What is the difference between correlation and association?

Authored by

stats writer

April 25, 2024

Correlation and association are two statistical measures used to describe the relationship between two variables. While both of these terms are often used interchangeably, there are subtle differences between them.

Correlation refers to the strength and direction of the linear relationship between two variables. It measures how closely the two variables are related to each other, with values ranging from -1 to +1. A correlation of +1 indicates a perfect positive relationship, while a correlation of -1 indicates a perfect negative relationship. A correlation of 0 indicates no relationship between the two variables.

On the other hand, association is a broader term that encompasses any type of relationship between two variables, not just a linear one. It can include non-linear relationships, such as quadratic or exponential, as well as qualitative relationships, such as a categorical variable. Association is typically measured using measures like chi-square or odds ratio.

In summary, the main difference between correlation and association is that correlation measures the strength and direction of a linear relationship between two variables, while association is a broader term that includes any type of relationship between two variables.

Correlation vs. Association: What's the Difference?

Two terms that are sometimes used interchangeably are correlation and association. However, in the field of statistics these two terms have slightly different meanings.

In particular, when we use the word correlation we're typically talking about the . This is a measure of the linear association between two *X* and *Y*. It has a value between -1 and 1 where:

-1 indicates a perfectly negative linear correlation between two variables0 indicates no linear correlation

between two variables1 indicates a perfectly positive linear correlation between two variables

Conversely, when statisticians use the word association they can be talking about *any* relationship between two variables, whether it's linear *or* non-linear.

To illustrate this idea, consider the following examples.

Visualizing Correlation vs. Association with Scatterplots

We use two words to describe the correlation between two random variables:

1. Direction

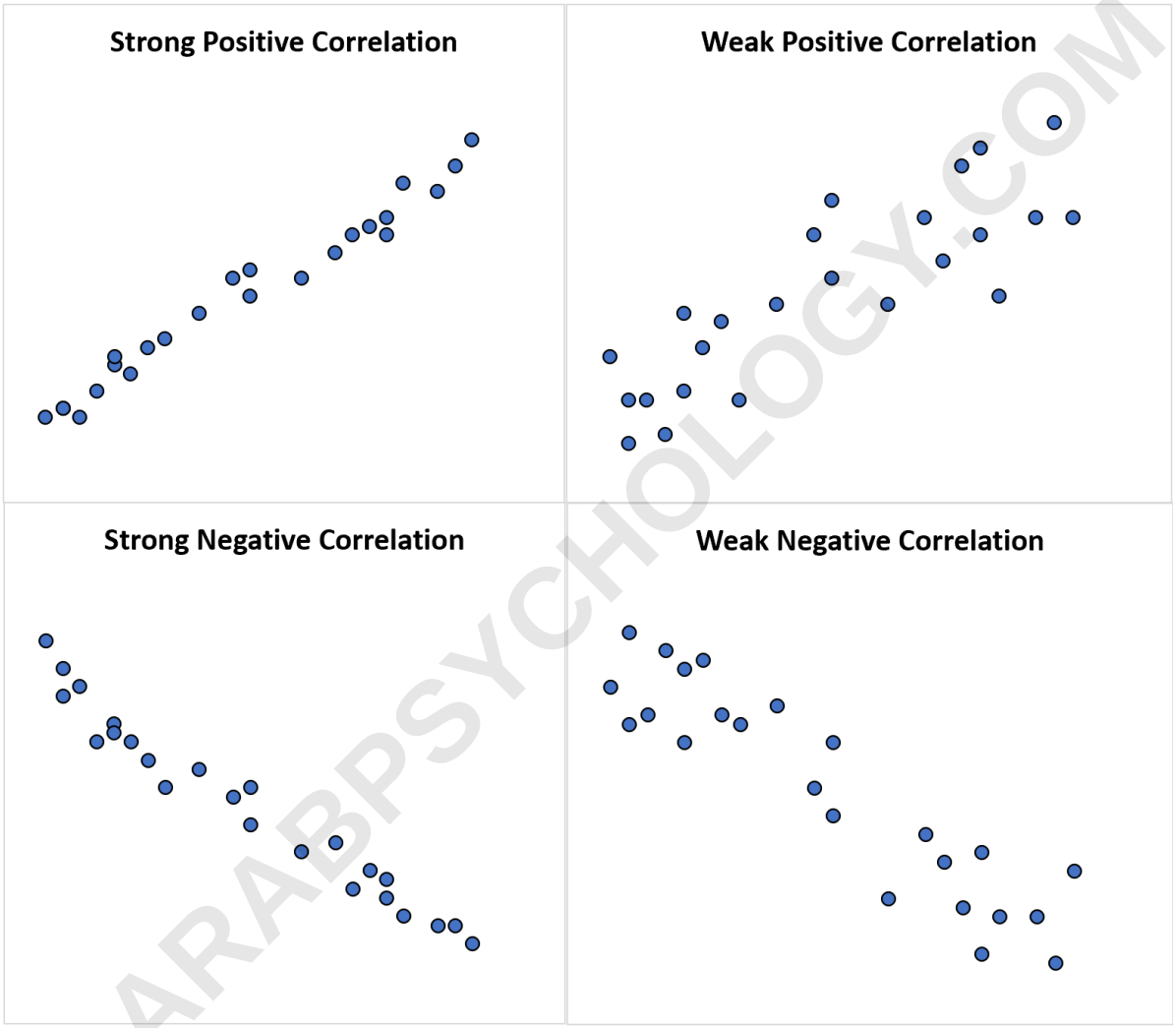
Positive: Two random variables have a positive correlation if *Y* tends to increase as *X* increases.Negative: Two random variables have a negative correlation if *Y* tends to decrease as *X* increases.

2. Strength

Weak: Two random variables have a weak correlation if the points in a scatterplot are loosely scattered.Strong: Two random variables have a strong correlation if the

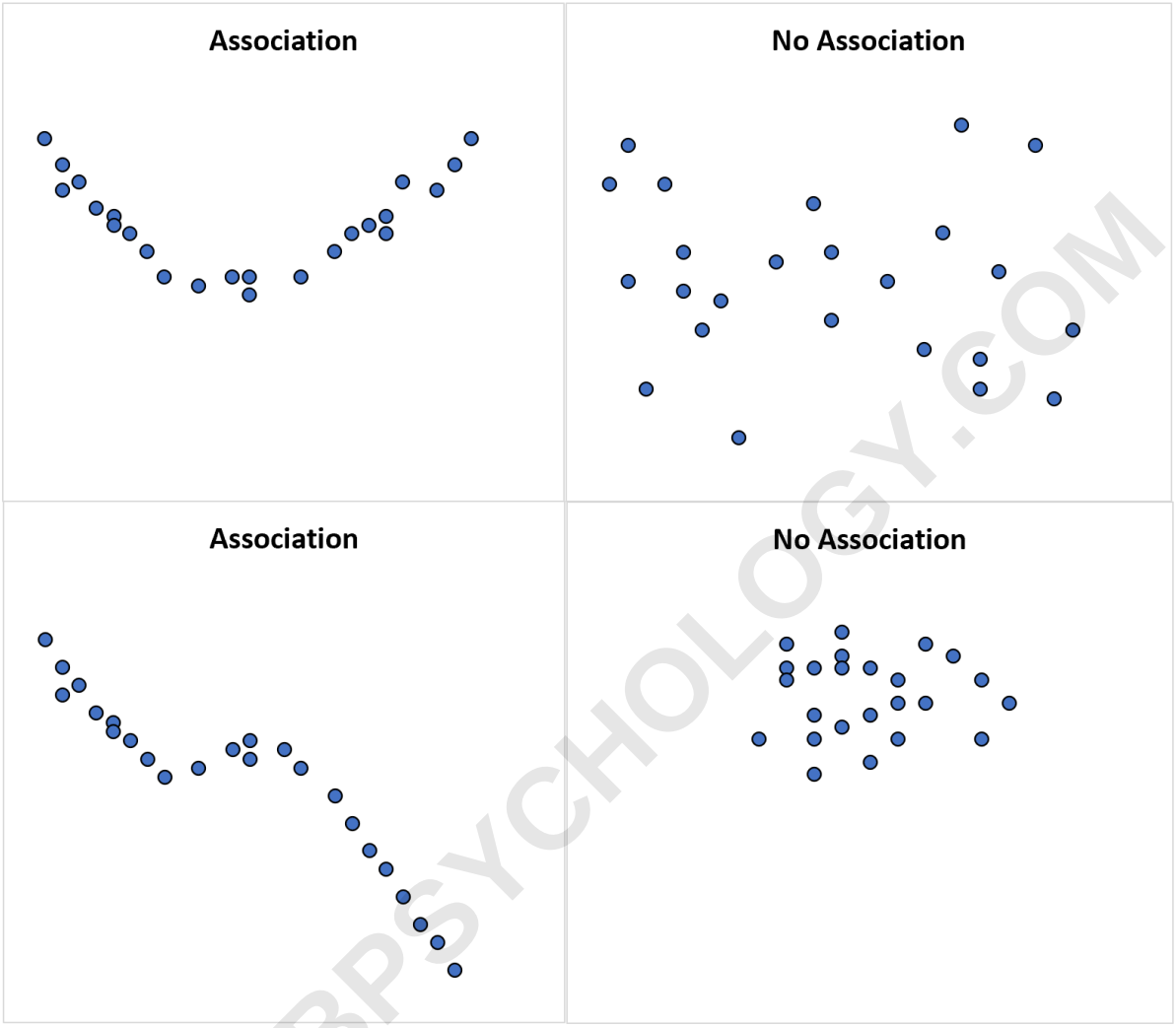
points in a scatterplot are tightly packed together.

The following illustrate examples of each type of correlation:



Compared to correlation, the word association can tell us whether or not there is *any* relationship between two random variables: linear *or* non-linear.

The following scatterplots illustrate some examples:



The scatterplot in the top left corner illustrates a quadratic relationship between two random variables, which means there *is* an association between the two variables but it's not a linear one.

However, just knowing that the correlation between the two variables is zero can be misleading because it

hides the fact there there exists a non-linear relationship instead.

Correlation vs. Association: A Summary

The terms correlation and association have the following similarities and differences:

Similarities:

Both terms are used to describe whether or not there is a relationship between two random variables.Both terms can use scatterplots to analyze the relationship bewteen two random variables.

Differences:

Correlation can only tell us if two random variables have a linear relationship while association can tell us if two random variables have a linear *or* non-linear relationship.Correlation quantifies the relationship between two random variables by using a number between -1 and 1, but association does not use a specific number to quantify a relationship.