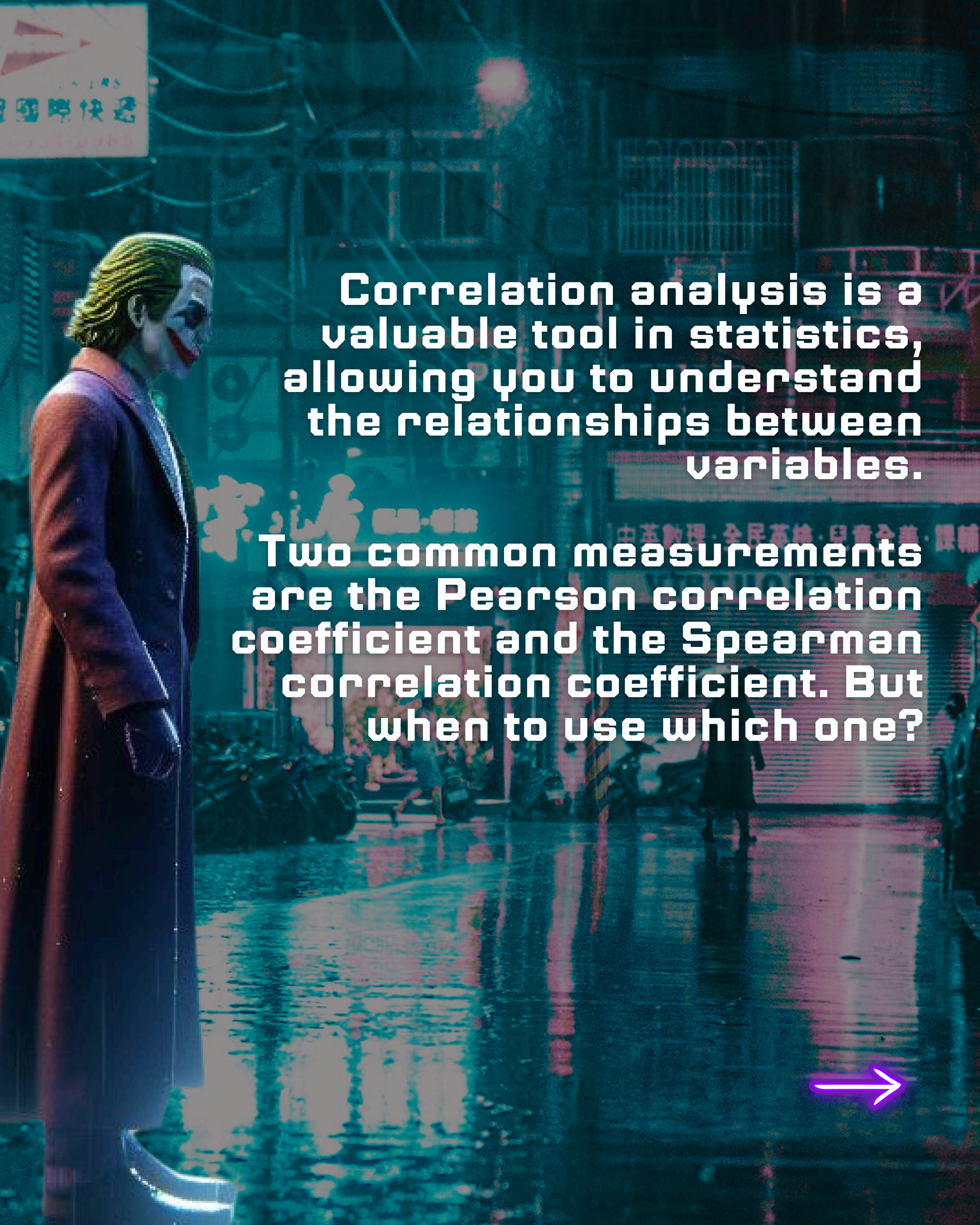
CORRELATION

PEARSON

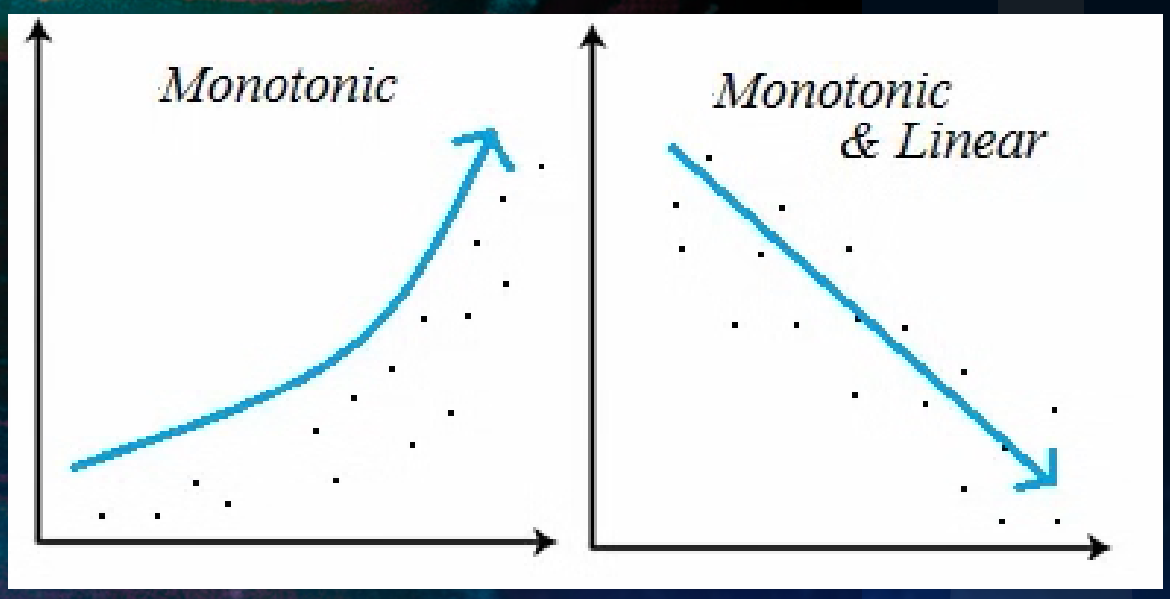
SPEARMAN


Correlation analysis is a valuable tool in statistics, allowing you to understand the relationships between variables.
Two common measurements are the Pearson correlation coefficient and the Spearman correlation coefficient. But when to use which one?

LINEAR MONOTONIC RELATIONSHIP

Pearson's correlation coefficient is ideal when seeking to evaluate linear relationships between continuous variables. It assumes that the relationship between variables follows a straight line.

Spearman's correlation coefficient is appropriate when the relationship is monotonic, that is, not necessarily linear, but consistently increasing or decreasing.



TYPE OF VARIABLES

The Pearson coefficient is more suitable for continuous variables, while the Spearman coefficient is robust to continuous and ordinal data, measuring relationships on an order scale.

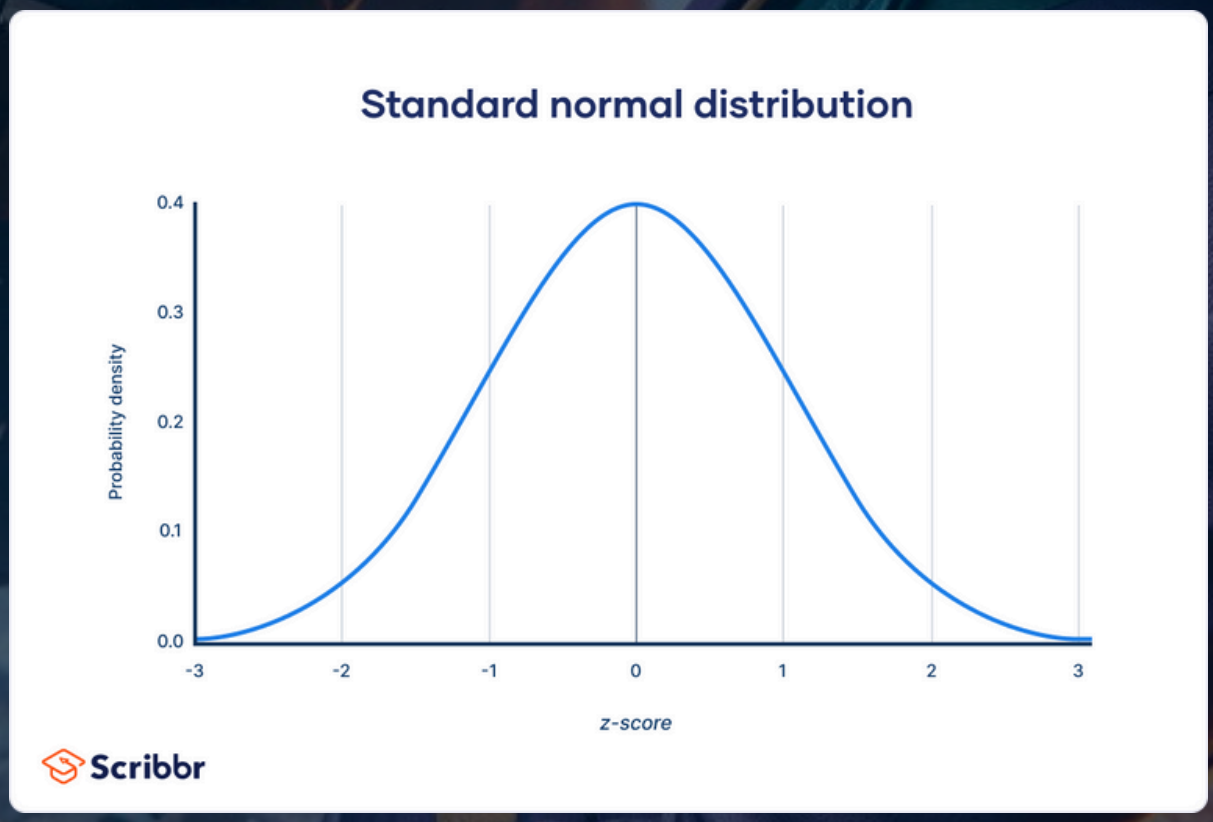
Continuous data -> Score: [85, 92, 78, 89, 95, 80, 88, 93, 87, 91]

Ordinal data -> Classification: ["Excellent", "Good", "Fair", "Very Good", "Poor", "Good", "Very Good", "Fair", "Excellent", "Very Good"]

DATA DISTRIBUTION

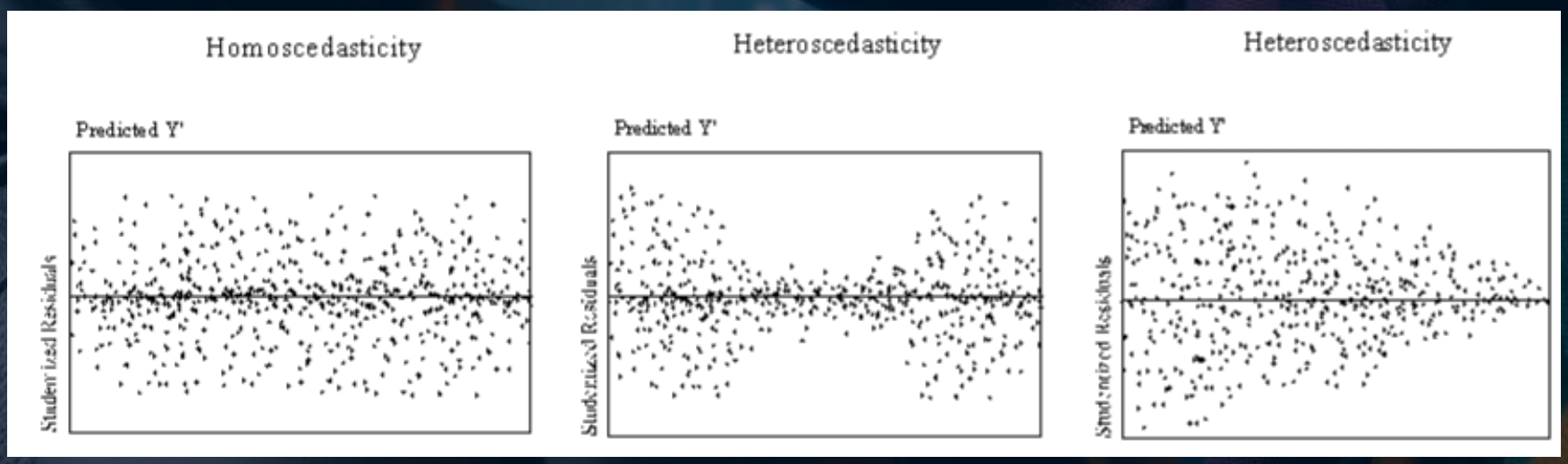
The Pearson coefficient assumes a normal distribution of variables.

If the data does not follow this distribution, the use of the Spearman coefficient, which is non-parametric, becomes preferable, as it does not require such an assumption.



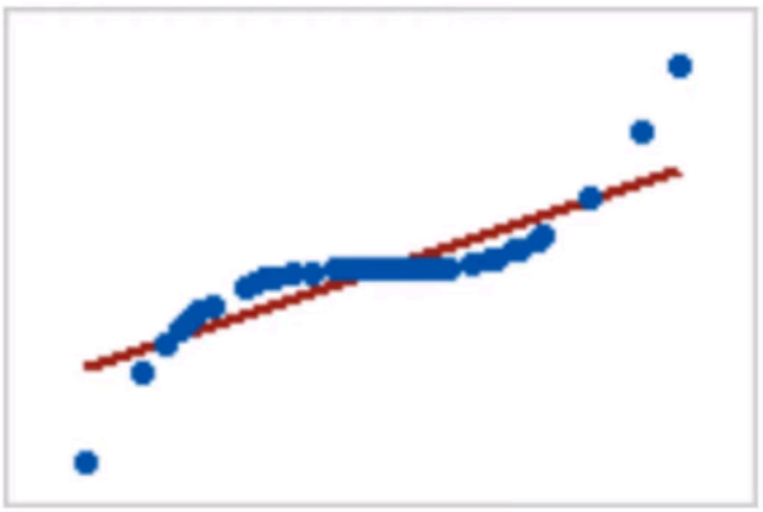
HOMOSCEDASTICITY

The Pearson coefficient assumes homoscedasticity, i.e., that the variability of observations is constant for all possible values of the variables. If this assumption is violated, the Spearman coefficient, being less sensitive to discrepancies, may be more appropriate.



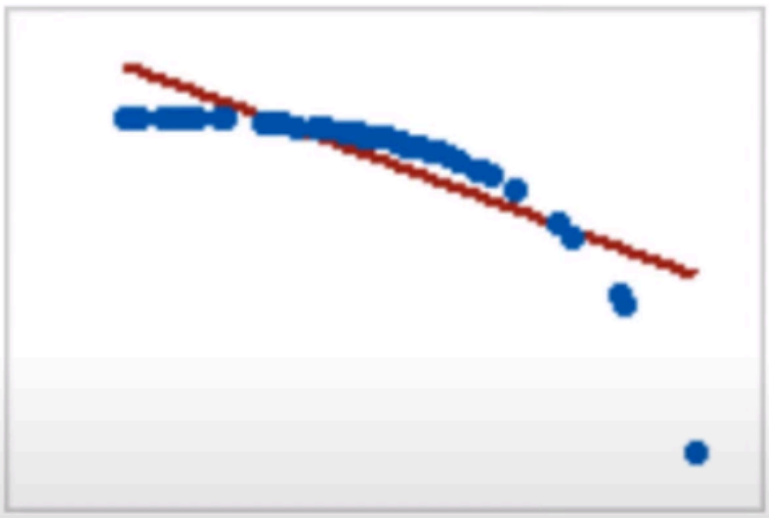
EXAMPLES

Pearson = +1, Spearman = +1



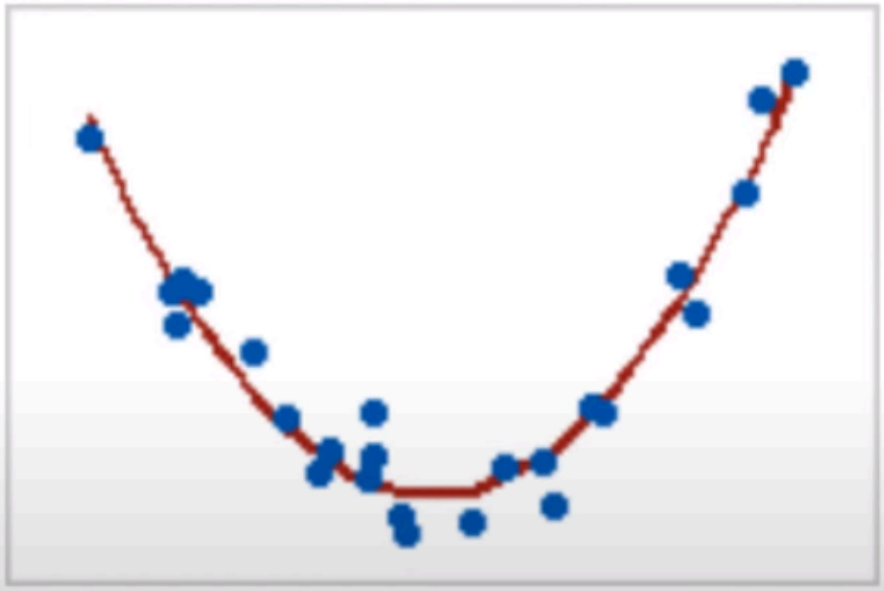
Pearson = +0.851, Spearman = +1

EXAMPLES



Pearson = −0.799, Spearman = −1

For non-linear and non-monotonic relationships none of them are suitable.



Coefficient of 0

IF YOU'VE MADE IT THIS FAR, LIKE IT!

TELL ME WHAT ELSE YOU WANT TO KNOW ABOUT AI OR DATA SCIENCE.

@simplific.ai