THE PEARSON CORRELATION COEFFICIENT

The Pearson correlation coefficient, named after Karl Pearson, measures the linear relationship between two continuous variables. It's a statistical metric that calculates the strength and direction of the correlation.

Pearson Correlation Coefficient Formula;

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / (\sqrt{\Sigma(x_i - \bar{x})^2} * \sqrt{\Sigma(y_i - \bar{y})^2})$$

where;

- x_i and y_i are individual data points
- $\bar{x}$ and $\bar{y}$ are means of the variables
- Σ denotes summation
- r is the Pearson correlation coefficient

Interpretation of Pearson Correlation Coefficient;

The coefficient (r) ranges from -1 to 1:

- *1*: Perfect positive linear relationship (as one variable increases, the other increases)
- *-1*: Perfect negative linear relationship (as one variable increases, the other decreases)
- *0*: No linear relationship (variables are uncorrelated)

Coefficient Strength;

- *$0.7 \leq |r| \leq 1$*: Strong correlation
- *$0.5 \leq |r| < 0.7$*: Moderate correlation
- *$0.3 \leq |r| < 0.5$*: Weak correlation
- *$|r| < 0.3$*: Very weak correlation

What the Coefficient Indicates;

1. Direction: Positive (direct) or negative (inverse) relationship.
2. Strength: Magnitude of the relationship (strong, moderate, weak).
3. Linearity: Measures linear relationships, not non-linear.

Example;

Height (x) and Weight (y) have a Pearson correlation coefficient of 0.8.

- Interpretation: There's a strong positive linear relationship between height and weight.

Limitations;

1. Assumes linear relationship.
2. Sensitive to outliers.
3. Doesn't imply causation.

Software Packages;

1. Python: `numpy.corrcoef()` or `pandas.DataFrame.corr()`
2. R: `cor()` or `corr()`
3. Excel: `CORREL()` function

By understanding the Pearson correlation coefficient, you can analyze relationships between variables and make informed decisions.