What is Somers' D? -Definition & Example

Authored by

stats writer

December 16, 2025

RECOMMENDED CITATION

stats writer (2025). *What is Somers' D? -Definition & Example*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=107575

ARABPSYCHOLOGY.COM

Somers' D, formally known as Somers' Delta ($delta$), is a specialized measure of association within nonparametric statistical methods. Its fundamental purpose is to quantify both the strength and the direction of the relationship between two variables, provided that both are measured on an ordinal scale. This statistic is critical when analyzing ranked data, where observations have a clear, inherent sequence--such as satisfaction scores, quality rankings, or Likert scale responses--but where the interval between ranks cannot be assumed to be equal. Unlike symmetric measures like Kendall's Tau-b, Somers' D is inherently asymmetric, requiring the analyst to explicitly designate one variable as the independent predictor (X) and the other as the dependent outcome (Y). This directional property makes Somers' D indispensable in predictive modeling contexts, allowing researchers to assess how effectively the rank ordering of X predicts the rank ordering of Y.

The standardized value for Somers' D always falls within the tight range of -1 to 1. This range allows for immediate and clear interpretation: a value near 1 signifies a strong positive agreement in ranks, while a value near -1 indicates a strong inverse relationship. Values approaching zero suggest that the rank of the independent variable provides minimal predictive power regarding the rank of the dependent variable. Due to its robustness and directional nature, Somers' D is often used in generalized linear models, particularly for evaluating the discriminatory power of ordered response models, and serves as a vital component in assessing the predictive accuracy of various classification algorithms.

Understanding Ordinal Variables

The appropriate application of Somers' D hinges entirely on the nature of the variables being analyzed, which must be ordinal. An ordinal variable is characterized by categories that can be meaningfully ordered or ranked, yet the numerical difference between those categories is not necessarily consistent or quantifiable. For example, ranking a service as "Poor (1)," "Average (2)," or "Excellent (3)" establishes an order, but we cannot assume the improvement from "Poor" to "Average" is the exact same magnitude as the improvement from "Average" to "Excellent." This contrasts sharply with continuous variables, such as income or temperature, where intervals are precise and consistent.

The challenge posed by ordinal data is that standard parametric correlation techniques, such as Pearson's correlation coefficient, assume underlying interval data and often rely on assumptions about the normal distribution of data. When these assumptions are violated by ranked data, the resulting correlation coefficient can be misleading. Somers' D circumvents these issues by focusing solely on the relative ranking of pairs of observations. By using ranks rather than the assumed interval properties of the assigned numerical codes, it provides a statistically sound and more conservative measure of association appropriate for ordered categorical data in research across social sciences, psychology, and epidemiology.

ARABPSYCHOLOGY.COM

The Purpose and Utility of Somers' D

The core utility of Somers' D lies in its specific design for assessing predictive relationships within ranked data. While measures like Kendall's Tau-b are symmetric and simply describe the overall association, Somers' D provides a directed measure of association, $delta_{Y|X}$, which explicitly asks: "Given the difference in rank on X, what is the probability that the rank on Y differs in the same direction?" This focus on prediction makes it highly valuable in analytical scenarios where the causal direction is hypothesized or established. For instance, in clinical trials, one might use Somers' D to evaluate how well a ranked dosage (X) predicts a ranked recovery outcome (Y).

Furthermore, Somers' D possesses a significant theoretical link to binary classification models. For a binary dependent variable (Y) and a continuous or ordinal independent variable (X), Somers' D is directly related to the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve via the formula: $D = 2 times AUC - 1$. This relationship solidifies its role in evaluating model discrimination--that is, the model's ability to correctly order or rank observations according to their likelihood of being in the positive category. As a member of the family of nonparametric statistical methods, its utility is expanded by its minimal requirements for distributional assumptions, making it a robust choice when population distribution parameters are unknown or complex.

Defining Concordant and Discordant Pairs

The calculation of Somers' D is rooted in the systematic comparison of all possible distinct pairs of observations in the dataset. For any two observations, i and j, with corresponding values (x_i, y_i) and (x_j, y_j), they are classified into one of three categories based on the relationship between their ranks. This rigorous pairwise comparison is fundamental to generating the N_C and N_D counts needed for the final coefficient calculation.

Two pairs are defined as concordant if the relative ranking between the two observations is the same for both variables. Specifically, if x_i is greater than x_j, and y_i is also greater than y_j, the pair is concordant. Similarly, if both x_i and y_i are smaller than their counterparts, they are also concordant. Concordant pairs (N_C) support the existence of a positive relationship: as the independent variable increases in rank, the dependent variable also increases in rank. A dataset exhibiting a very high number of concordant pairs suggests a strong positive association between the ordinal rankings.

In contrast, two pairs are classified as **discordant** if the relative ranking is inverse between the two variables. For example, if x_i is greater than x_j, but y_i is less than y_j, the pair is discordant. Discordant pairs (N_D) indicate a negative or inverse relationship: an increase in the rank of the independent variable corresponds to a decrease in the rank of the dependent variable. Finally, pairs are considered **tied** if they share the same rank on either the X variable, the Y

ARABPSYCHOLOGY.COM

variable, or both. The handling of these tied pairs--especially those tied on the independent variable--is what gives Somers' D its directional specificity, ensuring that the denominator only includes pairs that could potentially differentiate the ranks of X.

The Mathematical Formula for Somers' D

The formalized structure of the Somers' D formula ensures that the resulting coefficient is a normalized measure of the predictive relationship, ranging from -1 to 1. The formula is a ratio where the numerator quantifies the net agreement in rank movement, and the denominator standardizes this agreement by dividing by the total number of non-tied pairs based on the independent variable (X).

The calculation is structured as follows:

Somers' D = (NC - ND) / (NC + ND + NT, Y)

Where the components are rigorously defined based on the pairwise comparisons of the observations:

NC: The total number of concordant pairs, reflecting directional agreement.
ND: The total number of **discordant** pairs, reflecting directional disagreement.
NT, Y: The total number of pairs that are tied on the dependent variable (Y) but are **not** tied on the independent variable (X).

The denominator, encompassing $N_C + N_D + N_{T, Y}$, effectively represents the total number of pairs that are not tied on the independent variable X. The exclusion of pairs tied on X from the denominator is the mathematical mechanism that enforces the asymmetry of Somers' D (specifically, $\delta_{Y|X}$). This ensures that the denominator represents the maximum possible number of pairs where X could potentially distinguish the rank of Y, thus providing a measure of the predictive power of X relative to Y.

Interpreting the Range of Somers' D

Understanding the meaning behind the specific values obtained for Somers' D is crucial for drawing valid conclusions from the statistical analysis. Since the coefficient is standardized between -1 and +1, its interpretation is straightforward and immediately accessible.

A calculated D value of **+1.0** signifies a state of perfect monotonic agreement. In this extreme case, every single pair of observations is concordant, meaning that whenever the independent variable X increases in rank, the dependent variable Y increases correspondingly, without exception. This indicates a deterministic, positive relationship between the two ordinal variables. Conversely, a value of **-1.0** represents perfect negative (inverse) agreement, where all non-tied pairs are discordant, signifying that as the rank of X increases, the rank of Y reliably decreases.

Values closer to **0** indicate a weak relationship. For example, a D of 0.05 implies that the number of concordant pairs is nearly equal to the number of discordant pairs. This suggests that the rank of the independent variable has little systematic influence on the rank of the dependent variable. Researchers typically interpret values between 0.1 and 0.3 as a weak association, 0.3 to 0.5 as moderate, and above 0.5 as a strong association. The interpretation always remains relative to the context of the research, but the coefficient provides a precise and unbiased quantification of the observed rank correlation.

Practical Application: A Statistical Example in R

To illustrate the application of Somers' D in a real-world context, we revisit the scenario of a grocery store assessing the relationship between staff interaction and customer satisfaction. This example demonstrates how to apply nonparametric statistical methods to ordered survey responses.

The store defined two ordinal variables, both scored from 1 (lowest) to 3 (highest):

Cashier Niceness (Independent variable, X)
Customer Satisfaction (Dependent variable, Y)

The management wants to confirm their hypothesis that better cashier ratings lead to better satisfaction ratings. They collected ten paired ratings, which are visually represented in the table below, showing distinct patterns of association that must be quantified:

	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

ARABPSYCHOLOGY.COM

Using the R statistical environment, we can efficiently calculate Somers' D. We define the raw data vectors, load the necessary package (`DescTools`), and execute the function, with the independent variable listed first and the dependent variable second, maintaining the directional integrity of the measure. The following code performs the necessary counting of concordant and discordant pairs and computes the final 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 (Delta Y|X)
SomersDelta(nice, satisfaction)

0.6896552
```

Interpreting the R Output

The computation yields a Somers' D value of **0.6896552**. This single numerical output synthesizes the complex relationship embedded within the ten pairs of ordinal data points.

Since the resulting coefficient is positive and substantially greater than zero (0.69), we conclude there is a strong positive monotonic relationship between Cashier Niceness and Customer Satisfaction. This indicates that the observations demonstrate a high degree of concordance: when a customer rates the cashier as nicer (higher rank on X), they are highly likely to also provide a better overall satisfaction rating (higher rank on Y). This result strongly supports the grocery store's hypothesis and provides compelling evidence that improvements in cashier niceness directly translate to statistically significant improvements in customer satisfaction ranks. The high magnitude of the coefficient suggests a reliable and actionable correlation for operational decision-making.

Comparison to Other Nonparametric Measures

While Somers' D is often the preferred choice for directed ordinal association, researchers frequently encounter two related statistics: Kendall's Tau-b (tau_b) and Goodman and Kruskal's Gamma ($gamma$). All three originate from the counts of concordant and discordant pairs, but they diverge in their treatment of tied ranks, profoundly affecting their interpretation and magnitude.

Kendall's Tau-b is a symmetric measure that adjusts its denominator for ties on both X and Y. It

ARABPSYCHOLOGY.COM

addresses the overall association without designating a dependent variable. Goodman and Kruskal's Gamma, in contrast, completely excludes all tied pairs (on both X and Y) from its denominator, resulting in a coefficient that often appears larger in magnitude than both tau_b and Somers' D, potentially inflating the perceived strength of the association.

The defining feature of Somers' D ($delta_{Y|X}$) is its asymmetric denominator, which excludes only pairs tied on the independent variable (X). This directional adjustment ensures that Somers' D directly measures the reduction in prediction error when predicting Y using the ranks of X, making it the most appropriate and statistically rigorous choice for predictive modeling involving ordinal variables within the suite of nonparametric statistical methods.

ARABPSYCHOLOGY.COM