

Correlation Coefficients: Understanding When to Use What

1. Pearson Correlation (r):

- Measures linear relationship between two continuous variables.
- Assumes normality, linearity, no significant outliers, and homoscedasticity.
- Suitable when data shows a linear trend and is normally distributed.

2. Spearman Rank Correlation (rho):

- A non-parametric measure based on ranked values.
- Captures monotonic relationships (increasing or decreasing, not necessarily linear).
- Use when data is not normally distributed or the relationship is nonlinear but monotonic.

3. Kendall's Tau (tau):

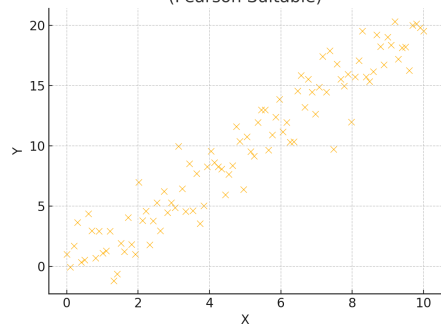
- A non-parametric correlation based on the number of concordant and discordant pairs.
- More robust in small datasets or when there are many tied ranks.

Summary of When to Use:

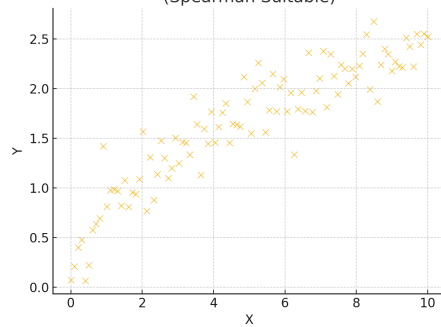
Method	Use When	Handles Non-Normal?	Monotonic?	Linear?
-----	-----	-----	-----	-----
Pearson (r)	Linear + normal	No	No	Yes
Spearman (rho)	Monotonic + non-normal	Yes	Yes	No
Kendall (tau)	Small/tied datasets + ordinal data	Yes	Yes	No

Below is a visual example comparing different types of relationships:

Linear Relationship
(Pearson Suitable)



Monotonic Nonlinear
(Spearman Suitable)



Nonlinear & Non-Monotonic
(Pearson Not Suitable)

