

Capítulo 6: Medidas de diversidad comunitaria y cálculos

20 septembre 2022

Contenido

- 1 **Introducción a las diversidades comunitarias**
 - Diversidades Alfa, Beta y Gamma
 - Medidas y cálculo de la diversidad Alfa

La mayoría de los métodos de diversidad asumen que los datos son recuentos de individuos. Tres niveles de diversidad (alfa, beta y gamma) se volvieron fundamentales en ecología comunitaria.

Whittaker introdujo y dividió la diversidad en varios componentes, los más conocidos son la diversidad alfa y beta.

- **Alfa** : Se define como la diversidad en un punto o muestra. Actúa como estadística resumida de una sola población. En general se utiliza para la diversidad local. En el estudio de microbioma, la diversidad alfa se conoce como diversidad dentro de una sola muestra o dentro de una comunidad.

La mayoría de los métodos de diversidad asumen que los datos son recuentos de individuos. Tres niveles de diversidad (alfa, beta y gamma) se volvieron fundamentales en ecología comunitaria.

Whittaker introdujo y dividió la diversidad en varios componentes, los más conocidos son la diversidad alfa y beta.

- **Alfa** : Se define como la diversidad en un punto o muestra. Actúa como estadística resumida de una sola población. En general se utiliza para la diversidad local. En el estudio de microbioma, la diversidad alfa se conoce como diversidad dentro de una sola muestra o dentro de una comunidad.

La mayoría de los métodos de diversidad asumen que los datos son recuentos de individuos. Tres niveles de diversidad (alfa, beta y gamma) se volvieron fundamentales en ecología comunitaria.

Whittaker introdujo y dividió la diversidad en varios componentes, los más conocidos son la diversidad alfa y beta.

- **Alfa** : Se define como la diversidad en un punto o muestra. Actúa como estadística resumida de una sola población. En general se utiliza para la diversidad local. En el estudio de microbioma, la diversidad alfa se conoce como diversidad dentro de una sola muestra o dentro de una comunidad.

Un propósito importante del estudio del microbioma es determinar si las comunidades de microbioma pueden clasificarse juntas o deben separarse en sus bacterias para diferenciar el tratamiento de control, la deficiencia genética del tipo salvaje, entre otras características.

- **Beta** : Es una medida de cambio en la diversidad a través de gradientes ambientales ; es decir, la tasa de cambio en la composición de especies de una comunidad a otra a lo largo de gradientes. Refleja el reemplazo de especies a medida que uno se mueve a través del espacio o tiempo.

En general, la diversidad beta evalúa las diferencias entre dos o más conjuntos locales y regionales, lo que permite dilucidar cuánta diversidad es exclusiva de un conjunto local, o descubrir cómo muchos taxones son compartidos entre las comunidades.

Un propósito importante del estudio del microbioma es determinar si las comunidades de microbioma pueden clasificarse juntas o deben separarse en sus bacterias para diferenciar el tratamiento de control, la deficiencia genética del tipo salvaje, entre otras características.

- **Beta** : Es una medida de cambio en la diversidad a través de gradientes ambientales ; es decir, la tasa de cambio en la composición de especies de una comunidad a otra a lo largo de gradientes. Refleja el reemplazo de especies a medida que uno se mueve a través del espacio o tiempo.

En general, la diversidad beta evalúa las diferencias entre dos o más conjuntos locales y regionales, lo que permite dilucidar cuánta diversidad es exclusiva de un conjunto local, o descubrir cómo muchos taxones son compartidos entre las comunidades.

Un propósito importante del estudio del microbioma es determinar si las comunidades de microbioma pueden clasificarse juntas o deben separarse en sus bacterias para diferenciar el tratamiento de control, la deficiencia genética del tipo salvaje, entre otras características.

- **Beta** : Es una medida de cambio en la diversidad a través de gradientes ambientales ; es decir, la tasa de cambio en la composición de especies de una comunidad a otra a lo largo de gradientes. Refleja el reemplazo de especies a medida que uno se mueve a través del espacio o tiempo.

En general, la diversidad beta evalúa las diferencias entre dos o más conjuntos locales y regionales, lo que permite dilucidar cuánta diversidad es exclusiva de un conjunto local, o descubrir cómo muchos taxones son compartidos entre las comunidades.

En la literatura sobre microbiomas rara vez se usa la diversidad gamma.

- **Gamma** : Es la diversidad de una región o paisaje que contiene varias comunidades. En la práctica la diversidad Alfa se puede considerar como la diversidad de la muestra u observación individual a la diversidad, en cambio, la diversidad Gamma se considera como la diversidad de todas las muestras combinadas.

Los índices de diversidad comunitaria combinan la riqueza y abundancia de especies en un único valor de uniformidad. Las comunidades que están numéricamente dominadas por una o pocas especies exhiben baja uniformidad, mientras que las comunidades donde la abundancia se distribuye por igual entre las especies exhiben alta uniformidad

Se tienen 3 índices particularmente importantes respecto a microbioma :

- 1 **Chao 1** : Medida cualitativa basada en especies,
- 2 **Shannon y Simpson** : Son medidas cuantitativas de especies.

Los índices de diversidad comunitaria combinan la riqueza y abundancia de especies en un único valor de uniformidad. Las comunidades que están numéricamente dominadas por una o pocas especies exhiben baja uniformidad, mientras que las comunidades donde la abundancia se distribuye por igual entre las especies exhiben alta uniformidad

Se tienen 3 índices particularmente importantes respecto a microbioma :

- 1 **Chao 1** : Medida cualitativa basada en especies,
- 2 **Shannon y Simpson** : Son medidas cuantitativas de especies.

Índice de riqueza de Chao 1 y número de taxones

Los estimadores de riqueza de especies estiman el número total de especies presentes en una muestra o comunidad.

Anne Chao desarrolló dos estimadores no paramétricos de riqueza de especies para datos de presencia/ausencia, que son Chao1 y Chao2, respectivamente.

El índice Chao1 se basa en el número de OTUs encontrados en una muestra. Se usa usualmente en estudios de ecología y microbioma. La fórmula es :

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{n_1^2}{2n_2}, \quad (1)$$

donde :

Índice de riqueza de Chao 1 y número de taxones

Los estimadores de riqueza de especies estiman el número total de especies presentes en una muestra o comunidad.

Anne Chao desarrolló dos estimadores no paramétricos de riqueza de especies para datos de presencia/ausencia, que son Chao1 y Chao2, respectivamente.

El índice Chao1 se basa en el número de OTUs encontrados en una muestra. Se usa usualmente en estudios de ecología y microbioma. La fórmula es :

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{n_1^2}{2n_2}, \quad (1)$$

donde :

- 1 S_{Chao1} es el número estimado de especies,
- 2 S_{obs} es el número de especies observadas en total,
- 3 n_1 es el número de taxones únicos (taxones representados por una única lectura en esa comunidad),
- 4 n_2 es el número de taxones dobles, como el número de especies representadas sólo dos veces en las muestras.

Este índice es particularmente útil para conjuntos de datos sesgados hacia las clases de baja abundancia, como seguramente es el caso de los microbiomas.

- 1 S_{Chao1} es el número estimado de especies,
- 2 S_{obs} es el número de especies observadas en total,
- 3 n_1 es el número de taxones únicos (taxones representados por una única lectura en esa comunidad),
- 4 n_2 es el número de taxones dobles, como el número de especies representadas sólo dos veces en las muestras.

Este índice es particularmente útil para conjuntos de datos sesgados hacia las clases de baja abundancia, como seguramente es el caso de los microbiomas.

- 1 S_{Chao1} es el número estimado de especies,
- 2 S_{obs} es el número de especies observadas en total,
- 3 n_1 es el número de taxones únicos (taxones representados por una única lectura en esa comunidad),
- 4 n_2 es el número de taxones dobles, como el número de especies representadas sólo dos veces en las muestras.

Este índice es particularmente útil para conjuntos de datos sesgados hacia las clases de baja abundancia, como seguramente es el caso de los microbiomas.

- 1 S_{Chao1} es el número estimado de especies,
- 2 S_{obs} es el número de especies observadas en total,
- 3 n_1 es el número de taxones únicos (taxones representados por una única lectura en esa comunidad),
- 4 n_2 es el número de taxones dobles, como el número de especies representadas sólo dos veces en las muestras.

Este índice es particularmente útil para conjuntos de datos sesgados hacia las clases de baja abundancia, como seguramente es el caso de los microbiomas.

- 1 S_{Chao1} es el número estimado de especies,
- 2 S_{obs} es el número de especies observadas en total,
- 3 n_1 es el número de taxones únicos (taxones representados por una única lectura en esa comunidad),
- 4 n_2 es el número de taxones dobles, como el número de especies representadas sólo dos veces en las muestras.

Este índice es particularmente útil para conjuntos de datos sesgados hacia las clases de baja abundancia, como seguramente es el caso de los microbiomas.

A partir de la fórmula previa podemos ver si el número de taxones únicos n_1 es mayor; es decir, si tiene muchos, en cuyo caso es probable que existan más OTUs no detectados, y entonces el índice Chao 1 estimará una mayor riqueza de especies de lo que sería para una muestra sin OTUs extraños.

Chao también derivó una solución de forma cerrada para la varianza de S_{Chao1} :

$$\text{Var}(S_{\text{Chao1}}) = n_2 \left(\frac{m^4}{4} + m^3 + \frac{m^2}{2} \right) \quad (2)$$

donde $m = \frac{n_1}{n_2}$. La fórmula anterior estima la precisión del índice Chao1 a partir de múltiples muestras. La diversidad alfa se calcula en base a datos de abundancia sin procesar. Además la estructura de datos debe ser en filas para muestras y columnas para los taxones.

A partir de la fórmula previa podemos ver si el número de taxones únicos n_1 es mayor; es decir, si tiene muchos, en cuyo caso es probable que existan más OTUs no detectados, y entonces el índice Chao 1 estimará una mayor riqueza de especies de lo que sería para una muestra sin OTUs extraños.

Chao también derivó una solución de forma cerrada para la varianza de S_{Chao1} :

$$\text{Var}(S_{\text{Chao1}}) = n_2 \left(\frac{m^4}{4} + m^3 + \frac{m^2}{2} \right) \quad (2)$$

donde $m = \frac{n_1}{n_2}$. La fórmula anterior estima la precisión del índice Chao1 a partir de múltiples muestras. La diversidad alfa se calcula en base a datos de abundancia sin procesar. Además la estructura de datos debe ser en filas para muestras y columnas para los taxones.

A partir de la fórmula previa podemos ver si el número de taxones únicos n_1 es mayor; es decir, si tiene muchos, en cuyo caso es probable que existan más OTUs no detectados, y entonces el índice Chao 1 estimará una mayor riqueza de especies de lo que sería para una muestra sin OTUs extraños.

Chao también derivó una solución de forma cerrada para la varianza de S_{Chao1} :

$$\text{Var}(S_{\text{Chao1}}) = n_2 \left(\frac{m^4}{4} + m^3 + \frac{m^2}{2} \right) \quad (2)$$

donde $m = \frac{n_1}{n_2}$. La fórmula anterior estima la precisión del índice Chao1 a partir de múltiples muestras. La diversidad alfa se calcula en base a datos de abundancia sin procesar. Además la estructura de datos debe ser en filas para muestras y columnas para los taxones.

Índice de diversidad de Shannon-Wiener

Una de las medidas más populares de la diversidad de especies es el índice de diversidad de Shannon-Wiener, denominado H' . Considera las diferencias en proporción o abundancia de cada especie. Este índice se basa en la teoría de la información, mide la incertidumbre :

¿Qué tan difícil sería predecir correctamente la especie del próximo individuo recolectado ?

La fórmula del índice de Shannon es :

$$H' = - \sum_{i=1}^S p_i \ln(p_i), \quad (3)$$

donde :

- 1 p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad,
- 2 S es el número total de especies presentes, de modo que $\sum_{i=1}^S p_i = 1$.

Índice de diversidad de Shannon-Wiener

Una de las medidas más populares de la diversidad de especies es el índice de diversidad de Shannon-Wiener, denominado H' . Considera las diferencias en proporción o abundancia de cada especie. Este índice se basa en la teoría de la información, mide la incertidumbre :

¿Qué tan difícil sería predecir correctamente la especie del próximo individuo recolectado ?

La fórmula del índice de Shannon es :

$$H' = - \sum_{i=1}^S p_i \ln(p_i), \quad (3)$$

donde :

- 1 p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad,
- 2 S es el número total de especies presentes, de modo que $\sum_{i=1}^S p_i = 1$.

Índice de diversidad de Shannon-Wiener

Una de las medidas más populares de la diversidad de especies es el índice de diversidad de Shannon-Wiener, denominado H' . Considera las diferencias en proporción o abundancia de cada especie. Este índice se basa en la teoría de la información, mide la incertidumbre :

¿Qué tan difícil sería predecir correctamente la especie del próximo individuo recolectado ?

La fórmula del índice de Shannon es :

$$H' = - \sum_{i=1}^S p_i \ln(p_i), \quad (3)$$

donde :

- 1 p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad,
- 2 S es el número total de especies presentes, de modo que $\sum_{i=1}^S p_i = 1$.

Índice de diversidad de Shannon-Wiener

Una de las medidas más populares de la diversidad de especies es el índice de diversidad de Shannon-Wiener, denominado H' . Considera las diferencias en proporción o abundancia de cada especie. Este índice se basa en la teoría de la información, mide la incertidumbre :

¿Qué tan difícil sería predecir correctamente la especie del próximo individuo recolectado ?

La fórmula del índice de Shannon es :

$$H' = - \sum_{i=1}^S p_i \ln(p_i), \quad (3)$$

donde :

- 1 p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad,
- 2 S es el número total de especies presentes, de modo que $\sum_{i=1}^S p_i = 1$.

Índice de diversidad de Shannon-Wiener

Una de las medidas más populares de la diversidad de especies es el índice de diversidad de Shannon-Wiener, denominado H' . Considera las diferencias en proporción o abundancia de cada especie. Este índice se basa en la teoría de la información, mide la incertidumbre :

¿Qué tan difícil sería predecir correctamente la especie del próximo individuo recolectado ?

La fórmula del índice de Shannon es :

$$H' = - \sum_{i=1}^S p_i \ln(p_i), \quad (3)$$

donde :

- 1 p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad,
- 2 S es el número total de especies presentes, de modo que $\sum_{i=1}^S p_i = 1$.

Algunas observaciones importantes :

- La teoría de la información mide la cantidad de incertidumbre, de modo que cuanto mayor sea valor de H' , mayor es la incertidumbre.
- La medida de Shannon-Wiener H' aumenta con el número de especies en la comunidad y en teoría puede alcanzar valores muy grandes. En la práctica, para las comunidades biológicas H' no parece exceder 5.0.
- El índice de Shannon “da más importancia” a categorías menos comunes (por ejemplo, especies raras en estudios de microbioma).
- Estrictamente hablando, la medida de contenido de información de Shannon-Wiener debe usarse sólo en muestras aleatorias extraídas de una gran comunidad en la que se conoce el número total de especies.

Algunas observaciones importantes :

- La teoría de la información mide la cantidad de incertidumbre, de modo que cuanto mayor sea valor de H' , mayor es la incertidumbre.
- La medida de Shannon-Wiener H' aumenta con el número de especies en la comunidad y en teoría puede alcanzar valores muy grandes. En la práctica, para las comunidades biológicas H' no parece exceder 5.0.
- El índice de Shannon “da más importancia” a categorías menos comunes (por ejemplo, especies raras en estudios de microbioma).
- Estrictamente hablando, la medida de contenido de información de Shannon-Wiener debe usarse sólo en muestras aleatorias extraídas de una gran comunidad en la que se conoce el número total de especies.

Algunas observaciones importantes :

- La teoría de la información mide la cantidad de incertidumbre, de modo que cuanto mayor sea valor de H' , mayor es la incertidumbre.
- La medida de Shannon-Wiener H' aumenta con el número de especies en la comunidad y en teoría puede alcanzar valores muy grandes. En la práctica, para las comunidades biológicas H' no parece exceder 5.0.
- El índice de Shannon “da más importancia” a categorías menos comunes (por ejemplo, especies raras en estudios de microbioma).
- Estrictamente hablando, la medida de contenido de información de Shannon-Wiener debe usarse sólo en muestras aleatorias extraídas de una gran comunidad en la que se conoce el número total de especies.

Algunas observaciones importantes :

- La teoría de la información mide la cantidad de incertidumbre, de modo que cuanto mayor sea valor de H' , mayor es la incertidumbre.
- La medida de Shannon-Wiener H' aumenta con el número de especies en la comunidad y en teoría puede alcanzar valores muy grandes. En la práctica, para las comunidades biológicas H' no parece exceder 5.0.
- El índice de Shannon “da más importancia” a categorías menos comunes (por ejemplo, especies raras en estudios de microbioma).
- Estrictamente hablando, la medida de contenido de información de Shannon-Wiener debe usarse sólo en muestras aleatorias extraídas de una gran comunidad en la que se conoce el número total de especies.

Índice de diversidad de Simpson

Simpson en 1949 propuso un nuevo concepto de diversidad que combina dos ideas separadas, la riqueza de especies y la uniformidad. La nueva medida no paramétrica de diversidad de Community Diversity Measures and Calculations establece que la diversidad está inversamente relacionada con la probabilidad de que dos individuos elegidos al azar pertenezcan a la misma especie.

Para una población infinita la fórmula del índice de Simpson viene dada por :

$$D = 1 - \sum_{i=1}^S p_i^2, \quad (4)$$

donde p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad.

Índice de diversidad de Simpson

Simpson en 1949 propuso un nuevo concepto de diversidad que combina dos ideas separadas, la riqueza de especies y la uniformidad. La nueva medida no paramétrica de diversidad de Community Diversity Measures and Calculations establece que la diversidad está inversamente relacionada con la probabilidad de que dos individuos elegidos al azar pertenezcan a la misma especie.

Para una población infinita la fórmula del índice de Simpson viene dada por :

$$D = 1 - \sum_{i=1}^S p_i^2, \quad (4)$$

donde p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad.

Índice de diversidad de Simpson

Simpson en 1949 propuso un nuevo concepto de diversidad que combina dos ideas separadas, la riqueza de especies y la uniformidad. La nueva medida no paramétrica de diversidad de Community Diversity Measures and Calculations establece que la diversidad está inversamente relacionada con la probabilidad de que dos individuos elegidos al azar pertenezcan a la misma especie.

Para una población infinita la fórmula del índice de Simpson viene dada por :

$$D = 1 - \sum_{i=1}^S p_i^2, \quad (4)$$

donde p_i es la proporción de individuos (o abundancia relativa) de la especie i en la comunidad.

Observaciones importantes :

- El índice de Simpson va de 0 (baja diversidad) a casi 1,
- A diferencia del índice de Shannon, el índice de diversidad de Simpson “da más importancia” a las especies más comunes,
- La hipótesis nula de uniformidad es que todas las especies en una comunidad hipotética son igualmente comunes. Sin embargo, la mayoría de las comunidades contienen unas pocas especies dominantes y muchas especies que son relativamente poco comunes,
- Las medidas de uniformidad intentan cuantificar esta representación desigual frente a la hipótesis nula.

La definición del índice de uniformidad de Simpson se obtiene del recíproco del índice de Simpson. El índice original de Simpson está dado por :

Observaciones importantes :

- El índice de Simpson va de 0 (baja diversidad) a casi 1,
- A diferencia del índice de Shannon, el índice de diversidad de Simpson “da más importancia” a las especies más comunes,
- La hipótesis nula de uniformidad es que todas las especies en una comunidad hipotética son igualmente comunes. Sin embargo, la mayoría de las comunidades contienen unas pocas especies dominantes y muchas especies que son relativamente poco comunes,
- Las medidas de uniformidad intentan cuantificar esta representación desigual frente a la hipótesis nula.

La definición del índice de uniformidad de Simpson se obtiene del recíproco del índice de Simpson. El índice original de Simpson está dado por :

Observaciones importantes :

- El índice de Simpson va de 0 (baja diversidad) a casi 1,
- A diferencia del índice de Shannon, el índice de diversidad de Simpson “da más importancia” a las especies más comunes,
- La hipótesis nula de uniformidad es que todas las especies en una comunidad hipotética son igualmente comunes. Sin embargo, la mayoría de las comunidades contienen unas pocas especies dominantes y muchas especies que son relativamente poco comunes,
- Las medidas de uniformidad intentan cuantificar esta representación desigual frente a la hipótesis nula.

La definición del índice de uniformidad de Simpson se obtiene del recíproco del índice de Simpson. El índice original de Simpson está dado por :

Observaciones importantes :

- El índice de Simpson va de 0 (baja diversidad) a casi 1,
- A diferencia del índice de Shannon, el índice de diversidad de Simpson “da más importancia” a las especies más comunes,
- La hipótesis nula de uniformidad es que todas las especies en una comunidad hipotética son igualmente comunes. Sin embargo, la mayoría de las comunidades contienen unas pocas especies dominantes y muchas especies que son relativamente poco comunes,
- Las medidas de uniformidad intentan cuantificar esta representación desigual frente a la hipótesis nula.

La definición del índice de uniformidad de Simpson se obtiene del recíproco del índice de Simpson. El índice original de Simpson está dado por :

Observaciones importantes :

- El índice de Simpson va de 0 (baja diversidad) a casi 1,
- A diferencia del índice de Shannon, el índice de diversidad de Simpson “da más importancia” a las especies más comunes,
- La hipótesis nula de uniformidad es que todas las especies en una comunidad hipotética son igualmente comunes. Sin embargo, la mayoría de las comunidades contienen unas pocas especies dominantes y muchas especies que son relativamente poco comunes,
- Las medidas de uniformidad intentan cuantificar esta representación desigual frente a la hipótesis nula.

La definición del índice de uniformidad de Simpson se obtiene del recíproco del índice de Simpson. El índice original de Simpson está dado por :

$$D_O = \sum_{i=1}^S p_i^2, \quad (5)$$

donde p_i es la proporción de la especie i en la comunidad, y el inverso del índice de Simpson está dado por $1/D_O$. El índice de uniformidad de Simpson se define como :

$$E = \frac{1}{S \sum_{i=1}^S p_i^2} \quad (6)$$

donde S es el número de especies en la muestra. Este índice varía de 0 a 1 y es relativamente poco afectado por las especies raras de la muestra.

$$D_O = \sum_{i=1}^S p_i^2, \quad (5)$$

donde p_i es la proporción de la especie i en la comunidad, y el inverso del índice de Simpson está dado por $1/D_O$. El índice de uniformidad de Simpson se define como :

$$E = \frac{1}{S \sum_{i=1}^S p_i^2} \quad (6)$$

donde S es el número de especies en la muestra. Este índice varía de 0 a 1 y es relativamente poco afectado por las especies raras de la muestra.