



Sumber Gambar: <https://www.dreamstime.com/royalty-free-stock-photos-statistics-background-concept-wordcloud-illustration-print-concept-word-cloud-graphic-collage-image35558718>



STATISTIKA DASAR

< PERTEMUAN 04 >

PROGRAM PRAKTIKI MENGAJAR
BATCH 3 TAHUN 2023

TIM DOSEN

1. Putri Sukma Dewi, S.Pd., M.Pd. (*Dosen Pengampu*)
2. Eko Teguh Widodo, SST., M.Sc. (*Dosen Praktisi*)

EKO TEGUH WIDODO

Working Experience

- *Senior IT Specialist*, BPS-Statistics Indonesia (Provincial Branch Office)
- *IT Specialist*, BPS-Statistics Indonesia (Provincial Branch Office)
- *Statistician*, BPS-Statistics Indonesia (Municipal and Provincial Branch Office)
- *IT Engineering Intern*, Sekolah Tinggi Ilmu Statistik (Institute of Statistics)
- *Freelancer as Private Instructor/Tutor* (Statistics, Mathematics, Information Technology), Software Engineer (Web, Desktop)

Education

- *Master of Science (M.Sc.)* in Information and Telecommunication Technology, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea
- *Bachelor of Applied Science (B.A.Sc., SST.)* in Computational Statistics, Sekolah Tinggi Ilmu Statistik (Institute of Statistics), Jakarta

Core Competency

- Statistical Data Analysis, Big Data and Software Engineering, Natural Language Processing, Knowledge Graph, Digital Transformation, Information Technology Management and Policy





KONSEP

- Korelasi dan Regresi Linier
- Distribusi Sampling



PRAKTIK

- Penerapan Korelasi dan Regresi Linier di R
- Demo dan Praktik Coding

EKO TEGUH WIDODO
MARKED BY



STATISTIKA DASAR

KORELASI, REGRESI, DAN DISTRIBUSI SAMPLING

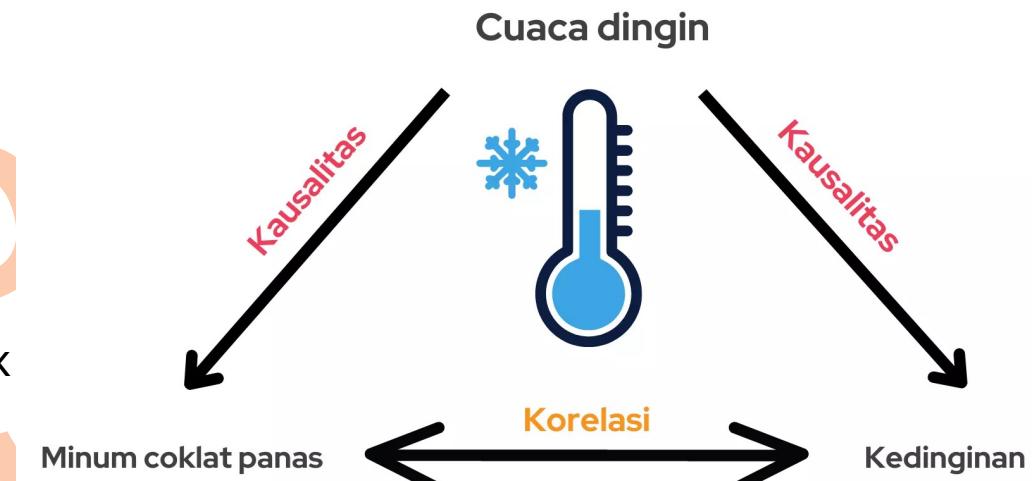


- ✓ Pencarian atau **pemeriksaan hubungan** antar variabel kuantitatif memberikan **insight** bagi peneliti untuk menentukan **metode analisis** (pemodelan) yang akan digunakan.
- ✓ Analisis hubungan antar variabel kuantitatif: menganalisis **pola data** antar variabel, **mengukur kuat/lehahnya** hubungan antar variabel, dan **pemodelan** hubungan antar variabel.
- ✓ **Korelasi** → ukuran statistik yang digunakan untuk mengetahui sejauh mana hubungan antara dua variabel atau lebih. Korelasi digunakan untuk mendeskripsikan hubungan sederhana antara variabel tanpa mempertanyakan sebab-akibatnya (kausalitas).

✓ Contoh Korelasi

- Korelasi antara kenaikan hasil belajar siswa dengan jumlah latihan soal yang diberikan
- Korelasi antara permintaan pada barang yang terbatas dengan harga dari barang tersebut

✓ **Korelasi bukan berarti kausalitas** → kausalitas berarti satu hal menyebabkan hal yang lain. Misalnya, tindakan A menyebabkan hasil B. Di sisi lain, korelasi hanyalah sebuah hubungan di mana tindakan A berhubungan dengan tindakan B, tetapi terjadinya A dan B bisa disebabkan hal-hal lain.



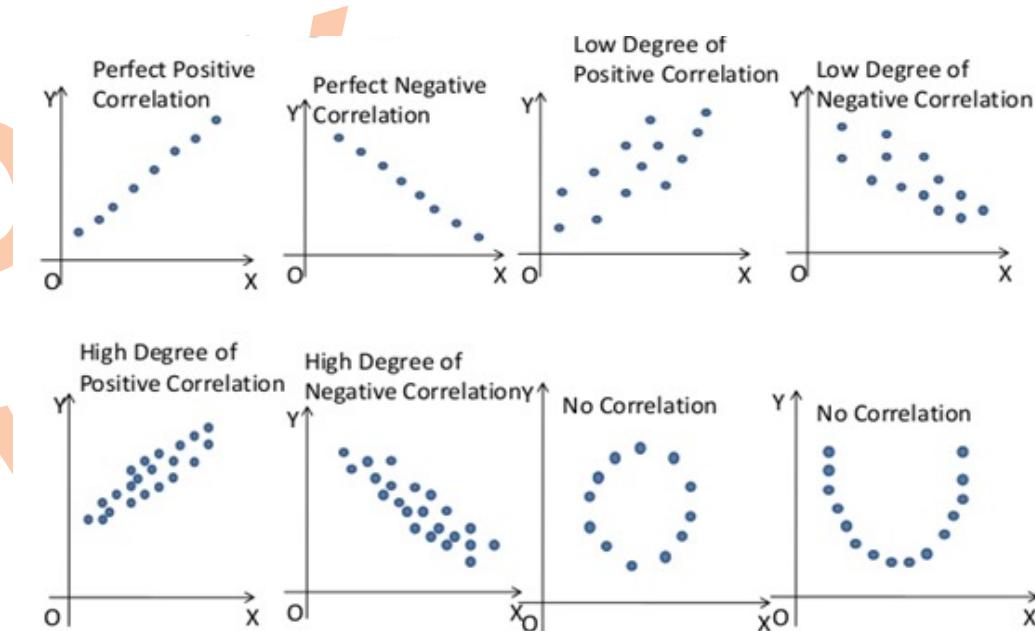
<https://revou.co/revoupedia/kosakata>
Sumber: Amplitude

Contohnya, meminum coklat panas dengan kedinginan adalah dua peristiwa yang saling berhubungan (ada korelasinya). Namun, sebenarnya kedua peristiwa tersebut tidak ada hubungan sebab-akibatnya.

Peristiwa yang menyebabkan terjadinya dua hal itu adalah adanya cuaca yang dingin.

Diagram Pencar (Scatter Plot)

- ✓ **Diagram Pencar (scatter plot)** digunakan untuk mengetahui pola hubungan antar variabel (secara visual) untuk mendeskripsikan data.
- ✓ Namun, diagram ini tidak dapat digunakan untuk mengambil kesimpulan tentang hubungan sebab-akibat (kausalitas) tanpa ada teori atau penelitian terdahulu yang menjadi dasar. Biasanya hanya untuk memperkuat bukti adanya hubungan sebab-akibat dari sebuah teori atau penelitian sebelumnya.
- ✓ Diagram ini digunakan sebagai salah satu **dasar penentuan model** hubungan antar dua variabel.
- ✓ Plot titik-titik (X, Y) sebanyak n observasi pada sumbu-X (horizontal) dan sumbu-Y (vertikal).
- ✓ **Pola dan arah** dari plotting tersebut secara visual dapat menentukan ada atau tidaknya hubungan dan juga seberapa kuat hubungan antar dua variabel kuantitatif.
- ✓ **Tidak selalu** plotting titik-titik (X, Y) menunjukkan adanya hubungan dengan **arah yang nyata** antara X dan Y .
- ✓ Seberapa kuat korelasi (hubungan) antara X dan Y ditentukan oleh besaran **koefisien korelasi (r)**.



✓ Nilai Koefisien Korelasi

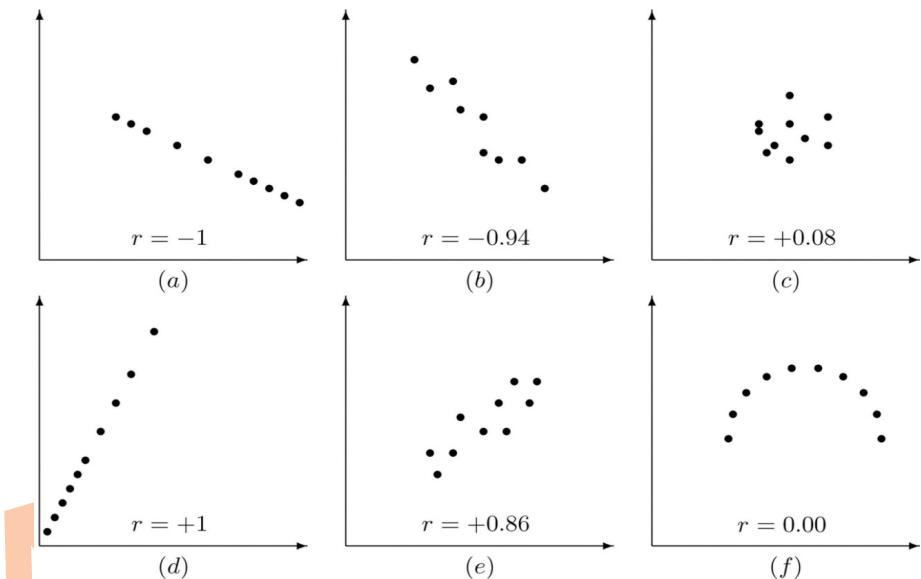
- Berada dalam interval -1 s.d. +1
- +1 menunjukkan tingkatan tertinggi hubungan searah, jika nilai kedua variabel bergerak pada arah yang sama; X dan Y meningkat/menurun.
- 1 menunjukkan tingkatan tertinggi hubungan berlawanan arah, jika nilai kedua variabel bergerak pada arah yang berlawanan; X meningkat dan Y menurun atau sebaliknya.

✓ Tipe Korelasi

- Pearson**, dikenal sebagai r , R , atau Pearson's r , mengukur kekuatan dan arah hubungan linier antara dua variabel; dikenal dan sering digunakan; merujuk pada *Pearson product-moment correlation coefficient*.
- Spearman Rank**, ukuran tentang hubungan antara ranking dari 2 variabel atau dua ranking dari variabel yang sama; mengukur bagaimana hubungan antar dua variabel yang digambarkan mengikuti fungsi *monotonic*.
- Korelasi Kendall Tau, Korelasi Ganda, Korelasi Parsial

✓ Korelasi vs Regresi

- Apabila dua variabel sudah diketahui mempunyai hubungan (korelasi) yang kuat dengan melihat nilai koefisien korelasinya, maka keterangan (nilai) dari variabel yang satu dapat dipergunakan untuk **memperkirakan/meramalkan nilai variabel lainnya**.
- Salah satu teknik atau metode untuk memperkirakan/meramalkan nilai suatu variabel (variabel dependen) berdasarkan nilai variabel lain (variabel independent) adalah **metode Regresi Linier**.



Nilai Koefisien korelasi Pearson:

- $0 < |r_{XY}| < 0,49$: Hubungan lemah
- $0,50 < |r_{XY}| < 0,79$: Hubungan sedang
- $0,80 < |r_{XY}| < 1$: Hubungan kuat

- $r_{XY} = +1$, jika hubungan X dan Y linier meningkat sempurna.
- $r_{XY} = -1$, jika hubungan X dan Y linier menurun sempurna.
- interval $(-1, 1)$, menunjukkan tingkat hubungan linear antara X dan Y.
- $r_{XY} \rightarrow 0$, jika hubungan X dan Y rendah atau tidak berkorelasi.
- $r_{XY} \rightarrow \pm 1$ hubungan antara X dan Y semakin kuat.

✓ Regresi digunakan untuk meneliti hubungan antara 2 atau lebih variabel.

✓ **Banyak variabel yang diteliti:**

- Satu X dan satu Y → sederhana, simple regression
- Banyak X dan satu T → ganda, multiple regression

✓ **Bentuk Regresi**

- Linier → plot pasangan (x, y) membentuk persamaan linier (garis lurus)
- Non-linier → pola dari diagram pencar dapat diperkirakan (dianggap) membentuk persamaan polinomial, eksponensial, logaritma, moving average, dll

✓ **Bentuk Kejadian**

Deterministik → kejadian yang memiliki derajat kepastian yang cukup tinggi sehingga fungsinya dapat dibuat dengan pasti.

Contoh:

Sebuah toko menjual buku tulis seharga Rp.2000 per buah. Jika ingin membeli 2 buku berarti harus membayar Rp.4000. Secara mudah fungsi deterministik dari uang yang harus dibayarkan untuk membeli buku tulis adalah:

$$Y = 2000 X,$$

dimana,

Y adalah uang yang harus dibayarkan,

X adalah banyaknya buku yang dibeli.

Stokastik → kejadian yang memiliki unsur ketidakpastian

Contoh:

Jika seorang mahasiswa yang rata-rata lama belajarnya 1 jam per hari mendapat IPK 3,0. yang 2 jam per hari mendapat IPK 3,1 dan yang 3 jam per hari mendapat IPK 3,5. Kalau ada mahasiswa yang rata-rata lama belajarnya **4 jam** per hari, berapa IPK-nya?

Apapun jawabannya tentu **mengandung ketidakpastian** karena tidak hanya lamanya belajar yang berpengaruh terhadap IPK.

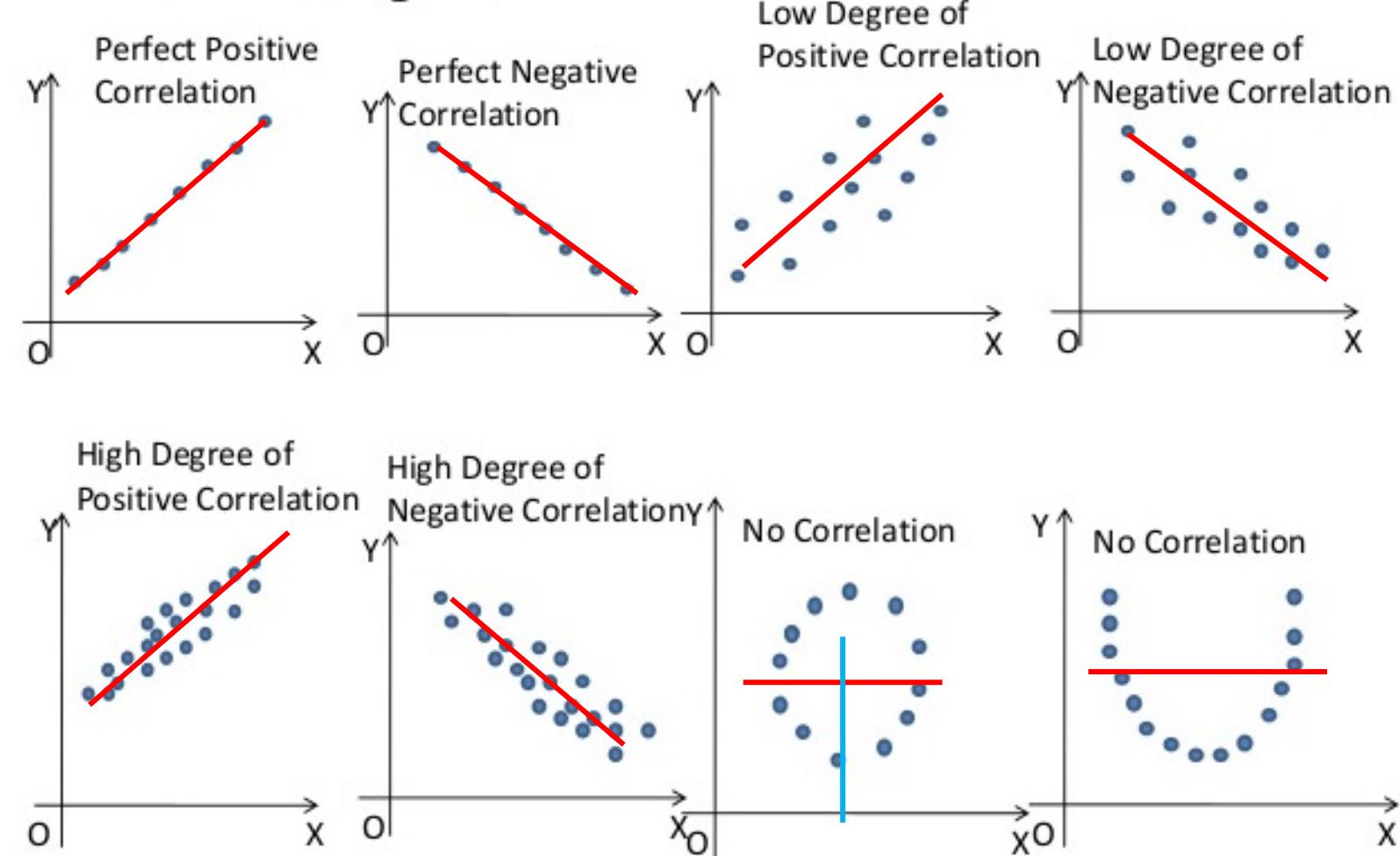
- ✓ Analisis Regresi meneliti bentuk hubungan antara variabel: variabel menentukan (**variabel bebas** → variabel yang menjelaskan, notasi X) dan variabel respons (**variabel tak bebas** → variabel yang dijelaskan, notasi Y).

- ✓ Dengan persamaan regresi, dapat dilakukan **prediksi besaran Y** untuk nilai X yang sudah diketahui, => X mempengaruhi Y, tidak sebaliknya

- ✓ Pembahasan regresi tidak bisa dipisahkan dari korelasi, yaitu (seberapa kuat) hubungan antara dua variabel tersebut.

EKO TE

Scatter Diagram



Bentuk regresi linier (garis lurus)

○ Bentuk umum regresi linier sederhana

$$Y = \alpha + \beta X + e$$

dimana,
 α dan β adalah parameter yang tidak diketahui dan akan diestimasi,
 e adalah komponen ketidakpastian (error atau residual).

○ Estimasi untuk regresi linier:

$$\hat{Y} = a + bX$$

dimana,
 b : regressor (estimator untuk β),
 a : konstanta (estimator untuk α),

E K U
Koefisien determinasi: $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

atau: $R^2 = (r_{xy})^2$

r_{xy} adalah koefisien korelasi Pearson

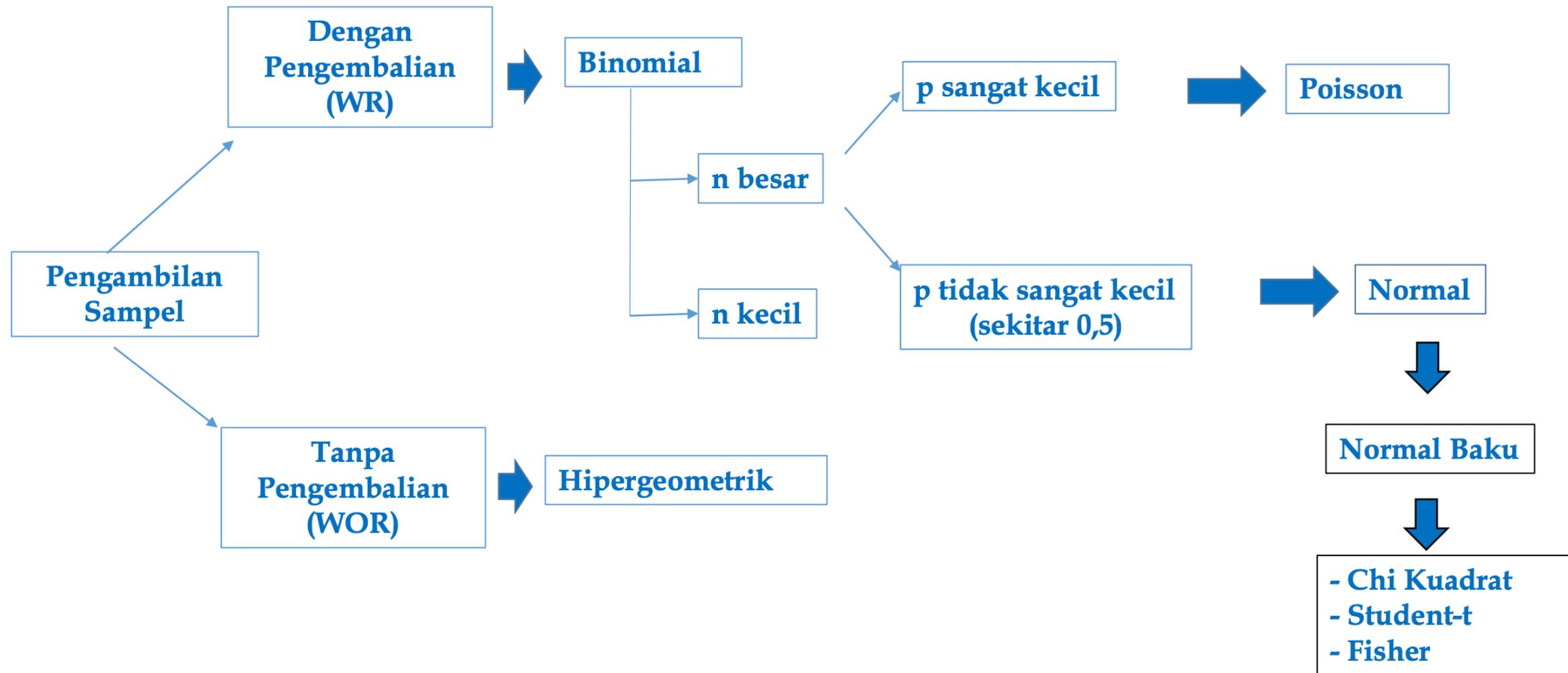
$$e = \hat{Y} - Y$$

Bagaimana cara menentukan atau mengestimasi nilai a dan b sedemikian rupa sehingga $e \rightarrow 0$?

Terdapat beberapa cara untuk mengestimasi nilai a dan b , yaitu:

1. Free-hand method (metode tangan bebas)
2. Metode setengah rata-rata
3. Metode kuadrat terkecil (MKT) atau Ordinary Least squared (OLS)

- ✓ **Koefisien Regresi** (a – intersep atau konstanta dan b – koefisien regresi yang dimaknai sebagai perubahan nilai sebesar b pada y jika x berubah satu unit)
- ✓ **Koefisien Determinasi** merupakan ukuran yang mengkaji **kemampuan model regresi memprediksi atau menjelaskan** suatu hasil (outcome) dalam regresi linear
- ✓ Lebih spesifik, R^2 menunjukkan **proporsi** dari varians dari variabel dependen (Y) yang dapat dijelaskan oleh model regresi linier dan variabel prediktor (X).
- ✓ **$R^2 = 0,98$** ==> 98 persen dari variasi (keragaman) dari Y dapat dijelaskan oleh model (keragaman dari variabel X), persentase tersebut sangat tinggi.





R Programming for Statistical Data Analysis

**KORELASI, REGRESI, DAN
NORMALITAS**



✓ Membaca RAW data excel dengan R

```
# Membaca raw data excel
install.packages("readxl") # install packages readxl

library(readxl) # load library readxl

data_ipm_hls <- read_excel(file.choose()) # memunculkan popup untuk memilih file excel
data_ipm_hls # menampilkan data

View(data_ipm_hls)
```

	Kabupaten/Kota	IPM	HLS
1	A1	67.25	12.27
2	A2	69.25	12.56
3	A3	66.23	11.51
4	A4	63.15	10.70
5	A5	65.67	12.07
6	A6	70.12	13.03
7	A7	65.20	11.34
8	A8	66.35	12.00
9	A9	70.34	11.65

END

```
> shapiro.test(data_ipm_hls$IPM)
Shapiro-Wilk normality test
data: data_ipm_hls$IPM
W = 0.94138, p-value = 0.06165

> shapiro.test(data_ipm_hls$HLS)
Shapiro-Wilk normality test
data: data_ipm_hls$HLS
W = 0.94315, p-value = 0.06981
```

Melakukan pengecekan normalitas ($z = \bar{x} / s_{\text{baku}}$) -> UJI NORMALITAS

```
install.packages("nortest") # install packages nortest

library(nortest) # load library nortest

# Shapiro-Wilk normality test
shapiro.test(data_ipm_hls$IPM)

# p-value = 0.06165, > 5% -> 0.05 --> variabel IPM berdistribusi normal

shapiro.test(data_ipm_hls$HLS)

# p-value = 0.06981, > 5% --> variabel HLS berdistribusi normal
```

nODO

✓ Menghitung korelasi

Melakukan pengujian KORELASI - PEARSON

```
cor.test(data_ipm_hls$IPM, data_ipm_hls$HLS, method = 'pearson')
```

nilai korelasi = 0.893947 -> korelasi kuat, positif -> searah = HLS naik, IPM naik

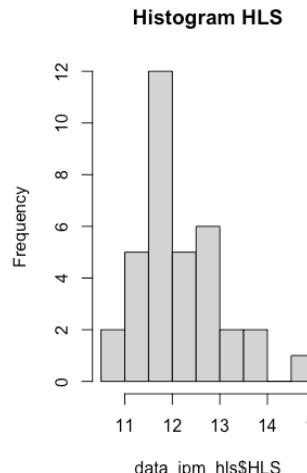
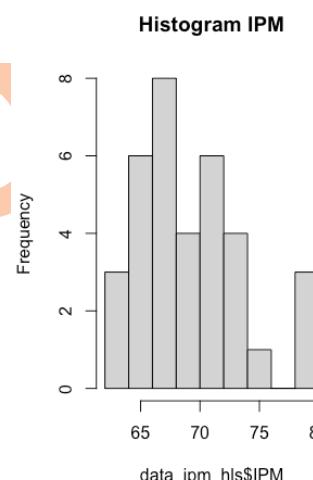
```
> cor.test(data_ipm_hls$IPM, data_ipm_hls$HLS, method = 'pearson')
```

Pearson's product-moment correlation

```
data: data_ipm_hls$IPM and data_ipm_hls$HLS
t = 11.458, df = 33, p-value = 4.816e-13
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7986102 0.9455189
sample estimates:
```

```
cor
0.893947
```

EKC



RKL

Visualisasikan

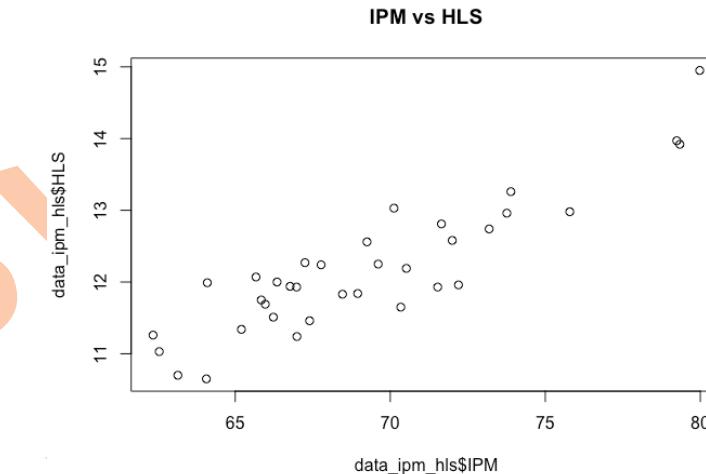
```
plot(data_ipm_hls$IPM, data_ipm_hls$HLS, main = "IPM vs HLS")
```

par membagi tampilan menjadi sekian kolom

```
par(mfrow = c(1, 2))
```

```
hist(data_ipm_hls$IPM, main = "Histogram IPM")
```

```
hist(data_ipm_hls$HLS, main = "Histogram HLS")
```



data_ipm_hls\$IPM

data_ipm_hls\$HLS

✓ Menghitung regresi

```
# REGRESI LINIER
# Menggunakan data curah hujan dan penjualan payung

# 1. Membaca data dari file excel
data_hujan_payung <- read_excel(file.choose())
data_hujan_payung

# Rename nama kolom menjadi curah_hujan dan penjualan_payung
colnames(data_hujan_payung)[colnames(data_hujan_payung) == "Curah Hujan (X)"] = "curah_hujan"
colnames(data_hujan_payung)[colnames(data_hujan_payung) == "Penjualan Payung (Y)"] = "penjualan_payung"
data_hujan_payung

# 2. Visualisasi data
plot(data_hujan_payung$curah_hujan, data_hujan_payung$penjualan_payung,
      main = "Hubungan antara Curah Hujan dan Penjualan Payung",
      xlab = "Curah Hujan (mm)", ylab = "Penjualan Payung")

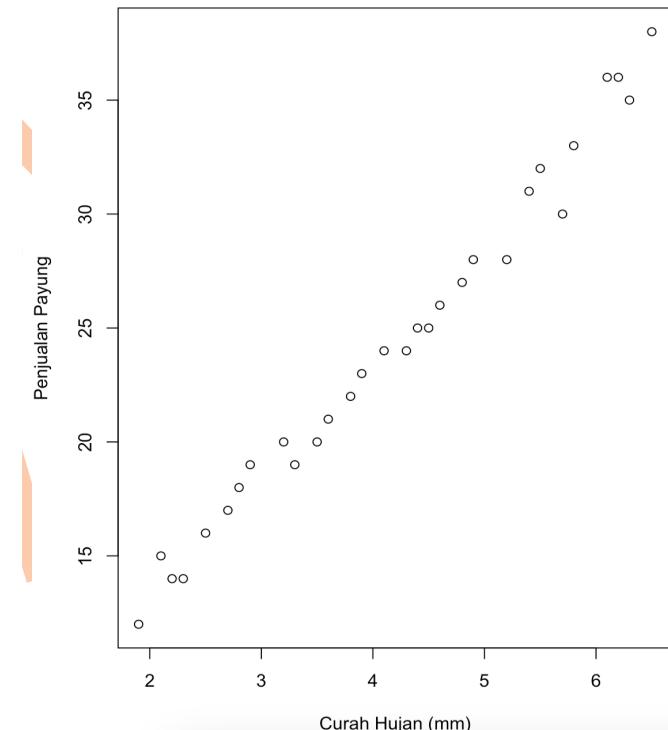
# 3. Uji Normalitas
shapiro.test(data_hujan_payung$curah_hujan)
shapiro.test(data_hujan_payung$penjualan_payung)
# --> kedua variabel p-value > 5% -> distribusi normal

# 4. Uji Korelasi
cor.test(data_hujan_payung$curah_hujan, data_hujan_payung$penjualan_payung, method = 'pearson')
# --> korelasinya 99% (sangat kuat) dan searah -> semakin tinggi curah hujan, penjualan payung semakin tinggi

# 5. Regresi Linier Sederhana
model_regresi <- lm(data_hujan_payung$penjualan_payung ~ data_hujan_payung$curah_hujan)
summary(model_regresi)
```



Hubungan antara Curah Hujan dan Penjualan Payung



```
Call:
lm(formula = data_hujan_payung$penjualan_payung ~ data_hujan_payung$curah_hujan)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.23880 -0.54954 -0.06048  0.66911  1.68151 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.6033    0.5589   4.658 7.06e-05 ***
data_hujan_payung$curah_hujan  5.1992    0.1274  40.808 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9567 on 28 degrees of freedom
Multiple R-squared:  0.9835,    Adjusted R-squared:  0.9829 
F-statistic: 1665 on 1 and 28 DF,  p-value: < 2.2e-16
```

Open the Link:

<https://github.com/ekotwidodo/UTI-Statistika-Dasar>

Download File:

04-Korelasi Regresi Normalitas.R



- Douglass A, Lind, William G, Marchal and Samuel A, Wathen. *Statistical Techniques in Business & Economics*. Mc. Graw Hill, NY, 2005.
- Supranto, J, MA. *Statistik-Teori dan Aplikasi*, Jilid I, edisi kedelapan. Erlangga. 2016. Jakarta
- Anderson, T.W, and Sclove, L. Stanley, *The Statistical Analysis of Data*, Second Edition, Houghton Mifflin Company, 1986, USA.
- Johnson R, and Bhattacharyya G. *Statistics, Principles and Methods*. John Wiley & Sons Inc. 1985. Canada.
- Asra A dan Sutomo S. *Pengantar Statistika I*. Rajawali Pers. 2016. Jakarta
- Asra A dan N.B. Parwanto. 2018. *1001 Soal Jawab Statistika Deskriptif*. Jakarta. In Media.
- ...

EKO TEGUH WIDODO
MARKED BY



lpdp
lembaga pengelola dana pendidikan

Kampus
Merdeka
INDONESIA JAYA

PRAKTIKI
MENGAJAR



THANK YOU

EKO

Link Materials:

<https://github.com/ekotwidodo/UTI-Statistika-Dasar>

Requests for Collaboration:

