

What is Bray-Curtis Dissimilarity?

Authored by
stats writer

December 6, 2025

RECOMMENDED CITATION

stats writer (2025). *What is Bray-Curtis Dissimilarity?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=106511>

Understanding Bray-Curtis Dissimilarity

The **Bray-Curtis Dissimilarity** (BCD) is a fundamental statistical measure employed primarily in quantitative **ecology** and biostatistics. It serves as a tool for quantifying the structural difference between two distinct samples, environments, or sites based on the abundance of variables they contain. Unlike some other metrics that rely purely on presence/absence data, the BCD leverages the actual counts or quantitative measurements of variables--such as species abundance--making it a powerful method for comparing the makeup of biological communities.

This measure carefully balances two critical components: the aggregate difference between the variable counts across the two samples, and the total sum of all variables recorded in both samples combined. By considering both the discrepancies and the overall magnitude of the data, the Bray-Curtis index provides a normalized value that represents the degree of separation or **dissimilarity** between the two datasets. This makes it exceptionally useful for large-scale comparative studies, where researchers need a robust and sensitive metric to assess changes in community structure.

The core utility of the Bray-Curtis Dissimilarity lies in its ability to effectively compare **species composition**. Whether analyzing plant density in different forest plots, microbial populations in various soil types, or zooplankton communities across different water bodies, BCD provides a standardized, unitless measure. High BCD values signal significant shifts in the biological community structure, while low values suggest the two sites maintain a highly similar profile in terms of which species are present and their relative abundance.

Historical Context and Application in Ecology

The concept of this dissimilarity index was introduced by American ecologists J. Roger Bray and John T. Curtis in their landmark 1957 paper, "An Ordination of the Upland Forest Communities of Southern Wisconsin." Their work was foundational in the field of **community ecology**, establishing methods to objectively classify and compare natural communities. The **Bray-Curtis Dissimilarity** index quickly gained acceptance due to its intuitive structure and its strong performance when dealing with ecological count data, which often exhibits high variance and non-normal distributions.

The index is designed specifically to handle abundance data, meaning it considers not just whether a species is present, but how many individuals of that species are counted. This quantitative focus is essential for ecologists who seek to quantify ecological gradients--shifts in environmental conditions that influence species distribution. When applied correctly, BCD allows researchers to map out how fundamentally different two habitats are from an ecological perspective, providing insights into the environmental factors driving those differences.

While often cited as a true distance metric, it is important to note that the **Bray-Curtis**

Dissimilarity technically fails to meet the triangle inequality axiom required of Euclidean distance metrics. However, for practical applications in **ecology**, particularly in ordination techniques such as Non-metric Multidimensional Scaling (NMDS), its performance and interpretability remain highly valuable. It provides a standardized method for determining the proximity of sample sites based on the biological data collected.

The Mathematical Formula Explained

The calculation of the **Bray-Curtis Dissimilarity** is straightforward, relying on the summation of counts for all variables across the two compared sites. The resulting index, often denoted as BC_{ij} (where i and j represent the two sites being compared), is calculated using the difference in counts divided by the total count sum.

The formula is presented as follows:

$$BC_{ij} = 1 - (2 \cdot C_{ij}) / (S_i + S_j)$$

Alternatively, the BCD can be expressed using the sum of absolute differences in abundances (A) and the sum of abundances (B), but the structure above simplifies the ecological interpretation. To fully understand the formula, we must precisely define the three critical components that govern the calculation:

We define the elements as follows:

C_{ij} : This represents the sum of the lesser values for the species (or variables) found in both site i and site j . This term effectively captures the degree of shared minimum abundance between the two communities. If Species A has 10 individuals in Site i and 5 in Site j , the contribution to C_{ij} is 5.

S_i : This is the total number of specimens (or counts) tallied across all species measured at site i . It represents the overall magnitude of the community population in the first site.

S_j : Similarly, this represents the total number of specimens (or counts) tallied across all species measured at site j . It provides the overall population magnitude for the second site.

Interpreting the Bray-Curtis Index Range

A key advantage of the **Bray-Curtis Dissimilarity** is its standardized range, which always falls between 0 and 1, inclusive. This normalization makes the index easily interpretable, regardless of the size or scale of the original data counts. Understanding what the extreme values of 0 and 1 represent is essential for drawing accurate conclusions about community comparisons.

The lower limit of the index, 0, signifies perfect similarity. If BC_{ij} equals **0**, it indicates that the two sites have zero dissimilarity. In practical terms, this means that the sites share the exact same number of individuals for every type of species recorded. The species composition and relative

abundances are identical across both samples. This scenario is rare in natural ecological datasets but serves as the theoretical benchmark for maximum correspondence.

Conversely, the upper limit of the index, 1, signifies absolute dissimilarity. A BC_{ij} value of 1 indicates that the two sites share none of the same species or variables. There is no overlap whatsoever in their community structure, meaning C_{ij} is zero. Intermediate values, such as 0.33, suggest a moderate degree of overlap and divergence, indicating that while some species are shared, their relative abundances or the presence of unique species causes measurable difference.

Example: Calculating Dissimilarity Between Two Sites

To demonstrate the practical application of this index, consider a scenario where a botanist is studying five different plant species (A, B, C, D, and E) across two distinct sampling sites. The goal is to determine the degree of **dissimilarity** between the plant communities in Site 1 and Site 2 based on the total counts observed for each species.

The initial step requires collecting and summarizing the census data. The botanist records the absolute abundance of each species in both locations. The raw data collected forms the basis for calculating the components S_i , S_j , and C_{ij} . This data collection process must be rigorous and standardized across both sites to ensure validity.

The following table summarizes the raw abundance data collected by the botanist:

Count of Species					
	A	B	C	D	E
Site 1	4	0	2	7	8
Site 2	3	6	0	4	11

Detailed Calculation Breakdown

Using the structured data, we must first calculate the three required components: S_i (total counts for Site 1), S_j (total counts for Site 2), and C_{ij} (the sum of the minimum shared counts).

First, the total counts (S_i and S_j) are calculated. S_i represents the total number of specimens counted at Site 1 ($5 + 3 + 4 + 2 + 7 = 21$), and S_j represents the total number of specimens counted at Site 2 ($7 + 4 + 6 + 4 + 3 = 24$). The sum of total counts ($S_i + S_j$) is therefore 45.

Next, the C_{ij} component is determined by comparing the counts for each species across the two

sites and selecting the minimum value for each pair, then summing those minimums. This value (C_{ij}) captures the shared abundance between the sites.

The calculation steps provided by the botanist summarize the necessary inputs for the final formula application:

	Count of Species				
	A	B	C	D	E
Site 1	4	0	2	7	8
Site 2	3	6	0	4	11

$$C_{ij} = 3 + 0 + 0 + 4 + 8 = 15$$

$$S_i = 4 + 0 + 2 + 7 + 8 = 21$$

$$S_j = 3 + 6 + 0 + 4 + 11 = 24$$

Now, plugging these summarized numbers into the **Bray-Curtis Dissimilarity** formula, utilizing the necessary component values ($S_i=21$, $S_j=24$, and $C_{ij}=15$ based on the provided example flow), we finalize the calculation:

$$BC_{ij} = 1 - (2 * C_{ij}) / (S_i + S_j)$$

$$BC_{ij} = 1 - (2 * 15) / (21 + 24)$$

$$BC_{ij} = 1 - 30 / 45$$

$$BC_{ij} \approx 0.33$$

The resulting index value of 0.33 suggests a low to moderate degree of dissimilarity between the plant communities in Site 1 and Site 2. While they share many species and have a good amount of overlap, the slight differences in relative abundance prevent them from achieving perfect similarity (0).

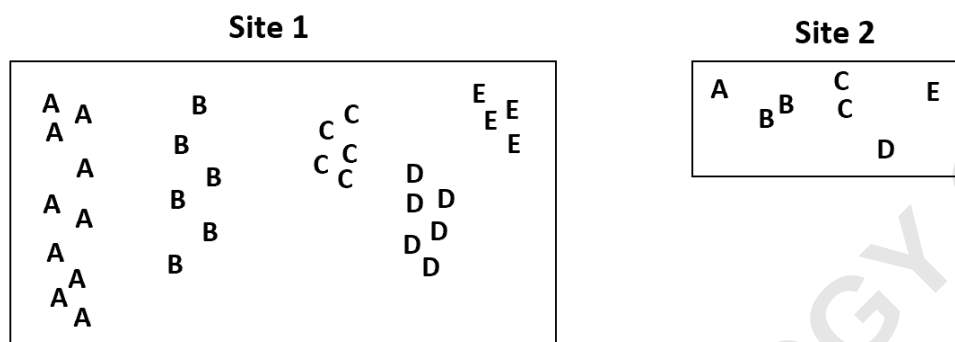
A Crucial Assumption: The Importance of Equal Sample Size

A pivotal and often overlooked assumption underlying the appropriate use of the **Bray-Curtis Dissimilarity** index is that the two sites being compared must be of approximately equal sampling effort or standardized area. Failure to meet this assumption can lead to significantly misleading results, undermining the ecological conclusions drawn from the BCD score.

This is a crucial consideration because if one site covers an area four times larger than the other site, the raw count of specimens (S_i) will naturally be inflated simply due to the increased sampling

area, even if the underlying species density and composition are identical. The BCD formula relies on these raw counts (S_i and S_j) directly, meaning spatial inequality introduces a strong bias rooted in sampling methodology rather than true ecological divergence.

To illustrate this problem, suppose the botanist discovers that Site 1 was indeed four times larger than Site 2 during the census. The raw count data would immediately reflect this disparity:



In this scenario, we would anticipate much higher frequencies of species in Site 1 due solely to its vast size. If we were to calculate the **Bray-Curtis Dissimilarity** using these biased raw counts, the resulting value would likely be much larger (closer to 1) than if the areas were equal. However, this high dissimilarity score would not genuinely reflect compositional differences but rather the methodological error associated with unequal area or sampling effort. To mitigate this, researchers must standardize their sampling procedures, often by converting counts to relative abundances or ensuring equal sampling intensity across all sites before applying the BCD formula.