

Anexo 1

Glosario de términos estadísticos

Bray Curtis: es un método estadístico, muy utilizado en ecología, que cuantifica la disimilitud entre dos muestras (u otros datos) diferentes, generando así una matriz de disimilitud o de distancia. La distancia Bray-Curtis hace referencia a la diferencia en la abundancia relativa de una especie (u otro nivel taxonómico) que existe entre dos muestras. Esta distancia toma valores que van del 0 al 1; así, la comparación de dos muestras idénticas dará un valor de 0, mientras que la de las dos muestras más diferentes dará 1. (Bray y Curtis, 1957; Espinosa, 2019).

Matriz de disimilitud o matriz de distancia: es una matriz de datos simétrica (mismo número de columnas que de filas) que compara un conjunto de muestras entre sí. La disimilitud indica cómo de diferentes son las muestras entre sí, devolviendo un valor para la comparación de cada muestra con respecto a las demás. Generalmente, la disimilitud se mide del 0 al 1, siendo 0 la distancia entre dos muestras idénticas, y 1 la distancia entre dos muestras completamente diferentes (Demey et al., 2011).

Normalización: transformación de los datos con el objetivo de corregir la forma de distribución de sus variables, obteniéndose así varianzas homogéneas para todas las variables. Estas condiciones permiten que los análisis multivariantes funcionen mejor (Ramette, 2007).

PCoA: el método de PCoA fue descrito por Gower (1966). Es un tipo de análisis de ordenación (es decir, que resume y proyecta un conjunto de datos multivariantes de una comunidad en el espacio dimensional o tridimensional). Se aplica a datos ya normalizados, para visualizar tendencias globales de cómo de similares o diferentes son las poblaciones bacterianas en ciertos tipos de muestras (p.e. pre- vs post-tratamiento). Esto permite visualizar fácilmente las diferencias entre los datos, así como agrupar las muestras en función de su similitud en composición bacteriana (Weiss *et al.*, 2017). La ventaja de utilizar un PCoA en lugar de un PCA es que este último se basa en distancia euclidiana, la cual se ve muy afectada cuando los datos tienen una gran cantidad de ceros, como es el caso de los datos de abundancia bacteriana. Por el contrario, el PCoA no se ve afectado por este exceso de ceros, lo cual lo convierte en un análisis de ordenación mucho más fiable que el PCA en lo que respecta a los análisis de la microbiota u otras comunidades bacterianas.