



Analisis Regresi

#9 Meeting

Heteroskedastisitas

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001

Regresi Linier - Review OLS

Sejauh ini, kita telah mempelajari metode Ordinary Least Squares (OLS) sebagai fondasi dari model regresi. Prinsip utama OLS adalah menemukan garis yang meminimalkan jumlah dari kuadrat residual ($\sum \text{error}^2$). Di balik kesederhanaan ini, OLS memiliki sebuah asumsi implisit yang sangat kuat: setiap observasi data sama pentingnya dan sama andalnya.

Bayangkan OLS sebagai sebuah sistem demokrasi yang sempurna, di mana setiap titik data memiliki hak suara yang sama persis (satu suara) dalam menentukan di mana garis regresi "terbaik" akan diletakkan. Untuk banyak kasus, pendekatan ini bekerja dengan sangat baik.

Regresi Linier - Heteroscedasticity

Heteroscedasticity adalah kondisi di mana tingkat penyebaran (varians) dari error tidak konstan di seluruh rentang data. Artinya, tingkat ketidakpastian prediksi model kita berubah-ubah.

Analogi Sederhana: Memprediksi Harga Rumah, Bayangkan membuat model untuk memprediksi harga rumah berdasarkan luas bangunannya.

1. Untuk Rumah Kecil (misal, 50 m²): Harga rumah di kisaran ini cenderung seragam. Prediksi mungkin meleset sekitar \pm Rp 50 juta. Data di area ini akan bergerombol rapat di sekitar garis regresi. Observasi ini bisa kita anggap sangat andal.
2. Untuk Rumah Mewah (misal, 500 m²): Harga di kisaran ini bisa sangat bervariasi tergantung pada kualitas marmer, desain interior, merek lift, dll. Prediksi bisa meleset sangat jauh, mungkin \pm Rp 2 Miliar. Data di area ini akan tersebar sangat lebar di sekitar garis regresi. Observasi ini kurang andal.

Regresi Linier - Heteroscedasticity

Dalam situasi ini, OLS akan menghadapi masalah. Karena OLS meminimalkan kuadrat error, observasi rumah mewah yang memiliki potensi error sangat besar akan memiliki pengaruh yang tidak proporsional (suara yang jauh lebih "bising"). Satu atau dua data rumah mewah yang outlier bisa "menarik" garis regresi secara signifikan, sehingga merusak akurasi prediksi untuk mayoritas data (rumah-rumah kecil dan menengah).

Solusi: **Weighted Least Squares (WLS)**
Akan dipelajari pada Supervised Learning
Peminatan Komputasi

Regresi Linier - Heteroscedasticity

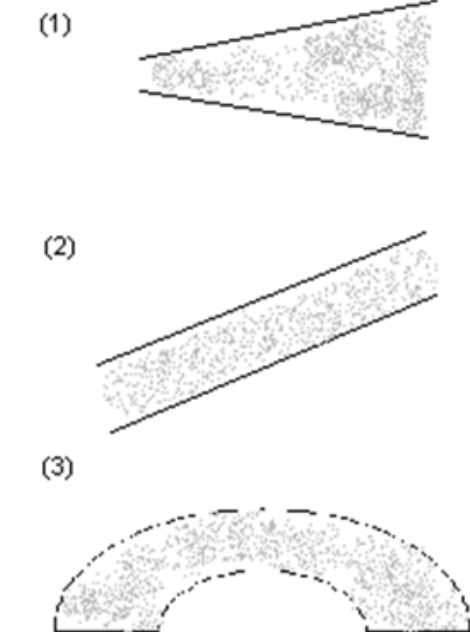
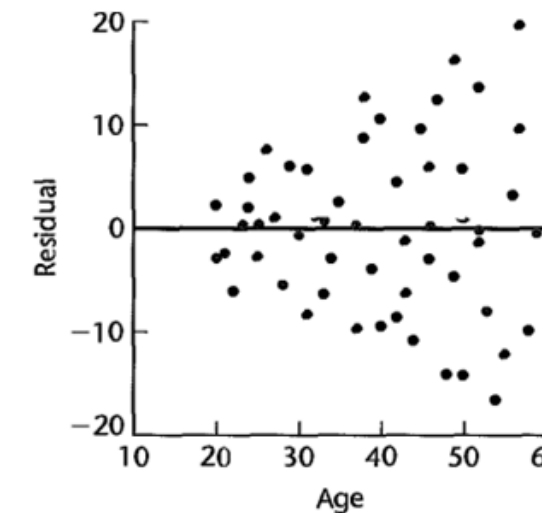
Ragam residual tidak berubah (membesar/mengecil) seiring dengan perubahan nilai prediktor X secara berurutan atau secara perubahan waktu karena $\varepsilon_i \sim (0, \sigma^2)$

Identifikasi homoskedastisitas

- Plot Residual \rightarrow tidak membentuk pola tertentu
- Uji Brown-Forsythe
- Uji Breusch-Pagan-Godfrey

Penyebab heteroskedastisitas

- Terdapat kesalahan input nilai peubah respon pada beberapa peubah penjelas
- Kasus ketidakhomogenan dapat terjadi secara alami pada peubah-peubah ekonomi, seperti kasus rumah tangga dengan pendapatan yang berbeda terkadang memiliki pengeluaran yang hampir sama
- Adanya manipulasi data yang menyebabkan sisaan memiliki ragam yang sistematis
- Kesalahan Spesifikasi Model



Regresi Linier - Heteroscedasticity

- **Homoskedastisitas** berarti varians dari sisaan (error) bersifat konstan ($Var(\epsilon_i) = \sigma^2$) pada setiap level dari variabel independen (X).
- **Heteroskedastisitas** (kebalikannya) terjadi ketika varians sisaan tidak konstan ($Var(\epsilon_i) = \sigma_i^2$). Artinya, lebar sebaran *error* berubah seiring perubahan nilai X .

Mengapa ini masalah? Jika terjadi heteroskedastisitas, estimator OLS ($\hat{\beta}$) kita tetap tidak bias (*unbiased*), namun:

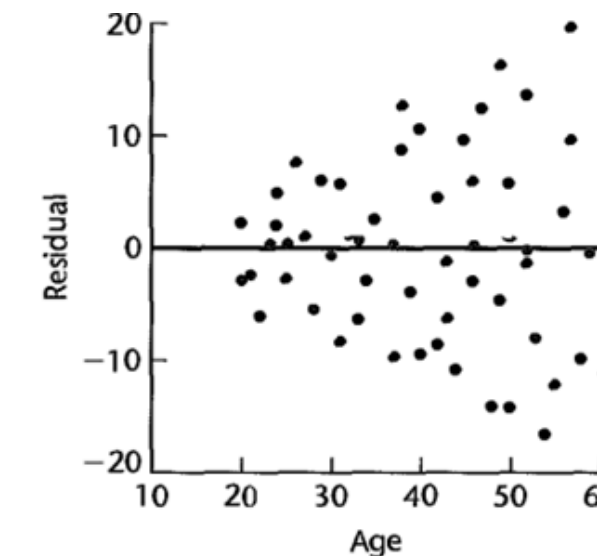
1. Estimator OLS tidak lagi efisien (bukan lagi BLUE - Best Linear Unbiased Estimator).
2. **Masalah terbesar:** *Standard error* dari koefisien regresi menjadi bias. Akibatnya, **Uji-t** dan **Uji-F** menjadi tidak valid atau tidak bisa dipercaya.

Identifikasi dengan Plot

Plot 1: Sisaan (Residuals) vs. Nilai Prediksi (Fitted Values)

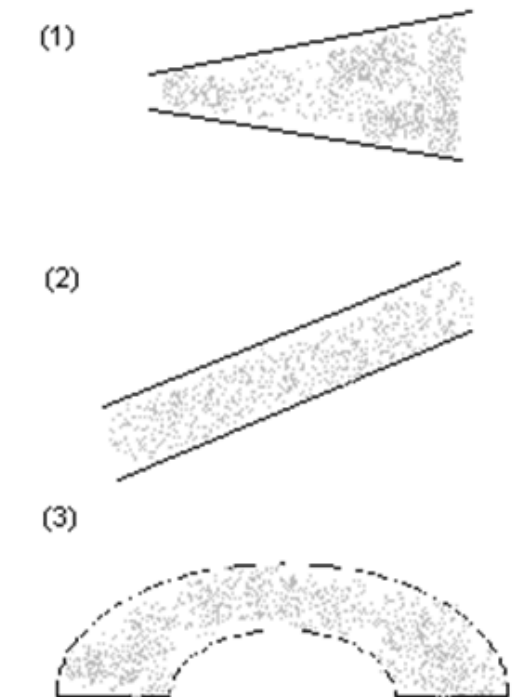
Ini adalah plot yang paling umum digunakan.

- **Sumbu Y:** Sisaan ($e_i = Y_i - \hat{Y}_i$)
- **Sumbu X:** Nilai prediksi (\hat{Y}_i)



Interpretasi:

- **Homoskedastisitas (Ideal):** Plot akan menunjukkan sebaran titik-titik yang **acak** di sekitar garis horizontal di titik 0, tanpa membentuk pola tertentu. Sebarannya terlihat konstan (seperti awan horizontal).
- **Heteroskedastisitas (Masalah):** Plot akan menunjukkan pola sistematis. Pola yang paling umum adalah **pola corong** atau **megafon** (semakin melebar atau menyempit seiring meningkatnya nilai prediksi).



Identifikasi dengan Plot

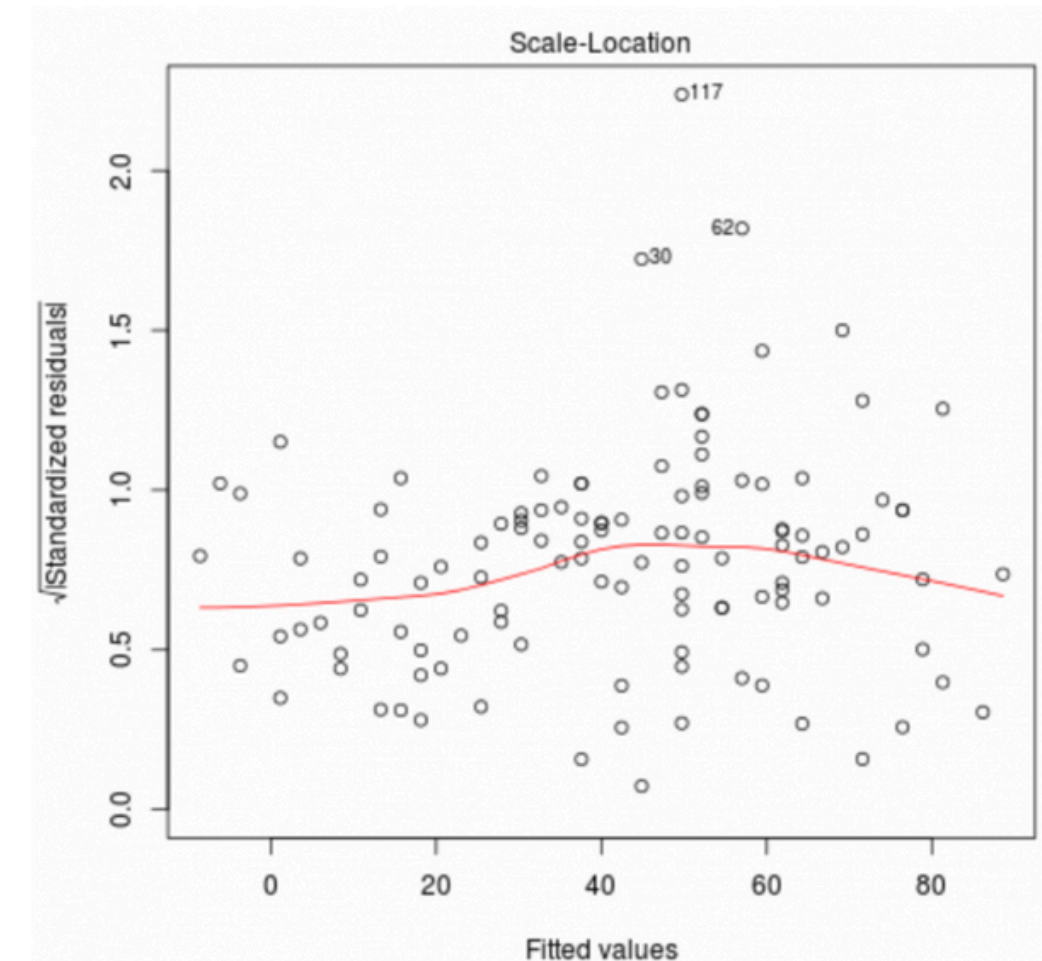
Plot 2: Scale-Location Plot (Standardized Residuals vs. Fitted Values)

Plot ini mirip dengan plot pertama, tetapi menggunakan akar kuadrat dari sisaan yang distandarisasi (Square Root of Standardized Residuals) pada sumbu Y.

- **Sumbu Y:** $\sqrt{|e_i^*|}$ (Akar dari sisaan terstandarisasi)
- **Sumbu X:** Nilai prediksi (\hat{Y}_i)

Interpretasi:

- **Homoskedastisitas (Ideal):** Garis tren yang menghubungkan titik-titik akan terlihat **horizontal atau datar**, menunjukkan bahwa varians sisaan konstan.
- **Heteroskedastisitas (Masalah):** Garis tren akan terlihat **miring (naik atau turun)**, menunjukkan bahwa varians sisaan berubah seiring perubahan nilai prediksi.



Identifikasi dengan Uji Formal

1. Uji Breusch-Pagan Godfrey (BP Test)
2. Uji White (White General Test)
3. Uji Goldfeld-Quandt (GQ Test)
4. Uji Brown Forsythe



Deteksi - Uji Breusch Pagan Godfrey

Tahapan-tahapan Uji Breusch-Pagan-Godfrey

- Estimasi $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k}$
- Bentuk peubah $p = \frac{e^2}{\hat{\sigma}^2}$
- Buat model regresi $p = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \dots + \alpha_k Z_k + v$
- Hitung JKR dari model tersebut
- Lakukan uji hipotesis

Hipotesis

H_0 : Ragam Residual Konstan

H_1 : Ragam Residual Tidak Konstan

Statistik Uji

$$\chi^2 = \frac{JKR}{2}$$

Kriteria Penolakan H_0

Tolak H_0 jika $\chi^2 > \chi^2_{\alpha; k}$

Deteksi - Uji White

- **Hipotesis:**

- H_0 : Homoskedastisitas
- H_1 : Heteroskedastisitas

- **Tahapan Uji:**

1. Jalankan model regresi utama (OLS)
2. Hitung kuadrat dari sisaan (e_i^2).
3. Buat **model regresi auxilier (pembantu)**: Regresikan e_i^2 pada variabel X asli, X kuadrat, dan (jika $k > 1$) interaksi antar X .
 - Contoh 1 X : $e_i^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_1^2 + v_i$
 - Contoh 2 X : $e_i^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1^2 + \alpha_4 X_2^2 + \alpha_5 X_1 X_2 + v_i$
4. Dapatkan nilai R^2 dari regresi auxilier ini.
5. Hitung statistik uji (LM Statistic): $LM = n \times R_{aux}^2$
6. Statistik uji LM ini mengikuti distribusi Chi-Square (χ^2) dengan $df = p$ (jumlah prediktor di regresi auxilier).
7. **Keputusan:** Tolak H_0 jika $p - value < \alpha$.

Deteksi - Uji Goldfeld-Quandt (GQ Test)

- **Hipotesis:**

- $H_0: \sigma_1^2 = \sigma_2^2$ (Varians kedua kelompok sama / Homoskedastisitas)
- $H_1: \sigma_1^2 \neq \sigma_2^2$ (Varians kedua kelompok berbeda / Heteroskedastisitas)

- **Tahapan Uji:**

1. **Urutkan** data berdasarkan salah satu variabel X yang diduga menjadi sumber heteroskedastisitas (misal, dari X terkecil ke terbesar).
2. **Bagi data:** Buang sejumlah c pengamatan di bagian tengah (misal, 20% data). Ini menyisakan dua kelompok: Grup 1 (nilai X rendah) dan Grup 2 (nilai X tinggi).
3. Jalankan **dua model regresi OLS terpisah**: satu untuk Grup 1 dan satu untuk Grup 2.
4. Dapatkan *Residual Sum of Squares* (RSS) dari masing-masing model, yaitu RSS_1 dan RSS_2 .

Deteksi - Uji Goldfeld-Quandt (GQ Test)

5. Hitung statistik uji F:

$$F = \frac{RSS_2/df_2}{RSS_1/df_1}$$

(Dimana $df_1 = n_1 - k - 1$ dan $df_2 = n_2 - k - 1$. n adalah jumlah observasi di grup tsb, k adalah jumlah prediktor).

6. Statistik uji F ini mengikuti distribusi F.

7. **Keputusan:** Tolak H_0 jika $p - value < \alpha$ (atau jika $F_{hitung} > F_{tabel}$).

Deteksi - Uji Brown Forsythe

Hipotesis

H_0 : Ragam Residual Konstan

H_1 : Ragam Residual Tidak Konstan

Statistik Uji

$$t^* = \frac{\bar{d}_1 - \bar{d}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Kriteria Penolakan H_0

Tolak H_0 jika $t^* > t_{\frac{\alpha}{2}; n-2}$

$$d_{i1} = |e_{i1} - \tilde{e}_1|$$

$$d_{i2} = |e_{i2} - \tilde{e}_2|$$

$$S^2 = \frac{\sum (d_{i1} - \bar{d}_1)^2 + \sum (d_{i2} - \bar{d}_2)^2}{n - 2}$$

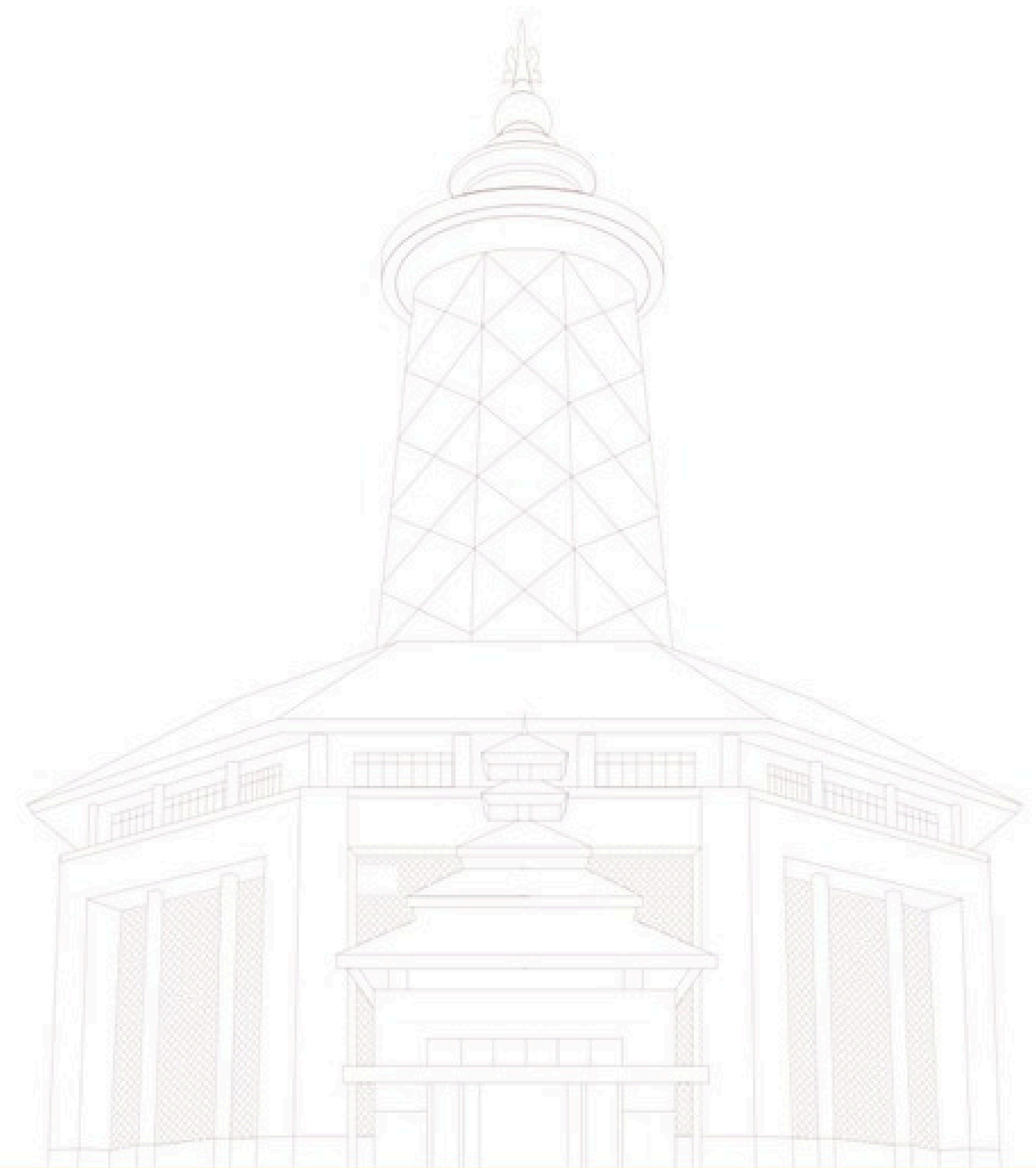
Regresi Linier - Heteroscedasticity

Efek heteroskedastistas

- Ragam dugaan parameter yang dihasilkan tidak minimum, sehingga pendugaan dengan OLS menjadi tidak efisien

Penanganan heteroskedastistas

- Transformasi peubah respon
- MKT Terboboti/WLS (Ragam tidak homogen)
- Mengubah bentuk model regresinya





SEE YOU NEXT WEEK !

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001
ferdian.bangkit@untirta.ac.id