



RUANG BELAJAR DATA EPS. 6

Metode Analisis **Regresi**

Klik untuk memulai proyek!





Perkenalkan.. Saya Muhammad Iqbal Pillie!

Analyst • Presenter • Content Creator

Pekerjaan:

- Data Analytics Consultant, **StarCore Analytics**
- Instructor, **RevoU**

Pendidikan:

- Statistika, **IPB University**

Pengalaman Trainer:

- Statistika UNJ
- Jakarta Smart City
- BAPPEDA DKI Jakarta
- Bank BJB Syariah
- Mitra Mandiri Informatika
- **PUPR Ruang Belajar Data Eps. 5** 🔥

#Throwback: Eps. 5 (EDA)

Screenshot of a Microsoft Teams video call interface showing a presentation slide titled "Data Collection".

The slide features a hand holding a tablet displaying a "Data Collection" icon. The text "Data Collection" is written in large bold letters.

Below the icon, the text "Atau proses pengumpulan data." is displayed.

Sumber Data:
Direktorat Jenderal Bina Marga, Sumber Daya Air, Cipta Karya, Perumahan, Bina Konstruksi, dll. → [Open Data PUPR](#)

Contoh data:
Data progres proyek, anggaran, penggunaan sumber daya, kondisi lapangan.

The presentation slide is framed by a yellow border. On the right side of the slide, there is a screenshot of the "Open Data PUPR" website, which has a blue and yellow color scheme with various icons representing different data categories.

On the far right of the slide, there is a video feed showing three participants: M. Iqbal Pillie (Unverified), Meyta Utari, and iman (Unverified). Below the video feed, there is a slide with the letter "I" and a pink circle containing the letter "NA".

The Teams interface shows the following details:

- Activity sidebar on the left.
- Top bar with search, activity, and team names.
- Header bar with time (01:19:23), participant count (346), and various interaction buttons (Chat, Q&A, People, Raise, React, View, Notes, YouTube, Custom Str..., Apps, More).
- Bottom bar with camera, mic, share, and leave buttons.



Rute Perjalanan



Anda di sini!

**Konsep Dasar
Analisis Regresi**

**Miskonsepsi
Regresi Linier**

**Regresi Ft.
Open Data PUPR**

What's Next?



Kasus Umum



- Usia
- Asupan kalori harian
- Tingkat aktivitas fisik

TINGGI BADAN



- Waktu
- Investasi di R&D
- Tingkat pendidikan

GDP



- Dosis pupuk
- Penggunaan air
- Suhu

HASIL PANEN



- Luas bangunan
- Jumlah kamar tidur
- Usia bangunan

HARGA RUMAH



- Durasi belajar
- Tingkat kehadiran
- Keaktifan di kelas

HASIL UJIAN



- Pengeluaran iklan (*ads*)
- Jumlah *salesman*
- Harga produk

REVENUE

#CONTOH

Sebuah studi dilakukan oleh sebuah *real-estate agency* untuk **memodelkan harga rumah** berdasarkan luas bangunan, jumlah kamar, dan usia bangunan.

Dalam kasus ini:

Variabel dependen:

Harga rumah

Variabel independen:

- Luas bangunan
- Jumlah kamar tidur
- Usia bangunan





Mengapa Regresi?

1. **Prediksi**
 - Kita dapat **memprediksi nilai variabel dependen** berdasarkan model regresi (satu atau lebih variabel independen) yang dibentuk.
2. **Inferensi** (menyimpulkan hubungan) antara variabel independen & dependen
 - Kita dapat **menduga besaran dan arah hubungan** antara variabel independen dan dependen.



“Prediksi = Forecasting?”



Prediksi vs Forecasting

PREDIKSI:

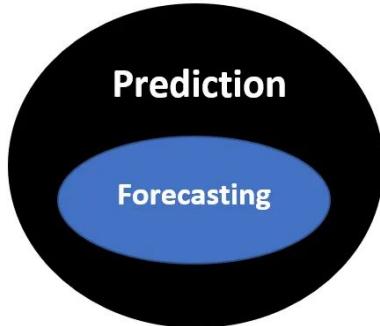
Menduga nilai Y menggunakan variabel lain

→ **Metode:** Regresi linier, Regresi logistik, Decision Tree, Random Forest, Support Vector Machine (SVM)), dll.

FORECASTING:

Menduga nilai Y di masa depan, biasanya melibatkan faktor waktu → Mempertimbangkan pola (Seasonality, Trend: Up or Down)

→ **Metode:** Exponential Smoothing, ARIMA, Vector Autoregressive (VAR), Long Short-Term Memory (LSTM), dll.



Forecasting is a sub-discipline of prediction in which we are making predictions about the future, on the basis of time-series data. Thus, the only difference between prediction and forecasting is that we

[Towardsdatascience.com](#)

| In summary, all forecasts are predictions but not all predictions are forecasts.

[Datascienceblog.net](#)



#SADARKAHANDA?



TINGGI BADAN

- Usia
- Asupan kalori harian
- Tingkat aktivitas fisik

Hubungan:
Eksponensial/Logaritmik



GDP

- Waktu
- Investasi di R&D
- Tingkat pendidikan

Hubungan:
Eksponensial



HASIL PANEN

- Dosis pupuk
- Penggunaan air
- Suhu

Hubungan:
Kuadratik (Polinomial)

NON LIN- EAR



HARGA RUMAH

- Luas bangunan
- Jumlah kamar tidur
- Usia bangunan

Hubungan:
Linier



HASIL UJIAN

- Durasi belajar
- Tingkat kehadiran
- Keaktifan di kelas

Hubungan:
Eksponensial



REVENUE

- Pengeluaran iklan (*ads*)
- Jumlah *salesman*
- Harga produk

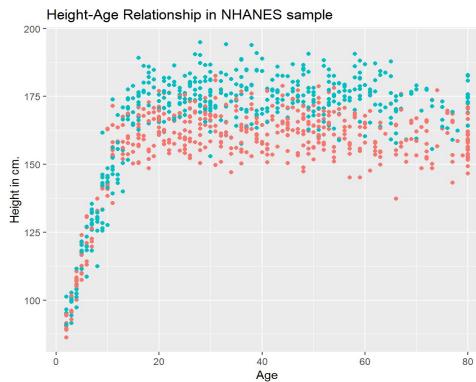
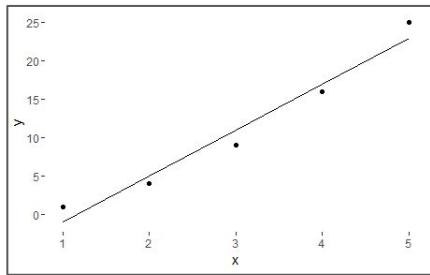
Hubungan:
Kuadratik (Polinomial)

LIN- EAR

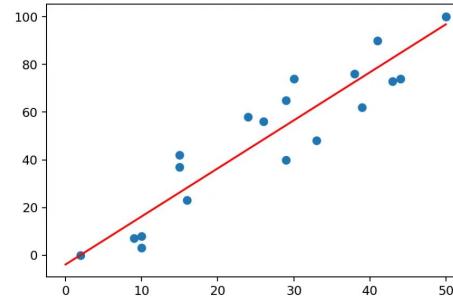


Linier vs Non-Linier

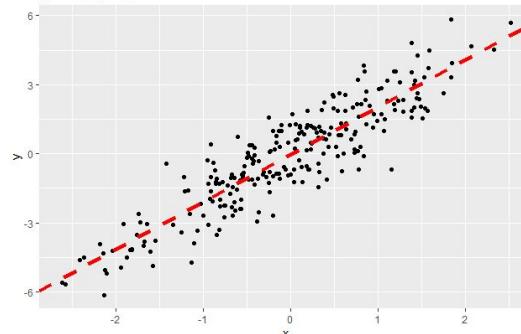
Contoh Regresi **Non-Linier**:



Contoh Regresi **Linier**:



geeksforgeeks



**FOKUS
HARI INI!**

#AgakTeknis

Konsep **Regresi Linier** adalah **mencari garis (line) yang paling mendekati pola data**, mengikuti persamaan matematis tertentu.

Sebagai ilustrasi, pada gambar di samping terlihat bahwa setiap titik data memiliki pasangan nilai X dan Y. Dari sekumpulan titik data tersebut akan terbentuk sebuah garis (**di gambar berupa garis merah**) yang menggambarkan posisi titik secara rata-rata. Garis tersebut merupakan persamaan regresi yang dapat ditulis sebagai berikut:

Regresi Linier Sederhana

$$Y = b_0 + b_1 X_1 + e$$

Y : variabel dependen
bo : intersep
e : error atau residual

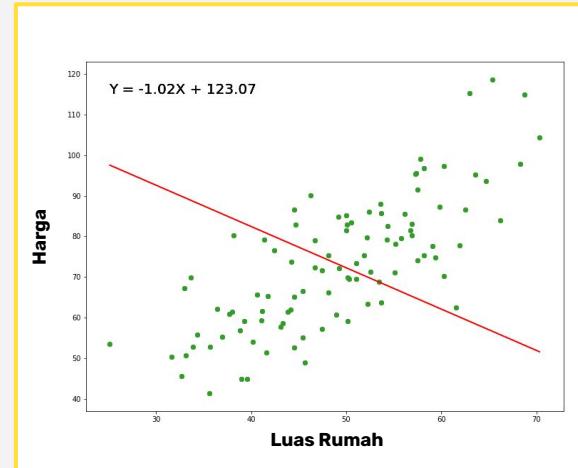
Regresi Linier Berganda

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e$$

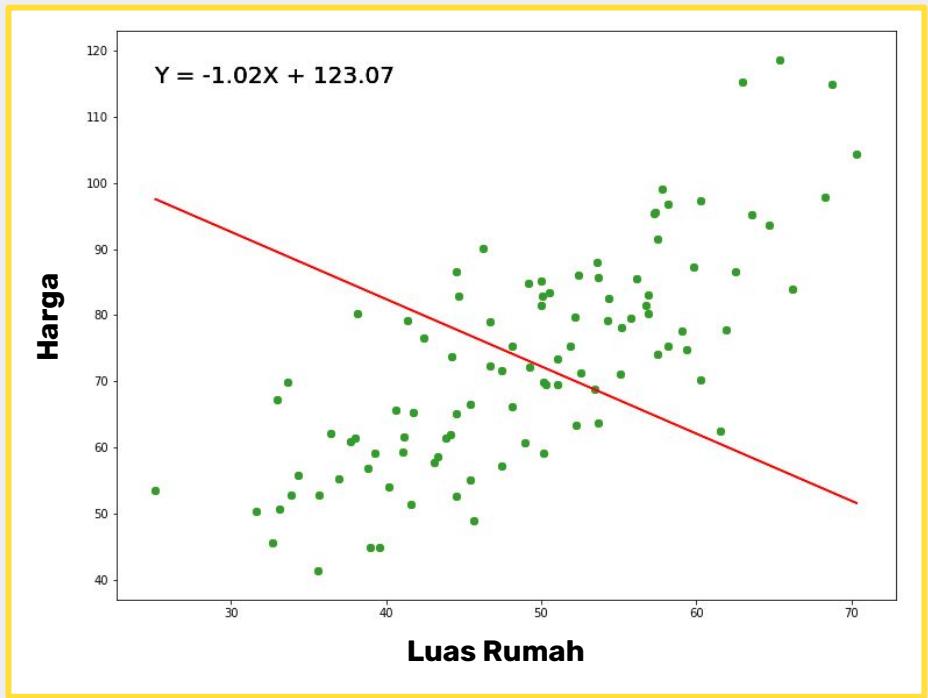
X : variabel independen
b1 : slope untuk variabel X1
b2 : slope untuk variabel X2
bn : slope untuk variabel ke-n

Contoh:

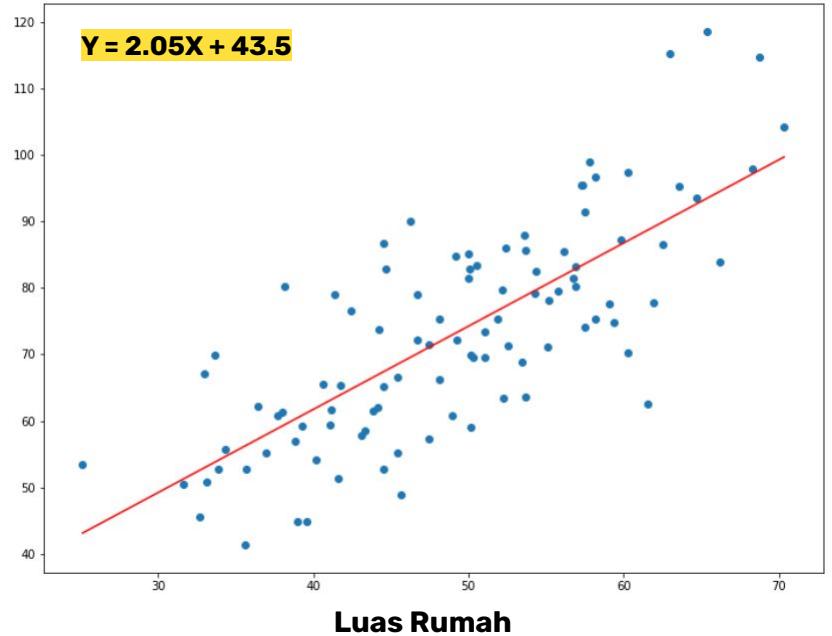
Hubungan harga vs luas rumah

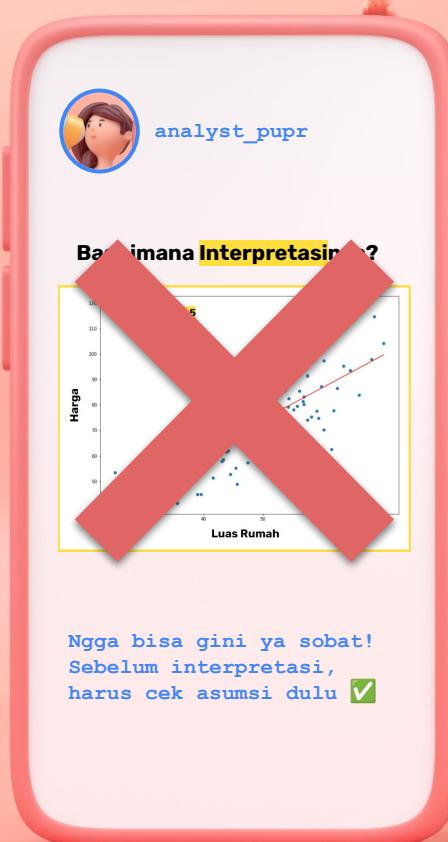


Ucapkan “Stop!”



Bagaimana Interpretasinya?





Asumsi Regresi Linier:

Linearity

Non multicollinearity

Homoscedasticity

Normality

Non auto-correlation



BLUE
(Best Linear Unbiased Estimators)

1. Linearity

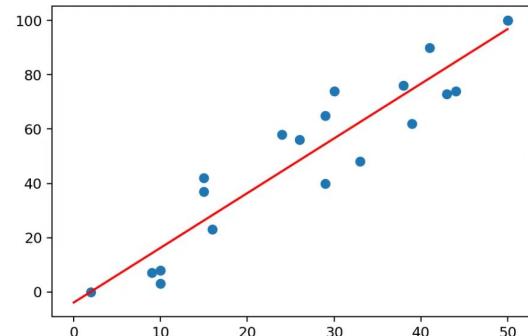
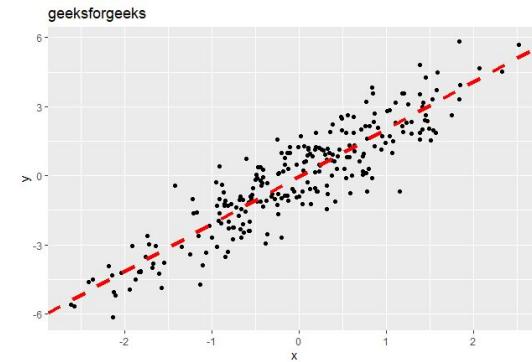
Merupakan kondisi **hubungan yang linier antara variabel independen (X) dengan variabel dependen (Y)**.

Cara mendekripsi:

- Kita akan pakai ini
- 1. Jika hanya ada **1 variabel X**:
→ Buat **Scatter Plot** antara **X** dan **Y**
 - 2. Jika ada ≥ 2 variabel X:
→ Buat **Scatter Plot** antara **residuals** dan **prediksi Y**

Asumsi terpenuhi jika **hubungan yang terlihat berbentuk linier**.

*Ilustrasi Linearity:



2. Non Multicollinearity

Merupakan kondisi hubungan yang **independen*** antar **sesama variabel independen (X)**.

*Independen: Berkorelasi relatif rendah.

Menurut [Dormann et al. \(2012\)](#) :



*Folk-lore threshold:
|Korelasi| > 0.7*

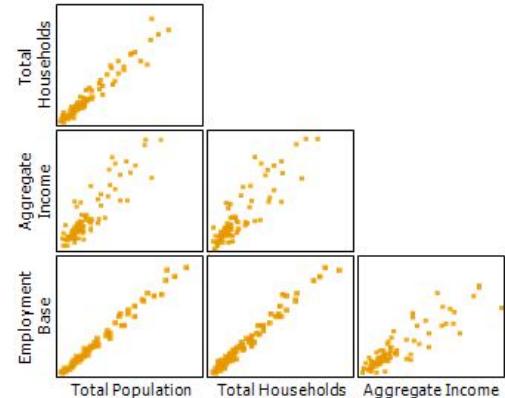
Dianggap korelasi tinggi (kondisi multikolinearitas)

final interpretation. However, all approaches tested yielded degraded predictions under change in collinearity structure and the 'folk lore'-thresholds of correlation coefficients between predictor variables of $|r| > 0.7$ was an appropriate indicator for when collinearity begins to severely distort model estimation and subsequent prediction. The

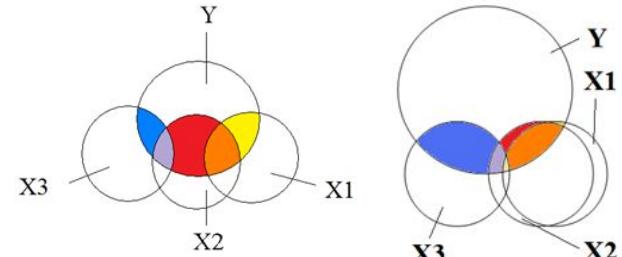
Cara mendekati:

1. Cek nilai korelasi menggunakan **Folk-lore threshold**.
Asumsi terpenuhi jika nilai korelasi: **$-0.7 \leq r \leq 0.7$** .
2. Cek nilai VIF. Kondisi terpenuhi jika nilai **VIF < 10**. Ada pendapat lain juga menggunakan $VIF < 5$ dan $VIF < 2.5$.

***Ilustrasi Multicollinearity:**



Moderate Multicollinearity Extreme Multicollinearity



3. Homoscedasticity

Merupakan kondisi **keragaman/variansi homogen*** antara **standard residuals sepanjang nilai prediksi Y**.

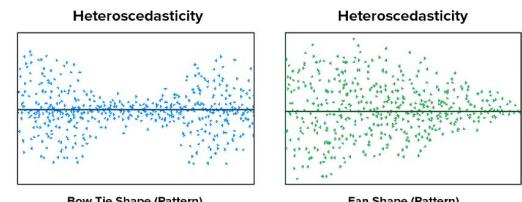
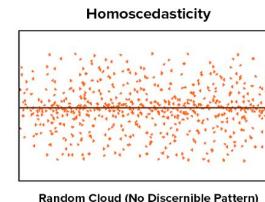
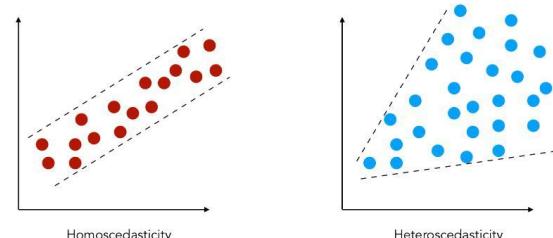
*Homogen: Bentuk lebar sebaran (corong) relatif sama dari ujung ke ujung.

Kita
akan
pakai
ini

Cara mendekksi:

1. Buat **Scatter Plot** antara **standard residuals vs prediksi Y**. Semakin homogen sebaran titik dari ujung ke ujung, semakin dianggap memenuhi asumsi homoscedasticity.
2. Lakukan uji **Bartlett's Test**, **Box's M Test**, atau **Levene's Test**.

*Ilustrasi Homoscedasticity vs Heteroscedasticity:



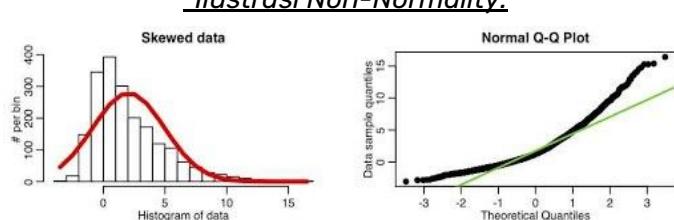
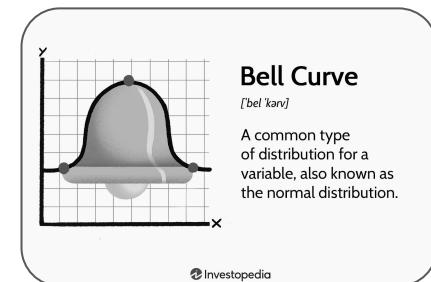
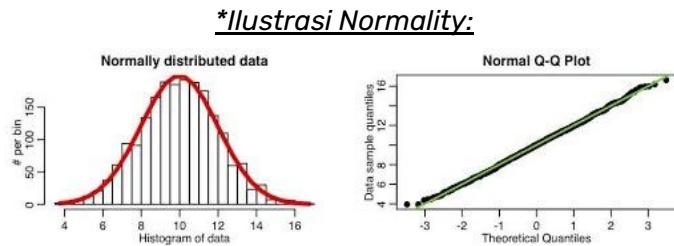
4. Normality

Merupakan kondisi **nilai residuals yang berdistribusi/menyebar normal**.

Kita akan pakai ini

Cara mendekripsi:

1. Buat **Scatter Plot** antara **residuals vs z-score** pada sebuah **Q-Q Plot**.
2. Buat **Histogram** nilai residuals. Semakin berbentuk lonceng, semakin dianggap memenuhi asumsi normality.
3. Lakukan uji **Jarque Bera test, atau Kolmogorov Smirnov Test**.



5. Non Autocorrelation

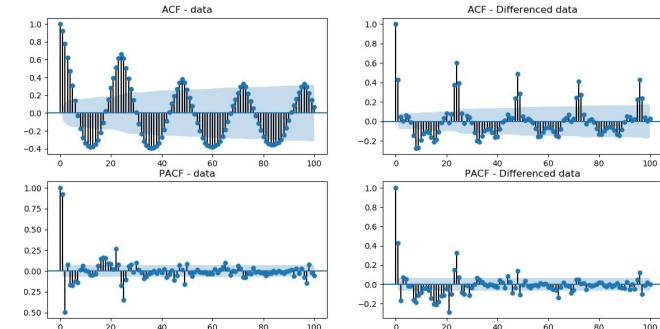
Merupakan kondisi **nilai residuals yang independen (tidak ada korelasi serial) sepanjang waktu.**

Umumnya, asumsi ini dicek untuk **data time-series**. **Pada banyak kasus data cross-sectional, asumsi ini diabaikan**, kecuali ada potensi korelasi spasial atau faktor lainnya menyesuaikan studi kasus.

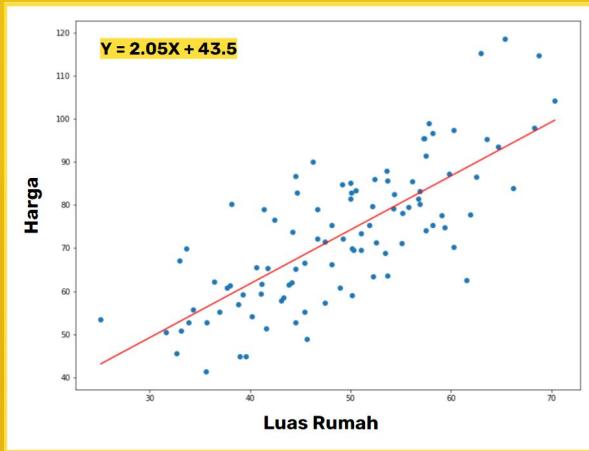
Cara mendekripsi:

1. Buat **Scatter Plot** antara **residuals vs urutan waktu**. Semakin tidak berpola, semakin dianggap memenuhi asumsi non-autocorrelation.
2. Buat **Correlogram Plot**.
3. Lakukan uji **Durbin Watson test, atau Ljung-Box Q Test**.

***Ilustrasi Non-Autocorrelation:**



ASUMSI ✓ >> INTERPRETASI ✓



Misalnya:

- **R-Square = 92%**

→ Model yang dibentuk oleh variabel luas rumah dapat menjelaskan 92% keragaman harga rumah. Sedangkan 8% lainnya merupakan faktor lain yang tidak ada di dalam model.

- **b1 (slope) = 2.05**

→ Setiap kenaikan 1 satuan luas rumah (misal: m²), terjadi peningkatan 2.05 satuan harga rumah (misal: juta Rupiah).

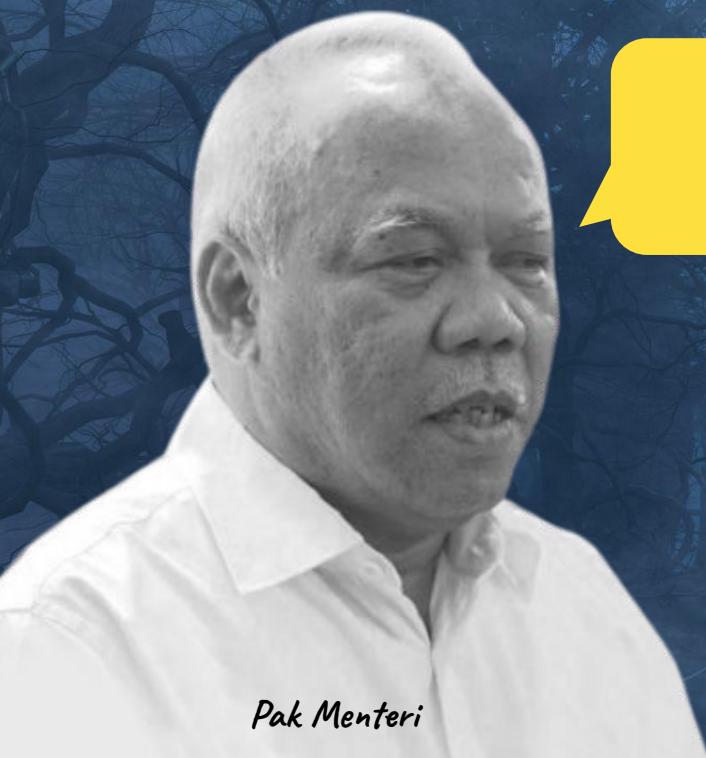
- **b0 (intercept) = 43.5**

→ Harga rumah minimum apabila luas rumah = 0 m² adalah 43.5 juta Rupiah.

**Namun, kondisi ini kurang relevan untuk diinterpretasikan karena tidak mungkin luas rumah 0 m².*

Bayangkan ...

*"Mas, Mba..
udah paham?"*



Pak Menteri

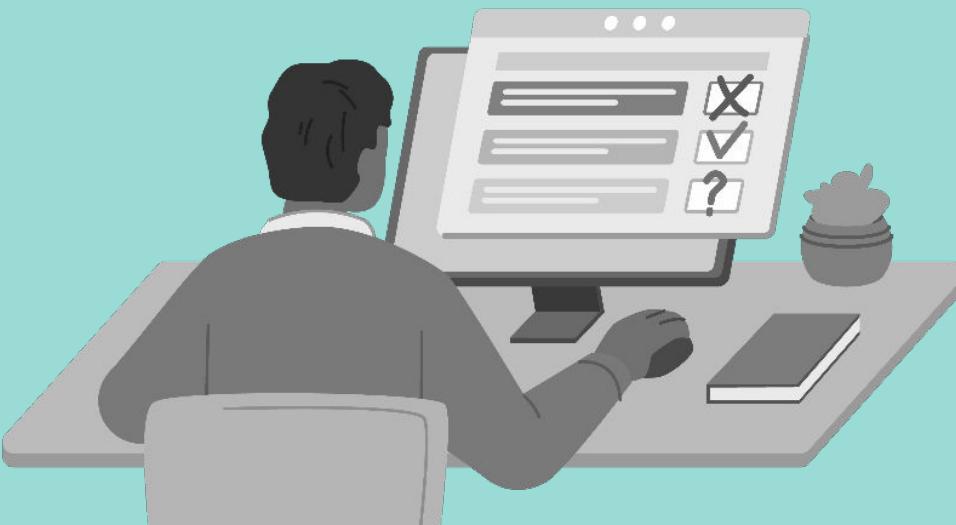


Ini anda

Hands-on

Regresi Linier

Sumber Data: [BPS](#) dan [Open Data PUPR](#)



Silakan akses dataset di sini:
bit.ly/DataRBD6

Latar Belakang

BPS:



Sejak tahun **2019**, rumah tangga diklasifikasikan memiliki akses terhadap hunian/**rumah layak huni** apabila memenuhi **4 kriteria**, yaitu:

- 1) kecukupan **luas** tempat tinggal **minimal 7,2 m² per kapita** (sufficient living space)
- 2) memiliki **akses** terhadap **air minum layak**
- 3) memiliki **akses** terhadap **sanitasi layak**
- 4) **ketahanan bangunan** (durable housing), yaitu atap terluas berupa beton/genteng/seng/kayu/sirap; dinding terluas berupa tembok/plesteran anyaman bambu/kawat, kayu/papan dan batang kayu; dan lantai terluas berupa marmer/granit/keramik/parket/vinil/karpet/ubin/tegel/teraso/kayu/papan/semen/bata merah.

<https://www.bps.go.id/id/statistics-table/2/MTIOMSMY/persentase-rumah-tangga-yang-memiliki-akses-terhadap-hunian-yang-layak-dan-terjangkau-menurut-provinsi--persen-.html>



Silakan akses dataset di sini:
bit.ly/DataRBD6



Data

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

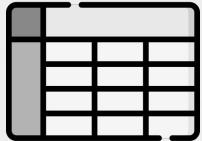
- 1. X1: % RT dengan Akses Layanan Air Minum
- 2. X2: % RT dengan Jenis Atap Terluas yang Layak
- 3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan
- 4. X4: % RT dengan Sumber Air Minum Layak
- 5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman
- 6. X6: % RT dengan Rumah Milik Sendiri
- 7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air
- 8. X8: % RT dengan Luas Hunian per Kapita $\leq 7,2 \text{ m}^2$
- 9. X9: % RT dengan Jenis Lantai Terluas yang Layak
- 10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat
- 11. X11: % RT yang Masih BAB sembarangan di Tempat Terbuka
- 12. X12: Jumlah SPAM



Level: Provinsi

Tahun: 2023





Data (Rule 1)

1) kecukupan luas tempat tinggal minimal 7,2 m² per kapita (sufficient living space)

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

1. X1: % RT dengan Akses Layanan Air Minum
2. X2: % RT dengan Jenis Atap Terluas yang Layak
3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan
4. X4: % RT dengan Sumber Air Minum Layak
5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman
6. X6: % RT dengan Rumah Milik Sendiri
7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air

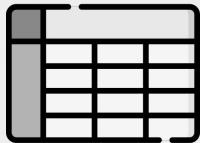
8. X8: % RT dengan Luas Hunian per Kapita <= 7,2 m²

9. X9: % RT dengan Jenis Lantai Terluas yang Layak
10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat
11. X11: % RT yang Masih BAB sembarangan di Tempat Terbuka
12. X12: Jumlah SPAM

Level: Provinsi

Tahun: 2023





Data (Rule 2)

2) memiliki akses terhadap air minum layak

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

1. X1: % RT dengan Akses Layanan Air Minum

2. X2: % RT dengan Jenis Atap Terluas yang Layak

3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan

4. X4: % RT dengan Sumber Air Minum Layak

5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman

6. X6: % RT dengan Rumah Milik Sendiri

7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air

8. X8: % RT dengan Luas Hunian per Kapita $\leq 7,2 \text{ m}^2$

9. X9: % RT dengan Jenis Lantai Terluas yang Layak

10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat

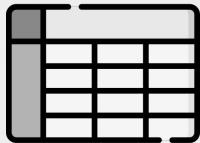
11. X11: % RT yang Masih BAB sembarangan di Tempat Terbuka

12. X12: Jumlah SPAM

Level: Provinsi

Tahun: 2023





Data (Rule 3)

3) memiliki akses terhadap sanitasi layak

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

1. X1: % RT dengan Akses Layanan Air Minum
2. X2: % RT dengan Jenis Atap Terluas yang Layak
3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan
4. X4: % RT dengan Sumber Air Minum Layak

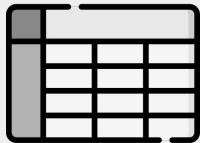
5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman

6. X6: % RT dengan Rumah Milik Sendiri
7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air
8. X8: % RT dengan Luas Hunian per Kapita <= 7,2 m²
9. X9: % RT dengan Jenis Lantai Terluas yang Layak
10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat
11. X11: % RT yang Masih BAB sembarangan di Tempat Terbuka
12. X12: Jumlah SPAM

Level: Provinsi

Tahun: 2023





Data (Rule 4)

4) ketahanan bangunan (durable housing), yaitu atap terluas berupa beton/ genteng/ seng/ kayu/ sirap; dinding terluas berupa tembok/ plesteran anyaman bambu/kawat, kayu/papan dan batang kayu; dan lantai terluas berupa marmer/ granit/ keramik/ parket/vinil/karpet/ ubin/tegel/teraso/ kayu/papan/ semen/bata merah.

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

1. X1: % RT dengan Akses Layanan Air Minum

2. X2: % RT dengan Jenis Atap Terluas yang Layak

3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan

4. X4: % RT dengan Sumber Air Minum Layak

5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman

6. X6: % RT dengan Rumah Milik Sendiri

7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air

8. X8: % RT dengan Luas Hunian per Kapita <= 7,2 m²

9. X9: % RT dengan Jenis Lantai Terluas yang Layak

10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat

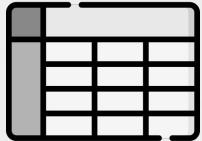
11. X11: % RT yang Masih BAB sembarangan di Tempat Terbuka

12. X12: Jumlah SPAM

Level: Provinsi

Tahun: 2023





Data (Additional)

- Dependent Variable

Y: % RT Memiliki Akses Hunian Layak & Terjangkau

- Independent Variable

1. X1: % RT dengan Akses Layanan Air Minum
2. X2: % RT dengan Jenis Atap Terluas yang Layak
3. X3: % RT dengan Sumber Air Minum Air Dalam Kemasan
4. X4: % RT dengan Sumber Air Minum Layak
5. X5: % RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman

6. X6: % RT dengan Rumah Milik Sendiri

7. X7: % RT dengan Fasilitas Cuci Tangan Dengan Sabun Dan Air

8. X8: % RT dengan Luas Hunian per Kapita $\leq 7,2 \text{ m}^2$
9. X9: % RT dengan Jenis Lantai Terluas yang Layak
10. X10: % RT dengan Ketahanan Bangunan (Atap, Lantai, Dinding) Memenuhi Syarat

11. X11: % RT yang Masih BAB sembarang di Tempat Terbuka

12. X12: Jumlah SPAM

Level: Provinsi

Tahun: 2023



Install XLMiner Analysis ToolPak dulu ya!

Extensions Help

Add-ons Get add-ons Manage add-ons View document add-ons

Macros Apps Script

AppSheet

← Google Workspace Marketplace X

XLMiner Analysis ToolPak

All filters (1) Works with Sheets Price

Search results for XLMiner Analysis ToolPak

Google doesn't verify reviews or ratings. [Learn more about reviews and results](#)

XLMiner Analysis ToolPak Installed

Frontline Systems Inc.

Statistical analysis in Google Sheets, with functions matching the Analysis ToolPak in desktop...

3.2 ⚡ 2M+

Logic Sheet

Automate your Sheets And save your precious time

Logic Sheet - Automate ... Logic Sheet

The best Google Sheets automation tool. Send emails, Slack messages automatically.

4.6 ⚡ 200K+

Power Formulas

=PowerFormulas()

Powerful formulas for Sheets

=API()

Import data from API/JSON

=SQL()

Talk to databases in Sheets

=AI()

Ask GPT-3 about everything

Power Formulas - Advan... Power Formulas

A set of powerful formulas that help you import data from API, SQL, AI (GPT) and Airtable.

5.0 ⚡ 2K+

Step 1: Join/Lookup Berbagai Tabel ke Dalam Satu Tabel Sendiri



Tips: Gunakan fungsi VLOOKUP

Y ▾ X1 ▾ X2 ▾ X3 ▾ X4 ▾ X5 ▾ X6 ▾ X7 ▾ X8 ▾ X9 ▾ X10 ▾ X11 ▾ X12 ▾

| PROVINSI | Y | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| ACEH | 65.91 | 39.36 | 98.00 | 43.52 | 89.74 | 17.19 | 84.12 | 72.59 | 7.64 | 97.68 | 95.45 | 9.00 | 24.00 |
| SUMATERA UTARA | 70.95 | 45.33 | 95.30 | 39.25 | 92.19 | 7.14 | 71.46 | 73.92 | 7.78 | 98.33 | 92.44 | 4.78 | 40.00 |
| SUMATERA BARAT | 59.85 | 40.98 | 98.79 | 43.76 | 85.59 | 8.57 | 72.61 | 88.77 | 6.24 | 99.49 | 97.70 | 9.31 | 30.00 |
| RIAU | 71.53 | 35.29 | 97.32 | 54.95 | 90.47 | 14.64 | 77.56 | 72.29 | 6.42 | 99.26 | 96.43 | 2.64 | 30.00 |
| JAMBI | 64.12 | 51.42 | 97.48 | 26.49 | 80.02 | 9.59 | 87.28 | 73.41 | 4.32 | 99.06 | 95.95 | 5.45 | 38.00 |
| SUMATERA SELATAN | 61.82 | 56.57 | 92.76 | 21.70 | 87.19 | 4.54 | 84.71 | 75.24 | 10.45 | 98.32 | 90.56 | 5.38 | 45.00 |
| BENGKULU | 54.74 | 48.85 | 97.36 | 14.59 | 73.08 | 4.05 | 88.38 | 81.41 | 6.08 | 98.83 | 95.36 | 5.12 | 22.00 |
| LAMPUNG | 63.15 | 56.07 | 90.88 | 22.95 | 82.78 | 3.72 | 92.40 | 79.39 | 2.06 | 97.64 | 87.70 | 1.64 | 59.00 |
| KEP. BANGKA BELITUNG | 32.57 | 23.01 | 43.23 | 64.33 | 81.64 | 6.12 | 88.65 | 88.17 | 3.18 | 99.84 | 43.08 | 1.85 | 30.00 |
| KEP. RIAU | 54.21 | 18.46 | 63.85 | 75.76 | 92.10 | 12.16 | 72.97 | 85.97 | 4.15 | 99.82 | 63.31 | 0.46 | 30.00 |
| DKI JAKARTA | 38.80 | 25.45 | 46.10 | 79.39 | 99.42 | 23.10 | 56.57 | 76.83 | 21.35 | 99.88 | 46.02 | 0.13 | 3.00 |
| JAWA BARAT | 54.17 | 40.74 | 84.91 | 49.03 | 93.86 | 10.49 | 83.38 | 82.03 | 7.60 | 97.62 | 77.68 | 2.52 | 117.00 |
| JAWA TENGAH | 68.85 | 53.88 | 93.86 | 29.78 | 93.76 | 10.83 | 91.05 | 86.15 | 1.65 | 91.46 | 85.11 | 2.90 | 34.00 |
| DI YOGYAKARTA | 85.79 | 68.16 | 97.27 | 22.76 | 96.69 | 16.23 | 86.43 | 84.29 | 3.81 | 97.99 | 94.26 | 0.59 | 84.00 |
| JAWA TIMUR | 70.74 | 53.04 | 93.94 | 35.75 | 96.01 | 10.44 | 90.92 | 83.37 | 3.19 | 94.10 | 87.32 | 5.30 | 74.00 |
| BANTEN | 63.06 | 37.6 | 81.11 | 56.82 | 92.95 | 15.39 | 85.67 | 82.79 | 4.82 | 96.14 | 76.24 | 4.95 | 68.00 |
| BALI | 84.26 | 44.62 | 94.41 | 40.69 | 98.31 | 15.42 | 85.24 | 90.54 | 6.14 | 99.35 | 93.66 | 2.61 | 25.00 |
| NUSA TENGGARA BARAT | 66.31 | 52.61 | 88.12 | 29.15 | 96.03 | 6.21 | 91.35 | 75.48 | 9.14 | 98.25 | 84.23 | 8.49 | 29.00 |
| NUSA TENGGARA TIMUR | 42.70 | 37.59 | 96.16 | 9.33 | 88.35 | 2.50 | 90.74 | 43.50 | 14.56 | 83.87 | 63.24 | 5.84 | 54.00 |
| KALIMANTAN BARAT | 64.16 | 56.53 | 98.26 | 23.70 | 82.08 | 3.38 | 91.43 | 77.59 | 4.62 | 99.70 | 97.92 | 6.08 | 71.00 |
| KALIMANTAN TENGAH | 56.49 | 27.98 | 94.04 | 50.07 | 77.72 | 11.16 | 81.92 | 75.67 | 5.73 | 99.52 | 93.27 | 2.53 | 40.00 |
| KALIMANTAN SELATAN | 57.50 | 41.28 | 91.28 | 37.59 | 76.29 | 8.70 | 83.09 | 83.31 | 6.96 | 99.49 | 90.10 | 1.93 | 31.00 |
| KALIMANTAN TIMUR | 75.82 | 16.51 | 96.54 | 79.19 | 87.90 | 9.14 | 75.14 | 78.87 | 5.78 | 99.74 | 96.22 | 1.06 | 32.00 |
| KALIMANTAN UTARA | 71.43 | 22.62 | 99.66 | 73.76 | 90.19 | 5.84 | 76.63 | 76.24 | 7.12 | 99.97 | 99.53 | 1.22 | 18.00 |
| SULAWESI UTARA | 72.39 | 37.1 | 98.13 | 51.40 | 94.37 | 6.22 | 79.47 | 86.48 | 7.42 | 97.01 | 93.78 | 5.67 | 61.00 |
| SULAWESI TENGAH | 59.60 | 36.95 | 95.86 | 44.20 | 86.85 | 7.16 | 88.44 | 78.81 | 11.17 | 97.76 | 93.46 | 10.40 | 37.00 |
| SULAWESI SELATAN | 71.88 | 45.07 | 97.47 | 38.23 | 92.12 | 12.83 | 87.68 | 85.52 | 5.71 | 99.24 | 86.73 | 2.26 | 49.00 |
| SULAWESI TENGGARA | 74.01 | 44.58 | 95.44 | 41.73 | 94.80 | 1.93 | 90.54 | 84.74 | 8.06 | 97.74 | 92.53 | 4.34 | 65.00 |
| GORONTALO | 70.75 | 26.66 | 99.21 | 62.64 | 96.00 | 4.14 | 85.14 | 82.44 | 11.59 | 99.26 | 97.50 | 9.42 | 20.00 |
| SULAWESI BARAT | 59.28 | 35.49 | 96.71 | 28.56 | 79.86 | 2.31 | 93.35 | 77.18 | 8.12 | 97.63 | 91.07 | 8.88 | 42.00 |
| MALUKU | 63.61 | 39.85 | 95.12 | 24.60 | 92.98 | 4.39 | 84.69 | 74.11 | 11.36 | 96.33 | 91.85 | 10.41 | 194.00 |
| MALUKU UTARA | 68.17 | 45.42 | 97.84 | 25.27 | 89.01 | 1.63 | 90.26 | 83.62 | 5.82 | 95.69 | 93.84 | 5.37 | 69.00 |
| PAPUA BARAT | 56.92 | 30.72 | 98.51 | 41.50 | 81.57 | 4.63 | 82.94 | 65.38 | 10.61 | 97.98 | 96.32 | 3.38 | 71.00 |
| PAPUA | 29.01 | 25.86 | 75.76 | 30.22 | 66.49 | 1.82 | 85.31 | 31.78 | 24.19 | 75.78 | 67.19 | 24.30 | 21.00 |

Step 2.1: Cek Non Multikolinearitas Antar Variabel Independen (X)



Tips: Gunakan fitur **Correlation** dari XLMiner Analysis Toolpak

XLMiner Analysis ToolPak

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation**
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Linear Regression
- Logistic Regression
- Moving Average
- Random Number Generation
- Rank and Percentile
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two-Sample for Means



| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|-----|------|------|------|------|------|------|------|------|------|-----|-----|-----|
| X1 | 1.0 | | | | | | | | | | | |
| X2 | 0.4 | 1.0 | | | | | | | | | | |
| X3 | -0.8 | -0.5 | 1.0 | | | | | | | | | |
| X4 | 0.1 | 0.0 | 0.3 | 1.0 | | | | | | | | |
| X5 | 0.0 | -0.3 | 0.4 | 0.5 | 1.0 | | | | | | | |
| X6 | 0.5 | 0.4 | -0.7 | -0.2 | -0.5 | 1.0 | | | | | | |
| X7 | 0.2 | 0.0 | 0.3 | 0.5 | 0.3 | 0.0 | 1.0 | | | | | |
| X8 | -0.4 | -0.3 | 0.0 | -0.1 | -0.1 | -0.3 | -0.7 | 1.0 | | | | |
| X9 | 0.0 | 0.0 | 0.4 | 0.3 | 0.3 | -0.3 | 0.8 | -0.5 | 1.0 | | | |
| X10 | 0.4 | 0.9 | -0.3 | 0.0 | -0.2 | 0.2 | 0.2 | -0.4 | 0.3 | 1.0 | | |
| X11 | -0.1 | 0.1 | -0.3 | -0.4 | -0.4 | 0.2 | -0.6 | 0.6 | -0.7 | 0.0 | 1.0 | |
| X12 | 0.3 | 0.2 | -0.3 | 0.2 | -0.2 | 0.3 | 0.0 | -0.1 | 0.0 | 0.1 | 0.0 | 1.0 |

Kesimpulan:

Variabel X3, X7, dan X9 dihapus karena muncul >1x

Variabel X10 (kelayakan hunian) juga dapat dihapus karena sudah terwakili oleh X2 (atap)

*Asumsi: Variabel dengan korelasi ≥ 0.7 atau ≤ -0.7 akan dihapus

Korelasi Tinggi:

| | | |
|------------|-----------|------|
| X3 | X1 | -0.8 |
| X6 | X3 | -0.7 |
| X8 | X7 | -0.7 |
| X9 | X7 | 0.8 |
| X10 | X2 | 0.9 |
| X11 | X9 | -0.7 |

Kemunculan:

| | | | | | | | | |
|----|----|-----------|----|-----------|----|-----------|-----|-----|
| X1 | X2 | X3 | X6 | X7 | X8 | X9 | X10 | X11 |
| 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 |

Step 2.2: Cek Non Multikolinearitas & Hapus Variabel Gejala Multikolinearitas



Tips: Gunakan fitur **Correlation** dari XLMiner Analysis Toolpak

XLMiner Analysis ToolPak

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation**
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Linear Regression
- Logistic Regression
- Moving Average
- Random Number Generation
- Rank and Percentile
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two-Sample for Means



| | X1 | X2 | X4 | X5 | X6 | X8 | X11 | X12 |
|-----|------|------|------|------|------|------|-----|-----|
| X1 | 1.0 | | | | | | | |
| X2 | 0.4 | 1.0 | | | | | | |
| X4 | 0.1 | 0.0 | 1.0 | | | | | |
| X5 | 0.0 | -0.3 | 0.5 | 1.0 | | | | |
| X6 | 0.5 | 0.4 | -0.2 | -0.5 | 1.0 | | | |
| X8 | -0.4 | -0.3 | -0.1 | -0.1 | -0.3 | 1.0 | | |
| X11 | -0.1 | 0.1 | -0.4 | -0.4 | 0.2 | 0.6 | 1.0 | |
| X12 | 0.3 | 0.2 | 0.2 | -0.2 | 0.3 | -0.1 | 0.0 | 1.0 |

Kesimpulan:

X11 (% RT masih BAB sembarangan di tempat terbuka) lebih diprioritaskan untuk dihapus
Karena X8 (% RT dengan luas hunian per kapita $\leq 7.2 \text{ m}^2$) lebih relevan untuk menilai Y (kelayakan hunian)

*Asumsi: Variabel dengan korelasi ≥ 0.7 atau ≤ -0.7 akan dihapus

**Ditemukan sepasang variabel dengan korelasi moderate, yakni 0.6, dan salah satunya dianggap lebih relevan dan representatif

Korelasi Tinggi:

X11 | X8 | 0.6

Step 2.3: Cek Non Multikolinearitas & Hapus Variabel Gejala Multikolinearitas



Tips: Gunakan fitur **Correlation** dari XLMiner Analysis Toolpak

The screenshot shows the 'XLMiner Analysis ToolPak' ribbon menu. The 'Correlation' option is highlighted with a yellow box and connected by a grey arrow to the correlation matrix below.

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation**
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Linear Regression
- Logistic Regression
- Moving Average
- Random Number Generation
- Rank and Percentile
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two-Sample for Means

| | X1 | X2 | X4 | X5 | X6 | X8 | X12 |
|-----|------|------|------|------|------|------|-----|
| X1 | 1.0 | | | | | | |
| X2 | 0.4 | 1.0 | | | | | |
| X4 | 0.1 | 0.0 | 1.0 | | | | |
| X5 | 0.0 | -0.3 | 0.5 | 1.0 | | | |
| X6 | 0.5 | 0.4 | -0.2 | -0.5 | 1.0 | | |
| X8 | -0.4 | -0.3 | -0.1 | -0.1 | -0.3 | 1.0 | |
| X12 | 0.3 | 0.2 | 0.2 | -0.2 | 0.3 | -0.1 | 1.0 |

Kesimpulan:

Bisa lanjut ke Regresi!

*Sudah tidak ada lagi variabel dengan korelasi ≥ 0.7 atau ≤ -0.7

*dan variabel tersisa dianggap sudah relevan dalam menentukan Y (kelayakan hunian)

Step 3: Regresi Linier (Pilih Residuals dan Standardized Residuals)

 **Tips:** Gunakan fitur **Linear Regression** dari XLMiner Analysis Toolpak

XLMiner Analysis ToolPak

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Linear Regression**
- Logistic Regression
- Moving Average
- Random Number Generation
- Rank and Percentile
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two-Sample for Means

Linear Regression

Input Y Range: B1:B35

Input X Range: C1:I35

Labels

Constant is Zero

Confidence Level: 95 %

Output Range: AE3

Residuals

Residual Plots

Standardized Residuals

Line Fit Plots

Normal Probability Plots

OK

SUMMARY OUTPUT

Regression Statistics

| | |
|-------------------|------|
| Multiple R | 0.93 |
| R Square | 0.86 |
| Adjusted R Square | 0.82 |
| Standard Error | 5.39 |
| Observations | 34 |

ANOVA

| | df | SS | MS | F | Significance F |
|------------|----|---------|--------|-------|----------------|
| Regression | 7 | 4518.72 | 645.53 | 22.22 | 0.00 |
| Residual | 26 | 755.47 | 29.06 | | |
| Total | 33 | 5274.19 | | | |

Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95% Upper 95%

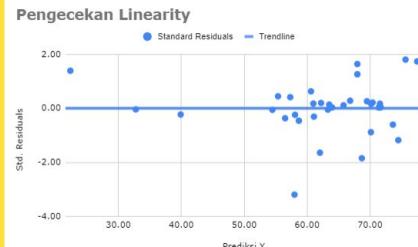
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95% | Upper 95% |
|-----------|--------------|----------------|--------|---------|-----------|-----------|-----------|-----------|
| Intercept | -41.25 | 21.75 | -1.90 | 0.07 | -85.96 | 3.46 | -85.96 | 3.46 |
| X1 | 0.05 | 0.10 | 0.50 | 0.62 | -0.16 | 0.26 | -0.16 | 0.26 |
| X2 | 0.58 | 0.08 | 7.21 | 0.00 | 0.41 | 0.75 | 0.41 | 0.75 |
| X4 | 0.76 | 0.15 | 5.15 | 0.00 | 0.45 | 1.06 | 0.45 | 1.06 |
| X5 | 0.15 | 0.26 | 0.60 | 0.55 | -0.37 | 0.68 | -0.37 | 0.68 |
| X6 | -0.13 | 0.18 | -0.69 | 0.49 | -0.50 | 0.25 | -0.50 | 0.25 |
| X8 | -0.86 | 0.23 | -3.80 | 0.00 | -1.32 | -0.39 | -1.32 | -0.39 |
| X12 | -0.03 | 0.03 | -0.99 | 0.33 | -0.09 | 0.03 | -0.09 | 0.03 |

Step 4: Cek Asumsi Tersisa (Selain Non Multicollinearity yang Sudah Dicek di Awal)

RESIDUAL OUTPUT

| Observation | Predicted Y | Residuals | Standard Residuals |
|-------------|-------------|-----------|--------------------|
| 1 | 70.14 | -4.23 | -0.88 |
| 2 | 70.18 | 0.77 | 0.16 |
| 3 | 68.69 | -8.84 | -1.85 |
| 4 | 71.39 | 0.14 | 0.03 |
| 5 | 63.97 | 0.15 | 0.03 |
| 6 | 60.99 | 0.83 | 0.17 |
| 7 | 56.50 | -1.76 | -0.37 |
| 8 | 62.21 | 0.94 | 0.20 |
| 9 | 32.76 | -0.19 | -0.04 |
| 10 | 54.48 | -0.27 | -0.06 |
| 11 | 39.90 | -1.10 | -0.23 |
| 12 | 62.03 | -7.86 | -1.64 |
| 13 | 74.49 | -5.64 | -1.18 |
| 14 | 77.49 | 8.30 | 1.74 |
| 15 | 73.63 | -2.89 | -0.60 |
| 16 | 63.30 | -0.24 | -0.05 |
| 17 | 75.64 | 8.62 | 1.80 |
| 18 | 65.79 | 0.52 | 0.11 |
| 19 | 57.99 | -15.29 | -3.20 |
| 20 | 63.51 | 0.65 | 0.14 |
| 21 | 58.69 | -2.20 | -0.46 |
| 22 | 55.37 | 2.13 | 0.44 |
| 23 | 67.98 | 7.84 | 1.64 |
| 24 | 70.40 | 1.03 | 0.21 |
| 25 | 71.57 | 0.82 | 0.17 |
| 26 | 61.08 | -1.48 | -0.31 |
| 27 | 71.70 | 0.18 | 0.04 |
| 28 | 67.98 | 6.03 | 1.26 |
| 29 | 69.50 | 1.25 | 0.26 |
| 30 | 57.31 | 1.97 | 0.41 |
| 31 | 60.61 | 3.00 | 0.63 |
| 32 | 66.83 | 1.34 | 0.28 |
| 33 | 58.08 | -1.16 | -0.24 |
| 34 | 22.36 | 6.65 | 1.39 |

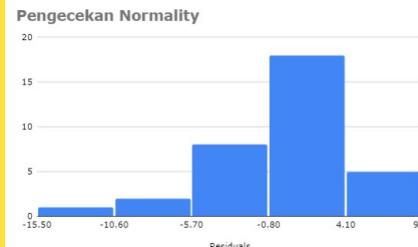
#1



Linearity ✗

Terlihat pola hubungan nilai residuals vs prediksi Y **tidak membentuk pola yang linear**. Sekilas, pola yang lebih tepat adalah Polinomial.

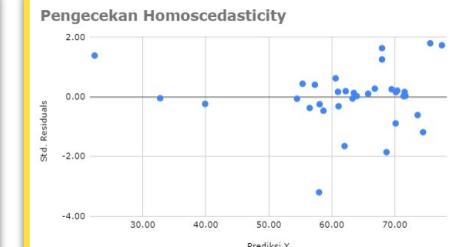
#2



Normality ✗

Terlihat sebaran residuals relatif **tidak normal**. Sekilas, sebaran agak menjulur ke kiri (left-skewed).

#3



Homoscedasticity ✗

Terlihat pola hubungan nilai residuals vs prediksi Y **tidak membentuk pola yang homogen dari ujung ke ujung**. Sekilas, pola melebar dari kiri ke kanan.

Step 5: Interpretasi Model (Anggap asumsi terpenuhi semua)

Asumsi:

Sebagai latihan, **kita anggap saja bahwa semua asumsi telah terpenuhi.**

Pada real case-nya, memang data kecil ($n = 34$) seperti yang sedang kita gunakan ini akan cenderung sulit memenuhi asumsi dari jenis analisis Parametrik seperti Regresi Linier ini.

Saran:

Untuk memperbesar kemungkinan asumsi regresi linier untuk terpenuhi, dapat gunakan **data dengan granularity lebih tinggi** (lebih detail), misalnya data level kota atau bahkan kelurahan.

Sesuai **Central Limit Theorem**, “*semakin besar jumlah sampel, semakin mendekati sebaran normal.*”



Step 5: Interpretasi Model (Anggap asumsi terpenuhi semua)

| SUMMARY OUTPUT | | | | | | | | |
|-----------------------|--------------|----------------|--------|---------|----------------|-----------|-----------|-----------|
| Regression Statistics | | | | | | | | |
| Multiple R | 0.93 | | | | | | | |
| R Square | 0.86 | | | | | | | |
| Adjusted R Square | 0.82 | | | | | | | |
| Standard Error | 5.39 | | | | | | | |
| Observations | 34 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 7 | 4518.72 | 645.53 | 22.22 | 0.00 | | | |
| Residual | 26 | 755.47 | 29.06 | | | | | |
| Total | 33 | 5274.19 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95% | Upper 95% |
| Intercept | -41.25 | 21.75 | -1.90 | 0.07 | -85.96 | 3.46 | -85.96 | 3.46 |
| X1 | 0.05 | 0.10 | 0.50 | 0.62 | -0.16 | 0.26 | -0.16 | 0.26 |
| X2 | 0.58 | 0.08 | 7.21 | 0.00 | 0.41 | 0.75 | 0.41 | 0.75 |
| X4 | 0.76 | 0.15 | 5.15 | 0.00 | 0.45 | 1.06 | 0.45 | 1.06 |
| X5 | 0.15 | 0.26 | 0.60 | 0.55 | -0.37 | 0.68 | -0.37 | 0.68 |
| X6 | -0.13 | 0.18 | -0.69 | 0.49 | -0.50 | 0.25 | -0.50 | 0.25 |
| X8 | -0.86 | 0.23 | -3.80 | 0.00 | -1.32 | -0.39 | -1.32 | -0.39 |
| X12 | -0.03 | 0.03 | -0.99 | 0.33 | -0.09 | 0.03 | -0.09 | 0.03 |

R-Square: 0.86 (86%)

Sebesar 86% keragaman tingkat akses hunian layak & terjangkau (Variabel Y) dapat dijelaskan oleh model ini—yang tersusun atas X1 (% RT dengan akses layanan air minum), X2 (% RT dengan Jenis Atap Terluas yang Layak), X4 (% RT dengan Sumber Air Minum Layak), X5 (% RT Menggunakan Layanan Sanitasi yang Dikelola Secara Aman), X6 (% RT dengan rumah milik sendiri), X8 (% RT dengan Luas Hunian per kapita <= 7,2 m²), X12 (Jumlah SPAM).

Significance F: 0.00 ≤ α = 5% (0.05)

Secara simultan, variabel-variabel independen pada model ini berpengaruh signifikan pada tingkat akses hunian layak & terjangkau (Variabel Y).

P-value:

- Signifikan (P-value ≤ α = 5% (0.05)): X2, X4, X8
- Tidak signifikan (P-value > α = 5% (0.05)): X1, X5, X6, X12

Coefficients:

- X1: 0.05 → Jika X1 naik 1%, maka RT dgn akses hunian layak & terjangkau **meningkat** 0.05%.
- X2: 0.58 → Jika X2 naik 1%, maka RT dgn akses hunian layak & terjangkau **meningkat** 0.58%.
- X4: 0.76 → Jika X4 naik 1%, maka RT dgn akses hunian layak & terjangkau **meningkat** 0.76%.
- X5: 0.15 → Jika X5 naik 1%, maka RT dgn akses hunian layak & terjangkau **meningkat** 0.15%.
- X6: -0.13 → Jika X6 naik 1%, maka RT dgn akses hunian layak & terjangkau **menurun** 0.13%.
- X8: -0.86 → Jika X8 naik 1%, maka RT dgn akses hunian layak & terjangkau **menurun** 0.86%.
- X12: -0.03 → Jika X12 naik 1%, maka RT dgn akses hunian layak & terjangkau **menurun** 0.03%.

Kita baru saja membuat

Deck Singkat* Regresi

*jauh lebih singkat dibandingkan Deck EDA kita di RBD Eps. 5 lalu ya!

Step 1: Join/Lookup Berbagai Tabel ke dalam Satu Tabel Sendiri

Tujuan: Mengintegrasikan berbagai tabel yang berhubungan ke dalam satu tabel yang mudah diolah.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Join Tables".
- Pilih tabel-tabel yang akan dijoin dan tentukan kriteria join.
- Klik "OK".

Hasil: Tabel yang telah dijoin dengan baik, memudahkan analisis selanjutnya.

Step 2.1: Cek Non Multikolinearitas Antar Variabel Independen (X)

Tujuan: Memastikan bahwa variabel independen tidak saling berkorelasi kuat.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Multicollinearity Test".
- Pilih variabel independen yang akan dianalisis.
- Klik "OK".

Hasil: Laporan yang menunjukkan nilai toleransi dan nilai VIF untuk setiap variabel, serta daftar variabel yang mungkin saling berkorelasi.

Step 2.2: Cek Non Multikolinearitas & Hapus Variabel Gejala Multikolinearitas

Tujuan: Mengidentifikasi dan menghapus variabel yang menjadi gejala multikolinearitas.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Multicollinearity Test".
- Pilih variabel independen yang akan dianalisis.
- Klik "OK".

Hasil: Laporan yang menunjukkan nilai toleransi dan nilai VIF untuk setiap variabel, serta daftar variabel yang mungkin saling berkorelasi.

Step 2.3: Cek Non Multikolinearitas & Hapus Variabel Gejala Multikolinearitas

Tujuan: Mengidentifikasi dan menghapus variabel yang menjadi gejala multikolinearitas.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Multicollinearity Test".
- Pilih variabel independen yang akan dianalisis.
- Klik "OK".

Hasil: Laporan yang menunjukkan nilai toleransi dan nilai VIF untuk setiap variabel, serta daftar variabel yang mungkin saling berkorelasi.

29

30

31

32

Step 3: Regresi Linier (Pilih Residuals dan Standardized Residuals)

Tujuan: Memilih tipe residual yang sesuai untuk analisis regresi linier.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Residuals".
- Pilih tipe residual yang akan digunakan.
- Klik "OK".

Hasil: Tabel yang menunjukkan hasil penghitungan residual.

Step 4: Cek Asumsi Terisa (Selain Non Multicollinearity yang Sudah Dicek di Awal)

Tujuan: Memeriksa asumsi terisa lainnya yang penting untuk regresi linier.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Assumptions".
- Pilih tipe analisis yang akan dilakukan.
- Klik "OK".

Hasil: Laporan yang menunjukkan hasil pemeriksaan terhadap asumsi terisa.

Step 5: Interpretasi Model (Anggap asumsi terpenuhi semua)

Tujuan: Menganalisis hasil regresi linier dan menarik kesimpulan.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Regression".
- Pilih variabel independen dan dependen.
- Klik "OK".

Hasil: Laporan yang menunjukkan hasil regresi linier.

Step 5: Interpretasi Model (Anggap asumsi terpenuhi semua)

Tujuan: Menganalisis hasil regresi linier dan menarik kesimpulan.

Alat: Microsoft Excel, XLMiner Add-in.

Proses:

- Buka Microsoft Excel.
- Pilih menu "XLMiner" > "Regression".
- Pilih variabel independen dan dependen.
- Klik "OK".

Hasil: Laporan yang menunjukkan hasil regresi linier.

33

34

35

36



Model Fitness



Variable Significance

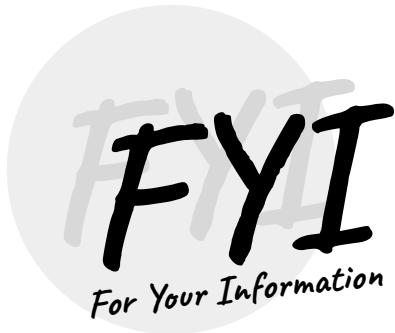


Business Decision

8 Miskonsepsi / Common Mistakes Ttg Regresi Linier



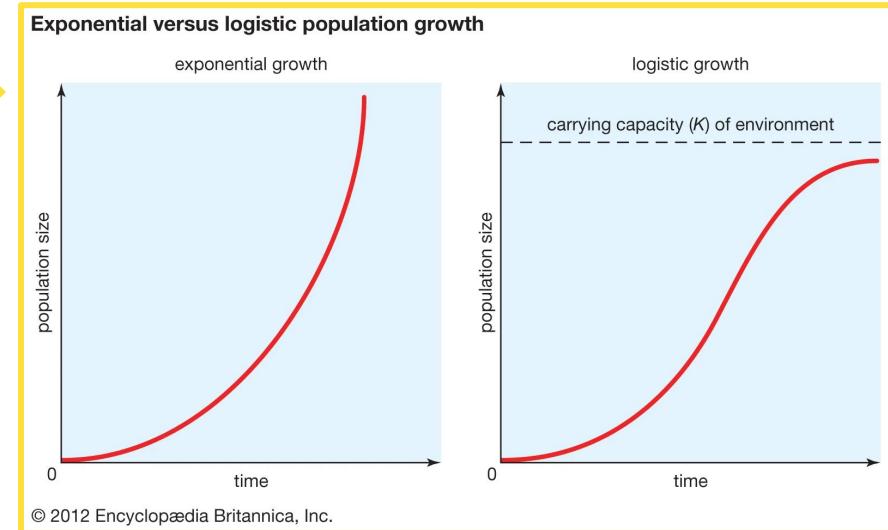
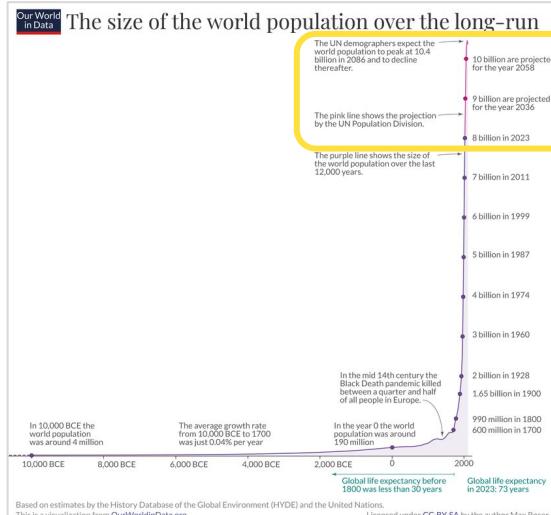
| Banyak orang lakukan.. | Padahal realitanya.. |
|--|--|
| Setiap model dipaksa linier | Tergantung hubungan X dan Y! |
| Asumsi normality & homoscedasticity diuji pada variabel (X dan/atau Y) | Seharusnya yang diuji adalah nilai residuals/error! |
| Asumsi multikolinearitas melibatkan Y | Seharusnya antar variabel X saja! |
| Asumsi harus dipenuhi semua | Boleh menggunakan persamaan regresi untuk tujuan prediksi, meskipun abaikan asumsi! |
| Asumsi berlaku untuk semua analisis regresi linier | Hanya berlaku jika menggunakan metode estimasi OLS (<i>default software</i>). Beda cerita jika menggunakan WLS, GLS! |
| Fokus interpretasi R-Square, mengabaikan Adjusted R-Square | Adjusted R-Square lebih objektif dalam menilai kebaikan model! |
| R-Square harus mendekati 100% | Utamakan penuhi asumsi dahulu. Belum tentu juga Y yang digunakan dapat diprediksi oleh X yang tersedia! |
| Semakin banyak X, semakin baik | Terlalu banyak X dapat meningkatkan potensi overfitting! |



"Hubungan antara 2 variabel tidak bersifat mutlak dan bertahan selamanya!"

→ Hubungan bisa berubah seiring perubahan waktu atau perbedaan sampel.

Contoh: Pertumbuhan populasi dunia (Exponential → Logistic)





Satyabrata Mishra

Data Science @ Tide

Published • 2y

21 articles

+ Follow

...

Parametric models have lost their sheen in the age of Deep Learning. But for smaller datasets, and when interpretability outweighs predictive power, models like linear and logistic regressions still hold the sway. But parametric models carry the weight of assumptions, and what happens when assumptions are violated...

LINEAR REGRESSION MODELS



Kill yourselves.

<https://www.linkedin.com/pulse/violating-linear-regression-assumptions-guide-what-do-mishra/>

What's Next?



Cara menanggulangi pelanggaran asumsi:

([https://www.linkedin.com/advice/1/what-most-common-misconceptions-linear-ai/#:-:text=One%20of%20the%20most%20common,%2C%20interactions%2C%20or%20polynomial%20terms](https://www.linkedin.com/advice/1/what-most-common-misconceptions-linear-ai/))



Alternatif Regresi Linier:

- Polinomial (*pangkat 2 atau lebih*)
- Logistic (*Y biner atau kategorik*)
- GLM (*asumsi normalitas dilanggar*)
- GAM: Generalized Additive Model

(*non-Parametrik, tidak butuh memenuhi asumsi, automatic feature engineering, but most probably complex → More data needed and difficult to interpret*)

- dll.



Decision Tree → *Lebih practical, interpretasi sederhana, robust terhadap outliers, tidak butuh asumsi linearity.*



Clustering Analysis → *Cek apakah terbentuk kelompok, misalnya level Provinsi/Wilayah Indonesia/ lainnya, jika iya, bisa dilakukan regresi di masing-masing grup.*



Post-Test

Klik link berikut!
bit.ly/PostTestRBD6



Jawablah dengan sejujur-jujurnya, nilai post-test ini tidak akan mempengaruhi gaji, tunjangan kinerja, apalagi dilaporkan pada bapak menteri. 🌟

CONTACT ME



Domicile:
Jababeka Residence,
Bekasi Regency



+62 852-1147-4471

M miqbalpillie18@gmail.com

THANK YOU & SEE YOU!