



IPB University
— Bogor Indonesia —

Study Program
Statistics and Data Science
Department of Statistics

PRAKTIKUM 6

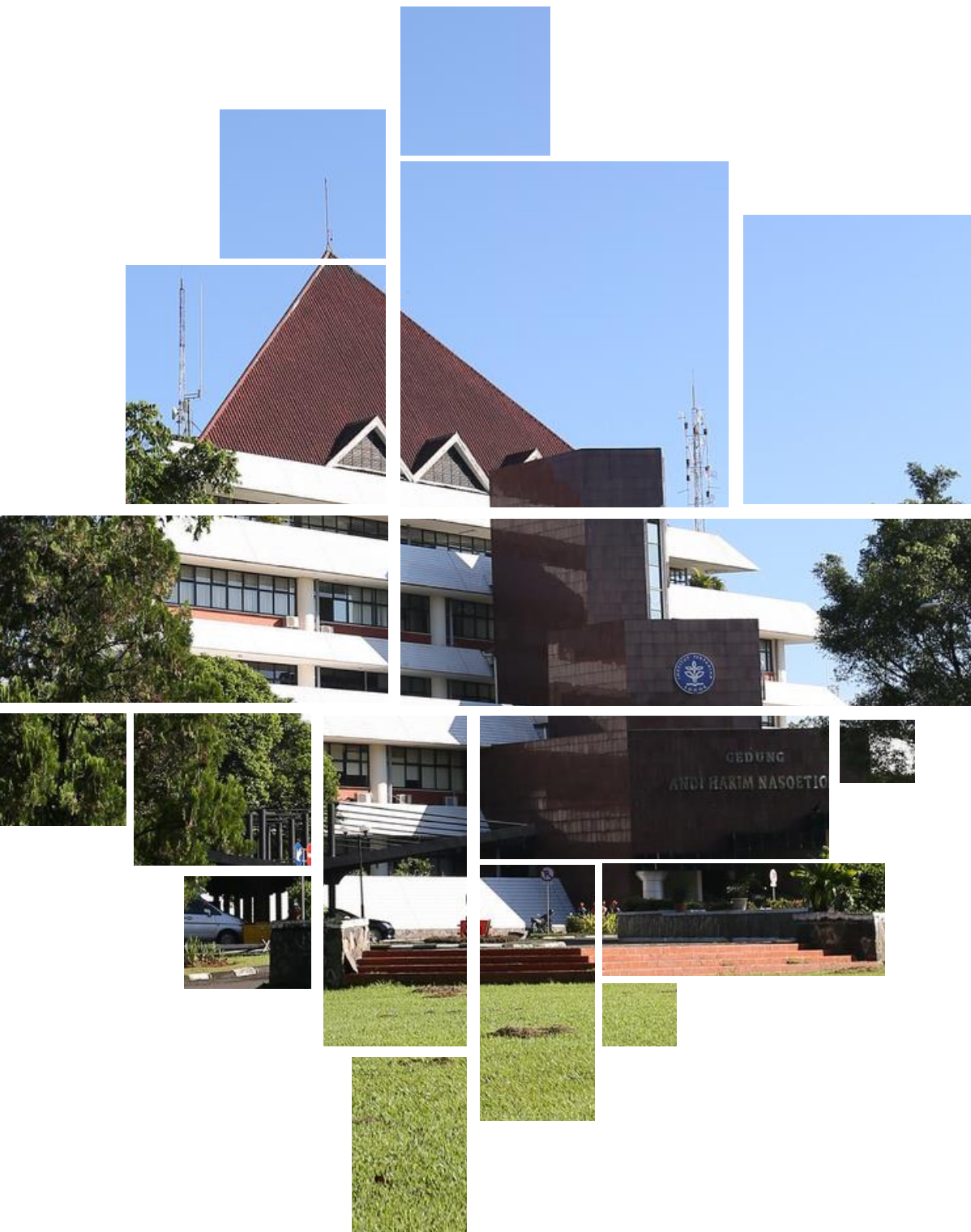
ANALISIS REGRESI

Diagnostik Regresi

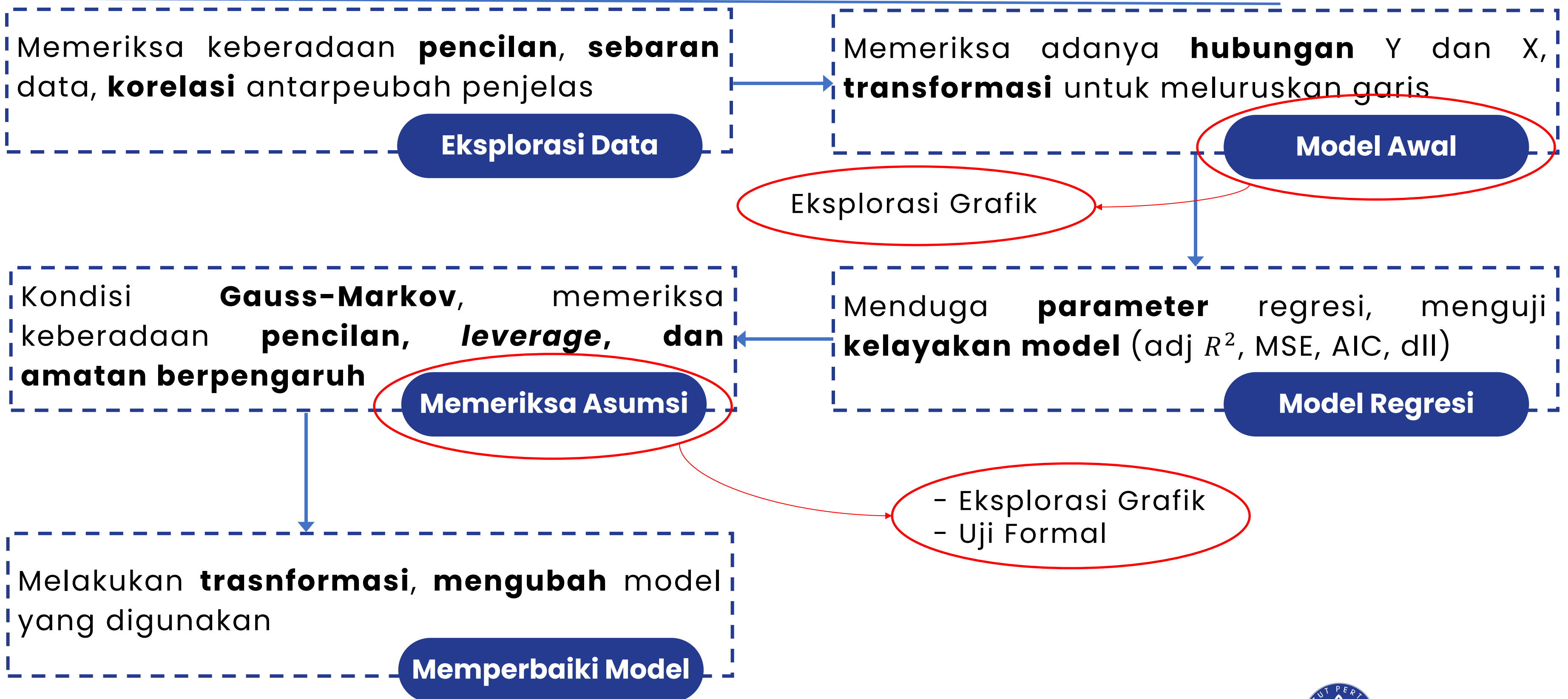
Oleh:

Nabil Bintang Prayoga – Farik Firsteady Haristiyanto

Selasa, 4 Maret 2025



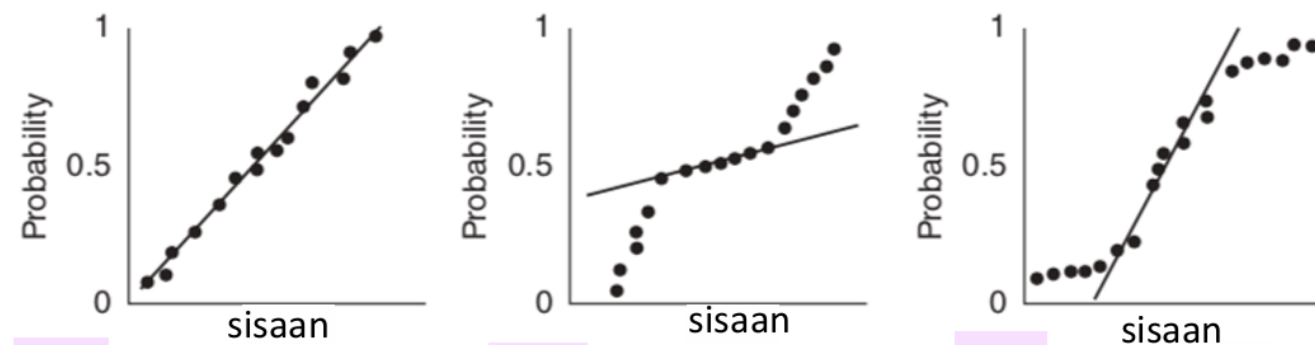
TAHAPAN ANALISIS REGRESI LINIER



$$\varepsilon_i \sim N(0, \sigma^2)$$

Normalitas galat atau galat menyebar normal

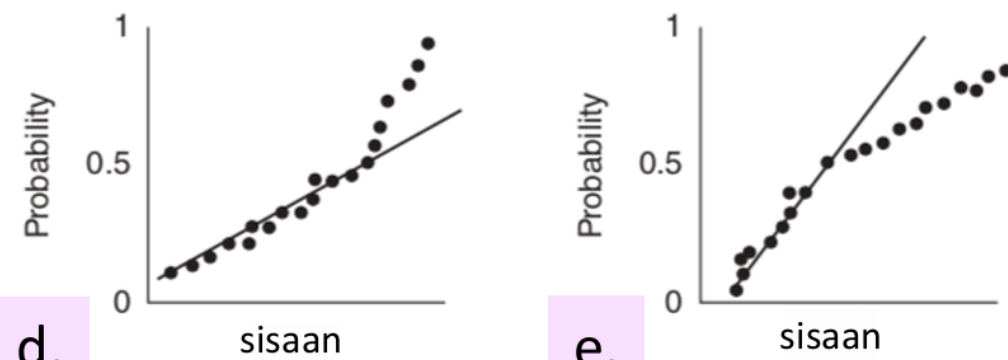
- Plot antara: **Sisaan vs Peluang Normal**
- Plot **Quantil-Quantil (QQ-Plot)**



a.

b.

c.

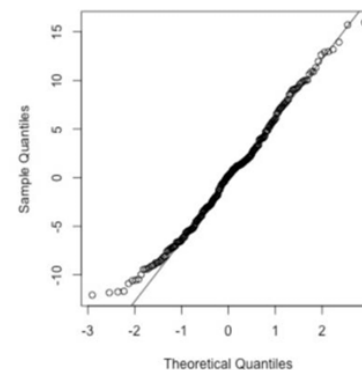


d.

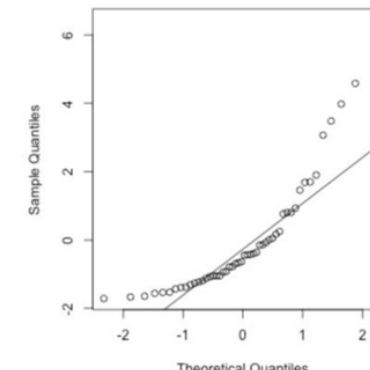
e.

- a. Simetri
- b. Simetri ujung menipis
- c. Simetri ujung menebal

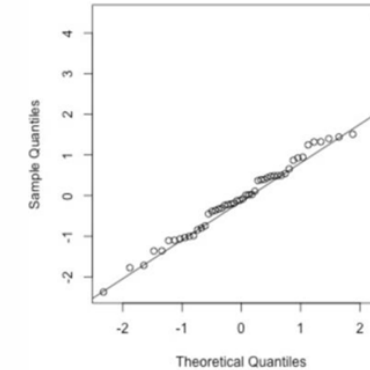
- d. Menjuler ke kiri
- e. Menjuler ke kanan



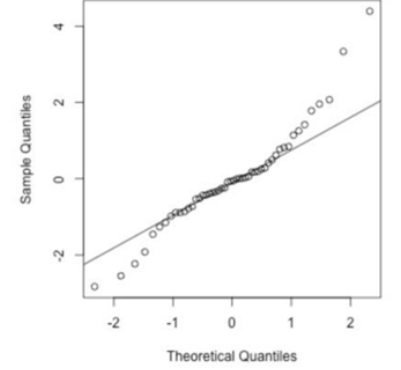
Menyebar normal



Menjuler ke kanan



Ada pencilan



Simetri ujung2 menipis

PLOT SISAAN VS Y DUGA (\hat{Y}) atau VS PEUBAH PENJELAS

- Nilai harapan galat sama dengan nol ($E[\varepsilon_i] = 0$) → Sisaan **menyebar di sekitar nol**
- Ragam galat homogen/homogenitas ($Var[\varepsilon_i] = \sigma^2$) → **Lebar pita** sisaan **sama**
- Galat saling bebas/tidak ada autokorelasi ($E[\varepsilon_i, \varepsilon_j] = 0$) → Tebaran sisaan tak berpola

Bentuk plot antara SISAAN vs Y-duga atau SISAAN vs peubah penjelas (X)

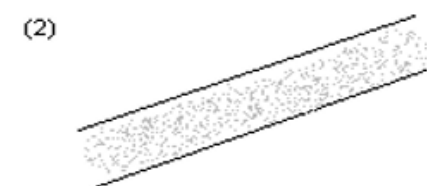


Pola tebaran sisaan memenuhi asumsi kondisi Gauss-Markov: berpusat di NOL, lebar pita sama, tidak berpola

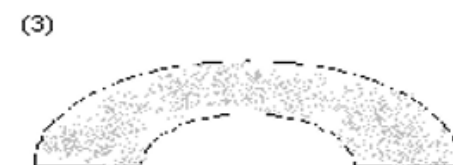
Pola tebaran sisaan yang tidak memenuhi asumsi kondisi Gauss-Markov:



➡ Ragam tidak homogen (perlu analisis kua-drat terkecil terboboti; atau transformasi thdp Y)



➡ Penyimpangan terhadap persamaan regresi bersifat sistematis (sisaan tidak bebas); atau karena tdk disertakannya β_0 ke dalam model



➡ Model tidak pas (perlu suku-suku lain dalam model atau transformasi thdp Y)

PLOT SISAAN VS URUTAN WAKTU

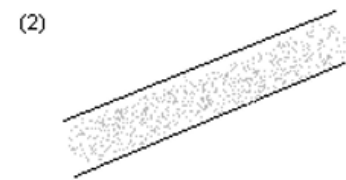
- Ragam galat homogen/homogenitas ($Var[\varepsilon_i] = \sigma^2$) → **Lebar pita** sisaan **sama**
- Galat saling bebas/tidak ada autokorelasi ($E[\varepsilon_i, \varepsilon_j] = 0$) → Tebaran sisaan tak berpola

Pengaruh waktu jangka panjang tidak mempengaruhi data.

Pola tebaran sisaan yang menginformasikan bahwa pengaruh waktu belum diperhitungkan



Ragam tidak homogen (perlu analisis kuadrat terkecil terboboti)

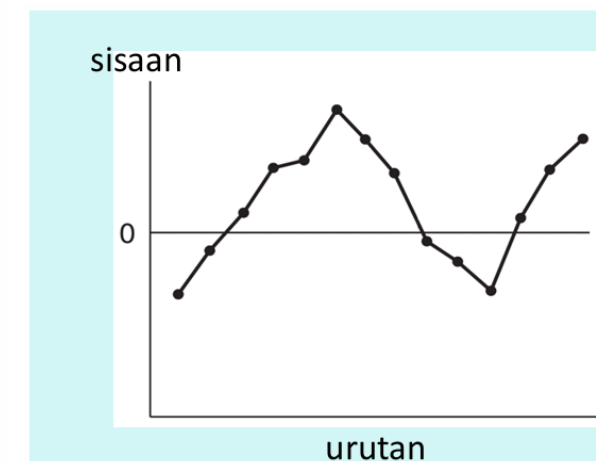


Suatu suku linier dalam waktu harus ditambahkan ke dalam model

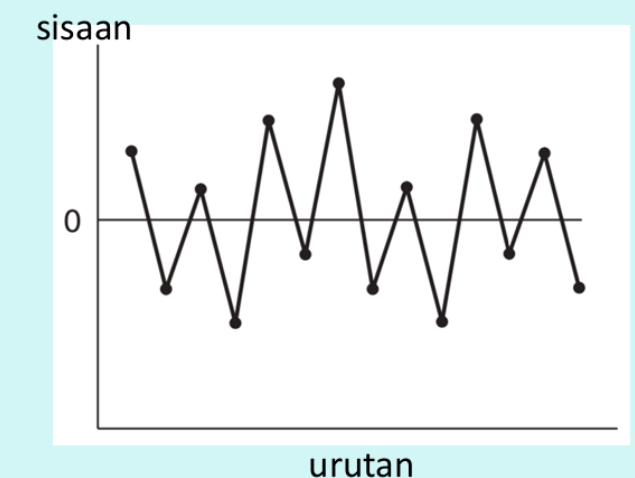


Suku linier dan kuadratik dalam waktu perlu ditambahkan ke dalam model

$$W_i = \frac{1}{x_i^2} \text{ atau } \frac{1}{s_i^2} \text{ jika setiap } x \text{ ada ulangan}$$



Ada outokorelasi POSITIF



Ada outokorelasi NEGATIF

PENCILAN

- Suatu amatan data dengan **nilai x** yang berada **dalam selang**, tetapi **y tidak mengikuti tren**
- Berpengaruh terhadap R^2 dan **galat baku** dugaan koefisien regresi
- Keberadaannya perlu diperiksa lebih lanjut: **salah input atau kombinasi keadaan tidak bisa**
- Dapat dideteksi dengan **nilai sisaan terbakukan r_i** dan/atau **plot sisaan e_i atau r_i vs \hat{Y}_i**
- Nilai $|r_i| > 2$ **atau 3** dicurigai sebagai pencilan
 h_{ii} adalah unsur diagonal ke- i matriks $H = X(X'X)^{-1}X'$

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}} \quad s = \sqrt{MS_{Res}}$$
$$h_{ii} = \frac{\frac{1}{n} + (x_i - \bar{x})^2}{\sum (x_k - \bar{x})^2}$$

TITIK LEVERAGE

- Suatu amatan data dengan **nilai x** tidak berada **dalam selang**, tetapi **y mengikuti tren** (hampir terletak dalam garis regresi titik contoh lain)
- Amatan memiliki **nilai leverage tinggi** jika memiliki **nilai peubah penjelas x yang ekstrem**
- Pada **satu peubah X**: nilai **sangat tinggi** atau **sangat rendah**, pada **beberapa peubah X**: nilai **sangat tinggi** atau **rendah** untuk **satu atau lebih X** atau **kombinasi** dari beberapa nilai ekstrem dari beberapa X
- Titik leverage ditandai dengan nilai $h_{ii} > 2\frac{p}{n}$

AMATAN BERPENGARUH

- **Penyisihannya** berkaitan dengan besarnya **perubahan** yang terjadi pada **dugaan parameter regresi, R^2 , dan uji hipotesis**. **Dugaan kemiringan garis** juga bisa berubah sehingga **berbahaya** jika pemanfaatan hasil analisis bertumpu pada **pemaknaan parameter**
- Bisa berasal dari **pencilan** dan **titik leverage**

Ukuran	Rumus	Kriteria
Jarak COOK	$D_i = \frac{e_i^2}{ps^2} \left[\frac{h_{ii}}{(1 - h_{ii})^2} \right] = \frac{(\hat{y}_{(i)} - \hat{y})'(\hat{y}_{(i)} - \hat{y})}{pMS_{res}}$	$D_i > F_{(p, n-p; 1-\alpha)}$
$DFFITs_i$	$DFFITs_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}}$	$ DFFITs_i > 2\sqrt{\frac{p}{n}}$
$DFBETAS_{j,i}$	$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 c_{jj}}}$	$ DFBETAS_{j,i} > \frac{2}{\sqrt{n}}$



MULTIKOLINEARITAS

- **Hubungan linier/korelasi** antarpeubah bebas, masalah jika **nilainya besar**
- **Tandanya** jika **dugaan parameter** regresi **berubah cukup besar** ketika peubah penjelasnya ditambah atau dikurang dan **perbedaan arah tanda**
- **Akibatnya** interpretasi menjadi sulit (**missinterpretasi**), ragam dan koragam **besar**, uji simultan dan parsial **tidak selaras**, penduga MKT dan galat baku **sensitif** terhadap perubahan tanda, $X'X$ **hampir atau bahkan singular**, *adjusted R²* **tinggi**
- **Cara mengatasinya** dengan **menyisihkan peubah penjelas** yang berkorelasi tinggi dengan peubah penjelas lain, **menambah data** amatan/contoh, **transformasi peubah** atau menggabungkannya, **regresi lain** seperti gulud/kuadrat terkecil parsial/komponen utama

MENDETEKSI MULTIKOLINEARITAS

- **Memeriksa koefisien korelasi** yang besar antarpeubah penjelas (biasanya lebih dari 0,5)
- **Nilai VIF** yang lebih dari 5 atau 10 $\rightarrow VIF_j = C_{jj} = \frac{1}{1-R_j^2} \rightarrow C = (X'X)^{-1}$
- **Berdasarkan nilai akar ciri $X'X$**

$$K = \frac{\lambda_{max}}{\lambda_{min}} \rightarrow \leq 100 \text{ (tidak ada), } 100 - 1000 \text{ (ada, lemah), } > 1000 \text{ (ada, kuat)}$$

$$K_j = \frac{\lambda_{max}}{\lambda_j} \rightarrow > 1000 \text{ (ada pada peubah penjelas ke-j)}$$

TERIMA KASIH



IPB University
— Bogor Indonesia —

Department of Statistics
Jl. Meranti W22 L4
Kampus IPB Dramaga Bogor 16680
Telp.: 0251-8624535
E-mail: statistika@apps.ipb.ac.id