Regression Analysis

EGI SAFITRI, S.MAT., M.SI

Sejarah Regresi

History of Regression

The term Regression was introduced by Fancis Galtom

“Although there is a tendency for tall parents to have tall children, and for short parents to have short children, the height distribution of a population does not change markedly from generation to generation”.

Regression = “Moderate regression”

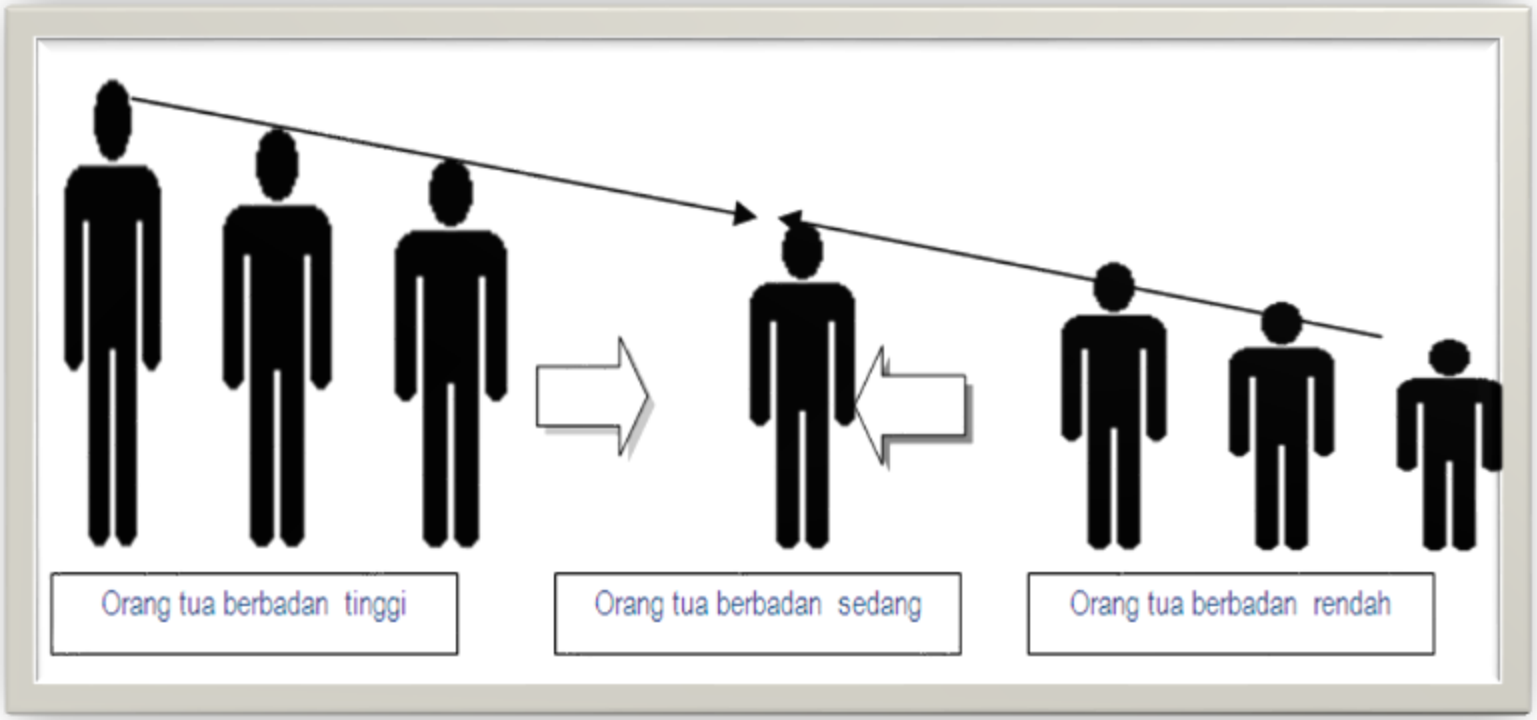
Regression Analysis

One of the tools that can be used in predicting future demand based on past data or to determine the effect of one independent variable on one dependent variable is using linear regression.

Simple linear regression is used for only one independent variable and one independent variable, while multiple linear regression is used for one independent variable and or more independent variables.

The purpose of applying these two methods is to forecast or predict the value of the independent variable which is influenced by the independent variable.

ILLUSTRATION



Example of Application of Regression Analysis

1. Regression analysis between parents' height and their children's height (Gultom).
2. Regression analysis between income and household consumption.
3. Regression analysis between price and sales of goods.
4. Regression analysis between wage rate and unemployment rate.
5. Regression analysis between bank interest rates and stock prices
6. Regression analysis between advertising costs and company sales volume.

Statistical Vs. Functional

Causal relationship (statistical dependency)

Consumption with income

Tenure with productivity

Advertising with sales

Functional/Identity relationship

Liquidity with current assets

Productivity with output

Employee wages with hours worked

The fundamental difference between correlation and regression?

Correlation shows only a relationship.

In the correlation of variables, there is no such thing as dependent and independent variables.

Regression shows an influence relationship.

In regression, there are dependent terms and independent variables.

Variable terms and notation in regression ?

Y	X
Dependent Variable	Independent variable
Explained Variable	Explanatory variable
Predicted Variable	Predictor variable
Regressand	Regressor variable
Response Variable	Stimulus or control variable

Simple Linear Regression Analysis

Introduction

Two numerical variables: want to know the relationship

Two numerical variables: one of the variables is considered as a variable that affects the other variable

Influencing variable: X, independent variable, explanatory variable

Affected variable : Y, independent variable, response variable

Carefully determine the independent and independent variables

Relationship between weight and height

Is there a relationship

If so, which is free and which is not free

Note the relationship

$$Y = a + bX$$

X independent variable

Y independent variable

Consider the following penomena

Yield and amount of fertilizer used

$$Y = a + bX$$

X Volume of fertilizer used

Y Yield

Other forms of multiple linear regression

Yields are not only determined by the volume of fertilizer applied, but also rainfall/irrigation

X1 = fertilizer volume (Free Var)

X2 = Rainfall/Irrigation (?Free Var)

$$Y = a_o + a_1 X_1 + a_2 X_2$$

Simple Linear Regression Form

How to determine the coefficient a and coefficient b?

Regression equation we make from sample data (estimator)

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

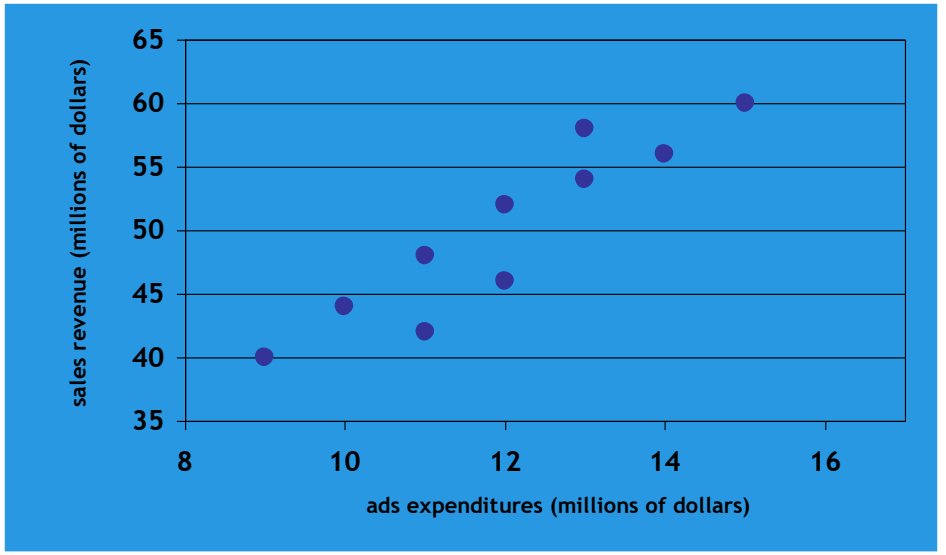
Determining the regression equation

Freehand method

Plot the sample points

Create the predictive regression

Then this method is invalid



Relationship between Variables

Unidirectional/Positive Relationship

Unidirectional/Positive RelationshipA unidirectional (positive) relationship occurs when changes in the independent variable result in unidirectional changes in the dependent variable. In other words, if the value of the independent variable increases, the value of the dependent variable also tends to increase, or conversely, if the value of the independent variable decreases, the value of the dependent variable tends to decrease

Example:

Suppose you are doing a regression analysis to understand the relationship between the number of study hours (X) and exam scores (Y). If the relationship between the number of hours studied and exam score is positive, then the more time spent studying (increase in X), the exam score (Y) tends to increase. Conversely, if the number of study hours decreases, test scores also tend to decrease.

Unidirectional/Negative Relationship

An unidirectional (negative) relationship occurs when changes in the independent variable result in changes in the opposite direction to the dependent variable. In other words, when the value of the independent variable increases, the value of the dependent variable tends to decrease, or vice versa.

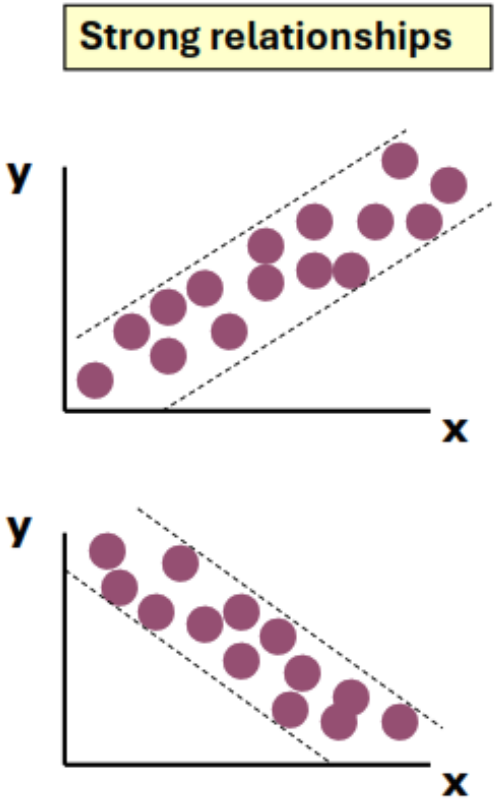
Example:

In a regression analysis between the amount of physical exercise (X) and body weight (Y), if the relationship is negative, the more physical exercise performed (increase in X), the weight (Y) tends to decrease. Conversely, if the amount of physical exercise decreases, body weight tends to increase.

Strong Relationship

When two variables have a strong relationship, the scatter plot shows a pattern where the data points tend to form a clear pattern or follow a strict straight line or curve pattern.

Strong Relationship



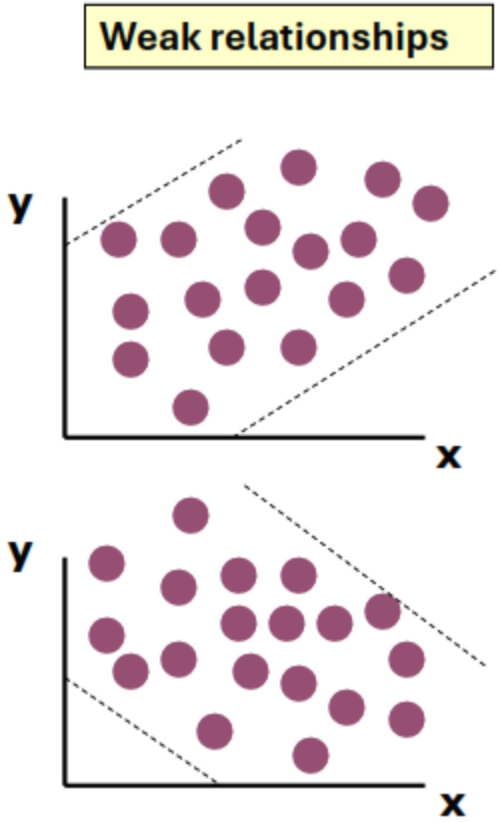
For example, if you have two variables that are positively correlated with each other with a strong relationship, the scatter plot will show dots that are regular
and follow a clear line pattern in an upward direction. Any change in the independent variable is consistently accompanied by a large change in the dependent variable.

Weak Relationship

A weak relationship between the variables can be seen on a scatter plot where the data points are widely scattered around the center line or do not follow a clear pattern. Although there is a relationship between the two variables, the pattern of the data points may not be very regular or clear.

Weak Relationship

For example, on a scatter plot with a weak relationship, the data points are scattered randomly without a consistent line or pattern. Changes in the independent variable may not necessarily result in large or consistent changes in the dependent variable.

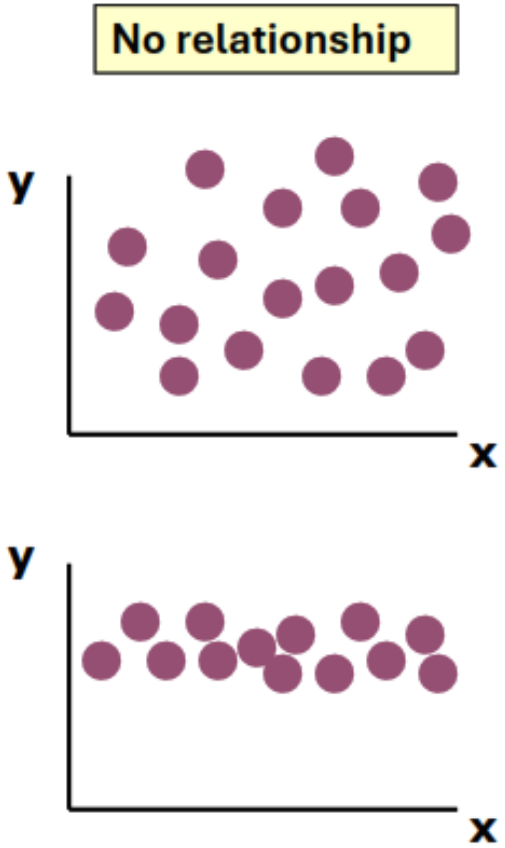


No Relationship

No relationship occurs when there is no clear pattern or relationship between the independent variable and the dependent variable. In the context of linear regression, this indicates that changes in the independent variable have no consistent influence on the dependent variable.

For example, if you analyze the relationship between a person's height (X) and intelligence (Y), and no relationship can be identified, this means that there is no predictable pattern where a person's height significantly affects their intelligence. In this case, there is no relationship that can be explained by a linear regression between the two variables.

No Relationship



For example, on a scatter plot with no relationship, the data points are randomly scattered along the axis with no identifiable direction or pattern. Changes in the independent variable do not provide definite clues about changes in the dependent variable.

Linear Regression Context

- Strong relationships can occur in both positive and negative relationships, where the independent variable has a large and consistent influence on the dependent variable.

- Weak relationships can occur in both positive and negative relationships, where the influence of the independent variable on the dependent variable is not so significant or consistent.