Analisis Korelasi

Sevi Nurafni
Data Sains, Fakultas Sains dan Teknologi
Universitas Koperasi Indonesia
Github.com/sevinurafni

Jatinangor, 15 Mei 2024

Contents:

Uji-t

Koefisien Korelasi

- Koefisien Korelasi untuk sample

$$r_{xy} = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{(n\sum X_i^2 - (\sum X_i)^2)(n\sum Y_i^2 - (\sum Y_i)^2)}}$$

- Koefisien Korelasi untuk populasi

$$\rho_{xy} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

Uji-t

- Uji t dapat digunakan untuk menguji signifikansi statistik dari Pearson correlation coefficient.
- Hipotesis yang diuji:
  - $H_0 : \rho = 0$, bahwa tidak ada korelasi antara dua variabel dalam populasi
  - $H_1 : \rho \neq 0$, bahwa ada korelasi yang signifikan
- Statistik ujinya:
  - $t = r\sqrt{\frac{N-2}{1-r^2}}$

terima H_0 jika $t \leq t_{\alpha}; n-2$

terima H_0 jika $t \geq -t_{\alpha}; n-2$

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

Contoh 1 - Data Hasil Penelitian

Diketahui data jumlah SKS dan IPK mahasiswa sebagai berikut.

Ujilah hipotesis yang berbentuk:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Jumlah SKS (X)	IPK (Y)
10	3.00
10	2.50
15	2.00
10	1.50
5	1.00

Contoh 3 - Perhitungan r

$$r_{xy} = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{(n\sum X_i^2 - (\sum X_i)^2)(n\sum Y_i^2 - (\sum Y_i)^2}}$$

$$r_{xy} = \frac{5(105) - (50)(10)}{\sqrt{(5(550) - (50)^2)(5(22.5) - (10)^2)}}$$

$$r_{xy} = \frac{25}{\sqrt{(250)(12.5)}}$$

$$r_{xy} = 0.447$$

Dari hasil ini didapat korelasi positif antara Jumlah SKS (X) dan IPK yang didapat (Y)

NO	Xi	Yi	XiYi	Xi^2	Yi^2
1	10	3.00	30	100	9.00
2	10	2.50	25	100	6.25
3	15	2.00	30	225	4.00
4	10	1.50	15	100	2.25
5	5	1.00	5	5	1.00
Sum	50	10	105	550	22.5

Contoh 3 - Perhitungan t

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

$$t = 0.447\sqrt{\frac{5-2}{1-(0.447)^2}}$$

$$t = 1.6646$$

$$t_{0.025;9} = 2.262$$

$$1.6646 \leq 2.262$$

$$1.6646 \geq -2.262$$

H_0 diterima

NO	Xi	Yi	XiYi	Xi^2	Yi^2
1	10	3.00	30	100	9.00
2	10	2.50	25	100	6.25
3	15	2.00	30	225	4.00
4	10	1.50	15	100	2.25
5	5	1.00	5	5	1.00
Sum	50	10	105	550	22.5

Fisherman Transformation

Fisher Transformation

- Sampel dengan ukuran n tertentu, ternyata r tidak terdistribusi secara normal ketika $\rho \neq 0$
- $Z = \frac{1}{2} ln(\frac{1+r}{1-r})$
  - Menghasilkan sebuah variabel normal asimtotik dengan rata-rata
    - $\mu_Z = \frac{1}{2} ln(\frac{1+\rho_0}{1-\rho_0})$ dan
    - Varians $\sigma_Z = \frac{1}{\sqrt{n-3}}$

https://en.wikipedia.org/wiki/Fisher_transformation

Fisher Transformation

- Hipotesis:
  - $H_0 : \rho = \rho_0 \neq 0$ dengan salah satu dari
    - $H_1 : \rho \neq \rho_0$
    - $H_1 : \rho > \rho_0$
    - $H_1 : \rho < \rho_0$
- Statistik uji:
  - $Z_{test} = (\dfrac{Z - \mu_Z}{\sigma_Z})$
- Kriteria uji, terima H_0 jika:
  - $-Z_{\alpha/2} < Z_{test} < Z_{\alpha/2}$
  - $Z_{test} \leq Z_{\alpha}$
  - $Z_{test} \geq - Z_{\alpha}$

https://real-statistics.com/correlation/one-sample-hypothesis-testing-correlation/correlation-testing-via-fisher-transformation/

Contoh 2 - Data Hasil Penelitian

Dengan menggunakan taraf signifikan 5%

Ujilah hipotesis yang berbentuk:

$$H_0 : \rho = 0.75$$

$$H_1 : \rho \neq 0.75$$

Test Kemampuan (X)	Prestasi Kerja (Y)
70	10
71	11
72	12
73	15
74	11
75	12
78	13
79	14
78	15
79	18

Contoh 2 - Perhitungan Fisher Tranformation

$$Z = \frac{1}{2} ln(\frac{1 + r}{1 - r})$$

$$Z = \frac{1}{2} ln(\frac{1 + 0.7262}{1 - 0.7262})$$

$$Z = \frac{1}{2} ln(6.3046)$$

$$Z = \frac{1}{2}(1.841)$$

$$Z = 0.9206$$

No	X	Y	X^2	Y^2	XY
1	70	10	4900	100	700
2	71	11	5041	121	781
3	72	12	5184	144	864
4	73	15	5329	225	1095
5	74	11	5476	121	814
6	75	12	5625	144	900
7	78	13	6084	169	1014
8	79	14	6241	196	1106
9	78	15	6084	225	1170
10	79	18	6241	324	1422
Sum	749	131	56205	1769	9866

Contoh 2 - Perhitungan Z_{test}

$$\mu_Z = \frac{1}{2} ln(\frac{1+\rho_0}{1-\rho_0})$$

$$\mu_Z = \frac{1}{2} ln(\frac{1+0.75}{1-0.75})$$

$$\mu_Z = 0.9730$$

$$\sigma_Z = \frac{1}{\sqrt{n-3}}$$

$$\sigma_Z = \frac{1}{\sqrt{10-3}}$$

$$\sigma_Z = 0.3780$$

Sehingga

$$Z_{test} = (\frac{Z - \mu_Z}{\sigma_Z})$$

$$Z_{test} = (\frac{0.9206 - 0.9730}{0.3780})$$

$$Z_{test} = -0.14$$

No	X	Y	X^2	Y^2	XY
1	70	10	4900	100	700
2	71	11	5041	121	781
3	72	12	5184	144	864
4	73	15	5329	225	1095
5	74	11	5476	121	814
6	75	12	5625	144	900
7	78	13	6084	169	1014
8	79	14	6241	196	1106
9	78	15	6084	225	1170
10	79	18	6241	324	1422
Sum	749	131	56205	1769	9866

Contoh 2 - Kesimpulan

Terima H_0 jika

$-Z_{\alpha/2} < Z_{test} < Z_{\alpha/2}$

$-1.96 < -0.14 < 1.96$

Sehingga H_0 diterima.

No	X	Y	X^2	Y^2	XY
1	70	10	4900	100	700
2	71	11	5041	121	781
3	72	12	5184	144	864
4	73	15	5329	225	1095
5	74	11	5476	121	814
6	75	12	5625	144	900
7	78	13	6084	169	1014
8	79	14	6241	196	1106
9	78	15	6084	225	1170
10	79	18	6241	324	1422
Sum	749	131	56205	1769	9866

Korelasi Berganda

Korelasi Linear Berganda

Alat ukur mengenai hubungan yang terjadi antara variabel terikat (Y) dan dua atau lebih variabel bebas ($X_1, X_2, ..., X_k$)

https://www.sciencedirect.com/topics/mathematics/multiple-correlation-coefficient

Koefisien Korelasi Berganda

- Digunakan untuk mengukur besarnya kontribusi variasi X_1 dan X_2 terhadap variasi Y
- Contoh
  - Fasilitas pendidikan dan kualitas dosen dengan prestasi belajar mahasiswa

Koefisien Penentu Berganda (R^2)

- Disebut juga dengan Koefisien Determinasi Berganda
- Menggambarkan ukuran kesesuaian garis linear berganda terhadap suatu data
- R-squared adalah presentase keragaman data yang mampu diterangkan oleh model
- R-squared tinggi adalah indikasi model yang baik
  - $R^2 = \dfrac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}$ atau
  - $R^2 = \dfrac{r_{Y.1}^2 + r_{Y.2}^2 - 2 r_{Y.1} r_{Y.2} r_{1.2}}{1 - r_{1.2}^2}$

Koefisien Penentu Berganda (R^2)

- $R^2 = \dfrac{r_{Y.1}^2 + r_{Y.2}^2 - 2r_{Y.1}r_{Y.2}r_{1.2}}{1 - r_{1.2}^2}$
  - $r_{Y1} = \dfrac{\sum x_1 y}{\sqrt{x_1^2 \cdot \sum y^2}}$
  - $r_{Y2} = \dfrac{\sum x_2 y}{\sqrt{x_2^2 \cdot \sum y^2}}$
  - $r_{12} = \dfrac{\sum x_2 x_2}{\sqrt{x_1^2 \cdot \sum x_2^2}}$

Koefisien Penentu Berganda

- Dengan:
- $\sum y^2 = \sum Y^2 - n\bar{Y}^2$
- $\sum x_1^2 = \sum X_1^2 - n\bar{X}_1^2$
- $\sum x_2^2 = \sum X_2^2 - n\bar{X}_2^2$

- Dengan:
- $\sum x_1 y = \sum X_1 Y - n\bar{X}_1\bar{Y}$
- $\sum x_2 y = \sum X_2 Y - n\bar{X}_2\bar{Y}$
- $\sum x_1 x_2 = \sum X_1 X_2 - n\bar{X}_1\bar{X}_2$

Koefisien Korelasi Berganda (R)

- Merupakan ukuran keeratan hubungan antara variabel terikat dengan semua variabel bebas secara bersama-sama.

- $R = \sqrt{\dfrac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}}$ atau $R = \sqrt{\dfrac{r_{Y.1}^2 + r_{Y.2}^2 - 2 r_{Y.1} r_{Y.2} r_{1.2}}{1 - r_{1.2}^2}}$

Koefisien Korelasi Berganda (R)

- Merupakan ukuran keeratan hubungan antara variabel terikat dengan semua variabel bebas secara bersama-sama.

- $R = \sqrt{\dfrac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}}$ atau $R = \sqrt{\dfrac{r_{Y.1}^2 + r_{Y.2}^2 - 2 r_{Y.1} r_{Y.2} r_{1.2}}{1 - r_{1.2}^2}}$

Korelasi Parsial

Korelasi Parsial

Mempelajari hubungan murni antara sebuah variable bebas (X1) dengan variable terikat (Y) dengan mengendalikan atau mengontrol variable-variabel bebas yang lain yaitu variable X2 yang diduga mempengaruhi hubungan antara variable X1 dengan Y.

https://www.ibm.com/docs/en/spss-statistics/beta?topic=features-partial-correlations

Koefisien Korelasi Parsial

“Ingin mengetahui hubungan antara Kepemimpinan Rektor (X_1) dan Motivasi Kerja Dosen (X_2) dengan Kinerja Dosen (Y) di suatu universitas.”

1. Hubungan antara kepemimpinan kepala sekolah (X_1) dengan kinerja guru (Y), X_2 tetap

$$r_{Y1.2} = \frac{r_{Y_1} - r_{Y_2} \cdot r_{12}}{\sqrt{(1 - r_{Y_2}^2)(1 - r_{12}^2)}}$$

https://www.ibm.com/docs/en/spss-statistics/beta?topic=features-partial-correlations

Koefisien Korelasi Parsial

2. Hubungan antara motivasi kerja (X_2) dengan kinerja dosen (Y), X_1 tetap

$$r_{Y2.1} = \frac{r_{Y_2} - r_{Y_1} \,.\, r_{12}}{\sqrt{(1 - r_{Y_1}^2)(1 - r_{Y_2}^2)}}$$

3. Hubungan antara kepemimpinan Rektor (X_1) motivasi kerja (X_2), Y tetap

$$r_{12.Y} = \frac{r_{12} - r_{Y_1} \,.\, r_{Y_2}}{\sqrt{(1 - r_{Y_1}^2)(1 - r_{Y_2}^2)}}$$

https://www.ibm.com/docs/en/spss-statistics/beta?topic=features-partial-correlations

- $r_{Y1} = \dfrac{\sum x_1 y}{\sqrt{x_1^2 \cdot \sum y^2}}$
- $r_{Y2} = \dfrac{\sum x_2 y}{\sqrt{x_2^2 \cdot \sum y^2}}$
- $r_{12} = \dfrac{\sum x_2 x_2}{\sqrt{x_1^2 \cdot \sum x_2^2}}$

- Dengan:
- $\sum y^2 = \sum Y^2 - n\bar{Y}^2$
- $\sum x_1^2 = \sum X_1^2 - n\bar{X}_1^2$
- $\sum x_2^2 = \sum X_2^2 - n\bar{X}_2^2$

- Dengan:
- $\sum x_1 y = \sum X_1 Y - n\bar{X}_1\bar{Y}$
- $\sum x_2 y = \sum X_2 Y - n\bar{X}_2\bar{Y}$
- $\sum x_1 x_2 = \sum X_1 X_2 - n\bar{X}_1\bar{X}_2$

Permasalahan

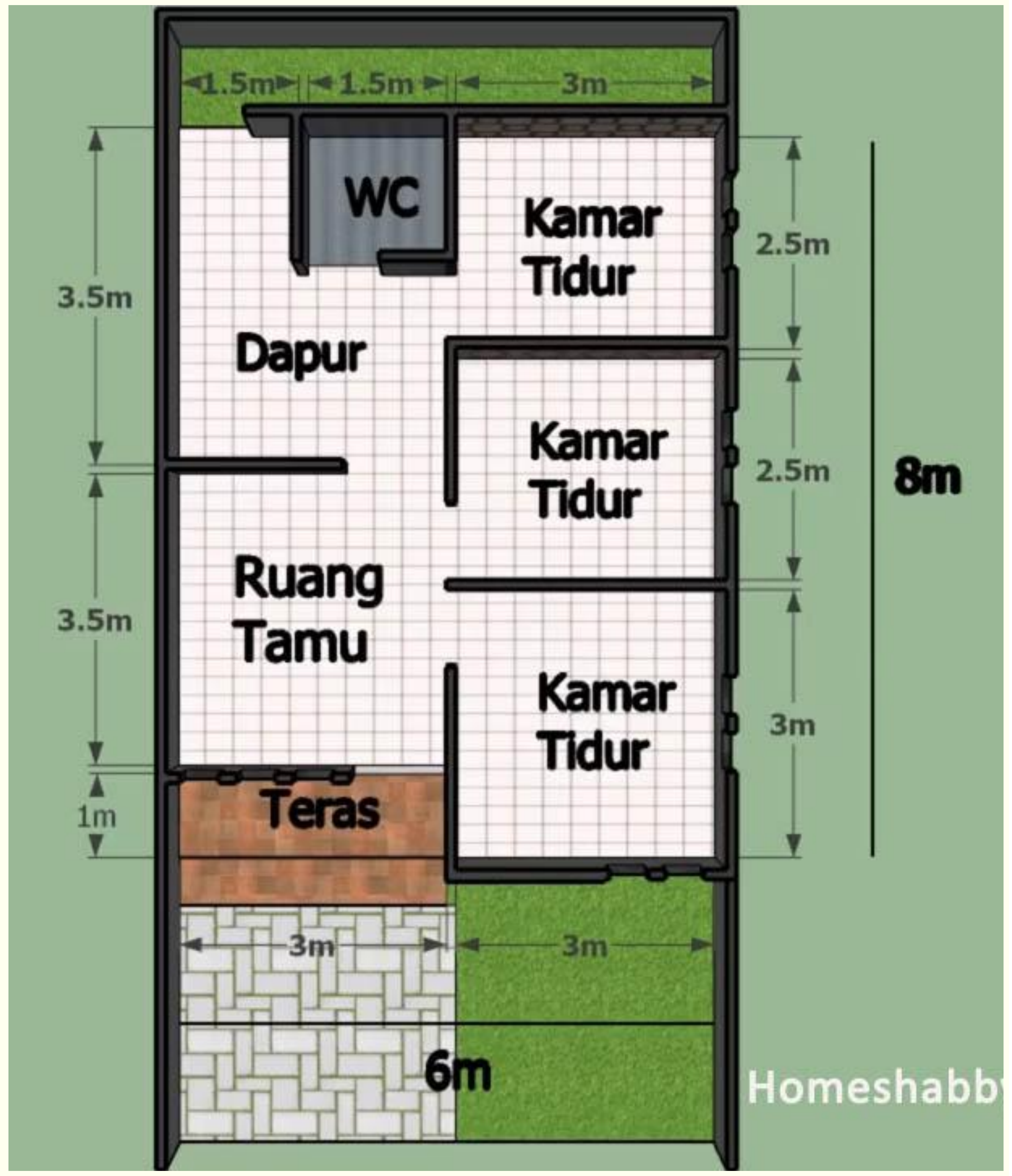
Beberapa Permasalahan

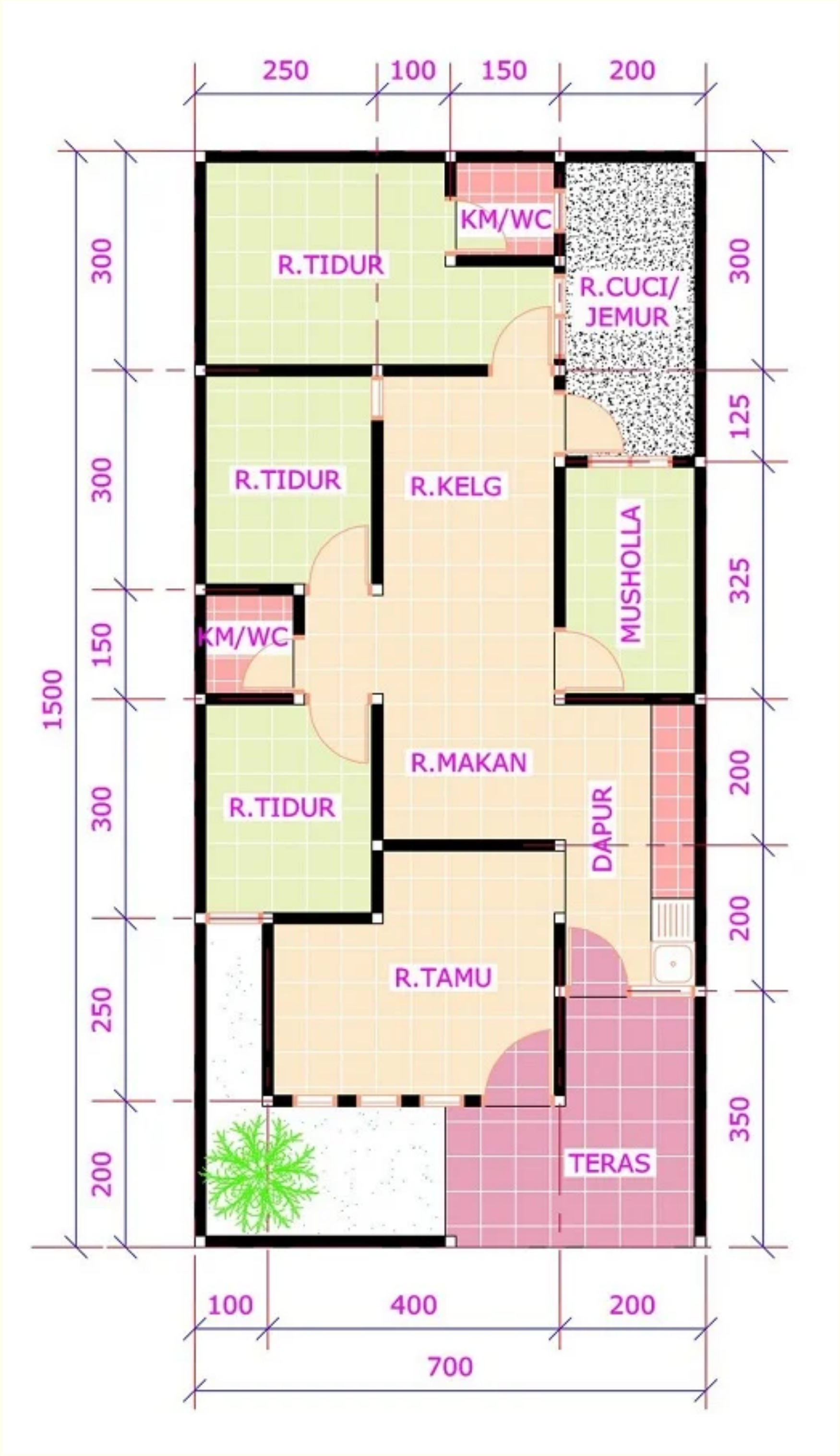
- Multikolinearitas, ketika dua atau lebih variabel independen dalam model regresi sangat berkorelasi satu sama lain

Misal: memprediksi harga rumah berdasarkan ukuran rumah dan jumlah kamar tidur

Menyebabkan: Interpretasi yang sulit dan
Model memberikan prediksi yang kurang akurat

Solusi: 1. Menghapus Variabel yang Berkorelasi Tinggi
2. Transformasi PCA
3. Menggunakan teknik regularisasi untuk mengurangi multikolinearitas.





Beberapa Permasalahan

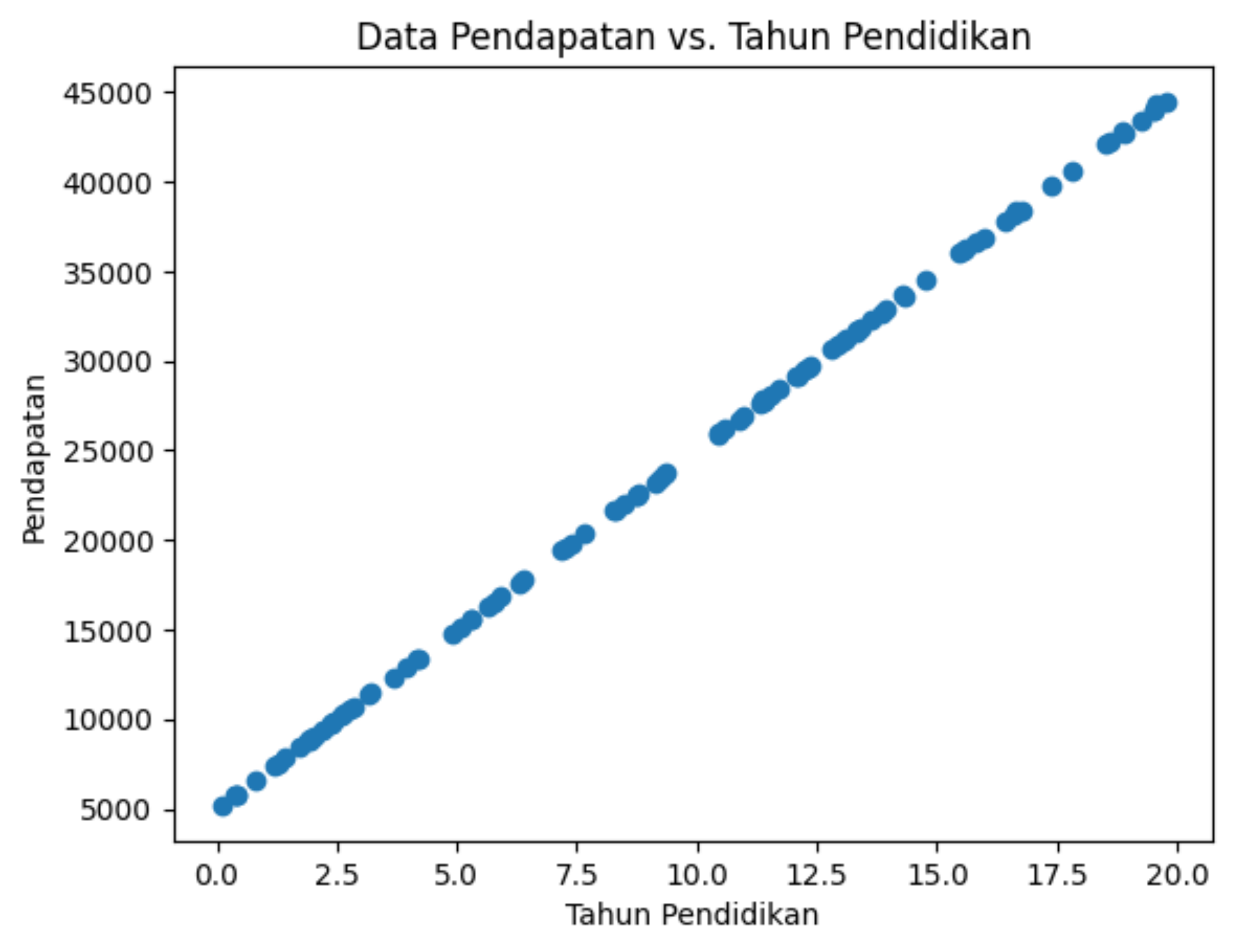
- Heteroskedastisitas, varians kesalahan (residual) tidak konstan pada semua tingkat variabel independen (X)

Menyebabkan: estimasi koefisien yang tidak efisien dan kesalahan standar yang bias

Solusi: 1. Menggunakan transformasi pada variabel dependen (Y) atau variabel independen (X) untuk menstabilkan varians.

2. Robust Standard Errors

3. Menambahkan Variabel yang Terlupakan

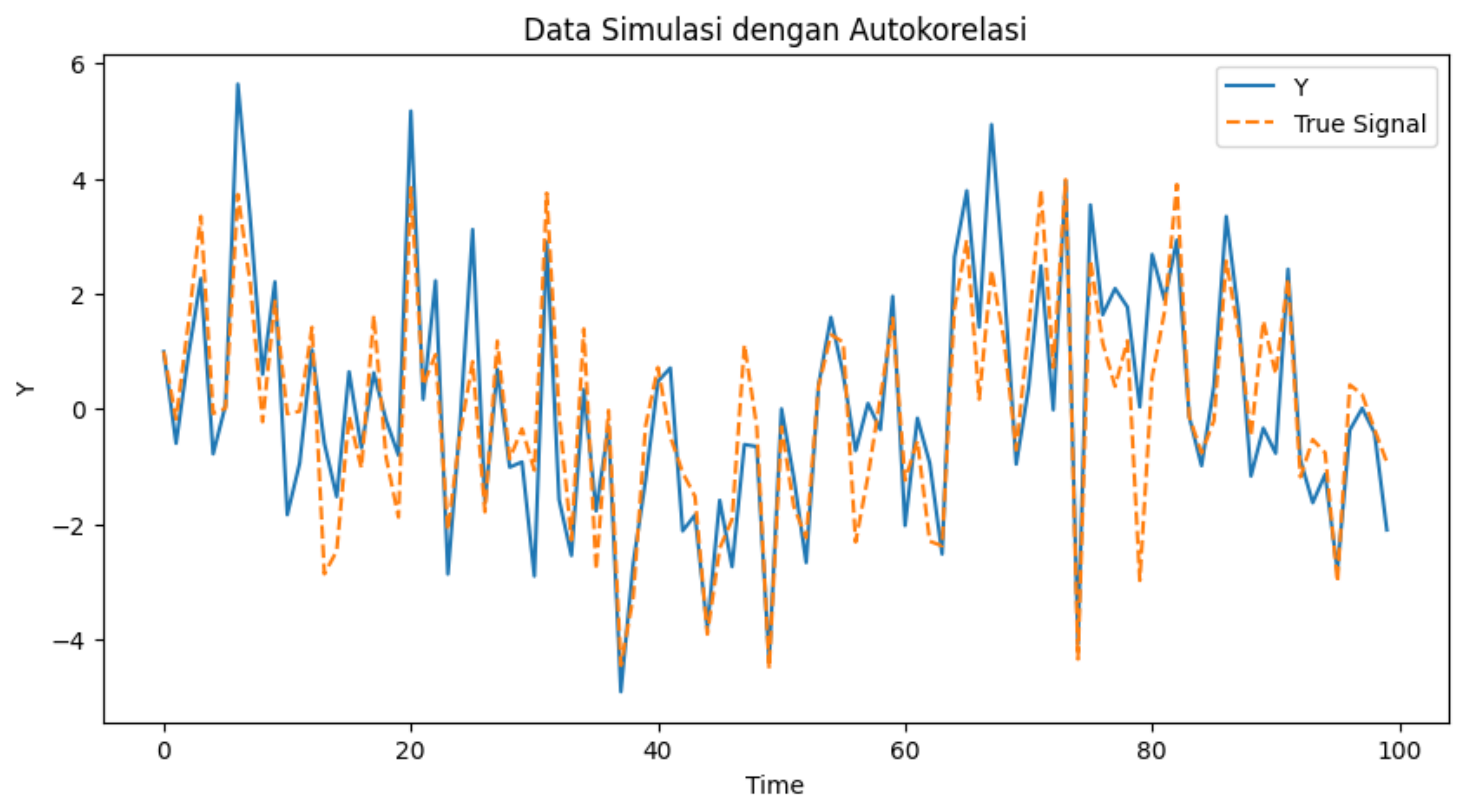


Beberapa Permasalahan

- Autokorelasi, residual (kesalahan prediksi) dari model tidak independen satu sama lain dan menunjukkan pola tertentu

Menyebabkan: estimasi koefisien regresi menjadi tidak efisien dan kesalahan standar menjadi bias.

Solusi: 1. Menambah lag variabel
2. Uji Durbin-Watson Test, untuk mendeteksi keberadaan autokorelasi.



Hal-hal Lain

Lakukan terlebih dahulu eksplorasi melalui plot XY:

- Mungkin ada data pencilan
- Mungkin perlu transformasi data (misal: model kuadratik)
- Mungkin perlu pemisahan model (misal: model untuk perusahaan swasta dalam negeri dan swasta asing tidak sama)

Hal-hal Lain

Pada kasus regresi berganda, terdapat teknik penyeleksian variabel bebas dalam model:

- Forward method
- Backward method
- Stepwise

Latihan Soal

Latihan Soal

Dilakukan suatu penelitian yang bertujuan untuk mempelajari tentang “Pengaruh Pendapatan Keluarga per Hari (X_1) dan Jumlah Anggota Keluarga (X_2) terhadap Pengeluaran Konsumsi Keluarga per Hari (Y)”. Penelitian tersebut menggunakan sampel sebanyak 10 keluarga. Hasil pengumpulan data diperoleh data sebagai berikut:

Carilah koefisien korelasi berganda dan parsial (jika jumlah anggota keluarga dianggap konstan)!

Responden	X_1 (ratusan ribu)	X_2 (orang)	Y (ratusan ribu)
1	100	7	23
2	20	3	7
3	40	2	15
4	60	4	17
5	80	6	23
6	70	5	22
7	40	3	10
8	60	3	14
9	70	4	20
10	60	3	19

https://www.ibm.com/docs/en/spss-statistics/beta?topic=features-partial-correlations

All models are wrong, but some are useful

G. E. P. Box

Terima Kasih

Jangan ragu untuk mengirim pesan kepada saya untuk mengajukan pertanyaan dan diskusi