

Understanding Somers' D: A Guide to Measuring Association Between Variables

Authored by
Mohammed looti

November 6, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding Somers' D: A Guide to Measuring Association Between Variables*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11529>

Defining Somers' D and Its Role in Predictive Modeling

Somers' D, often referred to as [Somers' Delta](#), is a highly valuable [statistical measure](#) designed to assess the strength and direction of the association between two variables. Its critical distinction lies in its asymmetric nature, making it ideal for scenarios where a clear predictive relationship exists between an independent variable (X) and a [dependent variable](#) (Y). Somers' D is especially favored when analyzing data where both X and Y are **ordinal**.

The primary utility of Somers' D stems from its directional focus. Unlike symmetric measures of correlation which treat X and Y equally, Somers' D specifically quantifies how effectively the independent variable predicts the rank ordering of the dependent variable. This makes it an indispensable tool across diverse fields, including social science, biostatistics, and for evaluating [predictive models](#) within machine learning, particularly when dealing with outcomes categorized into ordered levels.

An [ordinal variable](#) is characterized by values that possess an inherent, meaningful rank or order, such as customer satisfaction scales ("poor," "fair," "good," "excellent") or severity classifications (mild, moderate, severe). Because the analysis performed by Somers' D relies solely on these ranks rather than assuming uniform distances between values, it is classified as a robust component of [nonparametric statistical methods](#). This feature ensures its validity even when underlying assumptions about data distribution cannot be met.

The Mathematical Foundation of Somers' Delta Calculation

The calculation of Somers' D is fundamentally rooted in the analysis of every possible pair of observations within the dataset. By systematically comparing these pairs, the statistic determines whether their rankings agree (concordant) or disagree (discordant) across the two variables (X and Y). This comprehensive counting process forms the basis of the numerator, capturing the net agreement or disagreement in the data.

Given any two pairs of observations, (x_i, y_i) and (x_j, y_j) , they are categorized based on their relative ranks:

Two pairs are [concordant](#) if their rankings move in the same direction. Specifically, if $x_i < x_j$ and $y_i < y_j$, OR if $x_i > x_j$ and $y_i > y_j$.

Two pairs are **discordant** if their rankings move in opposite directions. Specifically, if $x_i < x_j$ and $y_i > y_j$, OR if $x_i > x_j$ and $y_i < y_j$.

Two pairs are **tied** if $x_i = x_j$ or $y_i = y_j$ (or both). For Somers' D, the treatment of ties is crucial for establishing directionality, focusing primarily on pairs tied on the independent variable (X) but not

on the dependent variable (Y).

The formula for Somers' D is explicitly designed to account for its asymmetry, ensuring the resulting coefficient accurately reflects the predictive power of X on Y. The difference between concordant and discordant pairs is normalized by a denominator that excludes pairs tied on the dependent variable (Y):

$$\text{Somers' D (Y|X)} = (\text{NC} - \text{ND}) / (\text{NC} + \text{ND} + \text{NT})$$

Where the components are precisely defined as:

NC: The total count of **concordant** pairs observed.

ND: The total count of **discordant** pairs observed.

NT: The number of pairs that are tied only on the independent variable (X) but are **not** tied on the dependent variable (Y). This specific definition ensures that the measure remains directional, focusing on the predictive marginal gain provided by X.

The resulting coefficient will always fall strictly between -1 and 1, providing a standardized measure of [rank correlation](#) that allows for meaningful comparison across different studies and datasets.

Interpreting the Direction and Strength of Somers' D

The range of values from -1 to 1 provides a clear, standardized framework for interpreting the relationship between two ordinal variables. Given that Somers' D is a directional statistic, both the sign and the magnitude carry specific implications regarding how effectively the independent variable (X) predicts the ordering of the dependent variable (Y).

The boundaries of the scale define perfect predictive outcomes:

-1: This indicates a perfect negative association. In this scenario, all comparable pairs are **discordant**. It means that an increase in the rank of X always corresponds to a decrease in the rank of Y, and vice-versa. The independent variable perfectly predicts the inverse order of the dependent variable.

1: This signifies a perfect positive association. Here, all comparable pairs are **concordant**. An increase in the rank of X consistently corresponds to an increase in the rank of Y. The independent variable perfectly predicts the exact order of the dependent variable.

0: This represents the absence of a linear relationship. The number of concordant pairs (NC) is exactly equal to the number of discordant pairs (ND). In practical terms, the independent variable

provides no useful information about the ordering of the dependent variable.

Intermediate values, such as 0.45 or -0.6, reflect the strength of the [monotonic relationship](#). For instance, a value of 0.6 suggests a strong positive trend, implying that in 60% of the non-tied pairs, the order of X aligns with the order of Y. Interpreting the precise magnitude requires contextual knowledge of the research area, but generally, coefficients closer to the extremes (1 or -1) indicate a powerful association, while those nearing zero suggest a weak or negligible predictive connection.

Somers' D Compared to Other Rank Correlation Measures

Although Somers' D belongs to the family of rank correlation statistics, it is often necessary to distinguish it from related nonparametric measures, most notably [Kendall's Tau](#) (specifically Tau-b) and Goodman and Kruskal's Gamma. Selecting the appropriate test hinges upon understanding how each measure handles the concepts of symmetry and ties.

The fundamental differentiation lies in directionality. Kendall's Tau-b is a symmetric measure, meaning that the correlation between X and Y is identical to the correlation between Y and X ($\text{Tau}(X, Y) = \text{Tau}(Y, X)$). Furthermore, its denominator excludes pairs tied on either X or Y, making it suitable when there is no defined dependent variable or assumed causality. In stark contrast, Somers' D is inherently **asymmetric** (directional); $D(Y|X)$ measures the predictive power of X over Y, and $D(Y|X)$ is typically not equal to $D(X|Y)$. This asymmetry is mathematically enforced by excluding only those pairs tied on the dependent variable (Y) from the denominator, thereby isolating the predictive capability of X. Researchers must choose Somers' Delta when the research question explicitly defines X as the predictor of Y.

Goodman and Kruskal's Gamma (G) also relies on the difference between concordant and discordant pairs ($NC - ND$). However, its denominator is simply $NC + ND$, meaning it completely ignores all tied pairs, regardless of whether they are tied on X, Y, or both. Because Gamma excludes ties entirely, its magnitude tends to be inflated compared to Somers' D or Kendall's Tau for the same dataset, often yielding a value closer to the maximum possible association (1 or -1). While Gamma can be useful for measuring raw association when tied ranks are considered noise, Somers' D provides a more conservative and robust assessment of predictive strength, especially in ordinal scales where ties are common and structural.

Practical Implementation: Calculating Somers' D in R

To illustrate the practical application of Somers' D, consider a retail scenario where management wants to evaluate whether the perceived friendliness of a cashier (X) influences the customer's overall satisfaction rating (Y). Both variables are measured using a three-point ordinal scale, making this a perfect use case for a directional rank correlation measure.

The independent variable (X) is the cashier's niceness (ranked 1 to 3, where 3 is the nicest). The dependent variable (Y) is the customer's overall satisfaction (ranked 1 to 3, where 3 is the highest satisfaction). This setup necessitates calculating Somers' D in the direction D(Satisfaction | Niceness).

The retailer collects ratings from ten customers, yielding a small dataset that includes several tied values:

	Nice	Satisfaction
Customer #1	1	2
Customer #2	1	2
Customer #3	1	1
Customer #4	2	2
Customer #5	2	3
Customer #6	2	2
Customer #7	3	2
Customer #8	3	3
Customer #9	3	3
Customer #10	3	3

To quantify the relationship between these two ordinal variables efficiently, we employ statistical software. In the R programming environment, the calculation is easily performed using the `DescTools` package, which contains the dedicated function `SomersDelta()`. The following code demonstrates the necessary steps to calculate the coefficient:

```
#enter data
nice <- c(1, 1, 1, 2, 2, 2, 3, 3, 3, 3)
satisfaction <- c(2, 2, 1, 2, 3, 2, 2, 3, 3, 3)

#load DescTools package
library(DescTools)

#calculate Somers' D
SomersDelta(nice, satisfaction)

0.6896552
```

Upon executing the command, the result yields a Somers' D value of **0.6896552**.

Since this value is relatively close to 1, it indicates a fairly strong and positive relationship between the cashier's perceived niceness (X) and the customer's overall satisfaction (Y). The positive sign confirms the intuitive hypothesis: as the ranking for cashier niceness increases, the ranking for customer satisfaction tends to increase significantly. This quantitative result provides strong evidence that cashier friendliness is a powerful positive predictor of customer experience satisfaction.

Strengths and Limitations of Using Somers' D

Somers' D is acknowledged as a robust statistical measure, yet researchers must weigh its specific strengths against its inherent limitations before application. Its primary strength lies in its tailored suitability for predictive modeling involving ordinal outcomes. Being explicitly directional, it offers a precise measure of the marginal contribution of the independent variable to establishing the order of the dependent variable. Furthermore, as a rank-based statistic, it is highly robust to the presence of **outliers** and avoids reliance on restrictive assumptions like normality or homoscedasticity, making it exceptionally reliable for analyzing complex, real-world data distributions.

However, a key limitation of Somers' D is its potential sensitivity to the structure of the ordinal variables, particularly the number of categories employed. If the ordinal scale contains very few categories (e.g., only "low" and "high"), the set of possible Somers' D values becomes restricted, which can lead to an underestimation of the true underlying correlation strength. Additionally, Somers' D inherently assumes that the relationship is **monotonic**--meaning the variables move together consistently in one direction (either both increasing or both decreasing). It is unable to detect or accurately quantify non-linear or complex non-monotonic relationships, which would require alternative statistical approaches.

In conclusion, Somers' Delta is an indispensable tool for analyzing ordinal data, especially in research situations that demand a directional assessment of predictive capability. Researchers should prioritize Somers' D over symmetric alternatives when the research design clearly distinguishes between the predictor and the outcome, ensuring the most accurate and context-appropriate interpretation of the relationship between ordered categories.

Additional Resources