat menimbulkan *specification bias* yaitu salah spesifikasi kalau variabel yang dibuang merupakan variabel yang sangat penting.
- Mentransformasikan variabel
- Mencari data tambahan

Heteroskedastisitas

Pola Data Heteroskedastis

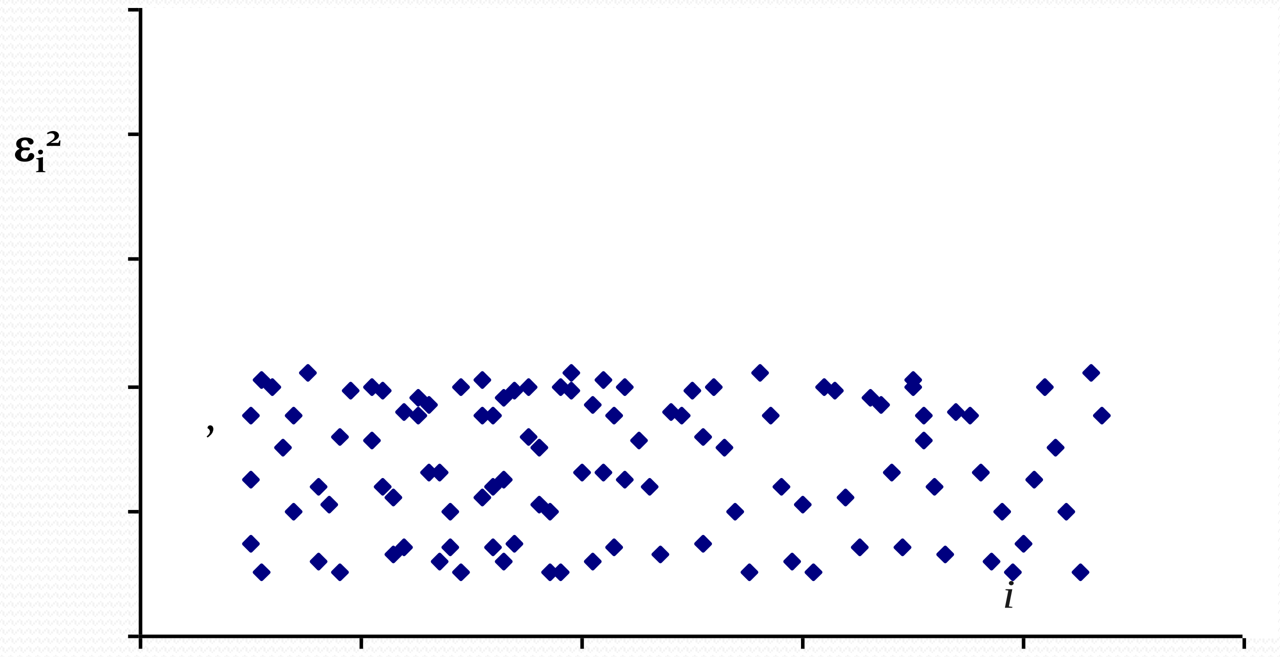


Data Heteroskedastisitas

Fakta:

- hubungan positif antara X dan Y, dimana nilai Y meningkat searah dengan nilai X.
- semakin besar nilai variabel bebas (X) dan variabel bebas (Y), semakin jauh koordinat (x,y) dari garis regresi (Error makin membesar)
- besarnya variasi seiring dengan membesarnya nilai X dan Y. Atau dengan kata lain, variasi data yang digunakan untuk membuat model tidak konstan.

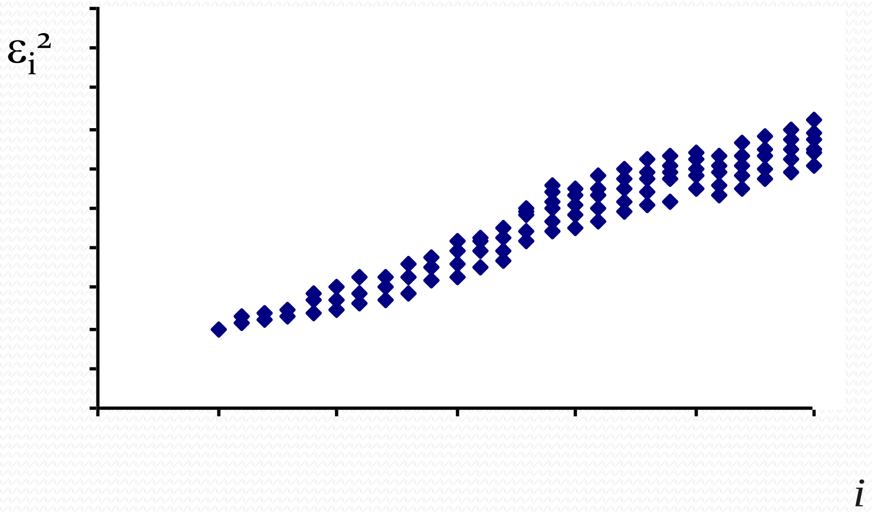
Pemeriksaan Heteroskedastisitas
1. Metode Grafik



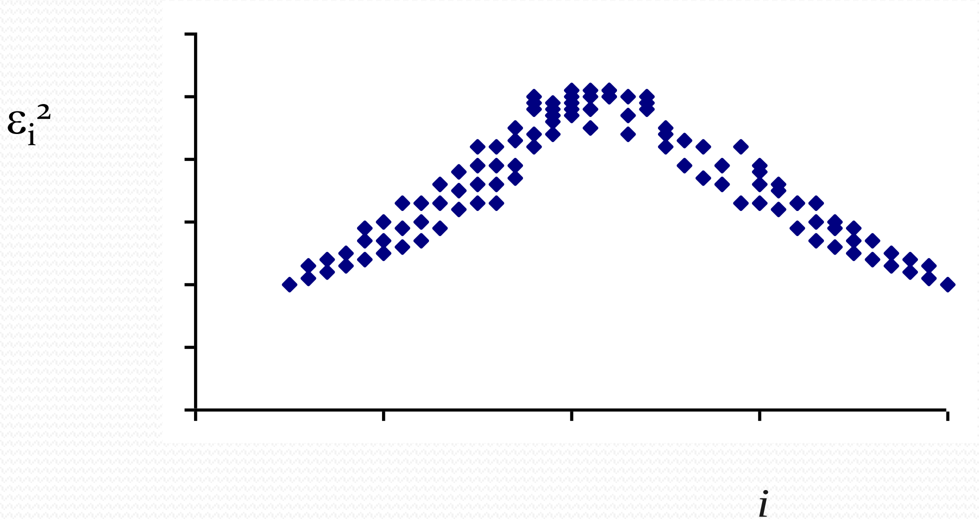
Pengamatan:

1. **Tidak adanya pola yang sistematis.**
2. **Berapapun nilai Y prediksi, residual kuadratnya relatif sama.**
3. **Variansi konstan, dan data homoskedastis.**

Pola Adanya Heteroskedastistisitas



Pola sistematis



Mengatasi heteroskedastisitas

1. Transformasi dengan Logaritma

Transformasi ini ditujukan untuk memperkecil skala antar variabel bebas. Dengan semakin 'sempitnya' *range* nilai observasi, diharapkan variasi error juga tidak akan berbeda besar antar kelompok observasi.

Adapun model yang digunakan adalah:

$$\text{Ln } Y_j = \beta_0 + \beta_1 \text{ Ln } X_j + u_j$$

Autokorelasi

- Autokorelasi: korelasi antara variabel regresor itu sendiri, pada pengamatan yang berbeda waktu atau individu. Umumnya kasus otokorelasi banyak terjadi pada data time series→ Kondisi sekarang dipengaruhi waktu lalu. Misal: Tinggi badan, upah, dsbnya.
- Salah satu alat deteksi:

→melihat pola hubungan antara residual (εi) dan variabel bebas atau waktu (X)

→ Durbin Watson

- Nilai d biasanya diantara 0 dan 4
- Contoh sebelumnya

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.991[a]	.982	.977	5.953	1.028

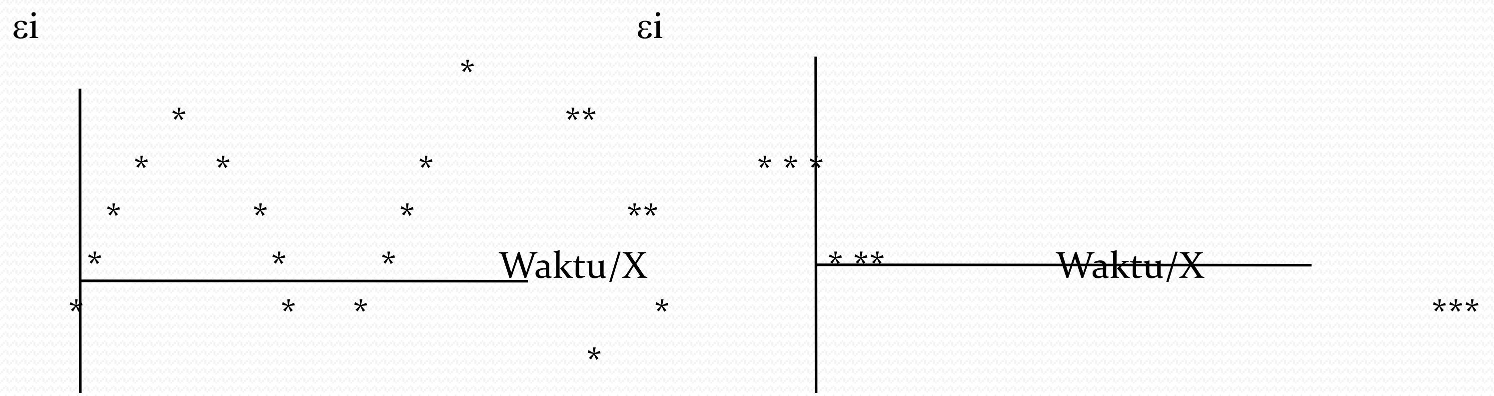
a. Predictors: (Constant), Kekayaan_X2, Pendapatan_X1

b. Dependent Variable: Konsumsi_Y

- d=1.028, du=(k=2,n=10)=1.6413 (tabel DW), 4-du=2.3587
- Karena du=1.6413<d=1.028<4-du=2.3587 → tidak berlaku maka ada autorelasi positif/ negatif

Mendeteksi Autokorelasi

- Pola Autokorelasi



- Gambar nomor (1) menunjukan adanya siklus, sedang nomor (2) menunjukan garis linier. Kedua pola ini menunjukan adanya autokorelasi.