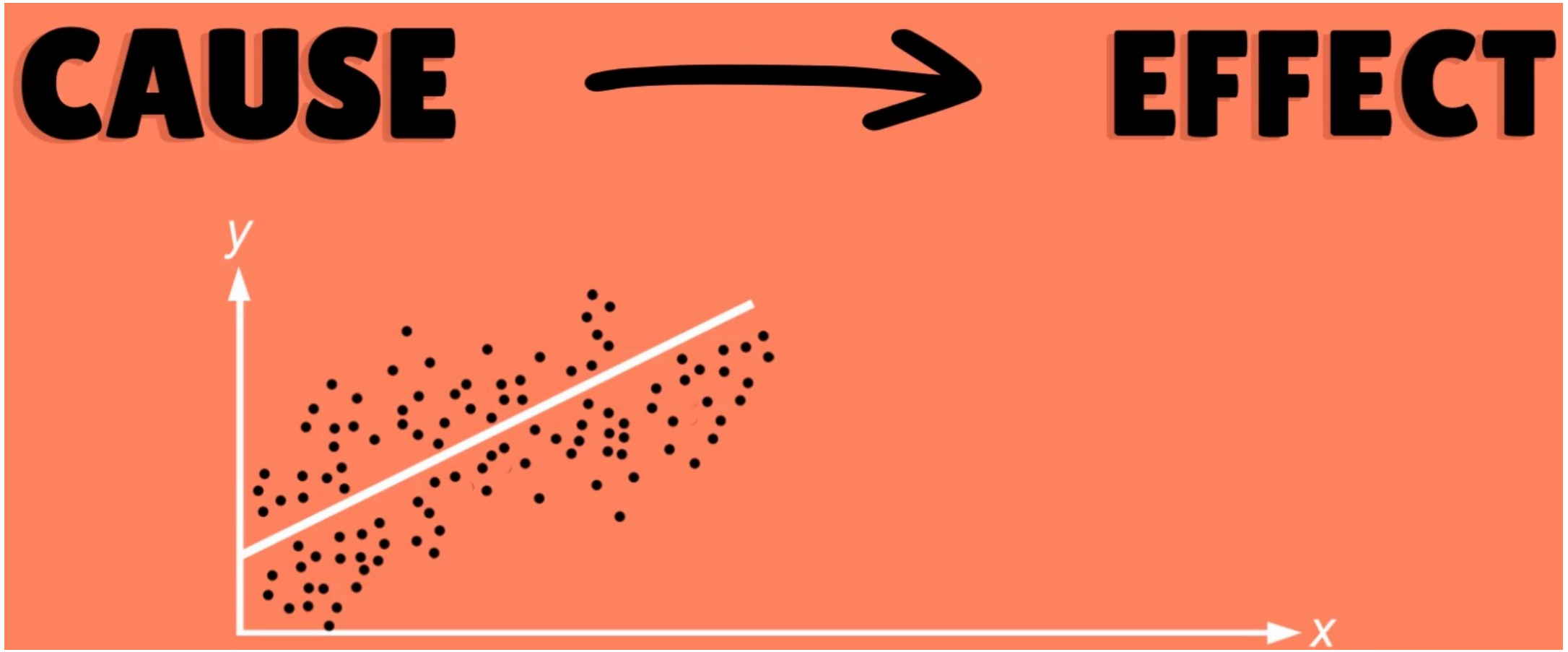
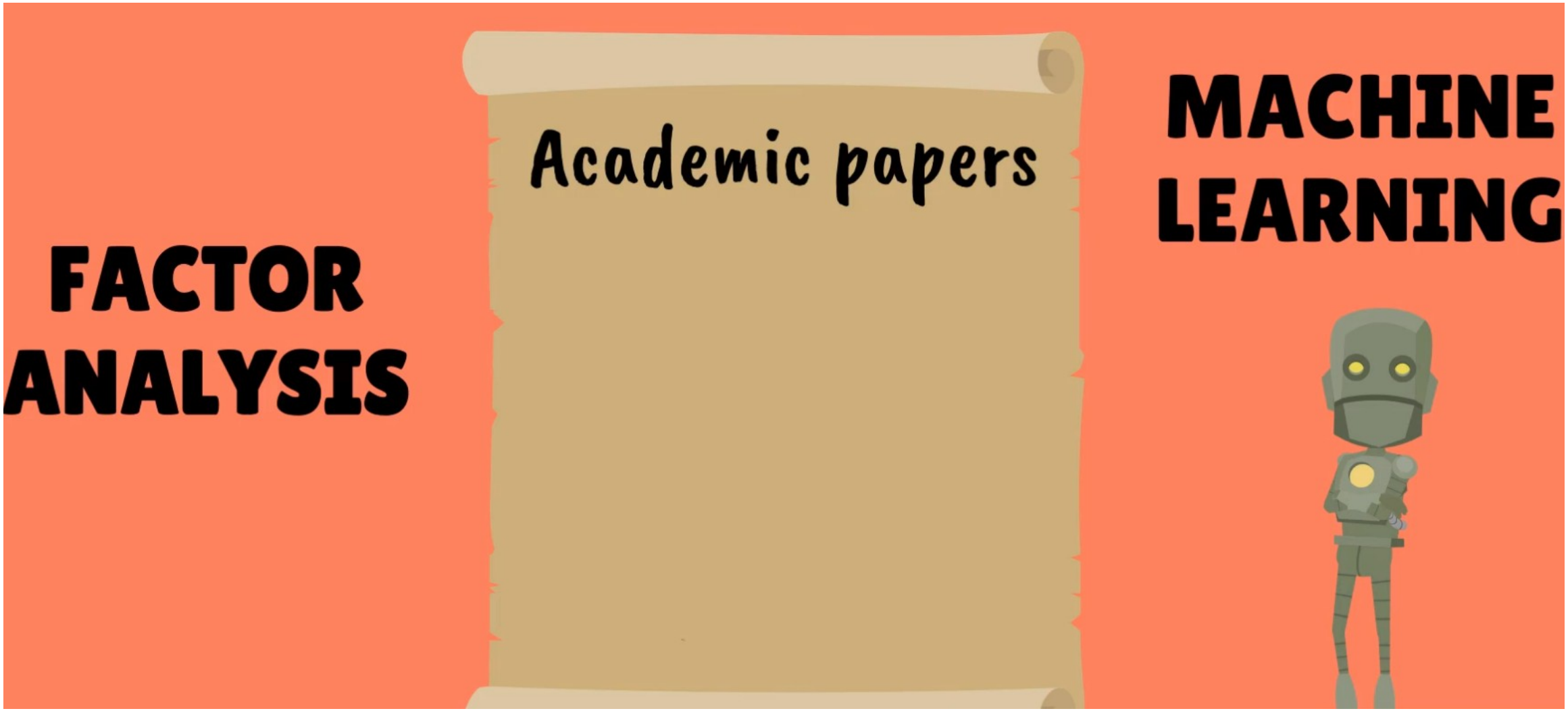
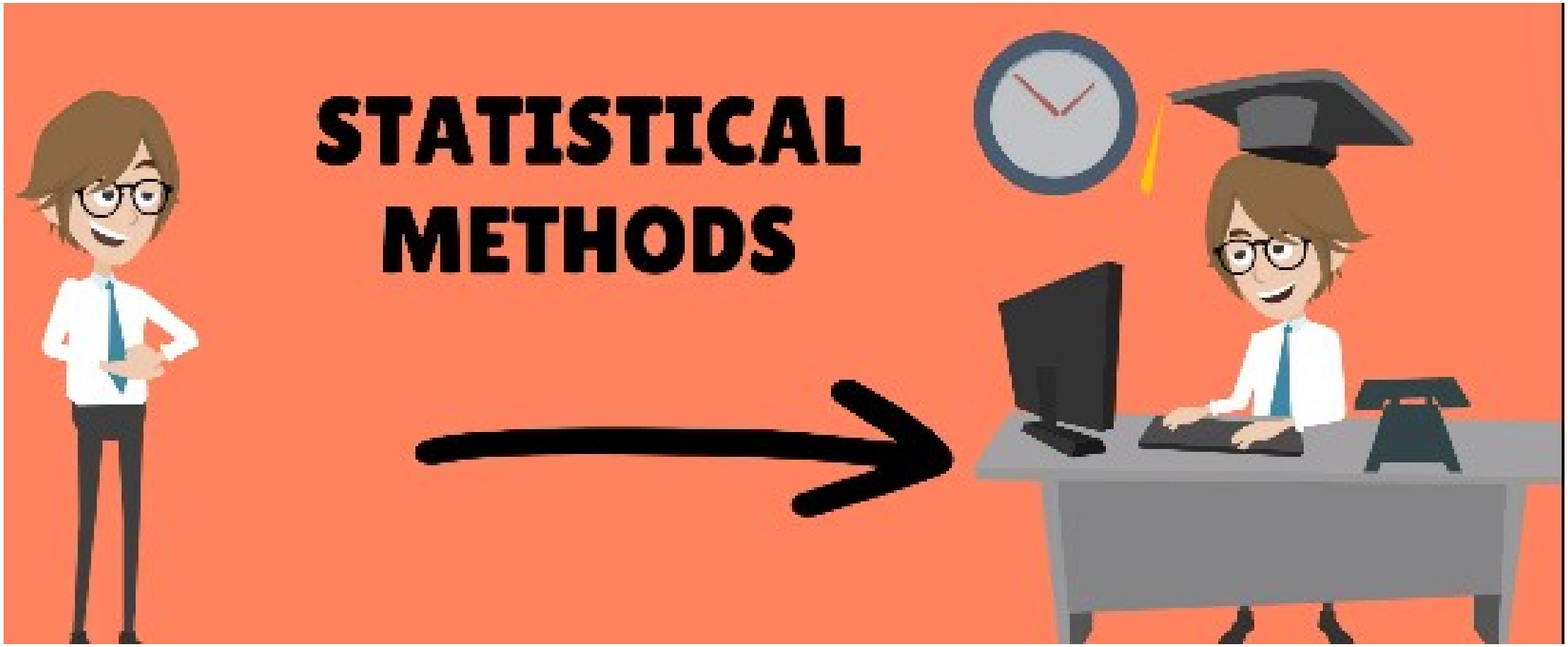
Regression Analysis



Regression Analysis



Regression Analysis



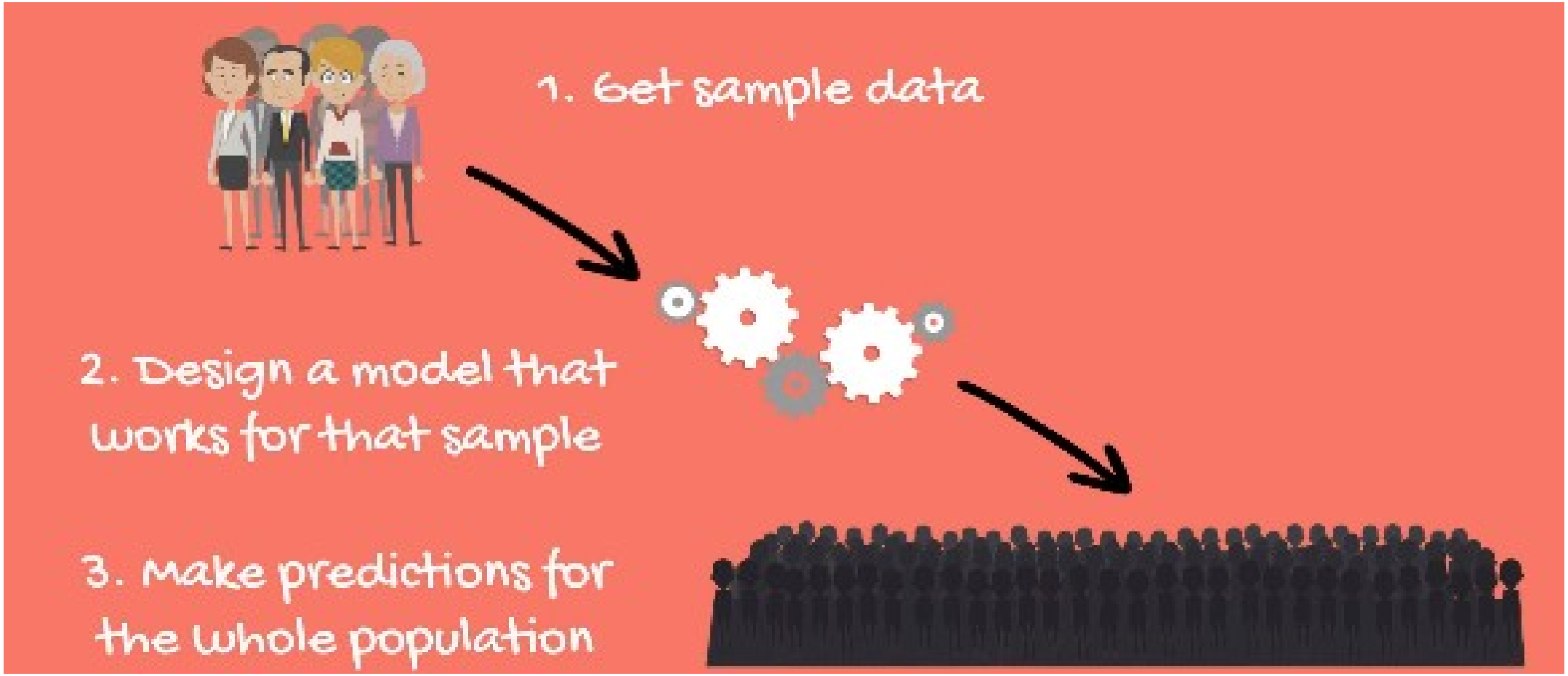
Regression Analysis

- Simple Linear Regression
- Multiple Linear Regression
- How to build a regression
- How to Interpret it
- How to compare different models

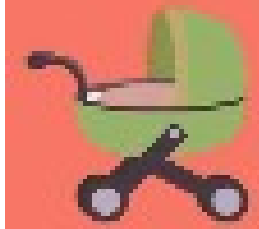
Linear Regression with StatsModels

- A linear regression is a linear approximation of a causal relationship between two or more variables.
- Regression models are highly valuable as they are one of the most common ways to make inferences and predictions.

The Process


1. Get sample data
2. Design a model that works for that sample
3. Make predictions for the whole population

The Process



DEPENDENT

/predicted/

Y

INDEPENDENT

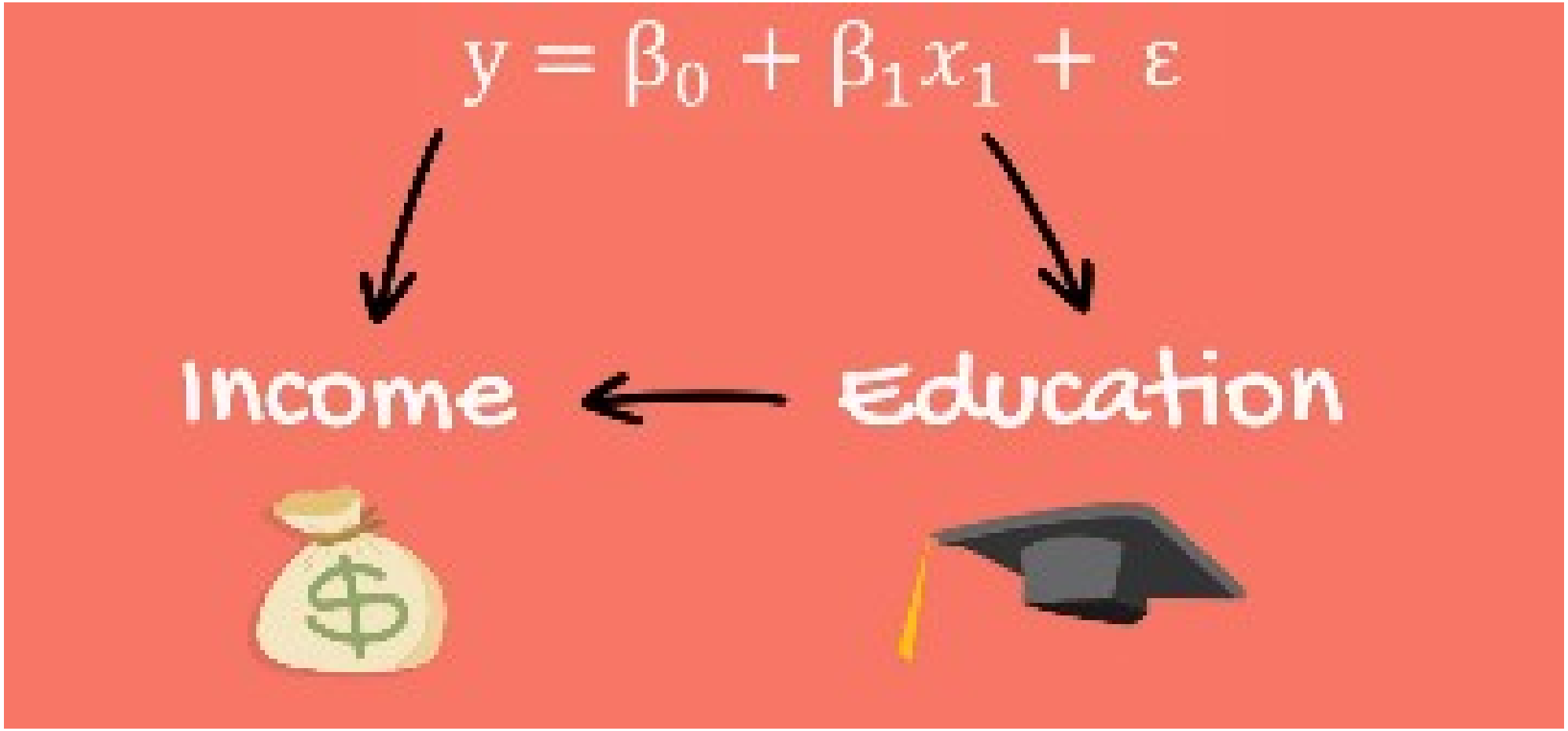
/predictors/

$x_1, x_2, \ldots, x_k$

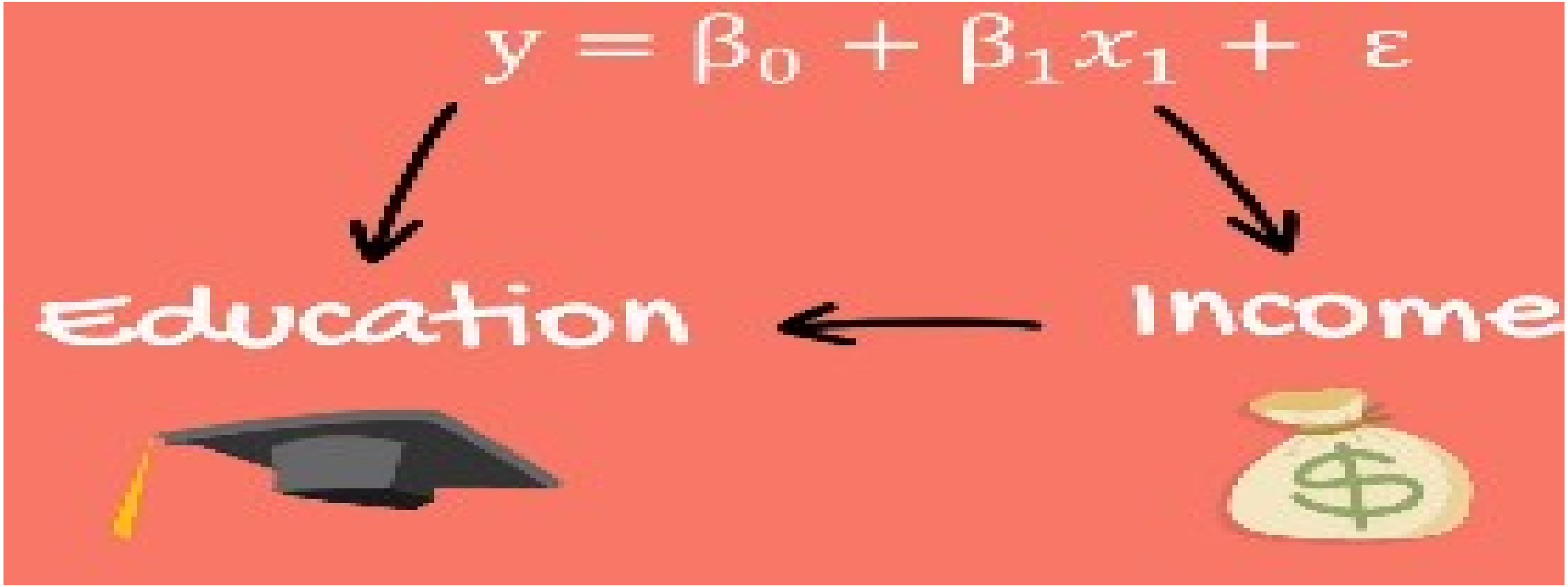
Simple Linear Regression

- Y = B0 + B1x1 + e
- Y is the dependent variable we are trying to predict
- X is the independent variable. When using regression analysis, we want to predict the value of y provided we have the value of x.
- To have a regression, y must depend on x in some causal way.
- Whenever there is a change in x, such change must translate in to a change in y.

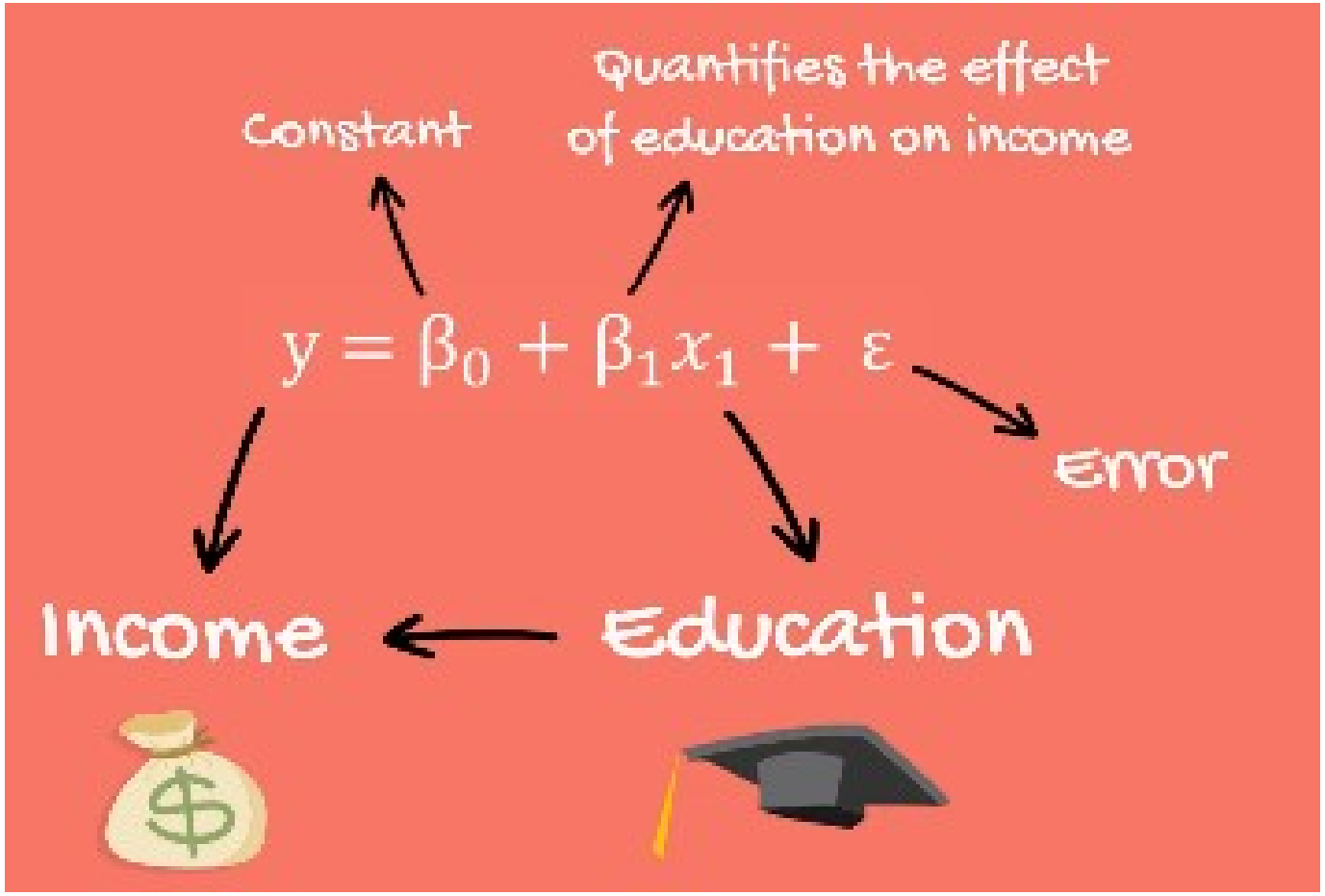
Simple Linear Regression - Example



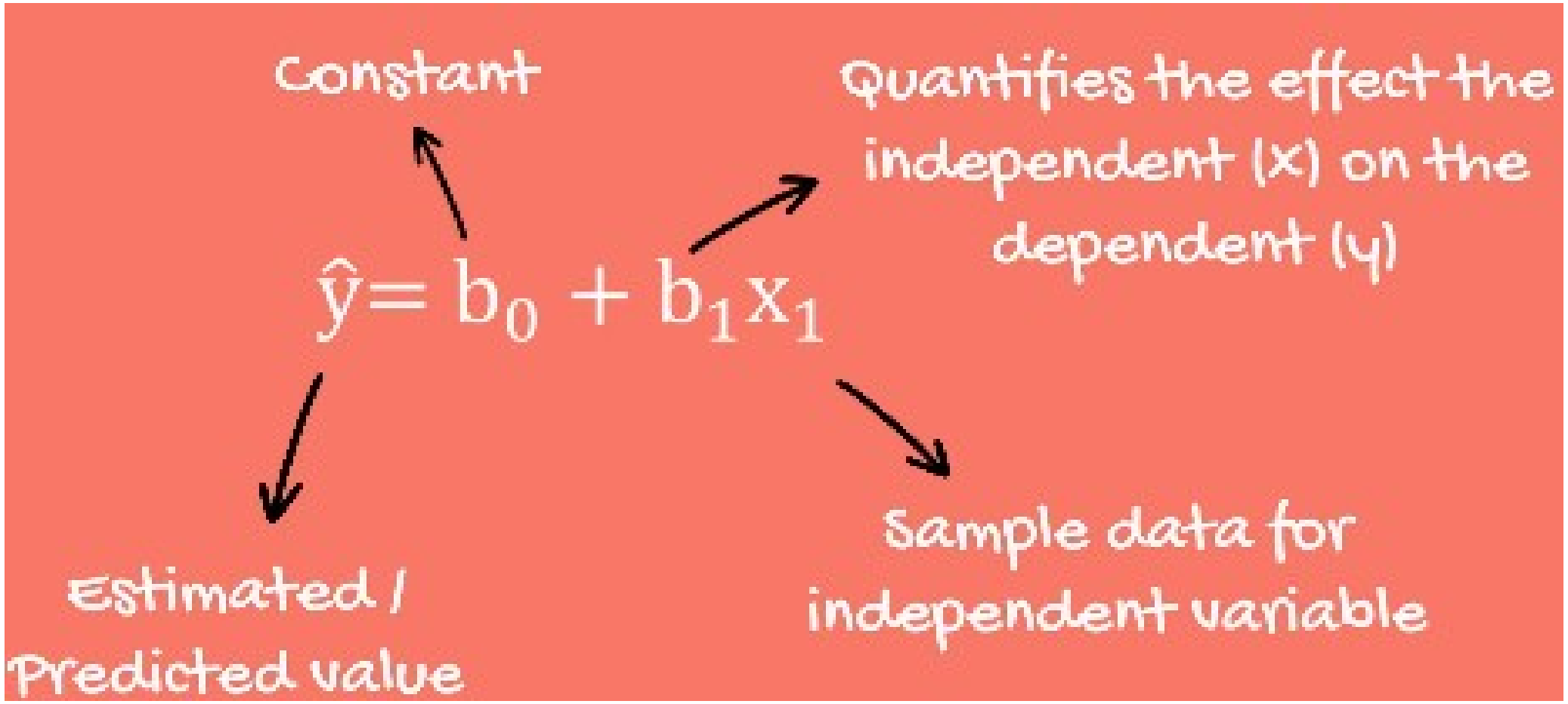
Simple Linear Regression - Example



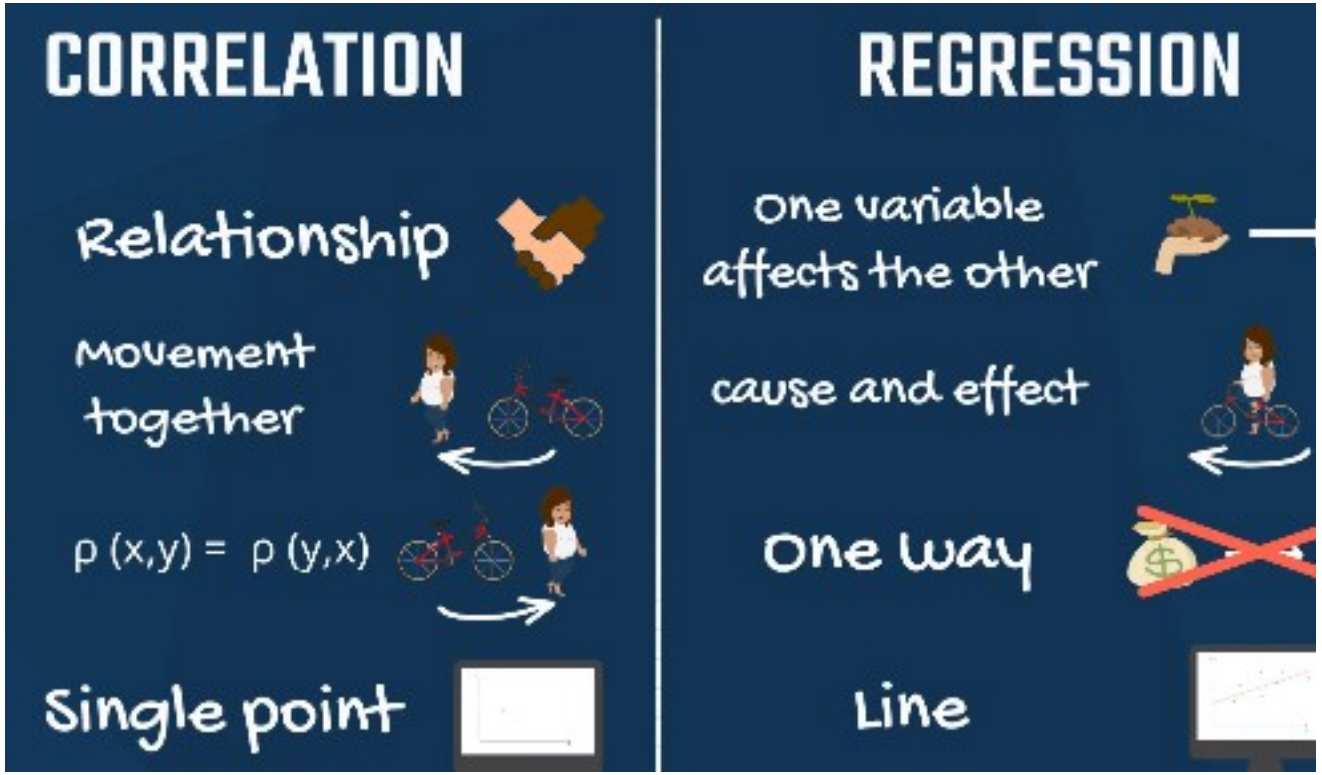
Simple Linear Regression



Simple Linear Regression Equation



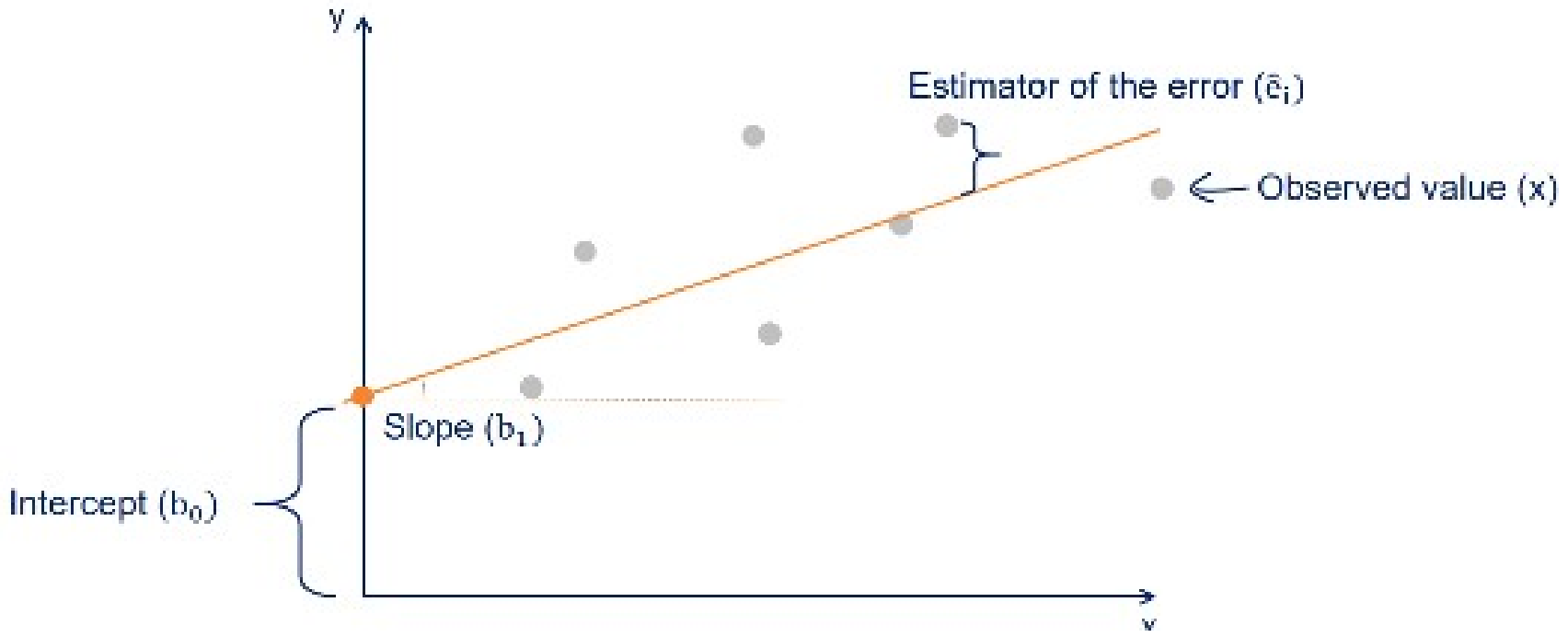
Correlation vs Regression



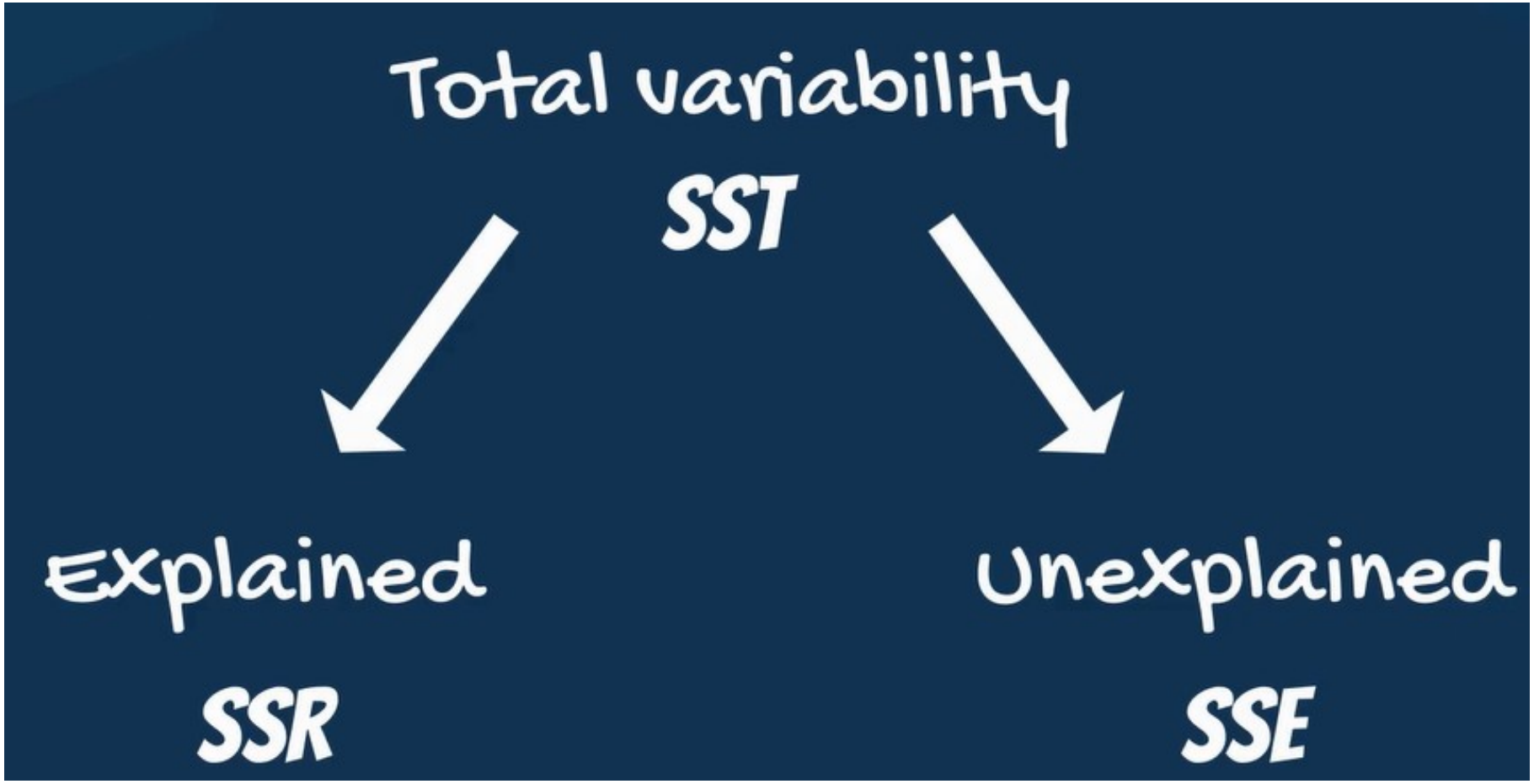
- Correlation does not apply causation
- Regression is based on causality
- Correlation measure the degree of relationship between 2 variables
- Regression analysis is about how one variable affects the other
- The correlation between x & y is the same between y & x
- Regression with x & y, y & x yields different results
- Graphically, correlation represented in a single point, whereas regression known for the best fitting line across all points

Linear regression model. Geometrical representation

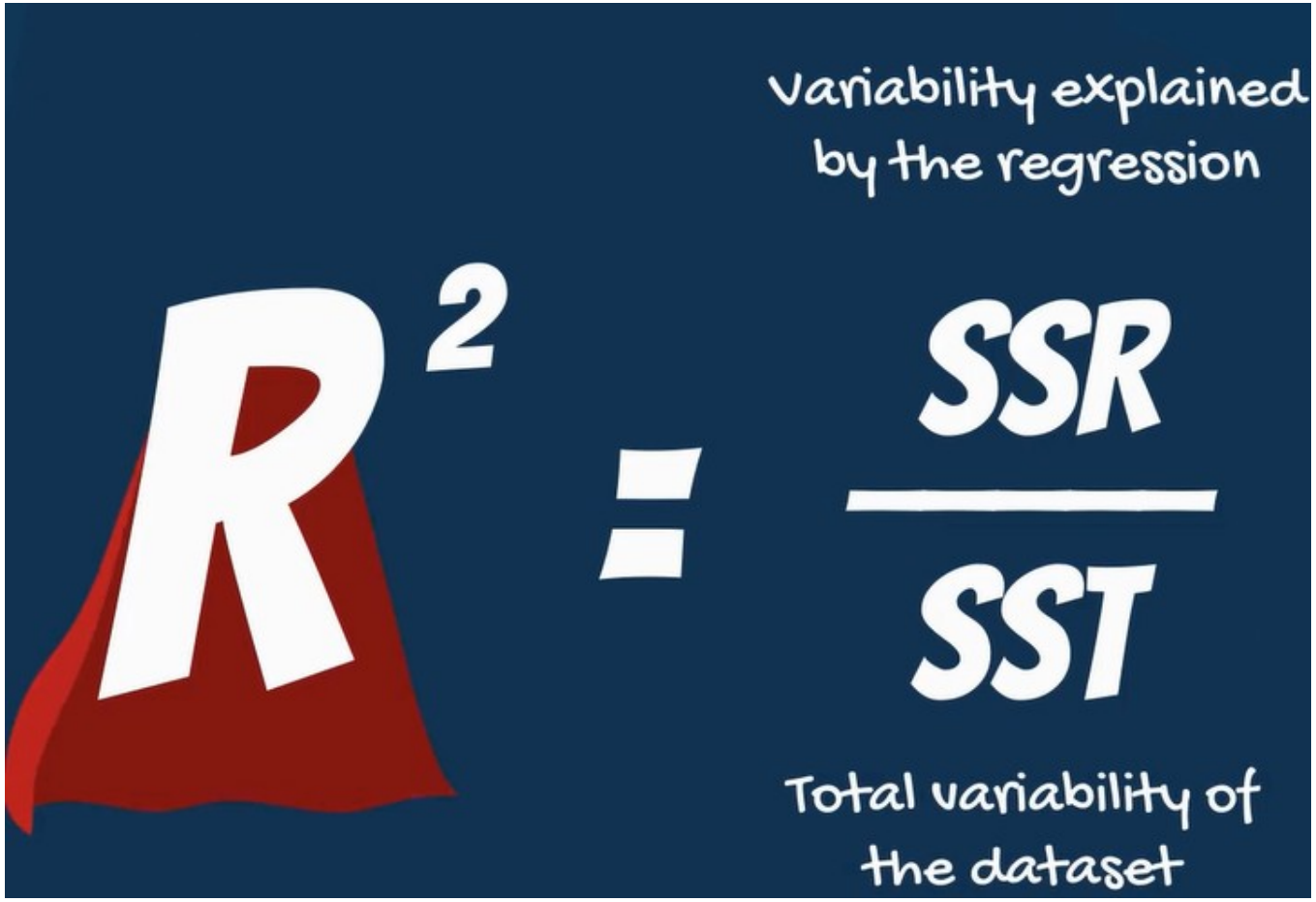
$$\hat{y}_i = b_0 + b_1 x_i$$



Goodness of fit: The R-Squared



Goodness of fit: The R-Squared



Goodness of fit: The R-Squared