

Pitfalls

While the correlation is widely used, there are a number of pitfalls that some are not aware of.

Correlation Applet

Correlation Applet

Pitfall 1

Correlation is a measure of **linear** relationship. It should only be used if the relationship is linear, and should never be used if the relationship is nonlinear.

Solution: produce a scatterplot of the data first, and visually assess if a linear relationship exists.

Pitfall 2

A high correlation is not “proof” of a strong linear relationship. You can have a relationship that is clearly not linear, yet have a high correlation.

Solution: produce a scatterplot of the data first, and visually assess if a linear relationship exists.

Pitfall 3

A low correlation is not “proof” of a weak linear relationship. A strong nonlinear relationship could actually exist.

Solution: produce a scatterplot of the data first, and visually assess if a linear relationship exists.

Pitfall 4

Correlation is sensitive to the presence of outliers. Their presence can severely distort the value of the correlation.

Solution: produce a scatterplot of the data first, and visually assess if a linear relationship exists.

Final Word

Long story short: produce a scatterplot of the data first, and visually assess if a linear relationship exists, or if outliers are present.