KUANTITATIF DAN KUALITATIF LINGKUNGAN

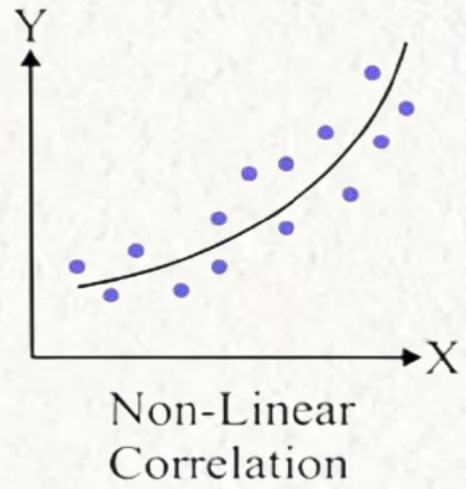
(STATISTIKA LINGKUNGAN)

Prof. Dr. Ir. Zulkifli Alamsyah, M.Sc.

PROGRAM STUDI MAGISTER ILMU LINGKUNGAN
PASCASARJANA UNIVERSITAS JAMBI

KUANTITATIF DAN KUALITATIF LINGKUNGAN

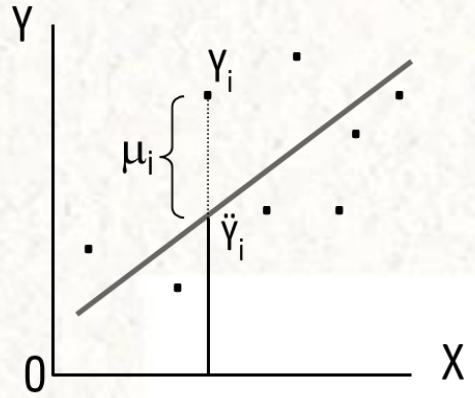
(STATISTIKA LINGKUNGAN)



Non-Linear
Correlation

STATISTIKA INFERENSIAL

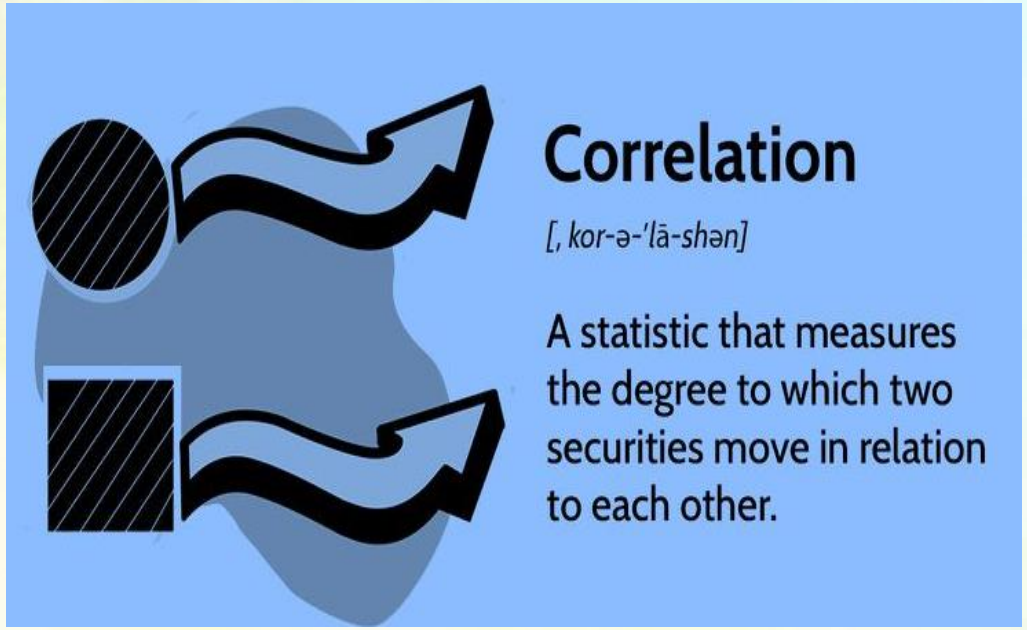
(Analisis Korelasi dan Regresi)



Prof. Dr. Ir. Zulkifli Alamsyah, M.Sc.

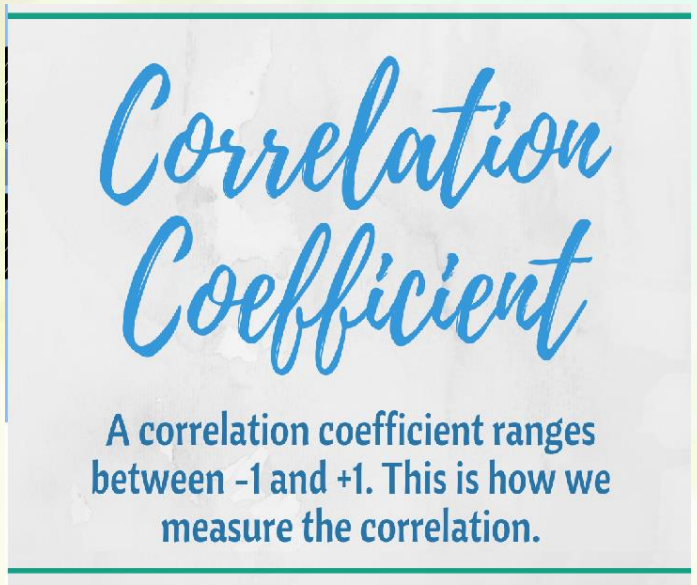
PROGRAM STUDI MAGISTER ILMU LINGKUNGAN
PASCASARJANA UNIVERSITAS JAMBI

PENGERTIAN KORELASI

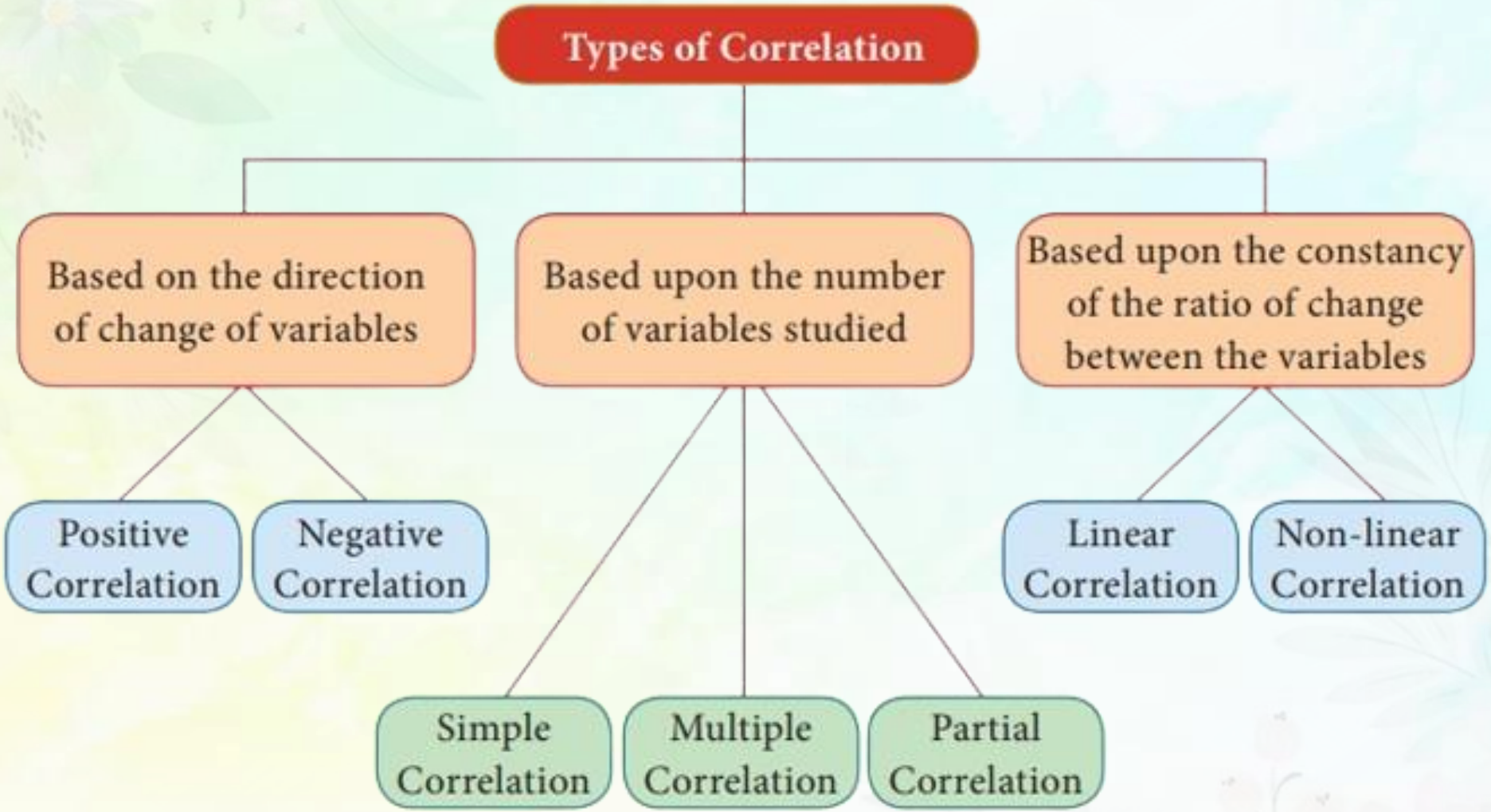


- Korelasi adalah ukuran statistik yang mengekspresikan sejauh mana dua variabel terkait secara linier (artinya mereka berubah bersama pada tingkat konstan).
- Mengukur derajat hubungan yang terjadi antar variabel-variabel ekonomi, tanpa membuat pernyataan tentang sebab dan akibat.
- **Scatterplot** dapat digunakan untuk memeriksa apakah dua variabel berhubungan atau tidak
- Hubungan antara 2 Variabel (Misal X dan Y) dapat **linear, non-linear, positif** atau **negatif**

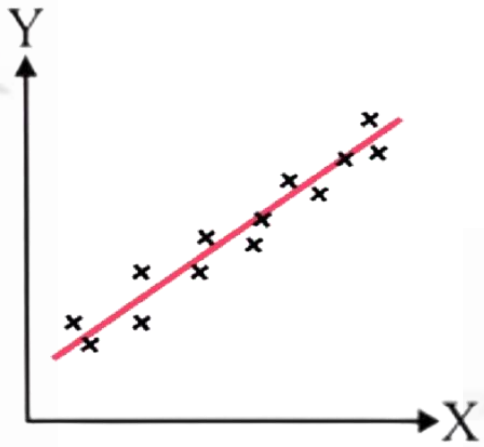
Koefisien Korelasi



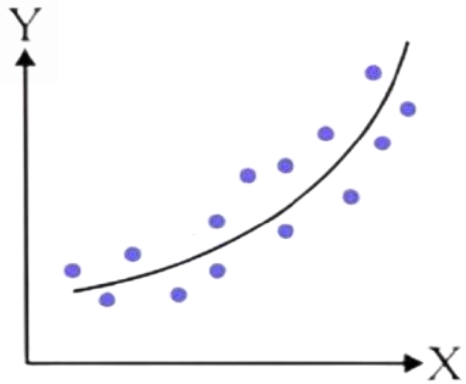
- Koefisien korelasi (disimbolkan dengan **r**) adalah nilai yang menunjukkan kekuatan hubungan yang ada di antara dua variabel.
- Nilai Koefisien korelasi berada antara angka -1 dan 1.
- -1: Korelasi negatif sempurna.
- 0: Tidak ada korelasi. Variabel tidak memiliki hubungan apa pun.
- +1: Korelasi positif sempurna.
- Korelasi antara dua variabel dikatakan positif atau langsung jika peningkatan (atau penurunan) pada satu variabel diikuti oleh peningkatan (atau penurunan) variabel yang lain.
- Korelasi antara dua variabel dikatakan negatif atau terbalik jika peningkatan (atau penurunan) pada satu variabel diikuti oleh penurunan (atau peningkatan) variabel yang lain.
- Semakin dekat r kepada nol, semakin lemah hubungan antara dua variabel


Types of Correlation
Based on the direction of change of variables
Based upon the number of variables studied
Based upon the constancy of the ratio of change between the variables
Positive Correlation
Negative Correlation
Simple Correlation
Multiple Correlation
Partial Correlation
Linear Correlation
Non-linear Correlation

Scatterplot Berbagai Bentuk Korelasi



Linear Correlation

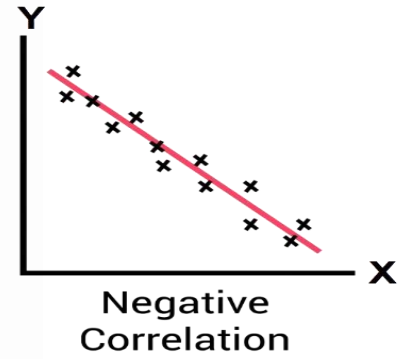


Non-Linear Correlation

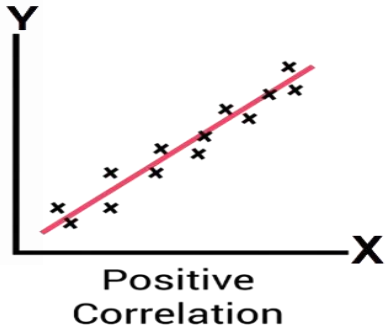
Korelasi Linear:
If semua titik (X,Y) pd diagram pencar mendekati bentuk garis lurus.

Korelasi Non-linear:
If semua titik (X,Y) pd diagram pencar tidak membentuk garis lurus.

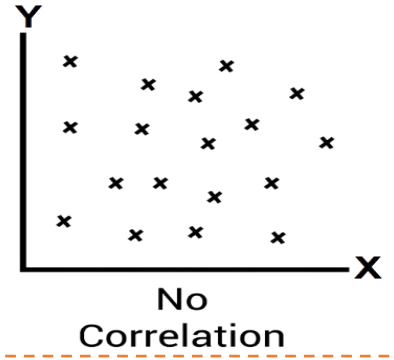
Scatterplot Berbagai Bentuk Korelasi



Korelasi Negatif:
If jika arah perubahan kedua variabel tidak sama ⇨ If X naik, Y turun.



Korelasi Positif:
If jika arah perubahan kedua variabel sama ⇨ If X naik, Y juga naik.



Dua Variabel Tidak Berkorelasi:
If semua titik (X,Y) pd diagram pencar tidak dapat ditentukan bentuknya

Multiple Correlation And Partial Correlation

- **Multiple correlation** measures the strength of the relationship between one dependent variable (Y) and two or more independent variables (X_1, X_2, …, X_k) simultaneously.
- It answers the question:
  → "How strongly do all independent variables together predict Y?"
- **Partial correlation** measures the relationship between two variables while controlling for (removing) the effect of one or more other variables.
- It answers the question:
  → "What is the pure relationship between X and Y after removing the influence of other variables?"

Mengukur Korelasi Linear

- Scatterplot (diagram pencar): Pencaran titik dari garis.
  Jika titik-titik mendekati garis ⇨ Korelasi yang kuat.

- Koefisien Korelasi (ρ atau **r**) ⇨ ρ untuk penduga populasi
  r koefisien korelasi dari sampel

$$r_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

r_{xy} = Koefisien korelasi antara X dan Y

$x_i = \bar{X} - X_i$; i = pengamatan

$y_i = \bar{Y} - Y_i$; n = jumlah pengamatan

Tanda garis diatas variabel menunjukkan nilai rata-rata.

Atau

$$r_{xy} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)\ (n \sum Y^2 - (\sum Y)^2)}}$$

Rumus koefisien korelasi diatas:

- Simetri terhadap X dan Y ⇨ $r_{xy} = r_{yx}$
- Hanya berlaku untuk hubungan yang linear
- Nilai r terletak antara -1 dan 1; $-1 < r < 1$.
  - r = -1: Korelasi negatif sempurna
  - r = 1: Korelasi positif sempurna
  - r = 0: Tidak berkorelasi

Beberapa catatan tentang nilai r:

- Secara empiris, hampir tidak pernah ditemukan korelasi sempurna (semua titik terpencar tepat pada garis).
- Nilai **r** yang mendekati nol menunjukkan derajat hubungan yang lemah.
- Koefisien **r** merupakan estimasi sampel terhadap koefisien korelasi populasi, ρ.
- Nilai **r** mengandung error, sehingga perlu diuji reliabilitasnya.

Contoh Penghitungan:
Korelasi Antara Harga (X) dan Suplai (Y)

n	X	Y	X^2	Y^2	XY
1	2	10			
2	4	20			
3	6	50			
4	8	40			
5	10	50			
6	12	60			
7	14	80			
8	16	90			
9	18	90			
10	20	120			
∑	**110**	**610**	1.540	47.700	8.520

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

Contoh Penghitungan:
Korelasi Antara Harga (X) dan Suplai (S)

n	X	Y	X^2	Y^2	XY
1	2	10	4	100	20
2	4	20	16	400	80
3	6	50	36	2.500	300
4	8	40	64	1.600	320
5	10	50	100	2.500	500
6	12	60	144	3.600	720
7	14	80	196	6.400	1.120
8	16	90	256	8.100	1.440
9	18	90	324	8.100	1.620
10	20	120	400	14.400	2.400
∑	110	610	1.540	47.700	8.520

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

n = 10
∑X = 110
∑Y = 610
$\sum X^2$ = 1.540
$\sum Y^2$ = 47.700
∑XY = 8.520

r_{xy} = + 0,97

Penarikan Kesimpulan Analisis Korelasi

Ada dua metode untuk pengambilan Kesimpulan dari analisis korelasi.

- Metode 1: Menggunakan nilai *p*
- Metode 2: Menggunakan tabel nilai kritis (t)

Beberapa Standar Interpretasi Koefisien Korelasi

r Range	Cohen (1988)	Evans (1996)	Hinkle et al. (2003)
0.00 – 0.10	Small (weak)	Very weak	Low
0.11 – 0.19	Small (weak)	Very weak	Low
0.20 – 0.29	Medium	Weak	Low
0.30 – 0.39	Medium	Weak	Moderate
0.40 – 0.49	Medium–Large	Moderate	Moderate
0.50 – 0.59	Large	Moderate	High
0.60 – 0.69	Large	Strong	High
0.70 – 0.79	Very large	Strong	Very high
0.80 – 1.00	Very large	Very strong	Very high

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences.

Evans, J. D. (1996). Straight forward Statistics for the Behavioral Sciences.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Applied Statistics for the Behavioral Sciences.

Menguji Signifikansi Koefisien Korelasi

- Jika koefisien korelasi untuk seluruh populasi (ρ) tidak diketahui, maka diestimasi menggunakan koefisien korelasi dari data sampel.(r).

  ρ = koefisien korelasi populasi (tidak diketahui)

  r = koefisien korelasi sampel (diketahui; dihitung dari data sampel)

- Uji signifikansi bertujuan untuk menguji apakah nilai **ρ** mendekati nol atau **berbeda secara signifikan dari nol**.

- Jika hasil uji menyimpulkan bahwa koefisien korelasi berbeda secara signifikan dari nol, maka disimpulkan bahwa kedua variabel yang dianalisis berkorelasi secara **signifikan** (nyata).

Langkah-Langkah Menguji Signifikansi Koefisien Korelasi

1. **Rumuskan hipotesis statistik.**

Hipotesis nol	H_0 : $\rho = 0$
Hipotesis alternatif	H_a : $\rho \neq 0$ atau $\rho > 0$ atau $\rho < 0$

2. **Gunakan alat uji statistik t untuk korelasi.**

$$t_c = \frac{r}{\sqrt{(1-r^2)/(n-2)}} \quad \text{atau} \quad t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

3. **Penetapan keputusan**

Tetapkan tingkat signifikansi (α) biasanya 5% (α=0.05) dan nilai kritis t (nilai t-tabel)

Bila t_c **> t-tabel** maka **tolak Ho**, artinya kedua variabel **berkorelasi signifikan**

$t_c \leq$ **t-tabel** maka **terima Ho**, artinya kedua variabel **tidak berkorelasi signifikan**

Cara cepat untuk Menguji Signifikansi Koefisien Korelasi

Cara cepat untuk menentukan apakah korelasi antara dua variabel signifikan atau tidak, dapat menggunakan formula:

$$|r| \geq \frac{2}{\sqrt{n}}$$

Makin banyak jumlah pengamatan (n), maka peluang terjadi korelasi yang signifikan antara dua varabel semakin besar.

Latihan:

Gunakan data berikut untuk menguji apakah terjadi korelasi yang signifikan antara X dan Y

OBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	2.3	2.8	3.1	3.6	4.2	4.8	5.3	5.7	6.2	6.8	7.1	7.5	7.8	8.3	8.7
Y	3.1	3.5	4	4.2	5.1	5.5	6.4	6.8	7.5	8	8.3	8.7	9.2	9.8	10.2

Metode Analisis Korelasi

Koefisien Korelasi Pearson

- **Deskripsi:** Mengukur kekuatan dan arah hubungan linear antara dua variabel kontinu.
- **Rentang:** -1 hingga +1.
- **Contoh Kasus Penggunaan:** Mempelajari hubungan antara luas lahan dan produksi pertanian

Korelasi Peringkat Spearman

- **Deskripsi:** Menilai hubungan monotonik antara dua variabel berperingkat (ordinal).
- **Rentang:** -1 hingga +1.
- **Contoh Kasus Penggunaan:** Membandingkan peringkat kinerja siswa dalam dua ujian.

Korelasi Tau Kendall

- **Deskripsi:** Mengukur kekuatan hubungan antara dua variabel dengan mempertimbangkan peringkat relatif data.
- **Rentang:** -1 hingga +1.
- **Contoh Kasus Penggunaan:** Cocok untuk ukuran sampel kecil atau ketika banyak peringkat yang sama.

Metode Analisis Korelasi

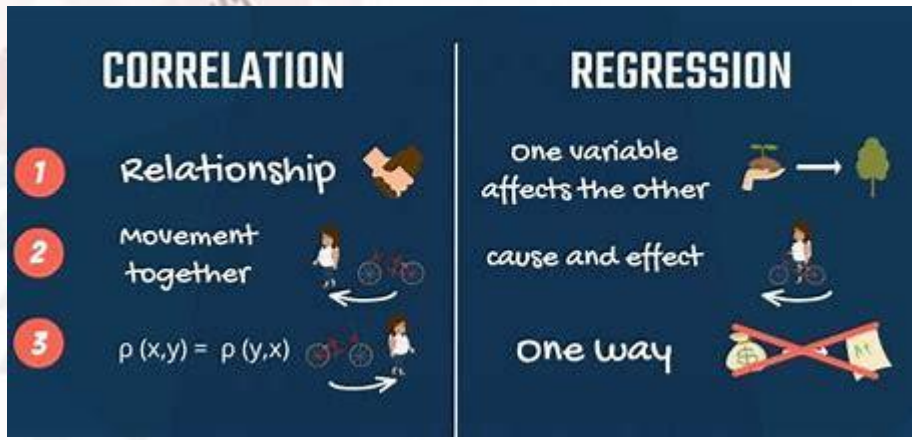
Korelasi Titik-Biserial

- **Deskripsi:** Digunakan ketika satu variabel bersifat kontinu dan variabel lainnya bersifat dikotomis (biner).
- **Rentang:** -1 hingga +1.
- **Contoh Kasus Penggunaan:** Memeriksa hubungan antara jenis kelamin (laki-laki/perempuan) dan nilai ujian.

Korelasi Parsial

- **Deskripsi:** Mengukur hubungan antara dua variabel sambil mengendalikan pengaruh satu atau lebih variabel tambahan.
- **Contoh Kasus Penggunaan:** Mempelajari korelasi antara olahraga dan kadar kolesterol sambil mengendalikan usia.

Perbedaan Antara Korelasi dengan Regresi

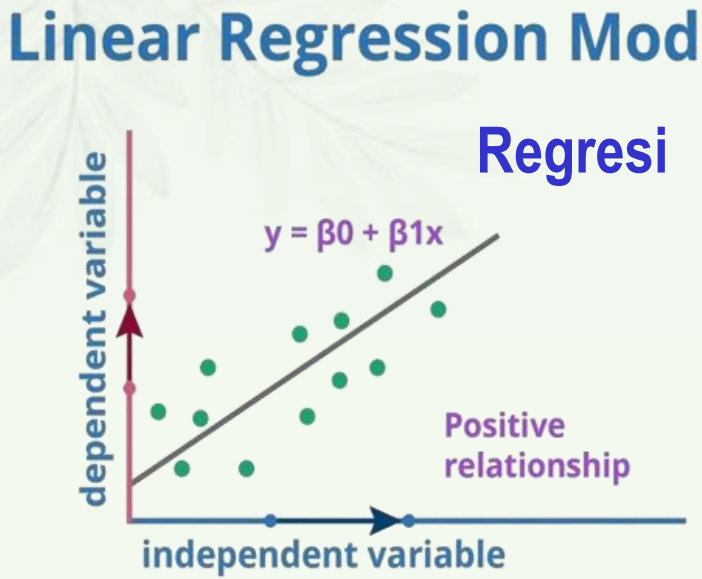


- Korelasi tidak menunjukkan hubungan sebab akibat (cause and effect). Korelasi antara dua variabel tidak berarti bahwa perubahan pada satu variabel menyebabkan perubahan pada variabel yang lain.
- Korelasi hanya mengevaluasi hubungan antara dua variabel, dan berbagai keadaan dapat menyebabkan terjadinya korelasi.

- Regresi mengukur ketergantungan suatu variabel (dependent/terikat) terhadap satu atau lebih variabel lain (explanatory/independent/bebas)
- Tujuan dari Regresi adalah:
  - mengestimasi nilai tengah variabel terikat dari nilai rata-rata variabel bebas.
  - Untuk menguji hipotesis mengenai sifat ketergantungan sesuai dengan teori

Linear Regression Model

Measuring the dependence of a dependent variable on 1 or more independent variables (explanatory/independent variables)



Purpose:

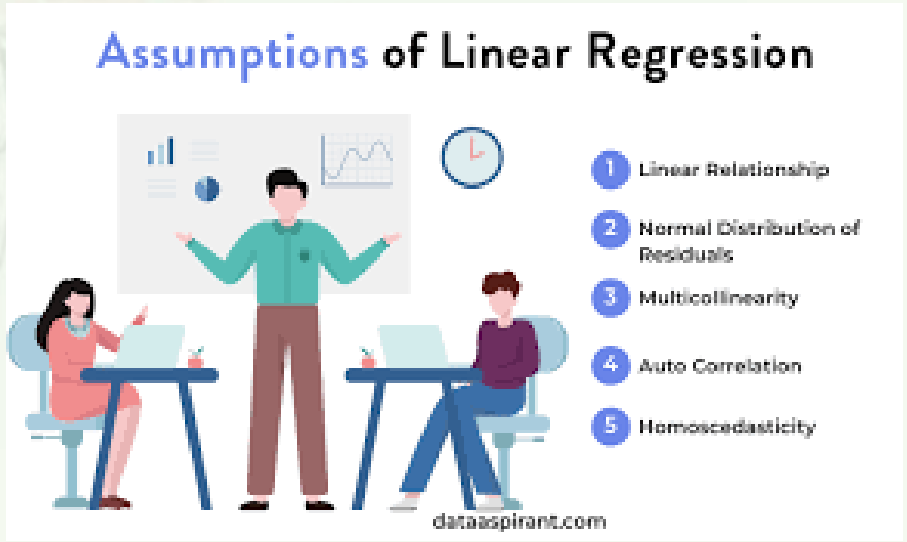
- Estimating the mean value of the dependent variable from the mean value of the independent variable.
- To test the hypothesis regarding the nature of dependence according to the theory.

Difference with Correlation

- Asymmetry in treating variables:
  - The independent variable is deterministic
  - The dependent variable is stochastic (random)

⇨ For each given value of X, it can give several values of Y with certain probabilities.

Asumsi-asumsi Mengenai μ_i:

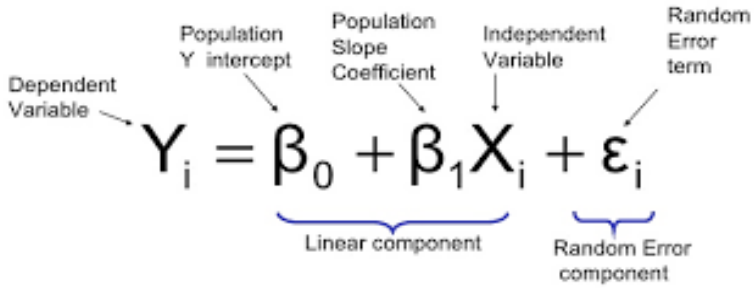


1. μ_i is a random variable and distributed normally.
2. Mean of $\mu_i = 0$, $e(\mu_i) = 0$.
3. No correlations among μ_i $cov(\mu_i,\mu_j) = 0$
4. Homoscedasticity, $var(\mu_i) = \sigma^2$
5. $Cov(\mu_i,X_i) = 0$
6. There is no bias in the model specifications
7. There is no multi-collinearity between explanatory variables

Model Regresi Sederhana

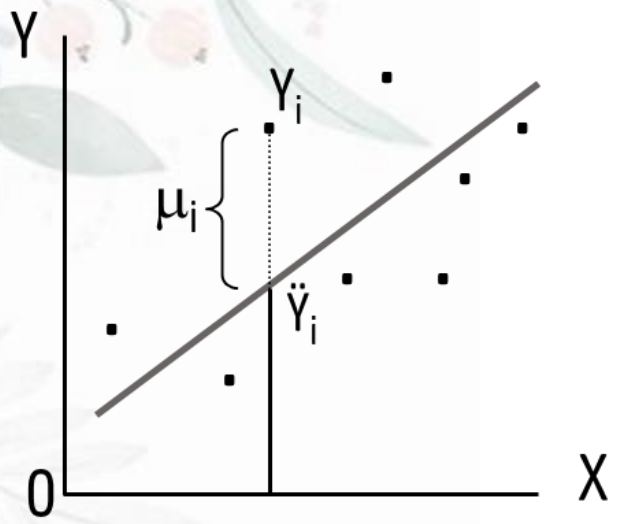
$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

- β_0 dan β_1 : parameters of the function whose value is to be estimated.
- Stochastic ⇨ for each value of X there is a probability distribution of all values of Y or the value of Y cannot be predicted with certainty because there is a stochastic factor μ_i which gives Y a random nature.



- The existence of the variable i is caused by:
  - Incomplete theory
  - Random human behavior
  - Incomplete model specifications
  - Errors in aggregation
  - Errors in measurement

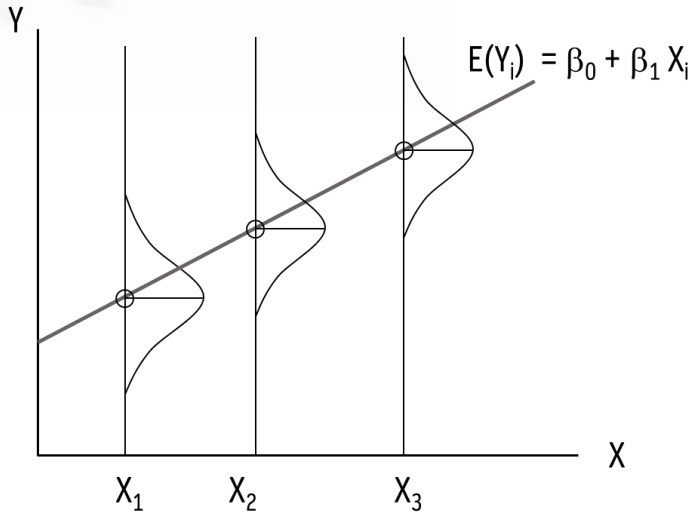
Zulkifli Alamsyah



$$\ddot{Y}_i = b_0 + b_1 X_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

{ Variation in Y } { Systematic Variation } { Random Variation }



$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

Mean of Y_i $(\bar{Y})$:

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

$$\mu_I = Y_i - E(Y_i)$$

Parameter Estimation in Simple Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

β_0 and β_1 = population parameters
β_0 = intercept / Constant
β_1 = Slope of the regression line

Method of Ordinary Least Squares

$$Y_i = \widehat{Y_i} + \widehat{u_i}$$

$$\widehat{u_i} = Y_i - \widehat{Y_i}$$

error or residual term $\widehat{u_i}$ is the difference between the actual Y_i and estimated $\widehat{Y_i}$

$$\widehat{u_i} = Y_i - (\widehat{\beta_1} + \widehat{\beta_2} X_i) = Y_i - \widehat{\beta_1} - \widehat{\beta_2} X_i$$

Ordinary Least Square (OLS) Method

Principle: Minimize the error value – find the smallest sum of squared deviations ($\Sigma \mu_i^2$).

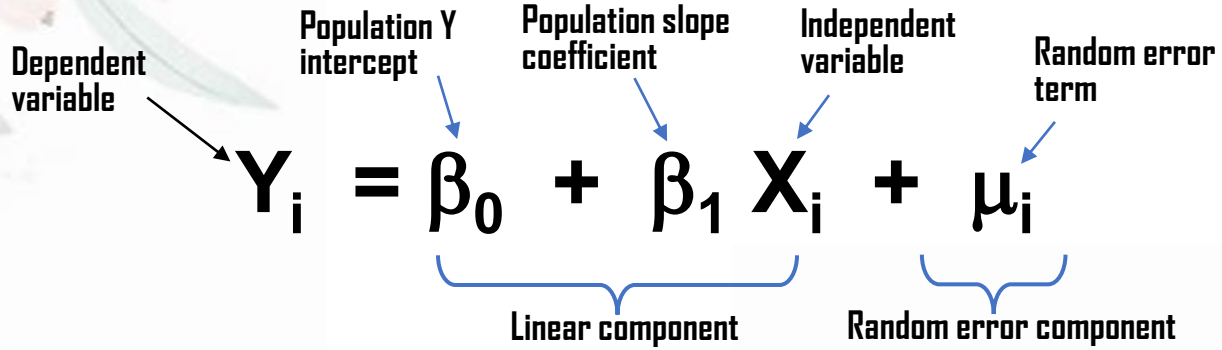
$$\mu_i = Y_i - \beta_0 - \beta_1 X_i$$
$$\mu_i^2 = (Y_i - \beta_0 - \beta_1 X_i)^2$$
$$\Sigma \mu_i^2 = \Sigma (Y_i - \beta_0 - \beta_1 X_i)^2$$

$\Sigma \mu_i^2$ minimum if:

$$\partial \Sigma \mu_i^2 / \partial \beta_0 = 0 \Rightarrow 2 \Sigma (Y_i - \beta_0 - \beta_1 X_i) = 0$$
$$\partial \Sigma \mu_i^2 / \partial \beta_1 = 0 \Rightarrow 2 \Sigma X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

Zulkifli Alamsyah

Prediction of Parameter Values



$$y_i = b_0 + b_1 X_i + \varepsilon_i$$

$$b_1 = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Where:

b_0 and b_1 are estimated values for β_0 and β_1.

$\bar{X}$ and $\bar{Y}$ are the averages of the X dan Y

Zulkifli Alamsyah

Prediction of Parameter Values

Standar error:

$$SE(b_1) = \left\{ \frac{\sigma^2}{\Sigma (X_i - \bar{X})^2} \right\}^{½}$$

$$SE(b_0) = \left\{ \frac{\Sigma X_i^2}{N \Sigma (X_i - \bar{X})^2} \right\}^{½} \sigma$$

σ Estimated from s, where:

$$s = (\Sigma \mu_i^2 / n-2)^2 \quad \text{and} \quad \mu_i^2 = (Y_i - \bar{Y})^2$$

Metode Ordinary Least Squares (OLS)

$Y_i = \beta_0 + \beta_1 X_i + \mu_i$ (1) General Linear Regression Equation

$Y_i = \beta_0 + \beta_1 X_i + \mu_i$ (2) β_0 and β_1 are estimated value for parameter

$\hat{Y}_i = \beta_0 + \beta_1 X_i$ (3) $\hat{Y}i$ = estimated value for model

$Y_i = \hat{Y}_i + \mu_i$ (4) Y_i = Actual value

$\mu_i = Y_i - \hat{Y}_i$ (5) μ_i = Residual value

$$\beta_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{n \sum x_i y_i}{\sum x_i^2}$$

$$\beta_0 = \frac{\sum (X_i)^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \bar{Y} - \beta_2 \bar{X}$$

Parameter coefficients for β_0 and β_1

Zulkifli Alamsyah

Standard error of the estimates (SEE)

$$Var(\beta_1) = \sigma^2 / \sum X_i^2$$

$$Se(\beta_1) = \sqrt{Var(\beta_1)} = \sqrt{\frac{\sigma^2}{\sum X_i^2}} = \frac{\sigma}{\sqrt{\sum X_i^2}}$$

$$Var(\beta_0) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

$$Se(\beta_0) = \sqrt{Var(\beta_1)} = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2} \sigma^2}$$

$$\sigma^2 = \frac{\sum \mu_i^2}{n-2} \qquad \sum \mu_i^2 = \sum y_i^2 - \beta_1^2 \sum x_i^2$$

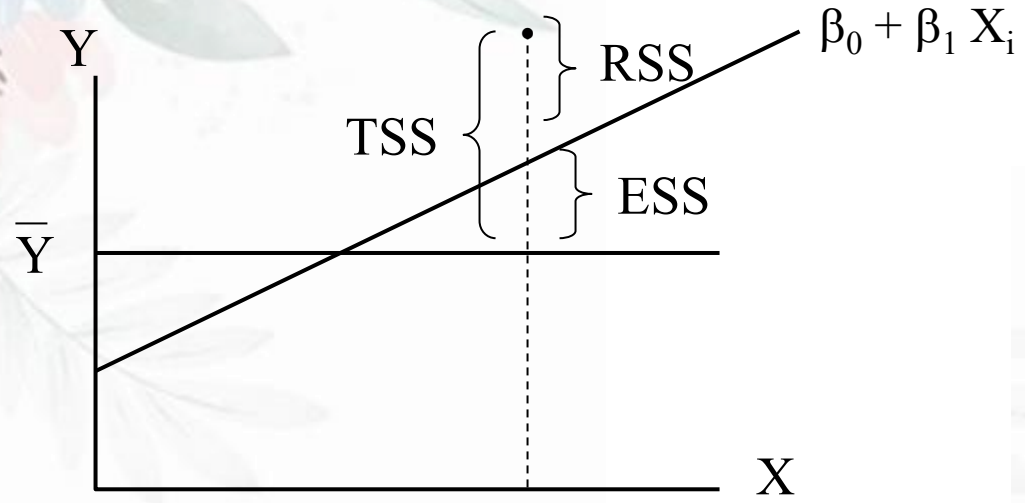
$$= \sum y_i^2 - \frac{\sum (x_i y_i)^2}{\sum x_i^2}$$

- Describes how much deviation or prediction error there is in a regression model.
- Measures how far the predicted value of the regression model deviates from the actual value in the same units as the dependent variable.

Koefisien Determinasi (R^2)

- R^2 is a statistical measure in regression that shows how much of the proportion of variation in the dependent variable can be explained by the independent variables in the regression model.
- R^2 ranges from 0 to 1, where the higher the value, the better the model can explain the data variability.
- R^2 cannot directly indicate the level of prediction error..

Koefisien Determinasi (R^2)



$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - Y)^2}{\sum (Y_i - Y)^2}$$

Or

$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum \mu_i^2}{\sum (Y_i - Y)^2}$$

TSS = RSS + ESS

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$= \frac{\sum (\hat{Y}_i - Y)^2}{\sum (Y_i - Y)^2} + \frac{\sum \mu_i^2}{\sum (Y_i - Y)^2}$$

Or:

$$R^2 = \beta_1^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right)$$

$$= \frac{\sum (x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$$

TSS = Total sum of squares
RSS = Residual sum of squares
ESS = Estimated/Explained sum of squares

Prof. Dr. Ir. Zulkifli Alamsyah, M.Sc.

PROGRAM STUDI MAGISTER ILMU LINGKUNGAN

PASCASARJANA UNIVERSITAS JAMBI