



Analisis Regresi

#10 Meeting

Autokorelasi

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

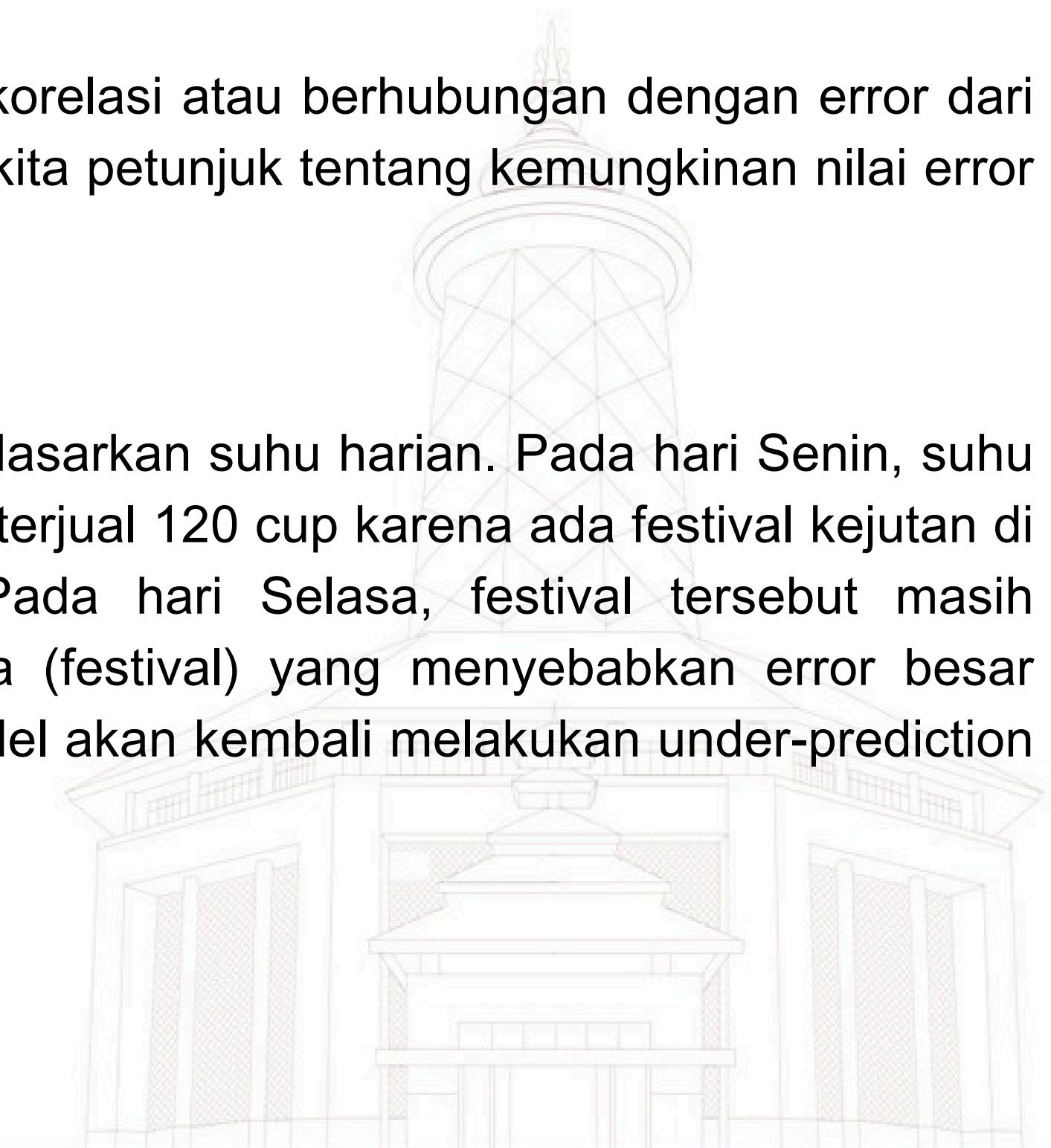




Regresi Linier - Autokorelasi

Autokorelasi adalah kondisi di mana error dari satu observasi berkorelasi atau berhubungan dengan error dari observasi lainnya. Artinya, nilai error pada satu titik data memberi kita petunjuk tentang kemungkinan nilai error pada titik data lain.

Analogi Sederhana: Memprediksi Penjualan Es Krim Harian, membuat model regresi untuk memprediksi penjualan es krim berdasarkan suhu harian. Pada hari Senin, suhu tinggi, dan model memprediksi penjualan 100 cup. Kenyataannya, terjual 120 cup karena ada festival kejutan di dekat toko. Model memiliki error positif yang besar (+20). Pada hari Selasa, festival tersebut masih berlangsung. Meskipun suhu sedikit berbeda, faktor tak terduga (festival) yang menyebabkan error besar kemarin kemungkinan besar masih ada. Jadi, sangat mungkin model akan kembali melakukan under-prediction (kesalahan positif lagi).



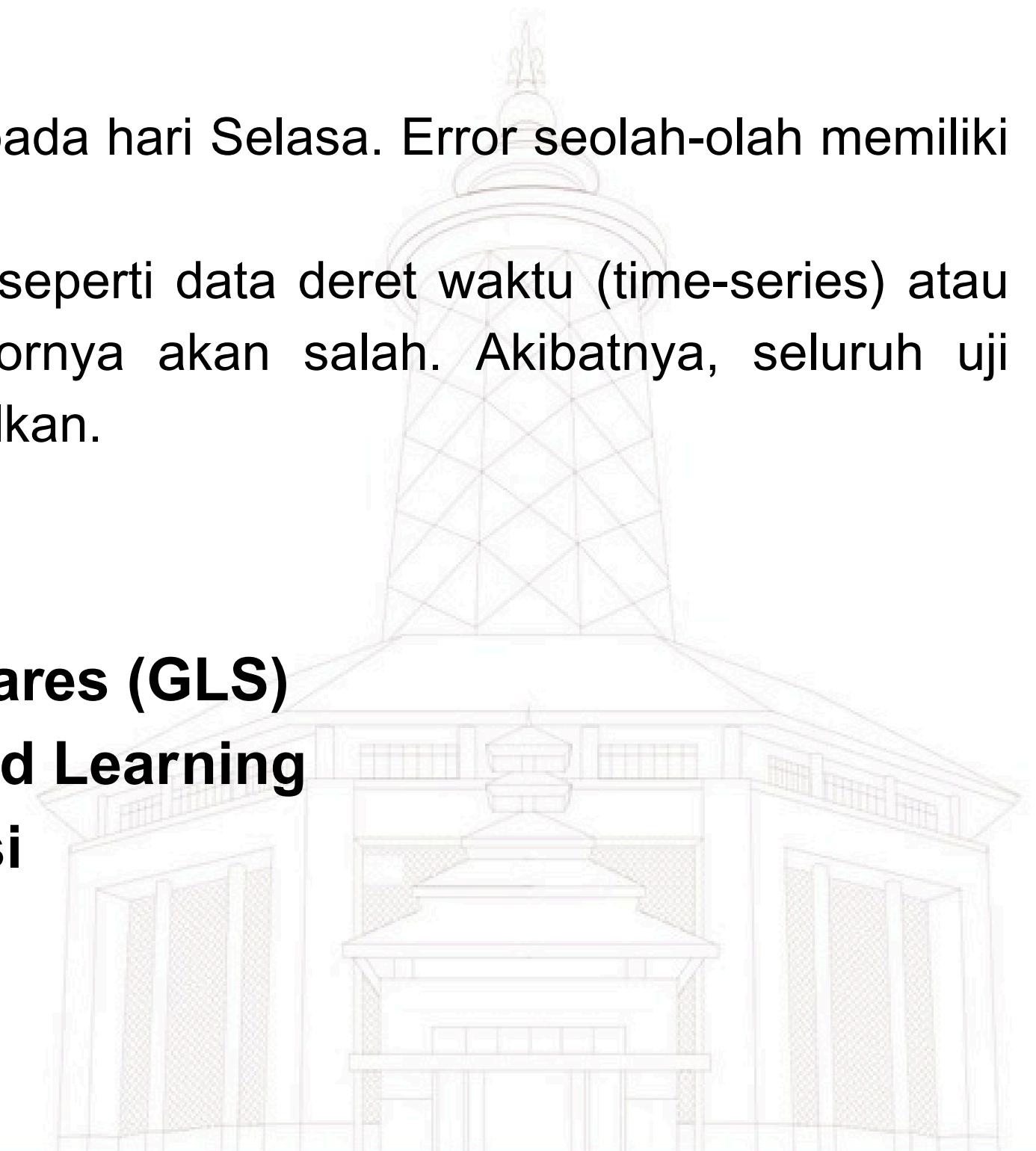


Regresi Linier - Autokorelasi

Dalam kasus ini, error pada hari Senin tidak independen dari error pada hari Selasa. Error seolah-olah memiliki "ingatan". Inilah yang disebut autokorelasi.

Masalah ini sangat umum terjadi pada data yang memiliki urutan, seperti data deret waktu (time-series) atau data geografis. Jika kita mengabaikan autokorelasi, standar errornya akan salah. Akibatnya, seluruh uji hipotesis (p-value, interval kepercayaan) menjadi tidak dapat diandalkan.

Solusi: Generalized Least Squares (GLS)
Akan dipelajari pada Supervised Learning
Peminatan Komputasi





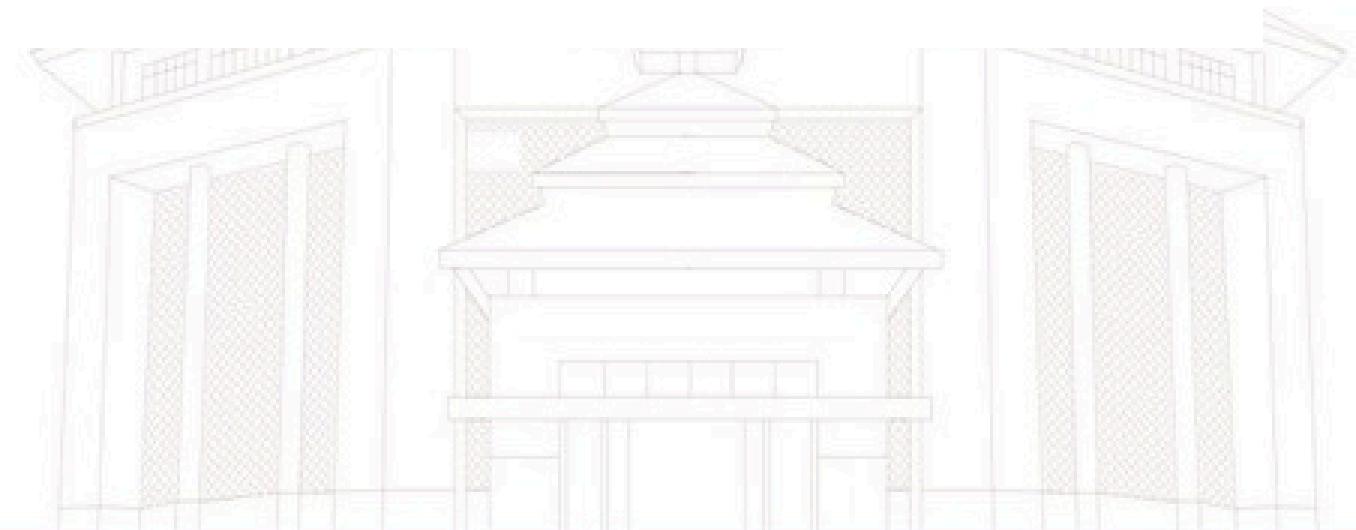
Regresi Linier - Non-Autokorelasi

Asumsi ini menyatakan bahwa sisaan (residual) dari satu observasi bersifat independen (saling bebas) dari sisaan observasi lainnya.

$$Cov(e_i, e_j) = 0 \text{ untuk } i \neq j$$

Secara sederhana, *error* pada satu titik data tidak memberikan informasi atau tidak memprediksi *error* pada titik data berikutnya.

Autokorelasi (pelanggaran asumsi) terjadi ketika sisaan tidak saling bebas. $Cov(e_i, e_j) \neq 0$





Regresi Linier - Sebab Autokorelasi

Konteks Utama:

- Asumsi ini paling krusial dan paling sering dilanggar pada data yang memiliki urutan, terutama data runut waktu (time series).
- Bisa juga terjadi pada data cross-section jika ada urutan spasial (misal, korelasi antar wilayah yang berdekatan).

Penyebab Umum Autokorelasi:

- Inersia: Fenomena ekonomi (seperti PDB, inflasi) cenderung memiliki momentum.
- Omitted Variables: Ada variabel penting yang (mungkin ber-autokorelasi) yang tidak dimasukkan ke dalam model.
- Manipulasi Data: Proses seperti interpolasi atau *averaging* dapat memunculkan pola autokorelasi buatan.



Regresi Linier - Akibat Autokorelasi

Jika autokorelasi terjadi (namun kita abaikan dan tetap menggunakan OLS):

1. Estimator Tetap Tidak Bias: Penduga parameter OLS ($\hat{\beta}$) tetap *unbiased*.
2. Kehilangan Efisiensi: Estimator OLS tidak lagi BLUE (*Best Linear Unbiased Estimator*).
Varians-nya bukan lagi yang minimum.
3. Standard Error Menjadi Bias: Ini adalah masalah terbesar.
 - Pada autokorelasi positif (paling umum), *standard error* cenderung di bawah estimasi (*underestimated*).
4. Uji Hipotesis Tidak Valid: Karena *standard error* bias, maka:
 - Nilai Uji-t akan terlihat lebih besar (membengkak).
 - Nilai Uji-F juga menjadi tidak bisa dipercaya.
 - Hasil: Kita cenderung terlalu sering menolak H_0 . Ini bisa mengarah pada kesimpulan yang salah, di mana variabel yang sebenarnya tidak signifikan secara statistik terlihat signifikan (*spurious regression*).





Identifikasi - Grafis

- Plot: Sisaan (e_i) pada sumbu Y vs. Waktu/Urutan (t) pada sumbu X.

Interpretasi Pola Plot:

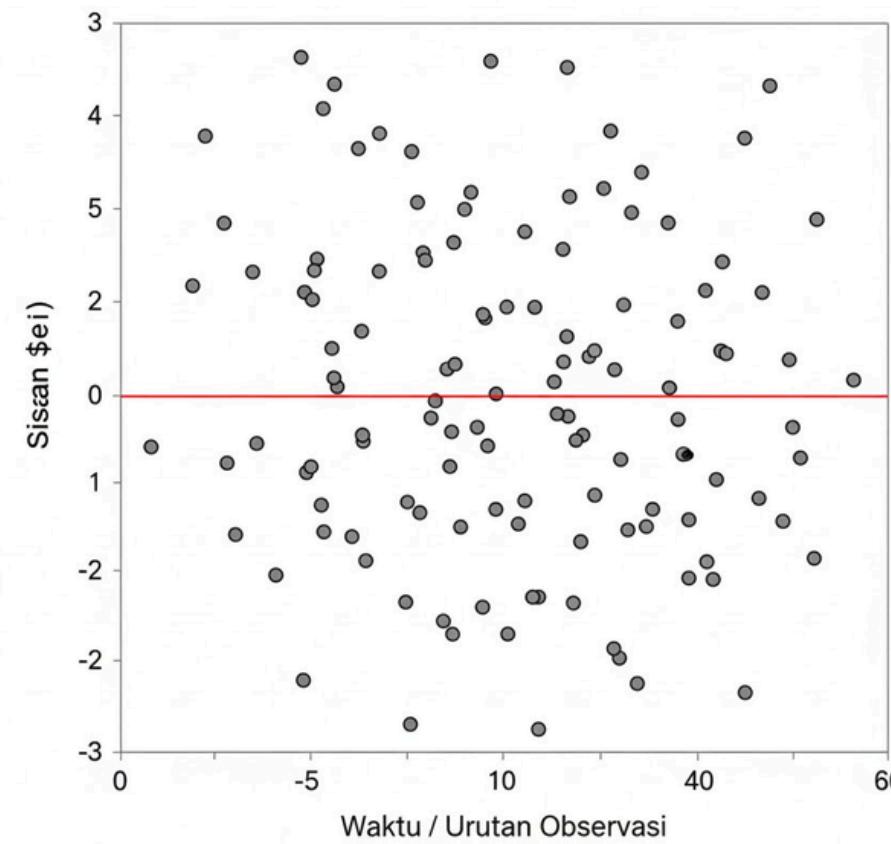
1. Ideal (Non-Autokorelasi): Titik-titik tersebar secara acak di sekitar garis nol, tanpa membentuk pola yang jelas.
2. Autokorelasi Positif (Masalah): Sisaan cenderung berkelompok. Sisaan positif diikuti oleh sisaan positif, dan sisaan negatif diikuti oleh sisaan negatif (membentuk "gugusan" di atas dan di bawah garis nol).
3. Autokorelasi Negatif (Masalah): Sisaan cenderung bolak-balik (osilasi). Sisaan positif diikuti oleh sisaan negatif, dan sebaliknya, secara cepat.



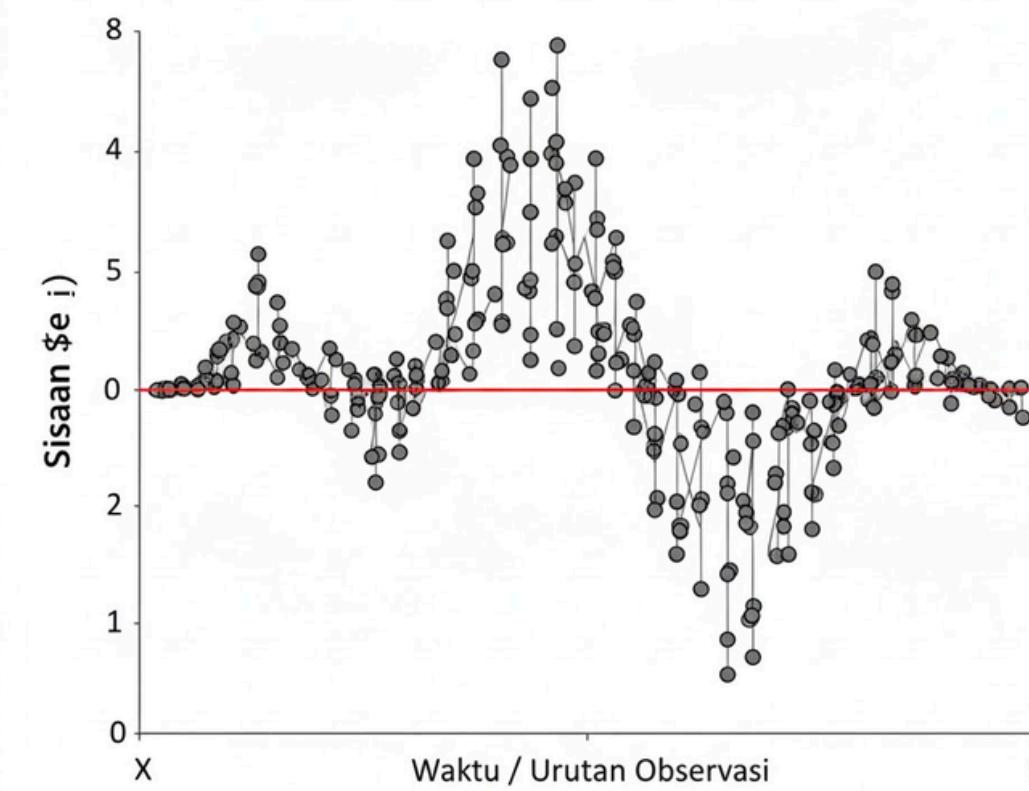


Identifikasi - Grafis

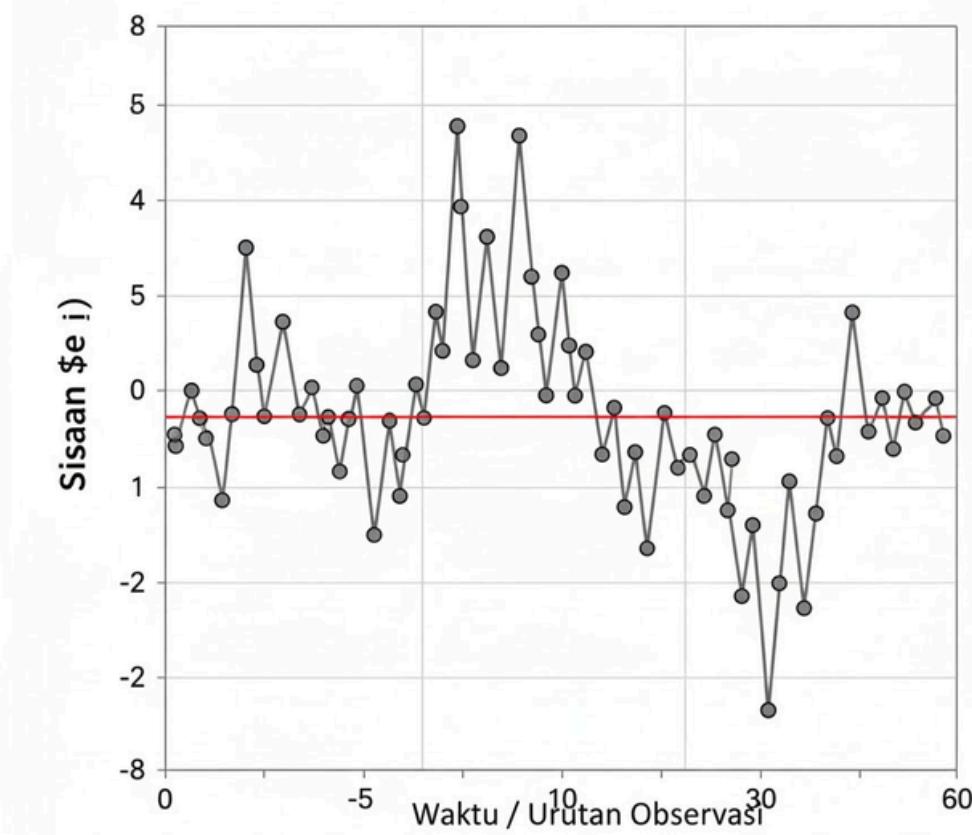
1. Plot Ideal (Non-Autokorelasi)



2. Plot Autokorelasi Positif



3. Plot Autokorelasi Negatif





Identifikasi - Uji Formal

Karena metode grafis bersifat subjektif, kita memerlukan uji statistik formal untuk mengkonfirmasi adanya autokorelasi. Tiga uji yang umum digunakan:

1. Uji Durbin-Watson (DW)
2. Uji Keacakan (Run Test)
3. Uji Breusch-Godfrey (LM Test)





Identifikasi - Uji Durbin Watson

Uji ini adalah uji klasik yang spesifik untuk mendeteksi autokorelasi orde pertama (AR(1)), yaitu

$$e_t = \rho e_{t-1} + v_t.$$

- Hipotesis (Uji Dua Sisi):

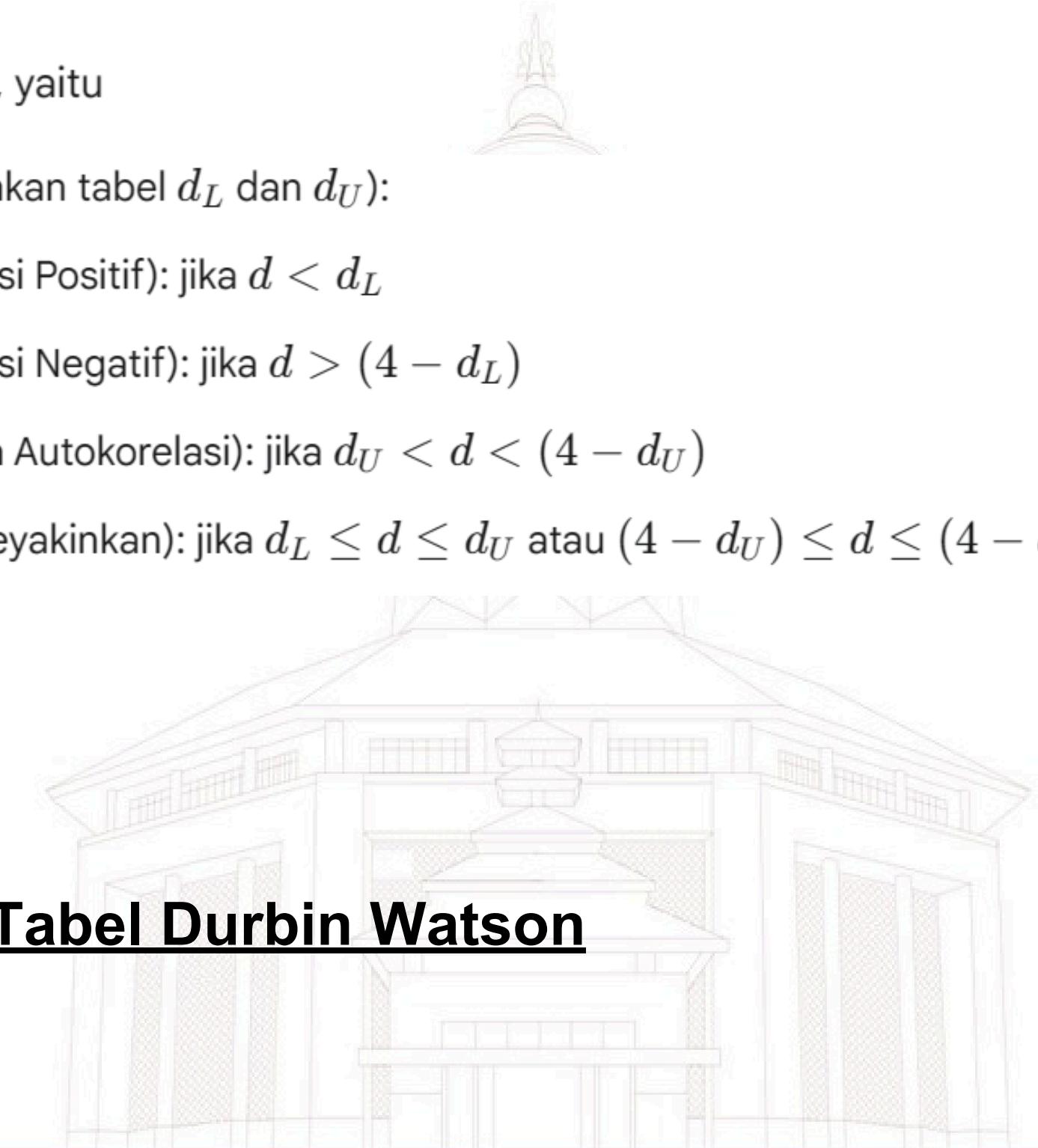
- $H_0 : \rho = 0$ (Tidak ada autokorelasi)
- $H_1 : \rho \neq 0$ (Ada autokorelasi)

- Statistik Uji:

- Kriteria Penolakan (Membutuhkan tabel d_L dan d_U):
 - Tolak H_0 (Ada Autokorelasi Positif): jika $d < d_L$
 - Tolak H_0 (Ada Autokorelasi Negatif): jika $d > (4 - d_L)$
 - Tidak Tolak H_0 (Tidak Ada Autokorelasi): jika $d_U < d < (4 - d_U)$
 - Area Inkonklusif (Tidak Meyakinkan): jika $d_L \leq d \leq d_U$ atau $(4 - d_U) \leq d \leq (4 - d_L)$

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

- Nilai d selalu berada antara 0 dan 4.
- $d \approx 2$ (artinya $\rho \approx 0$) → Tidak ada autokorelasi.
- $d \approx 0$ (artinya $\rho \approx 1$) → Autokorelasi positif.
- $d \approx 4$ (artinya $\rho \approx -1$) → Autokorelasi negatif.



Tabel Durbin Watson



Identifikasi - Uji Durbin Watson

keterbatasan dari Uji DW:

1. Hanya untuk AR(1): Uji ini tidak dirancang untuk mendeteksi pola autokorelasi orde tinggi (misal, AR(2), AR(4)/Musiman) yang lebih kompleks.
2. Tidak Valid untuk Model Dinamis: Uji DW TIDAK BOLEH digunakan jika model regresi menyertakan variabel dependen lag (misal, Y_{t-1}) sebagai salah satu prediktor X .
3. Membutuhkan Intersep: Model regresi harus memiliki konstanta (intersep).
4. Data Harus Lengkap: Uji ini mengasumsikan tidak ada data yang hilang (missing values) dalam urutan waktu.



Identifikasi - Uji Keacakan (Run Test)

Uji ini adalah uji non-parametrik yang mendekripsi apakah urutan sisaan (positif atau negatif) bersifat acak.

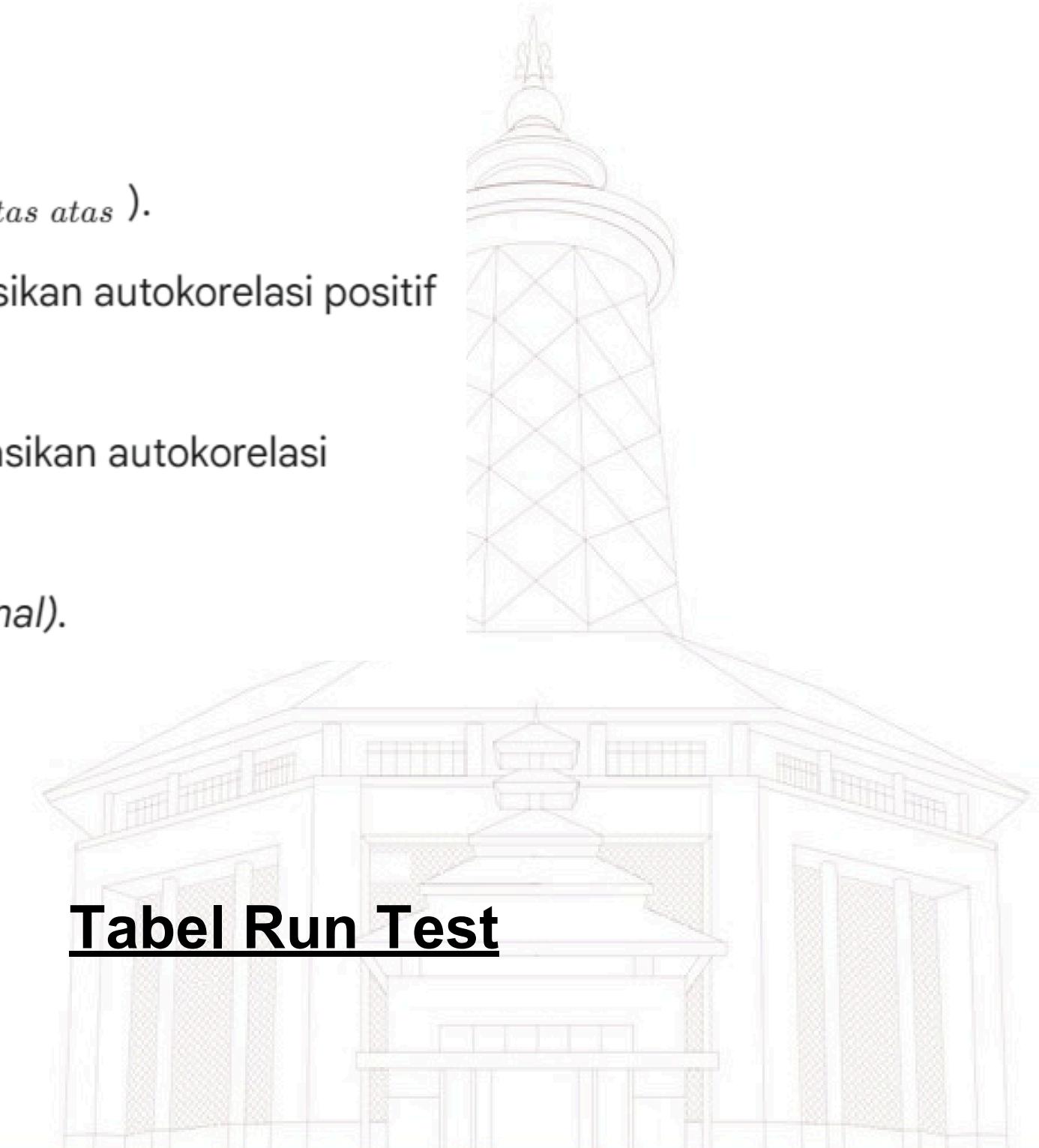
- Hipotesis:
 - H_0 : Urutan sisaan terjadi secara acak.
 - H_1 : Urutan sisaan terjadi tidak acak (ada pola/autokorelasi).
- Tahapan Uji:
 1. Jalankan regresi OLS dan dapatkan sisaan (e_i) dalam urutan aslinya (berdasarkan waktu atau urutan observasi).
 2. Klasifikasikan setiap sisaan berdasarkan tandanya: Positif (+) atau Negatif (-). (Abaikan nilai 0).
 3. Hitung jumlah sisaan positif (n_1) dan jumlah sisaan negatif (n_2).
 4. Hitung jumlah "run" (r). Sebuah "run" adalah serangkaian tanda yang sama yang tidak terputus.





Identifikasi - Uji Keacakan (Run Test)

- Kriteria Penolakan:
 - Bandingkan r dengan nilai kritis dari tabel ($r_{batas bawah}$ dan $r_{batas atas}$).
 - Tolak H_0 jika $r \leq r_{bb}$: Jumlah run *terlalu sedikit*. Ini mengindikasikan autokorelasi positif (pola berkelompok).
 - Tolak H_0 jika $r \geq r_{ba}$: Jumlah run *terlalu banyak*. Ini mengindikasikan autokorelasi negatif (pola bolak-balik).
 - *Jika n_1 atau $n_2 > 20$, dapat digunakan aproksimasi Uji-Z (Normal).*



Tabel Run Test



Identifikasi - Uji Keacakan (Run Test)

- Pola 1: + - + - + - + - + -
 - Jumlah Run (r) = 10
 - Interpretasi: Jumlah run sangat banyak (pola bolak-balik). Ini adalah indikasi kuat Autokorelasi Negatif.
- Pola 2: - - - - - + + + + +
 - Jumlah Run (r) = 2
 - Interpretasi: Jumlah run sangat sedikit (pola berkelompok). Ini adalah indikasi kuat Autokorelasi Positif.
- Pola 3: + + - - + - + + - -
 - Jumlah Run (r) = 6 (yaitu ++, --, +, -, ++, --)
 - Interpretasi: Jumlah run berada di "tengah-tengah", tidak terlalu banyak atau sedikit. Ini adalah pola yang diharapkan dari Urutan Acak (Tidak ada autokorelasi).





Identifikasi - Run Test jika n besar - Z Score

Langkah-langkahnya adalah mengubah jumlah run (r) yang kita amati menjadi skor-Z (Z_{hitung}).

1. Tentukan Variabel Dasar

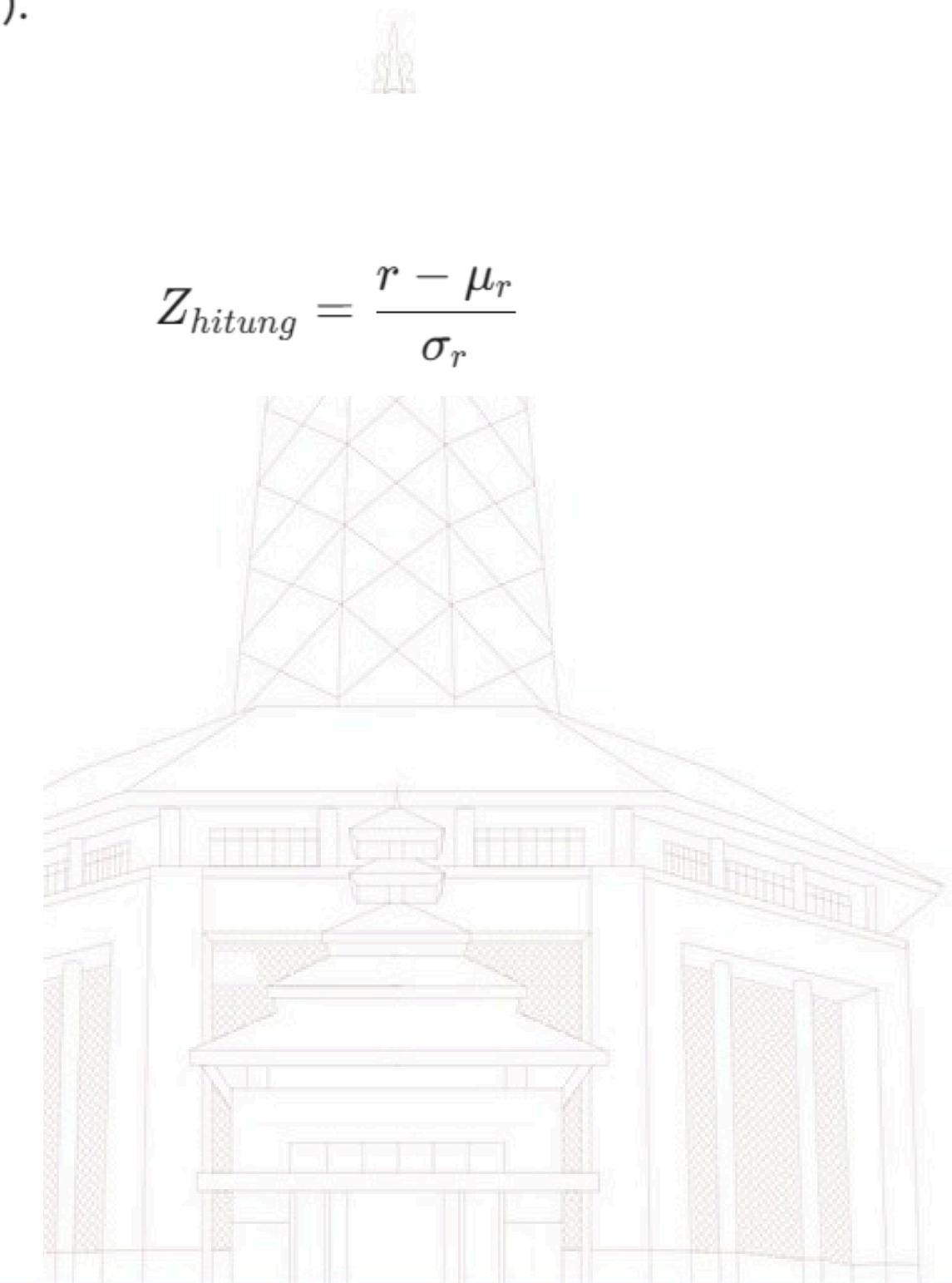
- n_1 = jumlah sisaan positif
- n_2 = jumlah sisaan negatif
- $N = n_1 + n_2$ (total sisaan non-nol)
- r = jumlah run yang diamati dari data

2. Hitung Mean (Nilai Harapan) dari r Ini adalah "rata-rata" jumlah run yang kita harapkan terjadi jika data benar-benar acak.

$$\mu_r = \frac{2n_1 n_2}{N} + 1$$

3. Hitung Standar Deviasi dari r Pertama, hitung variansnya:

$$\sigma_r^2 = \frac{2n_1 n_2 (2n_1 n_2 - N)}{N^2(N - 1)}$$



Identifikasi - Run Test jika n besar - Z Score

- H_0 : Urutan sisaan acak.
- H_1 : Urutan sisaan tidak acak (ada autokorelasi).
- Jika r lebih besar dari meannya (μ_r), kita kurangi 0.5:

$$Z_{hitung} = \frac{(r - 0.5) - \mu_r}{\sigma_r}$$

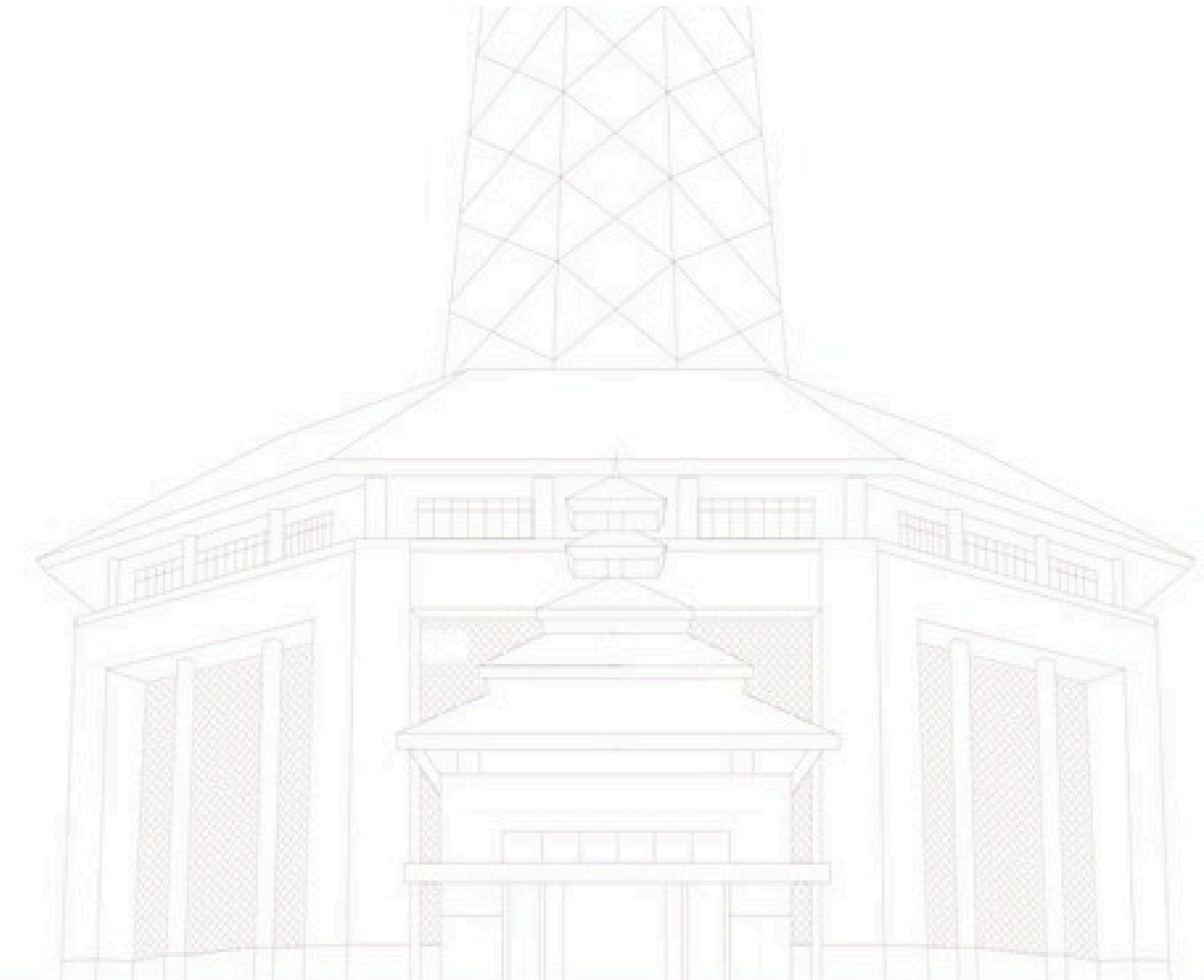
- Jika r lebih kecil dari meannya (μ_r), kita tambah 0.5:

$$Z_{hitung} = \frac{(r + 0.5) - \mu_r}{\sigma_r}$$

- **Tolak H_0 jika $|Z_{hitung}| > Z_{kritis}$** (misal, $|Z_{hitung}| > 1.96$).
- Atau, cara modern: **Tolak H_0 jika $p-value < \alpha$** .

Interpretasi Penting:

- Jika Z_{hitung} negatif signifikan (misal, -2.5): Ini berarti r (jumlah run) jauh lebih sedikit dari yang diharapkan (μ_r). Ini adalah tanda kuat **Autokorelasi Positif**.
- Jika Z_{hitung} positif signifikan (misal, +2.5): Ini berarti r (jumlah run) jauh lebih banyak dari yang diharapkan. Ini adalah tanda kuat **Autokorelasi Negatif**.





Identifikasi - Uji Breusch-Godfrey (BG / LM Test)

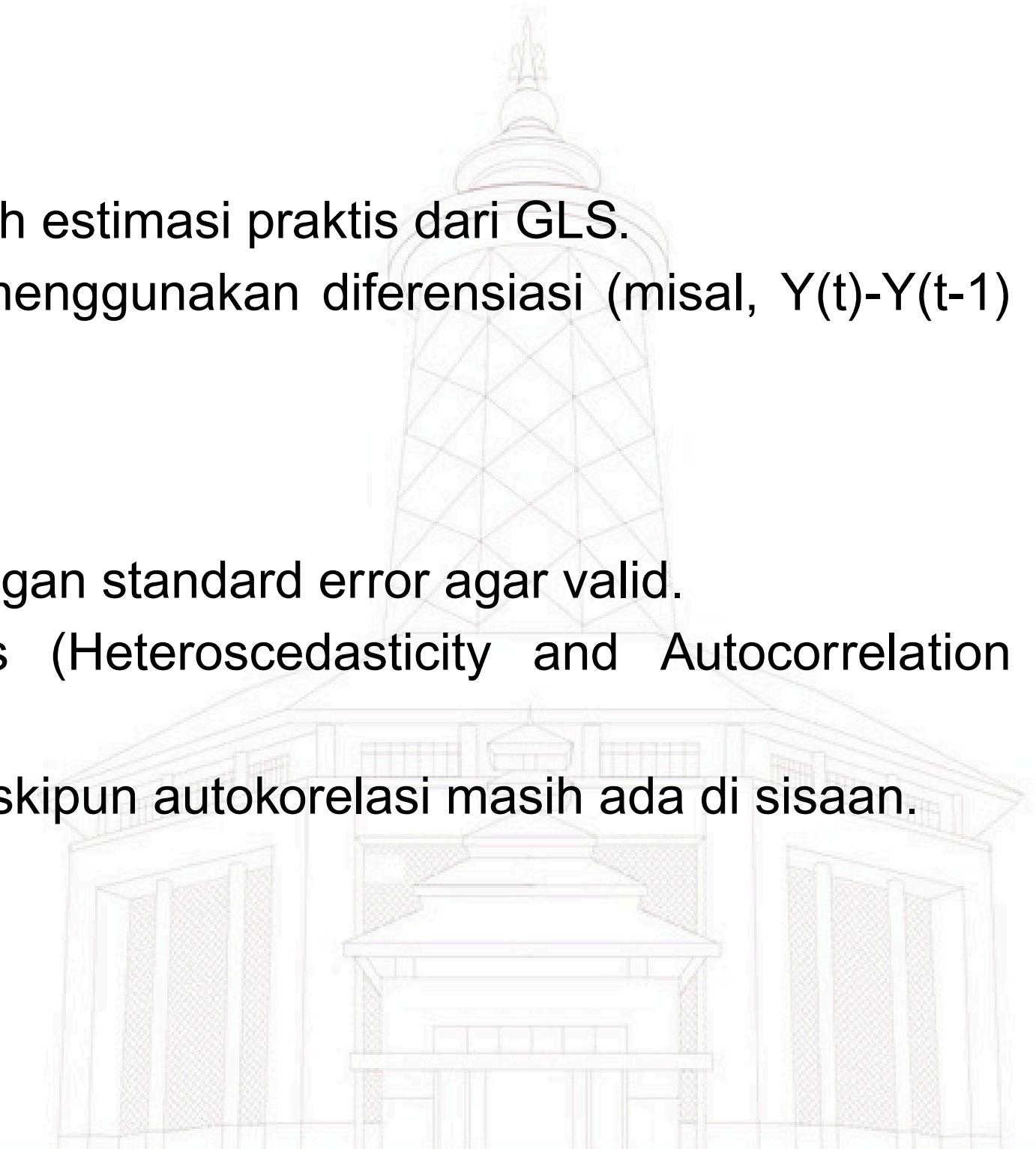
Keunggulan Uji BG:

1. Dapat menguji autokorelasi orde yang lebih tinggi (misal, $AR(p)$, bukan hanya $AR(1)$).
 2. Tetap valid digunakan meskipun model mengandung variabel dependen lag (Y_{t-1}).
- Hipotesis (untuk orde p):
 - $H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$ (Tidak ada autokorelasi sampai lag ke- p)
 - $H_1 : \text{Minimal ada satu } \rho_j \neq 0$
 - Tahapan Konseptual:
 1. Jalankan regresi OLS utama, dapatkan sisaan e_t .
 2. Jalankan regresi auxilier (pembantu): Regresikan e_t pada semua X asli DAN sisaan lag ($e_{t-1}, e_{t-2}, \dots, e_{t-p}$).
 3. Dapatkan R^2 dari regresi auxilier ini.
 4. Hitung statistik uji: $LM = (n - p) \times R_{aux}^2$
 5. Statistik LM ini mengikuti distribusi Chi-Square (χ^2) dengan $df = p$.
 - Keputusan: Tolak H_0 (berarti ada autokorelasi) jika $p - value < \alpha$.





Regresi Linier - Penanganan Autokorelasi



1. Metode Transformasi (Memperbaiki Model):

- Generalized Least Squares (GLS)
- Metode iteratif seperti Cochrane-Orcutt atau Prais-Winsten adalah estimasi praktis dari GLS.
- Diferensiasi Data: Untuk data time series yang non-stasioner, menggunakan diferensiasi (misal, $Y(t)-Y(t-1)$) seringkali dapat menghilangkan autokorelasi.

2. Koreksi Standard Error (Memperbaiki Uji):

- Model OLS tetap dipertahankan, namun kita mengoreksi perhitungan standard error agar valid.
- Metode yang paling terkenal adalah HAC Standard Errors (Heteroscedasticity and Autocorrelation Consistent), contohnya: Newey-West Standard Errors.
- Uji-t dan Uji-F yang dihasilkan menjadi lebih dapat dipercaya, meskipun autokorelasi masih ada di sisaan.



SEE YOU NEXT WEEK !

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

NIP. 199005202024061001

ferdian.bangkit@untirta.ac.id

