

# Análisis de diversidad

Autores

**Juan D. Vásquez-Restrepo**

*Department of Ecology & Evolutionary Biology, Princeton University*

## Tabla de contenidos

Introducción.....	1
Glosario y definiciones .....	2
1. Índices de diversidad tradicionales .....	3
2. El problema de los índices tradicionales .....	10
3. Índices de diversidad verdadera .....	11
4. Propiedad de duplicación de los números de Hill .....	14
5. Estimadores de riqueza .....	15
6. Curvas de acumulación .....	20
7. Curvas de rarefacción.....	24
8. Perfiles de diversidad .....	27
9. La diversidad a escala regional .....	28
10. Estructura de las comunidades .....	29
11. Aplicando la teoría usando R .....	34
Bibliografía.....	60

## Introducción

Este capítulo reúne los que, en mi experiencia, son los aspectos más relevantes en el análisis de la diversidad de comunidades biológicas. Se trata de una combinación entre fundamentos teóricos y mi trayectoria, tanto en investigación como en consultoría ambiental. Mi objetivo es brindarle al lector algunas herramientas básicas que le permitan calcular e interpretar datos biológicos de forma relativamente simple, como a mí me hubiese gustado que me lo explicaran. Espero que, en conjunto con las demás secciones de este libro, el lector logre obtener un panorama más amplio que le sirva para tomar decisiones mejor fundamentadas.

Buena parte del contenido fue originalmente publicada como un recurso en mi repositorio de GitHub (<https://github.com/VR-Daniel>), el cual ahora tengo la oportunidad de ampliar y

reorganizar en el contexto de este libro. Por motivos de extensión, exploraremos únicamente los índices más generales y ampliamente utilizados. Si el lector desea una lectura más profunda, puede consultar a Magurran (2004) para lo relacionado con índices y medidas de **diversidad taxonómica**, Tucker *et al.* (2017) para **diversidad filogenética**, Pla *et al.* (2012) para **diversidad funcional**, o Chao *et al.* (2014) para un marco unificado de las tres dimensiones basado en los números de Hill.

## Glosario y definiciones

A continuación, se incluye una lista de términos y definiciones que son de utilidad para facilitar la comprensión del presente capítulo:

Cobertura de muestra: proporción de los individuos de una comunidad que pertenecen a especies representadas en la muestra. Estima qué tan completa es una muestra respecto a la comunidad real.

Completitud del muestreo: porcentaje de especies esperadas que han sido detectadas en un muestreo, comparando la riqueza observada con la estimada.

Diversidad alfa ( $\alpha$ ): medida de la diversidad dentro de una sola comunidad o sitio. Puede expresarse como riqueza de especies o mediante otros índices que incorporan la abundancia.

Diversidad beta ( $\beta$ ): medida del recambio de especies entre comunidades o sitios. Representa las diferencias en la composición de especies entre unidades de muestreo.

Diversidad gamma ( $\gamma$ ): medida de la diversidad total a nivel regional o del conjunto de comunidades. Integra la diversidad alfa y beta.

Diversidad filogenética: dimensión de la biodiversidad que considera las relaciones evolutivas entre las especies.

Diversidad funcional: dimensión de la biodiversidad que incorpora las diferencias en los rasgos ecológicos o funciones de las especies.

Diversidad taxonómica: dimensión clásica de la biodiversidad basada en el número de taxones distintos (especies, géneros, familias).

Esfuerzo de muestreo: cantidad de unidades de muestreo o individuos recolectados en un estudio. Influye directamente en las estimaciones de diversidad.

Número Efectivo de Especies (ENS): medida de diversidad que representa el número de especies igualmente abundantes respecto a la comunidad real.

Orden de diversidad ( $q$ ): parámetro que determina la sensibilidad de un índice de diversidad a la abundancia de las especies. El orden 0 ignora las abundancias, el 1 favorece a las especies comunes y el 2 a las dominantes.

Rarefacción: método para estandarizar el esfuerzo de muestreo entre comunidades, estimando cuántas especies se esperarían encontrar si todas hubieran sido muestreadas con la misma intensidad.

Singleton /Doubleton: especies representadas por un solo individuo (*singleton*) o por dos individuos (*doubleton*) en una muestra. Son claves para estimar la riqueza no observada.

Tamaño de muestra: cantidad de unidades observadas en una muestra (p.e., individuos, trampas, parcelas).

## 1. Índices de diversidad tradicionales

Los índices de diversidad son una aproximación numérica que nos permite cuantificar y comparar la diversidad en comunidades biológicas, entendiendo como diversidad no solo el número de especies o taxones diferentes, sino también sus abundancias. Entre estos índices se encuentran aquellos que miden la diversidad a escalas locales o regionales, a lo que denominaremos **diversidad alfa** ( $\alpha$ ) y **gamma** ( $\gamma$ ), respectivamente, y los índices de diversidad **beta** ( $\beta$ ), relacionados con las diferencias o el recambio de especies entre comunidades Whittaker (1972).

En esta primera parte del capítulo abordaremos los conceptos teóricos detrás de los índices de diversidad alfa ( $\alpha$ ), tanto los tradicionales como los denominados diversidades verdaderas<sup>1</sup> o números de Hill. Vale la pena mencionar que los índices que veremos a continuación son índices de diversidad taxonómica (i.e., especies), la cual constituye solo una de las múltiples dimensiones de la biodiversidad. Por ejemplo, también existen índices de diversidad genética, diversidad filogenética o diversidad funcional, que igualmente pueden medirse a diferentes escalas, aunque no serán objeto del presente texto.

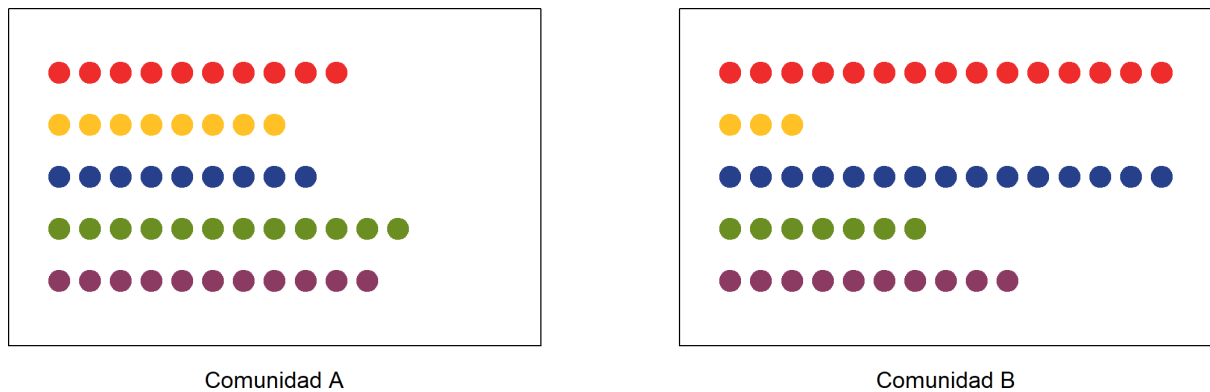
### 1.1. Riqueza de especies

El índice más simple de diversidad alfa ( $\alpha$ ) que podemos encontrar es la riqueza de especies ( $S$ ), es decir, el número de especies diferentes presentes en un sitio o una comunidad. Es importante tener presente que este valor no considera la abundancia, por lo que solo aporta información sobre la composición y no sobre la distribución (numérica, no geográfica). Además, la riqueza está influenciada por el **esfuerzo de muestreo**, de modo que menores

---

<sup>1</sup> En el término «diversidad verdadera», la palabra «verdadera» hace referencia a las propiedades matemáticas del índice, más que a su capacidad para representar la verdadera diversidad de las comunidades.

esfuerzos suelen capturar un menor número de especies, sin que esto sea un indicativo real de la diversidad presente en un sitio.



**Figura 1.** Diversidad en dos comunidades hipotéticas. Ambas tienen igual riqueza (5 especies) e igual tamaño (50 individuos), pero la Comunidad A presenta menor dominancia y, por tanto, mayor diversidad que la Comunidad B.

Recordemos que una comunidad es más diversa a medida que acumula más especies diferentes y estas, a su vez, tienden a estar representadas por un número similar de individuos. En otras palabras, son más equitativas y menos dominadas (Fig. 1). Sin embargo, es cierto que rara vez vamos a encontrar comunidades donde todas las especies estén igualmente representadas, ya sea porque las metodologías de muestreo tienen sesgos implícitos que no nos permiten detectar el 100% de los individuos, por más rigurosas que sean, o porque la propia biología de las especies, sus interacciones con otras y las condiciones ambientales influyen de manera diferencial en sus abundancias.

## 1.2. Índices de Margalef y Menhinick

Una opción para considerar la abundancia de las especies dentro de un indicador de riqueza es utilizar los índices de Margalef (1958) (Ec. 1) o Menhinick (1964) (Ec. 2), los cuales relacionan el número de especies con el total de individuos en la muestra analizada. Si bien estos índices sí incluyen un valor de abundancia, tampoco son informativos respecto a cómo esta está distribuida, dado que su intención es principalmente compensar el efecto que tiene el esfuerzo de muestreo sobre la riqueza.

$$Margalef = \frac{S - 1}{\ln(N)} \quad Ec. 1$$

$$Menhinick = \frac{S}{\sqrt{N}} \quad Ec. 2$$

donde  $S$  es la riqueza de especies y  $N$  el número total de individuos

La interpretación de ambos índices es directa: a mayor valor, mayor es la diversidad. No obstante, a pesar de que el índice de Margalef y el de Menhinick ayudan a reducir la inflación de los valores de diversidad, siguen siendo altamente sensibles al número de individuos, por lo que deben ser utilizados con precaución, ya que pueden llevarnos a conclusiones erradas cuando comparamos comunidades con diferentes esfuerzos de muestreo. Si bien son índices que se usan con menos frecuencia debido a sus limitaciones, he decidido incluirlos por razones históricas, además de que presentan ecuaciones simples que nos permitirán familiarizarnos con algunos términos y cómo estos se relacionan.

### 1.3. Índices de dominancia y equitatividad

Como se mencionó anteriormente, cuando hablamos de diversidad debemos considerar no solo el número de especies diferentes, sino también cómo están distribuidas sus abundancias. Para esto, existen algunos índices que incorporan de manera directa las abundancias relativas de las especies, como los índices de Simpson, Shannon o Pielou. Estos, si bien no son los únicos, están entre los más ampliamente utilizados, y entender su trasfondo teórico nos será útil para, posteriormente, abordar algunas de sus transformaciones. En cualquier caso, a la hora de aplicar estos índices no podemos ignorar la importancia del tamaño de muestra, ya que es bien sabido que las inferencias basadas en muestras pequeñas tienen mayor sesgo (Smith & Grassle 1977).

#### 1.3.1. Índice de Simpson

El índice de Simpson (Ec. 3), también conocido como índice de dominancia de Simpson (Simpson 1949), representa la probabilidad de que dos individuos tomados al azar de una muestra pertenezcan a la misma entidad o especie. Por lo tanto, es una medida de dominancia o desigualdad. Este índice es, quizás, uno de los que más confusión genera, no porque su cálculo o interpretación sean especialmente difíciles, sino porque existen diversas

transformaciones que reciben distintos nombres y que se aplican e interpretan de manera diferente.

$$\lambda = \sum_{i=1}^S p_i^2 \quad \text{Ec. 3}$$

*donde  $p_i$  es la abundancia relativa de cada especie*

Es importante mencionar que la probabilidad calculada con el índice de Simpson asume muestreo con reemplazo; es decir, una vez que se selecciona el primer individuo, este se reemplaza inmediatamente antes de seleccionar el segundo. Por otro lado, si se desea calcular la probabilidad sin reemplazo, se utiliza una variante conocida como índice de Simpson corregido (Ec. 4), el cual deriva de una modificación del índice original (Hunter & Gaston 1988).

$$D = \frac{\sum n_i(n_i - 1)}{N(N - 1)} \quad \text{Ec. 4}$$

*donde  $n_i$  es la abundancia de cada especie y  $N$  el total de individuos*

Ya sea que se utilice el índice de Simpson original o su versión corregida, ambos expresan la dominancia dentro de una comunidad, entendida como lo opuesto a la diversidad. En estos casos, la interpretación puede resultar contraintuitiva, ya que valores altos del índice indican baja diversidad. Por esta razón, se suele usar una variante conocida como índice de Gini-Simpson (Ec. 5) o índice de diversidad de Simpson, que se expresa como el complemento de la dominancia.

$$Gini = 1 - D \quad \text{Ec. 5}$$

*donde  $D$  corresponde a la dominancia de Simpson (original o corregida)*

Dado que la dominancia es una proporción que varía entre cero y uno, su complemento representa la probabilidad de que dos individuos seleccionados al azar no pertenezcan a la misma entidad, es decir, que sean diferentes. Aunque esto puede parecer trivial, este tipo de

detalles suele generar confusión, especialmente cuando se utilizan programas informáticos para calcular los índices sin saber con exactitud qué forma del índice está siendo empleada. Por ello, siempre es fundamental leer la documentación del software antes de interpretar los resultados. Por ejemplo, el paquete ‘vegan’ (Oksanen *et al.* 2024) de R calcula el índice de dominancia no corregido, pero regresa el resultado como índice de Gini; por su parte, ‘PAST’ (Hammer *et al.* 2001) devuelve tanto el índice de dominancia como el de diversidad, permitiendo elegir si se quiere la versión corregida o no corregida; mientras que ‘EstimateS’ (Colwell 2019), calcula el promedio del inverso del índice de dominancia en un número determinado de aleatorizaciones con o sin reemplazo.

### 1.3.2. Índice de Shannon

El índice de Shannon (Ec. 6), o índice de entropía de Shannon (Shannon 1948), fue originalmente propuesto como una medida de la entropía en cadenas de texto, en el marco de la teoría de la información. Posteriormente, fue adoptado en ecología debido a su aplicabilidad más allá del contexto para el cual fue desarrollado originalmente. Uno de los principales desafíos de este índice radica en su interpretación, ya que su unidad de medida son bits por especie<sup>2</sup>, lo que puede resultar poco intuitivo. No obstante, en la práctica, se asume que valores más altos del índice indican mayor diversidad dentro de una comunidad.

$$H' = - \sum_{i=1}^S p_i \cdot \log(p_i) \quad \text{Ec. 6}$$

donde  $p_i$  es la abundancia relativa de cada especie

La notación original de este índice utiliza el  $\log_2$  (logaritmo en base 2), pero también es común verlo expresado como  $\ln$  (logaritmo natural), lo cual no tiene efecto alguno en términos de la interpretación de los resultados, pero que sí debemos considerar a la hora de

---

<sup>2</sup> Un bit es una unidad ampliamente utilizada para cuantificar entropía en el contexto de la teoría de la información. La entropía representa el nivel promedio de incertidumbre asociado a los posibles resultados de un evento, o bien, la cantidad de información que se obtiene al conocer el resultado. Por ejemplo, al lanzar una moneda, la entropía es de 1 bit, ya que hay dos resultados posibles (cara o sello), y conocer uno nos informa completamente sobre cuál no ocurrió. Sin embargo, cuando se intenta aplicar esta unidad en contextos biológicos, como en estudios de diversidad ecológica, la interpretación de los bits se vuelve mucho menos intuitiva. A pesar de ello, se acepta que, a mayor entropía, mayor diversidad, ya que un sistema con más incertidumbre respecto a la identidad de un individuo tomado al azar (i.e., cuán difícil es predecir a qué especie pertenece un individuo) refleja una comunidad más equitativa y rica en especies (Rajaram *et al.* 2017).

comparar valores si estos están calculados con bases diferentes. Adicionalmente, el índice de Shannon se fundamenta en el supuesto de que todas las especies están representadas en la muestra y que fueron muestreadas aleatoriamente. Por lo tanto, debemos tener presente que este índice tiene un sesgo implícito, no solo cuando los esfuerzos de muestreo no son suficientes para capturar la totalidad de especies, sino también porque muchas metodologías no son completamente aleatorias y favorecen la detectabilidad de algunos individuos. El índice de Shannon no tiene un límite superior; puede tomar un valor mínimo de cero cuando solo hay una especie, o un valor máximo igual al logaritmo de la riqueza.

### 1.3.3. Índice de Brillouin

El índice de Brillouin (Ec. 7), al igual que el índice de Shannon, es una medida de entropía que los ecólogos tomamos prestada para el estudio de comunidades biológicas (Margalef 1958). En cierta medida, puede considerarse su contraparte, ya que, a diferencia del índice de Shannon, que se utiliza para muestras asumidas como aleatorias, el índice de Brillouin se recomienda para censos completos o, al menos, para muestreos exhaustivos en los que no se puede garantizar la aleatoriedad (Pielou 1975). Esto ocurre, por ejemplo, cuando se emplean métodos de muestreo con detectabilidad diferencial (e.g., trampas de luz, trampas de caída, redes de niebla). Sin embargo, tanto el índice de Shannon como el de Brillouin suelen dar valores similares o correlacionados.

$$\hat{H} = \frac{\log(N!) - \sum \log(n_i!)}{N} \quad \text{Ec. 7}$$

*donde N es el total de individuos en la colección y n<sub>i</sub> el número de individuos de cada especie*

Al igual que en el índice de Shannon, la base logarítmica utilizada en el cálculo del índice de Brillouin puede ser cualquiera, siempre que se mantenga consistente. Otro aspecto importante a considerar es su sensibilidad a la abundancia de las especies raras (al igual que Shannon), porque pequeños cambios en su abundancia pueden tener un gran impacto en el valor del índice, contrario a lo que pasa, por ejemplo, con el índice de Simpson, que es más sensible a cambios en la abundancia de las especies dominantes (Peet 1974).



#### 1.3.4. Índices de equitatividad

Finalmente, tenemos los denominados índices de equitatividad, los cuales representan una razón entre la diversidad observada (calculada a partir de alguno de los índices presentados anteriormente) y la diversidad máxima posible para la comunidad (Ec. 8). Estas transformaciones facilitan la interpretación de los índices, ya que adquieren un valor de cero cuando hay una única especie y de uno cuando todas las especies tienen la misma abundancia. Por lo tanto, nos proporcionan información sobre qué tan equitativas son las comunidades.

$$\text{Equitatividad} = \frac{\text{Diversidad}}{\text{Diversidad}_{\max}} \quad \text{Ec. 8}$$

Para calcular la equitatividad basada en el índice de Simpson (Ec. 9) comúnmente se utiliza su recíproco, que como veremos más adelante, corresponde a la diversidad verdadera de orden dos. En ese contexto, su máximo valor posible es igual al número de especies observadas.

$$E_{\text{Simpson}} = \frac{1/D}{S} \quad \text{Ec. 9}$$

*donde  $1/D$  es el índice recíproco de Simpson y  $S$  el número de especies*

Para el índice de Shannon, su equitatividad se conoce como índice de Pielou (Ec. 10), y de los tres índices presentados en esta sección, es el más ampliamente usado. Aquí, su valor máximo se calcula como el logaritmo de la riqueza. Para el cálculo de este índice, también es importante asegurarse de que la base del logaritmo utilizada en el índice de Shannon sea la misma que se utiliza en el denominador, ya que de lo contrario los resultados pueden ser inconsistentes.

$$J' = \frac{H'}{\log S} \quad \text{Ec. 10}$$

*donde  $H'$  es el índice de Shannon y  $S$  el número de especies*

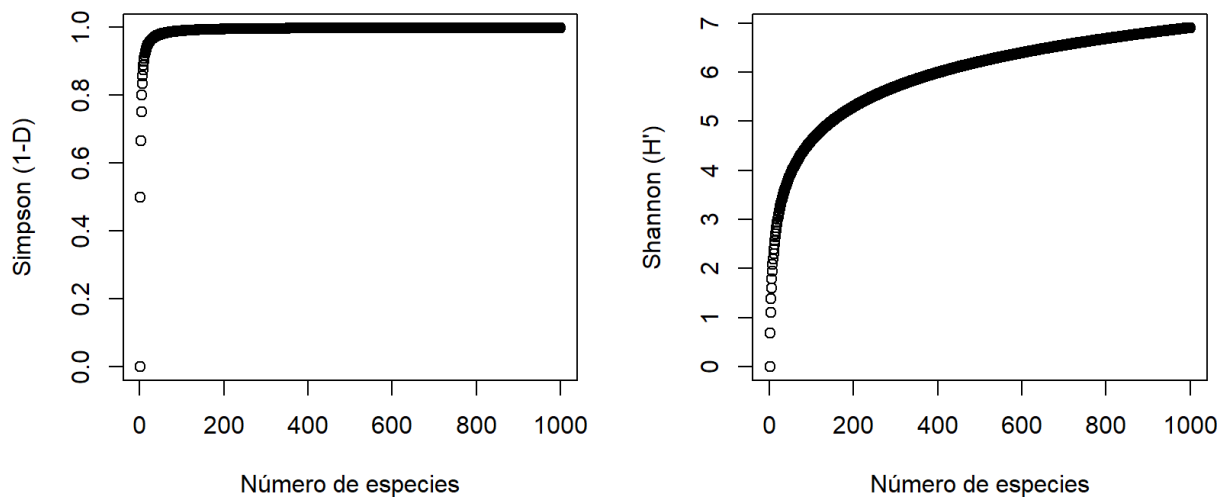
Finalmente, podemos aplicar la misma fórmula para calcular la equitatividad usando el índice de Brillouin (Ec. 11), aunque en este caso, la diversidad máxima es un poco más compleja de calcular.

$$E_{Brillouin} = \frac{\hat{H}}{\frac{1}{N} \log \frac{N!}{\left\{ \left[ \frac{N}{S} \right]! \right\}^{S-r} \cdot \left\{ \left( \left[ \frac{N}{S} \right] + 1 \right)! \right\}^r}} \quad Ec. 11$$

donde  $\hat{H}$  es el índice de Brillouin,  $N$  es el número de individuos en la comunidad,  $S$  el número de especies observadas,  $[N/S]$  la parte entera de división de  $N/S$ , y  $r$  el residuo de la división  $N/S$

## 2. El problema de los índices tradicionales

Como ya hemos visto, los diferentes índices utilizados tradicionalmente para cuantificar y comparar la diversidad biológica se fundamentan en el uso de la riqueza y la abundancia. No obstante, algunos autores han señalado ciertos problemas fundamentales subyacentes a estos (Jost 2006; Jost 2010; Daly *et al.* 2018), problemas que generalmente son subestimados, pero que tienen enormes implicaciones al momento de realizar inferencias y extraer conclusiones. Por ejemplo, además de que su interpretación puede resultar difícil debido a que emplean unidades diferentes o poco intuitivas (número de especies vs. bits por especie vs. probabilidades), muchos de estos índices no varían de forma lineal, una característica que a partir de ahora llamaremos «propiedad de duplicación». Tal como lo plantea Jost (2006), podemos intuir que una comunidad con 16 especies igualmente abundantes debería ser el doble de diversa que una con 8 especies igualmente abundantes. Sin embargo, esto no ocurre, ya que la mayoría de los índices de diversidad que utilizamos no responden proporcionalmente a los cambios en la riqueza. Para ilustrarlo, podemos calcular algunos índices simulando un número arbitrario de sitios o comunidades, en los que la riqueza aumente gradualmente de uno en uno, y donde todas las especies estén igualmente representadas (Fig. 2).



**Figura 2.** Relación entre los índices de Simpson y Shannon respecto al número de especies.

Para ambos índices, el incremento se da de manera logarítmica antes de alcanzar una asíntota, con grandes cambios en valores bajos de riqueza. Si en este ejemplo, donde tenemos una comunidad con 1000 especies imagináramos una catástrofe que eliminara el 80% de ellas, veríamos que la variación en el índice de Simpson es apenas perceptible, lo que podría llevarnos a concluir erróneamente que, a pesar de la pérdida de la mayor parte de la diversidad, la comunidad no sufrió grandes cambios. En el caso del índice de Shannon ocurre algo similar, aunque el cambio es un poco más marcado. Es por esto que el uso de estos índices ha sido ampliamente cuestionado, pues no solo no son comparables entre sí, sino que además pueden llevarnos a conclusiones erradas cuando queremos hacer inferencias basadas en extrapolaciones o interpolaciones. Más adelante, cuando abordemos el tema de las diversidades verdaderas, volveremos sobre esta relación y veremos cómo cambia.

### 3. Índices de diversidad verdadera

Los índices de diversidad verdadera o números de Hill no son más que una transformación de los índices de diversidad tradicionales, pero con ciertas ventajas interpretativas y metodológicas. Aunque la base matemática de estos índices se conoce desde la década de 1970 (Hill 1973), no tuvieron gran acogida hasta que Lou Jost y Anne Chao (Jost 2006; Chao *et al.* 2010; Jost 2010; Chao & Jost 2012; Chao *et al.* 2014) los retomaron como parte de un marco teórico unificado para el análisis de la diversidad en comunidades biológicas.

Los números de Hill permiten expresar los índices tradicionales en una misma unidad (número efectivo de especies), lo que facilita su interpretación. Además, cumplen con la propiedad de duplicación, lo que los hace comparables y extrapolables de forma más coherente. El **número efectivo de especies** (ENS, por sus siglas en inglés) es una forma de expresar la diversidad como si todas las especies fueran igual de abundantes. En otras palabras, es el número de especies que tendría una comunidad completamente equitativa para tener el mismo valor de diversidad que la comunidad real (Jost 2006).

Es importante tener presente que el número efectivo de especies varía en función de la distribución de las abundancias, por lo que tiende a ser menor a medida que aumenta la dominancia. Por ejemplo, imaginemos dos comunidades hipotéticas con ocho especies cada una, pero con distribuciones de abundancia desiguales (Tabla 1). Ambas tienen el mismo número total de individuos, pero la primera presenta menor dominancia. En este caso, aunque ambas tienen la misma riqueza, la comunidad con distribución más equitativa tendrá un mayor número efectivo de especies de orden uno (en breve aprenderemos qué son los órdenes y cómo calcularlos).

**Tabla 1.** Comunidades hipotéticas para el cálculo del número efectivo de especies de orden uno.

	A	B	C	D	E	F	G	H	ENS
Comunidad 1	40	35	25	60	12	20	45	25	7.25
Comunidad 2	88	12	5	32	15	2	46	62	5.41

El cálculo de la diversidad verdadera utiliza una fórmula generalizada (Ec. 12) para los diferentes índices, conocidos en este contexto como órdenes o números de Hill. En términos operacionales, es posible calcular  ${}^qD$  para cualquier valor de  $q$  (orden), ya sea un número entero o decimal, positivo o negativo. No obstante, los únicos órdenes con interpretación biológica clara son cero, uno y dos.

$${}^qD = \left( \sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad \text{Ec. 12}$$

donde  $q$  es el orden de diversidad que nos interesa y  $p_i$  es la abundancia relativa de cada especie

Para el caso de  $q = 0$ , el término  $p_i^0$  se hace uno, ya que, por las propiedades de las potencias, todo número distinto de cero elevado a cero es igual a uno. Asimismo, al resolver la operación del exponente general, este también se hace uno, de modo que la expresión es equivalente a sumar uno tantas veces como el número de especies. Es por esta razón que el número de Hill de orden cero es equivalente a la riqueza de especies (Ec. 13).

$${}^0D = \left( \sum_{i=1}^S p_i^0 \right)^{1/(1-0)} = \sum_{i=1}^S 1 = S \quad \text{Ec. 13}$$

Ahora, el primer orden es un caso especial, porque cuando  $q = 1$ , el exponente general se vuelve indeterminado (división por cero), y por lo tanto no es posible resolver directamente la ecuación general. Sin embargo, su límite sí existe, por lo que este valor se calcula aproximando  $q$  a un número muy cercano a uno ( $q = 0.999999$ ). Esta aproximación no afecta significativamente el valor del índice, ya que, tras simplificar la expresión, se obtiene que el orden uno es equivalente al exponencial del índice de Shannon (Ec. 14), el cual puede calcularse perfectamente sin necesidad de recurrir a la teoría de límites.

$${}^1D = \left( \sum_{i=1}^S p_i^1 \right)^{1/(1-1)} = \exp \left( - \sum_{i=1}^S p_i \cdot \log(p_i) \right) = \exp(H') \quad \text{Ec. 14}$$

Finalmente, para  $q = 2$ , el exponente general se convierte en uno negativo. Si acudimos nuevamente a las propiedades de las potencias, sabemos que un número elevado a una potencia negativa es equivalente a su recíproco elevado al opuesto de la potencia, por lo que podemos simplificar la expresión como su inverso. De este modo, se obtiene una fórmula equivalente a calcular el inverso del índice de Simpson (Ec. 15).

$${}^2D = \left( \sum_{i=1}^S p_i^2 \right)^{1/(1-2)} = \frac{1}{(\sum_{i=1}^S p_i^2)} = \frac{1}{\lambda} \quad \text{Ec. 15}$$

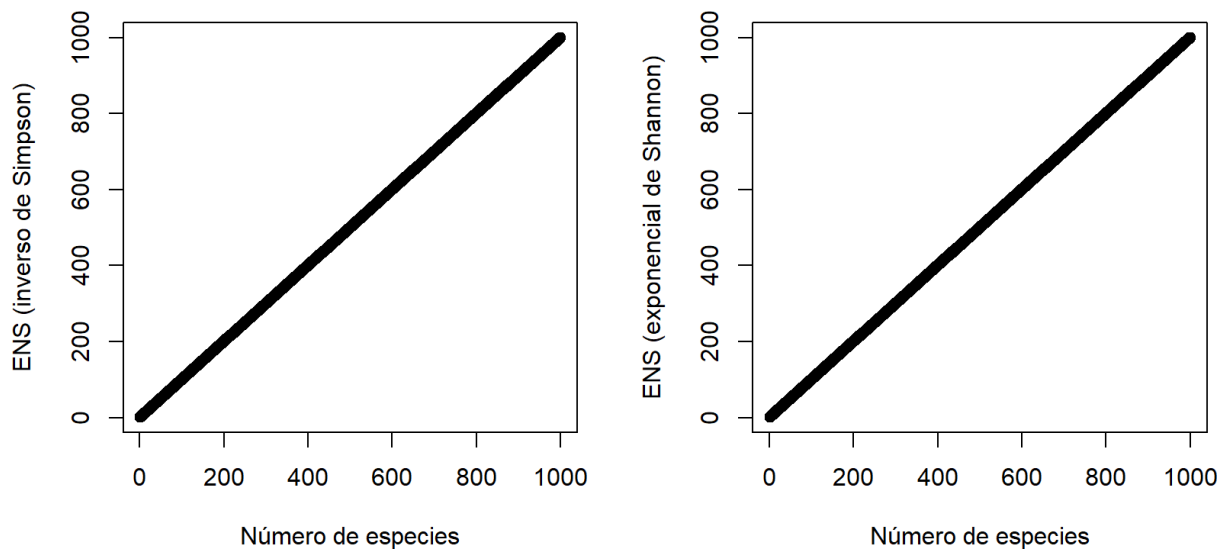
Dado que los índices de diversidad verdadera estandarizan las unidades en términos de especies, su interpretación es mucho más simple. Por ejemplo, ya sabemos que el orden cero es equivalente a la riqueza, es decir, representa el número total de especies presentes en la muestra. El orden uno, por su parte, puede interpretarse como el número de especies

comunes, dándonos también información indirecta sobre la equitatividad, en la medida en que su valor se acerca al del orden cero. Además, la diferencia entre  ${}^0D$  y  ${}^1D$  nos indica el número de especies raras. Por otro lado, el orden dos se puede interpretar como el número de especies dominantes, de modo que, mientras más especies dominantes haya, menor será la dominancia de la comunidad. Esta idea puede parecer contraintuitiva, pero tiene sentido si consideramos que, cuando todas las especies dominan por igual, es porque la comunidad es completamente equitativa. Finalmente, la razón entre  ${}^2D$  y  ${}^0D$  nos proporciona una proporción de dominancia o factor de inequitatividad (Jost 2010).

También es importante mencionar que los órdenes de diversidad indican el grado de sensibilidad del índice a las especies raras o comunes (Jost 2006; Jost 2010), ya que el cálculo de las diversidades verdaderas da un peso diferencial a las especies según sus abundancias relativas. En el caso de la diversidad de orden cero, esta es completamente insensible a las abundancias, ya que considera únicamente la presencia de especies, sin importar cuántos individuos haya de cada una. En cambio, las diversidades de orden menor que uno (pero distinto de cero) favorecen a las especies raras, mientras que las de orden mayor que uno otorgan más peso a las especies comunes. Por lo tanto, el único orden que pondera a las especies según su frecuencia, sin favorecer ni a las raras ni a las comunes, es la diversidad de orden uno.

#### **4. Propiedad de duplicación de los números de Hill**

Si bien Jost (2006, 2010) no definió de manera explícita algo denominado como propiedad de duplicación, sí hizo referencia directa a la relación lineal y monótonamente creciente que resulta de transformar los índices de diversidad tradicionales en índices de diversidad verdadera. Sin embargo, este término aparece posteriormente como un axioma de los índices de diversidad en el trabajo de Daly *et al.* (2018). Ahora que contamos con las bases conceptuales del número efectivo de especies, podemos graficar nuevamente la relación entre los índices de Simpson y Shannon (transformados a los números de Hill) y la riqueza a medida que esta aumenta. Para ello, vamos a simular nuevamente 1000 comunidades cuya riqueza incremente de uno en uno (Fig. 3).



**Figura 3.** Relación entre el número efectivo de especies (ENS) y la riqueza para los números de Hill de orden uno y dos.

Aquí podemos ver claramente cómo la relación que anteriormente no era lineal (Fig. 2), ahora sí lo es, lo que nos permite extrapolar de manera mucho más simple. Por ejemplo, si el número efectivo de especies de orden cero para una comunidad es 12, podemos concluir que otra con un número efectivo de especies de 18 es 1.5 veces (o un 50% más) diversa, o que una con 24 es el doble de diversa, y así sucesivamente. De forma análoga, para los órdenes uno y dos podemos interpretar cuántas veces una comunidad tiene más o menos especies comunes, raras o dominantes, así como cuántas veces es más o menos equitativa.

## 5. Estimadores de riqueza

Los estimadores de riqueza, como su nombre lo indica, son métodos que nos permiten hacer aproximaciones teóricas sobre la riqueza real esperada en un sitio o comunidad. Se entiende como riqueza esperada la suma de la riqueza observada más la riqueza no observada. Son una herramienta mediante la cual podemos evaluar el esfuerzo de muestreo.

### 5.1. Estimadores paramétricos

Los estimadores paramétricos son, en principio, funciones o distribuciones matemáticas conocidas que se pueden aplicar a nuestros datos con el objetivo de modelarlos teóricamente. Por el momento, no vamos a detallar la formulación matemática de estos estimadores, ya que

son los menos utilizados debido a que dependen de modelos que, por lo general, no se ajustan a datos reales (asumen la distribución de las abundancias).

## 5.2. Estimadores no paramétricos

A diferencia de los anteriores, los estimadores no paramétricos no asumen a priori ningún modelo conocido, en otras palabras, no ajustan una función existente a nuestros datos. A continuación, veremos solo dos de los varios tipos de estimadores no paramétricos que existen, dado que son los más ampliamente utilizados y funcionan relativamente bien para la mayoría de conjuntos de datos.

### 5.2.1. Chao

Los estimadores de Chao (1984, 1987) son estimadores de riqueza que se basan en el número de especies observadas y el número de especies raras o infrecuentes para calcular la cantidad de especies no observadas. Entre estos se encuentran el Chao1 (Ec. 16), que se utiliza cuando se dispone de datos de abundancia, y el Chao2 (Ec. 17), que se aplica a datos de presencia-ausencia (también llamados datos de incidencia).

$$Chao1 = S_{obs} + \frac{f_1^2}{2f_2} \quad Ec. 16$$

$$Chao2 = S_{obs} + \frac{Q_1^2}{2Q_2} \quad Ec. 17$$

donde  $S_{obs}$  es la riqueza observada y  $f_1/Q_1$  y  $f_2/Q_2$  el número singletons (especies con abundancia de uno o una única ocurrencia, respectivamente) y doubletons (especies con abundancias de dos o dos únicas ocurrencias, respectivamente)

Tanto Chao1 (Ec. 18) como Chao2 (Ec. 19) tienen una versión corregida (Chao 2005), la cual ayuda a reducir el sesgo de la estimación en conjuntos de datos pequeños y permite calcular el estimador cuando los *doubletons* son cero (lo que hace que la versión no corregida sea indeterminada).

$$Chao1 = S_{obs} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)} \quad Ec. 18$$



$$Chao2 = S_{obs} + \left(\frac{m-1}{m}\right) \left(\frac{Q_1(Q_1-1)}{2(Q_2+1)}\right) \quad Ec. 19$$

donde  $S_{obs}$  es la riqueza observada y  $f_1/Q_1$  y  $f_2/Q_2$  el número singletons (especies con abundancia de uno o una única ocurrencia, respectivamente) y doubletons (especies con abundancias de dos o dos únicas ocurrencias, respectivamente)

Estos estimadores suelen desempeñarse bien con la mayoría de los conjuntos de datos, pero es importante tener presente, al momento de interpretarlos, que Chao calcula riquezas mínimas, es decir, el límite inferior del número de especies que esperaríamos encontrar.

### 5.2.2. Jackknife y Bootstrap

Dentro de la familia de los estimadores no paramétricos, existen dos que se basan en un método estadístico de submuestreo: el Jackknife (Ec. 20 y Ec. 21) y el Bootstrap (Ec. 22). De manera similar a los estimadores de Chao, estos índices buscan corregir el sesgo de subestimación de la riqueza de especies a partir de los datos observados, ajustando el valor hacia arriba para compensar las especies no detectadas. Su principal diferencia radica en que, mientras el Jackknife utiliza la frecuencia de aparición de especies en subconjuntos de la muestra para estimar las no detectadas, el Bootstrap se basa en el remuestreo con reemplazo del conjunto observado para estimar la probabilidad de que algunas especies hayan pasado desapercibidas.

$$Jack1 = S_{obs} + Q_1 \left(\frac{m-1}{m}\right) \quad Ec. 20$$

$$Jack2 = S_{obs} + \left[ \frac{Q_1(2m-3)}{m} - \frac{Q_2(m-2)^2}{m(m-1)} \right] \quad Ec. 21$$

donde  $Q_1/Q_2$  es el número de especies que aparecen en una o dos unidades de muestreo (independientemente de su abundancia), y  $m$  es el número de unidades de muestreo (e.g., sitios, parcelas, trampas)

El estimador Jackknife tiene dos versiones, denominadas de primer y segundo orden. La diferencia entre ellas radica en que Jack1 considera como raras las especies que aparecen en

una única unidad de muestreo, mientras que Jack2 incluye también aquellas que aparecen en una o en dos unidades únicamente.

$$\text{Boot} = S_{obs} + \sum_{k=1}^{S_{obs}} (1 - p_k)^m \quad \text{Ec. 22}$$

donde  $S_{obs}$  es el número de especies,  $p_k$  la proporción de individuos que pertenecen a la especie  $k$ , y  $m$  el número de unidades de muestreo (e.g., sitios, parcelas, trampas)

Ahora, respecto a los estimadores Chao, el enfoque de Jackknife y Bootstrap se basa en la presencia-ausencia por unidad de muestreo, y no en el número de individuos. Esto reduce su sensibilidad a las abundancias, pero los hace dependientes de la cantidad de réplicas disponibles.

### 5.2.3. ACE e ICE

Los estimadores ACE (*Abundance-based Coverage Estimate*) e ICE (*Incidence-based Coverage Estimate*) (Colwell & Coddington 1994) son una modificación de los estimadores basados en la **cobertura de muestra** propuestos por Chao y Lee (1992), pero que, al igual que los anteriores, buscan aproximar el número de especies esperadas a partir de los datos observados. De manera análoga a los estimadores de Chao, el ACE (Ec. 23) es la versión para datos de abundancia, y el ICE (Ec. 24), para datos de presencia-ausencia. Matemáticamente, el cálculo del ACE y el ICE es un poco más complejo, sin embargo, aunque existen diversos programas y paquetes que nos facilitan el trabajo, es importante saber qué es lo que está calculando.

$$ACE = S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{f_1}{C_{ACE}} \cdot \gamma_{ACE}^2 \quad \text{Ec. 23}$$

$$C_{ACE} = 1 - \frac{f_1}{N_{rare}} \quad (\text{Ec. 23a})$$

$$\gamma_{ACE}^2 = \max \left\{ \frac{S_{rare}}{C_{ACE}} \cdot \frac{\sum_{i=1}^{10} i(i-1)f_i}{(N_{rare})(N_{rare}-1)} - 1 \right\} \quad (\text{Ec. 23b})$$

donde  $S_{abund}$  es el número de especies abundantes (por defecto  $> 10$ ),  $S_{rare}$  el número de especies raras (por defecto  $< 10$ ),  $f_1$  el número de singletons,  $N_{rare}$  el número total de individuos en las

*especies raras, y  $f_i$  el número de especies que tienen exactamente  $i$  individuos cuando se agrupan todas las muestras*

$$ICE = S_{frec} + \frac{S_{infr}}{C_{ICE}} + \frac{Q_1}{C_{ICE}} \cdot \gamma_{ICE}^2 \quad \text{Ec. 24}$$

$$C_{ICE} = 1 - \frac{Q_1}{N_{infr}} \quad (\text{Ec. 24a})$$

$$\gamma_{ICE}^2 = \max \left\{ \frac{S_{infr}}{C_{ICE}} \cdot \frac{N_{infr}}{(N_{infr} - 1)} \cdot \frac{\sum_{j=1}^{10} j(j-1)Q_j}{(N_{infr})^2} - 1 \right\} \quad (\text{Ec. 24b})$$

*donde  $S_{frec}$  es el número de especies frecuentes (por defecto  $> 10$  muestras),  $S_{infr}$  el número de especies infrecuentes (por defecto  $< 10$  muestras),  $Q_1$  el número de singletons,  $N_{infr}$  el número total de ocurrencias de las especies raras, y  $Q_j$  el número de especies que ocurren exactamente en  $j$  muestras.*

Este grupo de estimadores asume como rareza o infrecuencia a las especies que tienen menos de 10 individuos u ocurrencias, por lo que su uso se recomienda en conjuntos de datos donde las especies suelen ser abundantes (aunque el valor por defecto se puede modificar). En este caso, los estimadores ACE e ICE también son estimadores de riquezas mínimas.

### 5.3. ¿Cómo elegir el estimador adecuado?

Cuando se trata de elegir qué estimador de riqueza utilizar, no existe una regla universal, cada métrica tiene sus requisitos, ventajas y limitaciones. La primera pregunta que debemos hacernos es qué tipo de datos tenemos, si son datos de abundancias o de presencia-ausencia. Esta distinción es clave, ya que, si utilizamos un estimador basado en incidencia sobre datos de abundancia, o viceversa, estaremos asumiendo implícitamente que todas las especies tienen la misma frecuencia (es decir, abundancia igual a uno), lo cual distorsiona la estimación.

También es necesario ser realistas respecto a nuestros datos, ya que, si el muestreo no abarca una escala temporal amplia, los datos de abundancia tienden a estar más sesgados que los de incidencia. Esto se debe a que, en los conjuntos de datos basados en incidencia, basta con detectar una especie una sola vez para que quede representada, mientras que en los de abundancia, la frecuencia registrada puede estar afectada por la metodología de muestreo, la estacionalidad, o por factores externos o intrínsecos a las especies que influyen en sus

abundancias o detectabilidades durante el periodo de muestreo. Por otro lado, el esfuerzo de muestreo es fundamental. Un muestreo insuficiente o poco representativo puede afectar gravemente la calidad de los datos y, por ende, la precisión de cualquier estimación, sin importar qué estimador se utilice.

Además, los estimadores son sensibles a la escala espacial, es decir, al tamaño de las unidades de muestreo utilizadas. En general, al aumentar el tamaño del «grano» (p.e., de trampas a transectos o parcelas), los valores estimados de riqueza también aumentan. Sin embargo, no todos los estimadores responden igual ante esta variación, estimadores basados en abundancia como Chao1 y ACE, tienden a ser robustos frente a cambios en la escala, mientras que otros basados en incidencia, como Chao2 o ICE, son más sensibles a la heterogeneidad entre unidades de muestreo, especialmente cuando estas son grandes o no comparables. En este contexto, los estimadores Jackknife 1 y 2, aunque también basados en incidencia, han mostrado un desempeño más robusto que ICE o Chao2 en unidades de muestreo amplias o no homogéneas (Tabla 2).

**Tabla 2.** Estimadores de riqueza con el mejor desempeño según el tipo de datos y escala. Adaptado de Hortal et al. (2006).

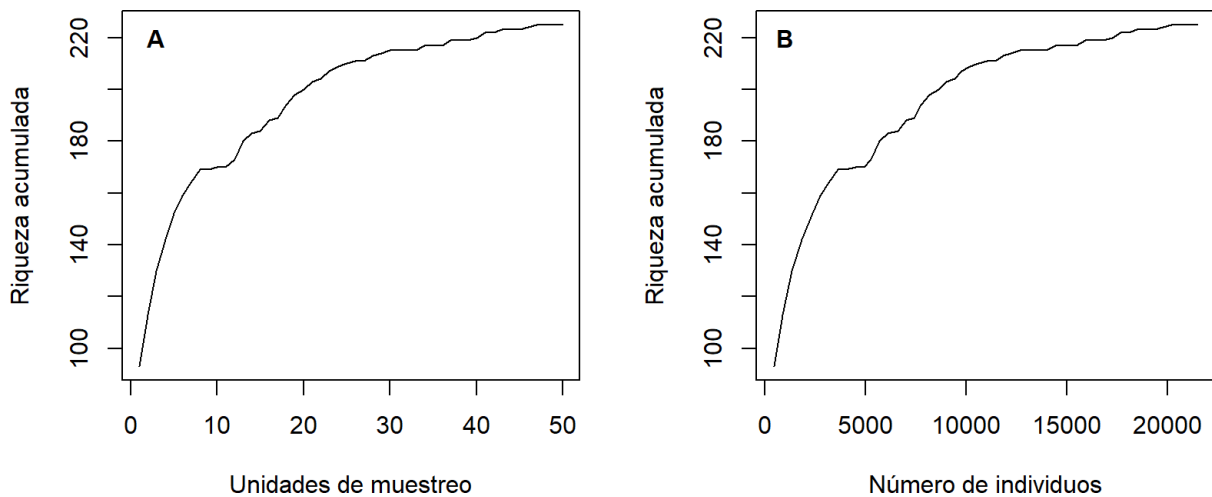
	<b>Abundancia</b>	<b>Incidencia</b>
Comparable y/o grano fino	ICE, Jackknife, Chao2	ICE, Jackknife, Chao2
No comparable y/o grano grueso	ACE o Chao1	Jackknife

Es importante notar que en esta tabla no se incluye el estimador Bootstrap, ya que, si bien es preciso (i.e., poca variación entre escalas), también tiende a subestimar consistentemente la riqueza de especies, en ocasiones, por debajo del número observado.

## 6. Curvas de acumulación

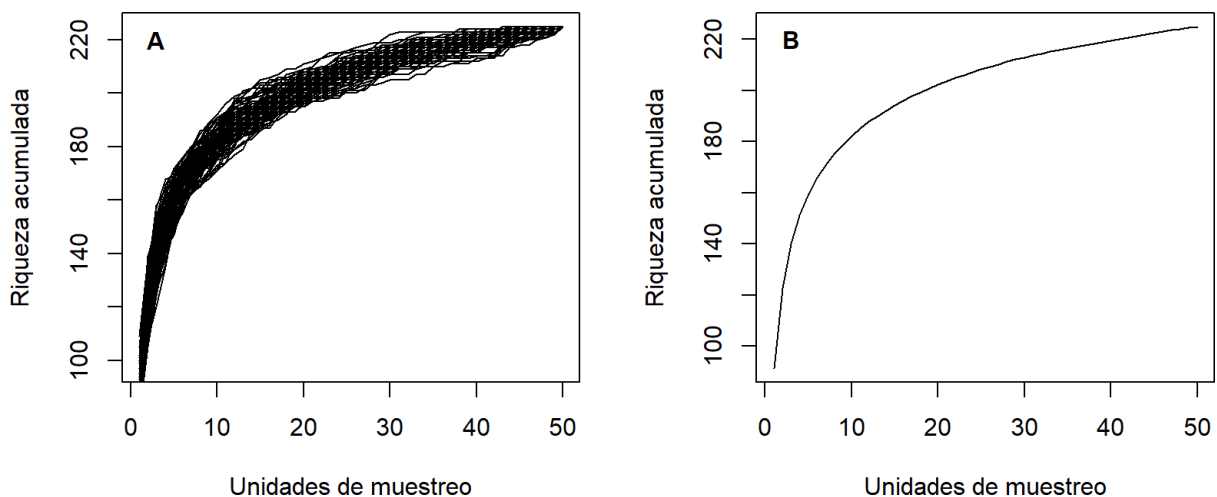
La curva de acumulación de especies es un método que nos permite visualizar gráficamente cómo varía la riqueza a medida que aumenta el esfuerzo de muestreo. El principal uso de las curvas de acumulación es evaluar qué tan completo es un muestreo, a partir de la riqueza que un determinado esfuerzo logra capturar. La forma más simple de construir una curva de acumulación es utilizando el número acumulado de especies por unidad de muestreo (Fig. 4A) o el número acumulado de especies por número de individuos (Fig. 4B), en el mismo

orden en que se encuentren en nuestra matriz, aunque esto nos dará una curva que no está suavizada.



**Figura 4.** Curva de acumulación de especies no suavizada (sin permutación) utilizando unidades de muestreo (A) o número de individuos (B).

Según el orden en que se dispongan los datos, la forma de las curvas de acumulación puede variar significativamente. Por esta razón, para suavizar las curvas realizamos permutaciones sobre nuestra matriz, con el objetivo de aleatorizar el orden de los sitios (Fig. 5A) y calcular el valor promedio de diversidad para cada unidad de muestreo (Fig. 5B). El número de permutaciones para construir una curva dependerá del criterio del investigador o de la cantidad de datos, no existe un valor mínimo ni máximo, aunque es común utilizar 100 o 1000.

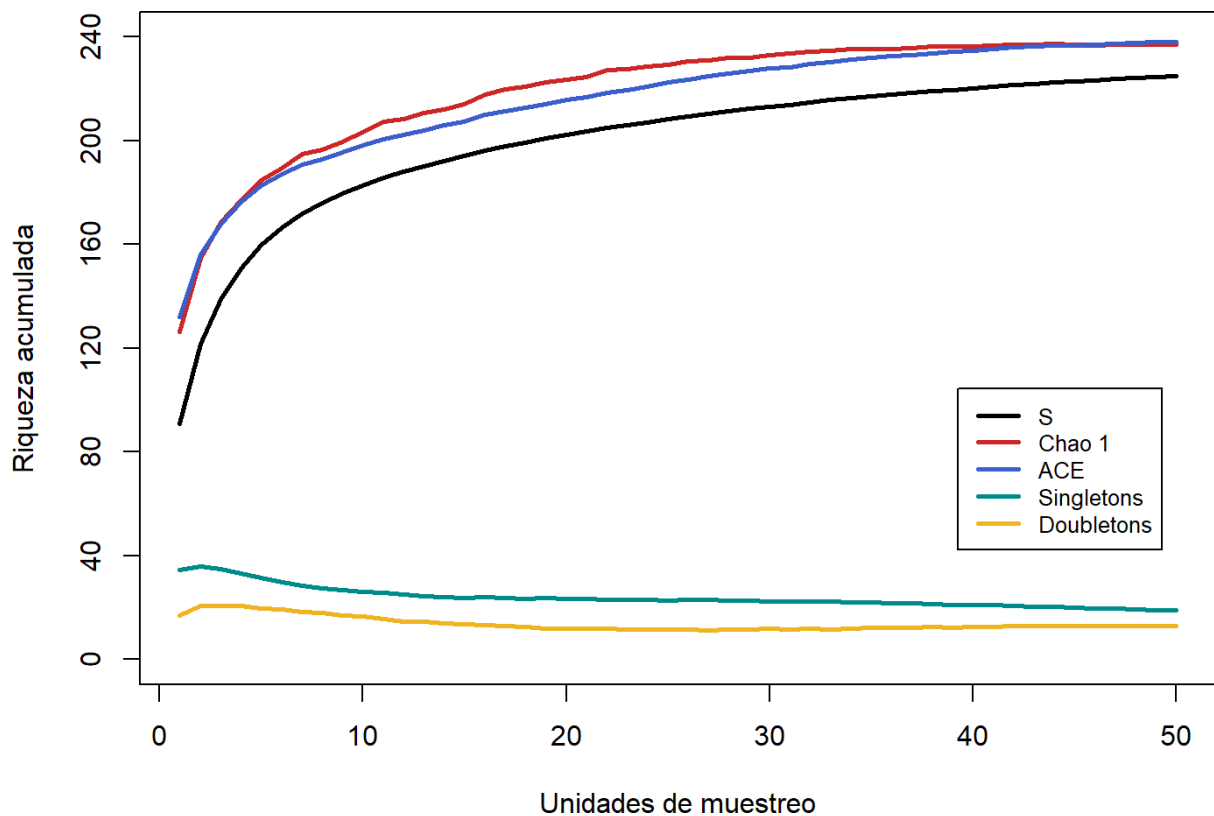


**Figura 5.** Posibles curvas de acumulación tras 100 permutaciones (A), curva de acumulación promedio de las 100 permutaciones (B).

Una de las ventajas de utilizar permutaciones es que también nos permiten calcular los intervalos de confianza para el promedio de la riqueza por sitio, lo cual no solo permite incorporar incertidumbre a nuestro análisis, sino que también nos proporciona información adicional sobre la comunidad. La amplitud de los intervalos está relacionada de manera indirecta con la equitatividad, ya que en comunidades más equitativas las permutaciones no generan matrices en las que el número de especies por sitio varíe demasiado al cambiar el orden.

Respecto a la interpretación de las curvas de acumulación, lo que idealmente buscamos es una curva asintótica, que indique que la cantidad de nuevas especies que aparecen a medida que aumenta el esfuerzo de muestreo es muy baja. En otras palabras, que nuestro muestreo capturó la mayor parte de la diversidad. En caso contrario, una curva no asintótica nos indica que el esfuerzo de muestreo no ha sido suficiente. La velocidad a la que la curva alcanza la asíntota también nos señala, de manera indirecta, si nuestra comunidad es equitativa o no. Por ejemplo, curvas que rápidamente se estabilizan son indicadoras de comunidades menos dominadas, y viceversa (Magurran 2004). Vale la pena mencionar que, una curva asintótica no significa necesariamente que se encontraron el 100% de las especies, sino que, de haber más, estas serán especies raras.

El uso combinado de los estimadores de riqueza y las curvas de acumulación nos permite calcular la **completitud del muestreo** (en inglés *completeness*), es decir, qué tantas especies de las esperadas logramos encontrar. Por ejemplo, para la comunidad hipotética de la Fig. 6 tenemos que, la diversidad observada ( $S = 225$ ) se mantiene por debajo de ambos estimadores, lo que significa que nuestro esfuerzo de muestreo aún no logra capturar el 100% de la diversidad esperada. Por su parte, los estimadores Chao ( $= 237$ ) y ACE ( $= 238$ ) convergen en un valor similar. Por lo tanto, podríamos decir que la completitud de nuestro muestreo es de aproximadamente un 94.5%. En otras palabras, con el esfuerzo de muestreo realizado encontramos el 94.5% de las especies que esperaríamos encontrar.



**Figura 6.** Ejemplo de una curva de acumulación de especies observadas y estimadas.

No existe un umbral mínimo para considerar si un esfuerzo de muestreo es bueno o malo. Lo ideal es alcanzar una curva asintótica con diversidades observadas cercanas a las estimadas. Sin embargo, que la completitud sea superior al 60%, 70%, 80% o 90% depende del criterio del investigador, de la comunidad que se esté analizando y del área cubierta en el espacio. Es cierto que valores demasiado bajos son un indicador de que se necesita mayor esfuerzo de muestreo, pero comunidades extremadamente diversas pueden requerir un esfuerzo muy grande para alcanzar la asíntota. Adicionalmente, al interpretar una curva de acumulación no solo nos interesa que esta tienda a ser asintótica, también debemos prestar atención a los *singletons* y *doubletons*, ya que estos nos indican si las especies raras son principalmente un artificio del muestreo (curvas de *singletons/doubletons* no asintóticas), o si lo más probable es que sean reales (curvas de *singletons/doubletons* asintóticas).

Por último, las curvas de acumulación son herramientas visuales que proporcionan mucha información de manera simplificada, lo que las convierte en un análisis bastante popular. No obstante, cuando se tienen muy pocas unidades de muestreo o individuos, la construcción de

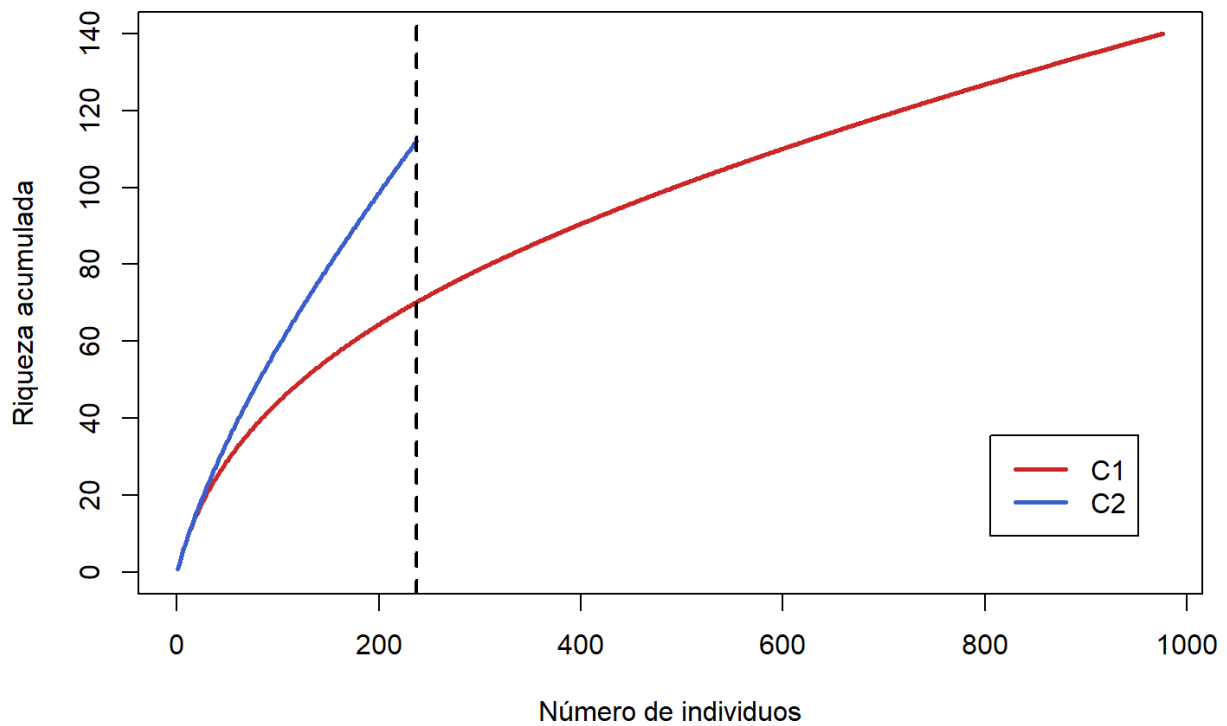
las curvas pierde todo sentido, ya que no se cuenta con un número mínimo de datos razonable que permita obtener una curva con cierta forma y poder interpretar tendencias.

## 7. Curvas de rarefacción

Como ya hemos visto, el esfuerzo de muestreo es un factor que debemos considerar al comparar la diversidad entre sitios o comunidades, y para esto se utiliza la **rarefacción**. La rarefacción es, en principio, una aplicación de la «Ley del Mínimo», y busca estimar el número esperado de especies que tendría una comunidad en relación con aquella que presenta el menor esfuerzo de muestreo. Rarificar es equivalente a **interpolar**, es decir, utilizar los valores conocidos para obtener un valor dentro del rango de los datos. Por el contrario, **extrapolar** consiste en utilizar los valores conocidos para tratar de estimar un valor fuera del rango de los datos.

Existen diversos métodos para rarificar, pero los más comunes son aquellos que se basan en el uso de abundancias o **tamaño de muestra** (*individual-based*, en inglés) y que estiman el número de especies que se esperaría si todas las muestras tuvieran el mismo número mínimo común de individuos. Por ejemplo, para dos comunidades hipotéticas con diferente esfuerzo de muestreo (Fig. 7), no podemos hacer comparaciones directas entre ambas, por lo que debemos rarificar. En este caso, vemos que al interpolar al número mínimo común de individuos (línea punteada), la Comunidad 1 tendría cerca de 70 especies, mientras que en la Comunidad 2 habría 112. Por lo tanto, concluiríamos que la Comunidad 2 es más diversa. No obstante, aunque el esfuerzo de muestreo, expresado como número de individuos, es el mismo, la cobertura de la muestra no lo es, por lo que nuestra conclusión podría estar errada.





**Figura 7.** Curva de rarefacción basada en tamaño de muestra para dos comunidades hipotéticas con diferente esfuerzo de muestreo.

La cobertura de muestra (*sample coverage*, en inglés) (Ec. 25) es una medida de qué tan completa es una muestra, basada en la proporción del número total de individuos en una comunidad que pertenecen a las especies representadas en la muestra (Chao & Jost 2012). En palabras más simples, la cobertura nos indica qué tanto muestreamos una comunidad respecto a lo que esperaríamos. Es, en parte, similar a lo que hacemos al calcular la completitud con base en las especies observadas y los estimadores de riqueza, pero en este caso no decimos que encontramos un porcentaje determinado de las especies que esperábamos encontrar, sino que muestreamos un porcentaje determinado de la comunidad.

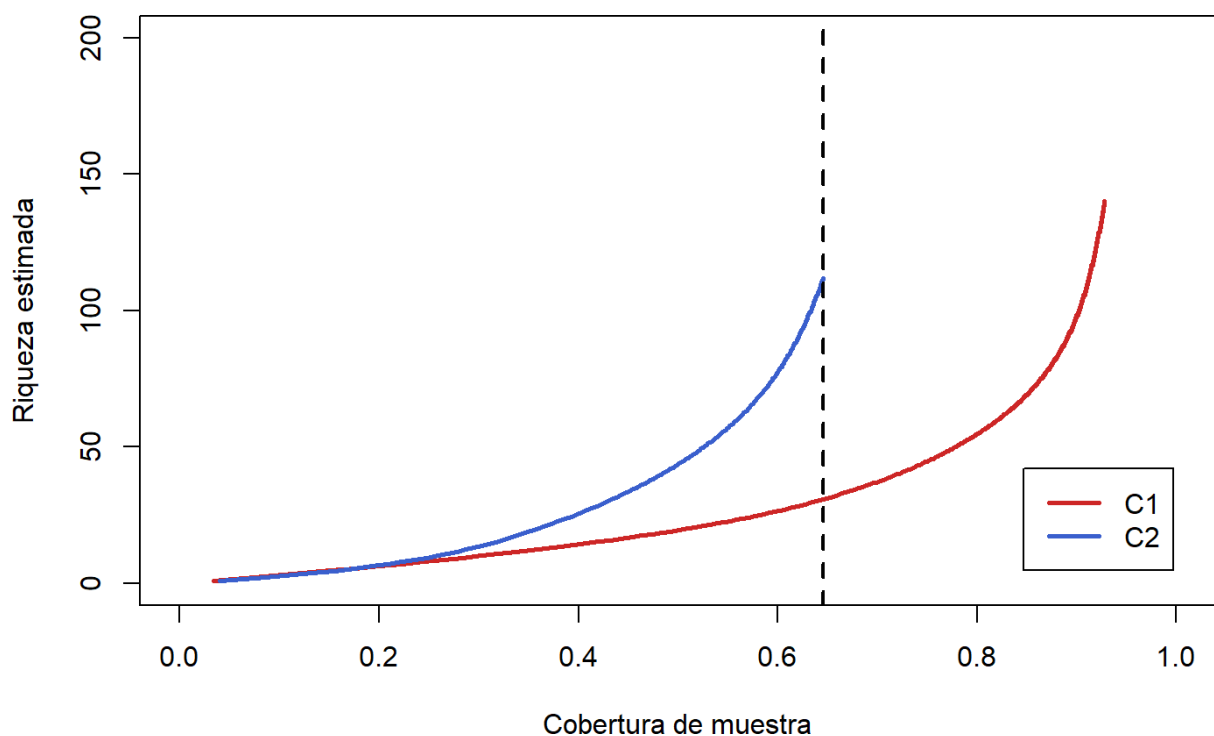
$$\hat{C}_n = 1 - \frac{f_1}{n} \left[ \frac{(n-1)f_1}{(n-1)f_1 + 2f_2} \right] \quad \text{Ec. 25}$$

donde  $n$  es el número de individuos, y  $f_1$  y  $f_2$  son el número singletons (especies con abundancia de uno) y doubletons (especies con abundancias de dos o dos únicas ocurrencias)

La cobertura de muestra es una aproximación más adecuada para comparar entre comunidades, porque, dependiendo del tipo de ecosistema o del método de muestreo, el

mismo número de individuos no implica necesariamente el mismo esfuerzo. Adicionalmente, si restamos la cobertura de muestra a uno, obtendremos el denominado déficit de muestreo, que es equivalente a la proporción de la comunidad que pertenece a especies no muestreadas. Este déficit también indica la probabilidad de que una nueva especie previamente no registrada sea encontrada si la muestra aumentara en un individuo (Chao & Jost 2012), por lo que el concepto de cobertura resulta mucho más informativo.

Si calculamos entonces la cobertura para nuestras dos comunidades hipotéticas de la Fig. 7, tendríamos que, para la Comunidad 1, esta es del 93%, mientras que para la Comunidad 2 es del 65%. Por lo tanto, al tener coberturas diferentes, no podemos asegurar cuál es más diversa, siendo necesario rarificar nuevamente, pero esta vez con base en la cobertura de muestra y no en el tamaño de muestra (Fig. 8).



**Figura 8.** Curva de rarefacción basada en cobertura de muestra para dos comunidades hipotéticas con diferente esfuerzo de muestreo.

Una vez rarificamos con base en la cobertura de muestra, podemos ver que, si bien la conclusión no cambió, los valores de riqueza sí. Usando la rarefacción basada en el tamaño de muestra, teníamos que para la Comunidad 1 se esperarían cerca de 70 especies, y para la

Comunidad 2 unas 112. Pero al rarificar usando la cobertura, tenemos que el número esperado si las hubiésemos muestreado en un 65% es de 31 para la Comunidad 1 y de 116 para la Comunidad 2.

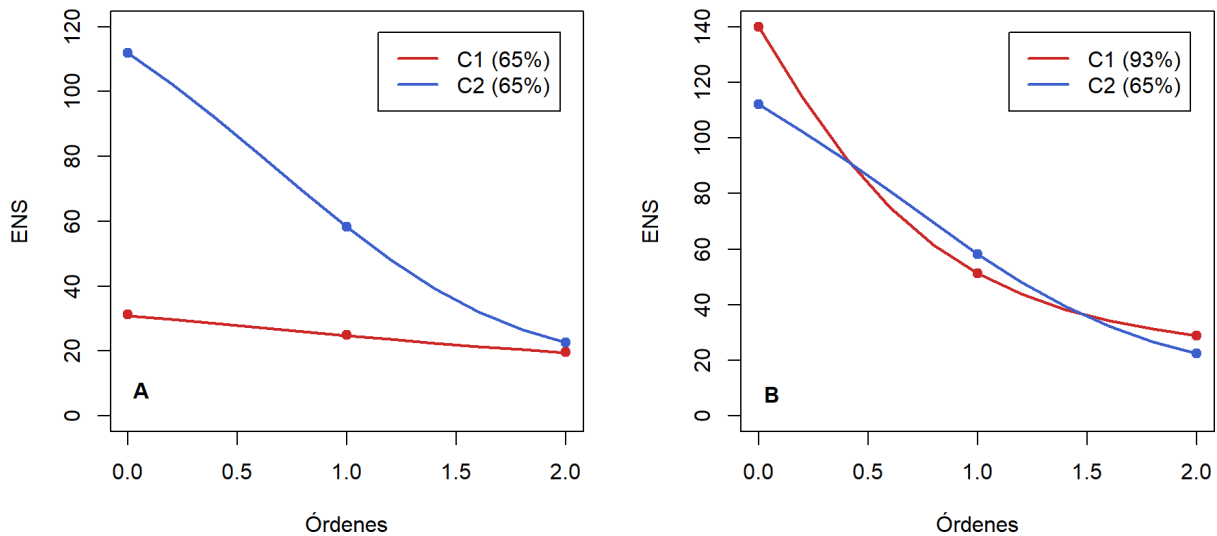
Dependiendo de nuestros datos, es posible que la conclusión cambie, por lo que lo ideal es siempre interpolar o extrapolar con base en cobertura de muestra. Para el caso de las extrapolaciones, se recomienda no hacerlo a más del doble del tamaño muestral, ya que más allá la incertidumbre se vuelve demasiado alta (Chao & Jost 2012). Para las curvas de rarefacción también es posible calcular intervalos de confianza, los cuales, al comparar entre comunidades, nos brindan información adicional sobre la significancia estadística de la diferencia de medias. En términos simples, los intervalos de confianza nos indican el rango dentro del cual es probable que se encuentre la verdadera riqueza esperada. Si los intervalos de dos comunidades no se superponen, podemos asumir que la diferencia es estadísticamente significativa, en cambio, si se superponen, la diferencia observada podría deberse al azar.

Conceptual y operativamente, la rarefacción puede generar algunas confusiones, dado que es un término que se aplica tanto en los estimadores de riqueza para una comunidad como en la metodología para comparar la riqueza entre varias comunidades. Lo que debemos tener presente es que, aunque en ambos casos se emplea el mismo método (la rarefacción), se utiliza con fines diferentes. En el primer caso, se usa para estimar el número de especies dentro de una comunidad si todas las muestras tuviesen un número mínimo común de individuos. En el segundo, se emplea para comparar la diversidad de dos o más comunidades diferentes, suponiendo que todas hubiesen tenido un mínimo esfuerzo de muestreo equivalente.

## **8. Perfiles de diversidad**

Los perfiles de diversidad son otra herramienta gráfica que nos ayuda a comparar la diversidad entre diferentes comunidades a partir de las diversidades verdaderas. A grandes rasgos, estos perfiles nos permiten evaluar visualmente aspectos como la riqueza y la equitatividad, siendo especialmente útiles si recordamos que la diversidad no se refiere solo a cuántas especies hay, sino también a cuántos individuos hay y cómo están distribuidos.

Construir un perfil de diversidad (Fig. 9) es bastante sencillo una vez que tenemos calculadas las diversidades verdaderas. La interpretación de los perfiles se basa en los órdenes de diversidad, de la misma manera que ya hemos mencionado, el orden cero indica la riqueza, el uno, las especies comunes (en términos de abundancia, no especies compartidas), y el dos, las especies dominantes. Los órdenes intermedios no tienen una interpretación directa, pero en este contexto se calculan para ayudar a suavizar las curvas.



**Figura 9.** Perfiles de diversidad para dos comunidades hipotéticas, con rarificación basada en cobertura (A), sin rarificación (B).

Así, para estas dos comunidades hipotéticas, al mismo esfuerzo de muestreo (Fig. 9A) tenemos que, la Comunidad 2 presenta mayor riqueza que la Comunidad 1, pero la Comunidad 1 tiene un menor número de especies raras ( ${}^0D - {}^1D$ ) y menor dominancia ( ${}^2D / {}^0D$ ). Por el contrario, si por alguna razón evitamos rarificar (Fig. 9B), vemos como nuestras inferencias se invierten.

## 9. La diversidad a escala regional

Cuando ampliamos la escala de observación y pasamos de las comunidades locales al nivel del paisaje, nos encontramos con la diversidad gamma ( $\gamma$ ). A diferencia de la diversidad alfa ( $\alpha$ ) o beta ( $\beta$ ), la diversidad gamma no representa una propiedad independiente de las comunidades biológicas, sino que integra las otras dos dimensiones. En palabras más simples, la diversidad gamma es la diversidad regional entendida como la contribución de diferentes

comunidades más efecto del recambio de especies. Desde un enfoque matemático, la diversidad gamma puede abordarse desde dos perspectivas: la formulación aditiva (Lande 1996) (Ec. 26) y la formulación multiplicativa (Whittaker 1972) (Ec. 27), cada una con implicaciones distintas para la interpretación ecológica.

$$\gamma = \alpha + \beta \quad \text{Ec. 26}$$

$$\gamma = \alpha \times \beta \quad \text{Ec. 27}$$

*donde  $\alpha$  es el valor promedio de la diversidad alfa de todas las comunidades y  $\beta$  el valor promedio de la diversidad beta entre comunidades*

La formulación aditiva se utiliza principalmente en el contexto de los índices de diversidad tradicionales, mientras que la formulación multiplicativa se asocia con el marco de las diversidades verdaderas o números efectivos de especies. Si bien la partición aditiva surgió como una alternativa mucho tiempo después que la multiplicativa, en la primera los componentes  $\alpha$  y  $\beta$  son matemáticamente dependientes, lo que puede generar patrones espurios que no reflejan procesos ecológicos reales, sino artefactos derivados de la forma en que se calcula la diversidad (Jost 2007; Baselga 2010a). Por ejemplo, la diversidad beta de comunidades poco diversas que comparten todas sus especies y que por definición debería ser reducida o nula, puede llegar a ser mayor que la de comunidades muy diversas que no comparten ninguna especie (Jost 2007), lo cual no tiene sentido. En cambio, la partición multiplicativa garantiza la independencia matemática entre  $\alpha$  y  $\beta$ , lo que permite interpretar los patrones observados como resultado de diferencias biológicas reales entre comunidades.

## **10. Estructura de las comunidades**

Hasta ahora, hemos entendido la diversidad en términos del número de especies y su composición. Sin embargo, las comunidades biológicas no son estáticas. Tradicionalmente, la diversidad beta ( $\beta$ ) se ha interpretado como el recambio de especies, es decir, la variación en la composición de especies entre comunidades (Whittaker 1960), tanto en términos de qué especies están presentes como cuántas. A partir de ello, podemos calcular algunas métricas que nos permiten estimar qué tan similares o diferentes son dos o más comunidades biológicas (Koleff et al. 2003; Baselga 2010b).

Dos de los índices más utilizados son el índice de similitud de Jaccard (Ec. 28) y el índice de similitud de Sørensen (Ec. 29), los cuales provienen de la teoría de conjuntos y son particularmente útiles para el análisis de datos de presencia-ausencia.

$$Jaccard = \frac{a}{a + b + c} \quad Ec. 28$$

$$Sørensen = \frac{2a}{2a + b + c} \quad Ec. 29$$

donde *a* es el número de especies compartidas entre dos comunidades, *b* el número de especies únicas de la comunidad uno, y *c* el número de especies únicas de la comunidad dos

La única diferencia entre ambos índices radica en el peso relativo que asignan a las especies compartidas. Así, el índice de Sørensen da mayor importancia a la similitud (es más conservador), por lo que tiende a arrojar valores más altos que el índice de Jaccard para las mismas comunidades, especialmente cuando hay pocas especies en común. Ambos índices varían entre 0 (sin especies compartidas) y 1 (composición idéntica), aunque en algunos casos se expresan como su complemento, es decir, como un índice de disimilitud. Por ello, es fundamental conocer qué tipo de métrica se está utilizando antes de hacer comparaciones.

Por otro lado, cuando disponemos de datos de abundancia, es útil emplear métricas que también reflejen la estructura cuantitativa de las comunidades. Entre los muchos índices disponibles, el índice de Bray-Curtis (Ec. 30) y el índice de Morisita-Horn (Ec. 31) se encuentran entre los más ampliamente utilizados. Al igual que los índices basados en presencia-ausencia, valores más altos indican una mayor similitud entre comunidades, tanto en términos de composición como de abundancia.

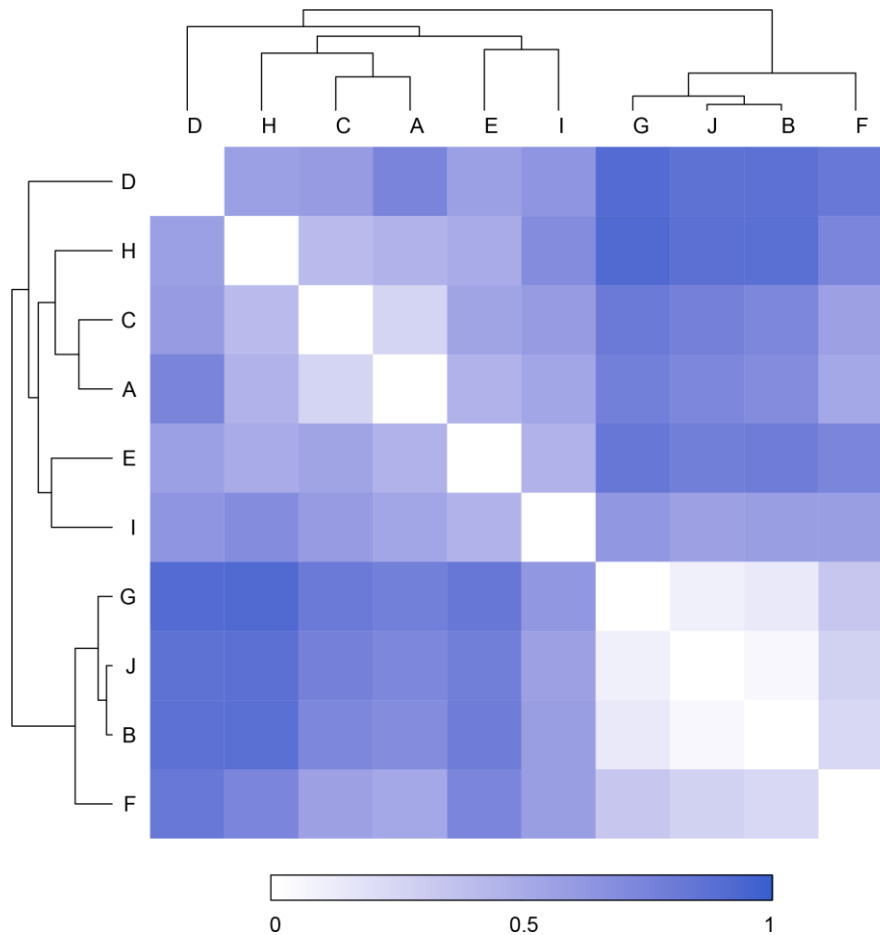
$$BC = \sum \frac{|x_{ij} - x_{ik}|}{(x_{ij} + x_{ik})} \quad Ec. 30$$

donde  $x_{ij}$  y  $x_{ik}$  son las abundancias de la especie *i* en las comunidades *j* y *k*

$$MH = \frac{2 \sum x_i y_i}{\left( \frac{\sum x_i^2}{N_x^2} + \frac{\sum y_i^2}{N_y^2} \right) \cdot N_x + N_y} \quad \text{Ec. 31}$$

donde  $x_i$  es la abundancia de la especie  $i$  en la comunidad uno,  $y_i$  la abundancia de la especie  $i$  en la comunidad dos,  $N_x$  el total de individuos en la comunidad uno y  $N_y$  el total de individuos en la comunidad dos

Cual sea el índice que usemos, estas métricas generan valores de distancia entre pares, lo que nos permite construir matrices de distancia cuando analizamos más de dos comunidades. Estas matrices pueden visualizarse mediante gráficas como dendrogramas (Fig. 10), los cuales ofrecen una representación jerárquica de las relaciones entre comunidades.



**Figura 10.** Ejemplo de un dendrograma y su correspondiente matriz de disimilitud usando el índice de Jaccard.

Otra forma sumamente útil para describir comunidades biológicas son los análisis multivariados, entre los que se encuentran el escalamiento no métrico multidimensional (NMDS) o el análisis de coordenadas principales (PCoA). Ambos análisis resumen la variación de múltiples variables en un espacio reducido de dimensiones. Esto permite representar gráficamente la estructura entre comunidades de forma más simple facilitando así su interpretación. Si el lector está interesado en profundizar en los análisis multivariados con aplicación biológica, puede referirse a Palacio *et al.* (2020). Los NMDS o PCoA pueden calcularse con cualquier métrica que genere una matriz de distancias, sin embargo, la elección depende del tipo de datos y de la pregunta ecológica (Tabla 3).

**Tabla 3.** Comparación de índices de similitud para su uso en análisis multivariados.

Índice	Tipo de datos	Pros	Contras
Jaccard	Incendencia	Simple, útil cuando solo interesa la composición	Ignora abundancia, sensible a especies raras
Sørensen	Incendencia	Favorece similitud, más conservador	Ignora abundancia, resultados similares a Jaccard
Bray-Curtis	Abundancia	Capta bien diferencias en dominancia	Sensible a especies dominantes y abundantes
Morisita-Horn	Abundancia	Robusto al tamaño de muestra, bueno para comparar comunidades desbalanceadas	Menos sensible a especies raras o cambios sutiles en abundancia
Gower	Mixtos	Permite combinar datos de distinta naturaleza	Más complejo de interpretar, requiere estandarización previa

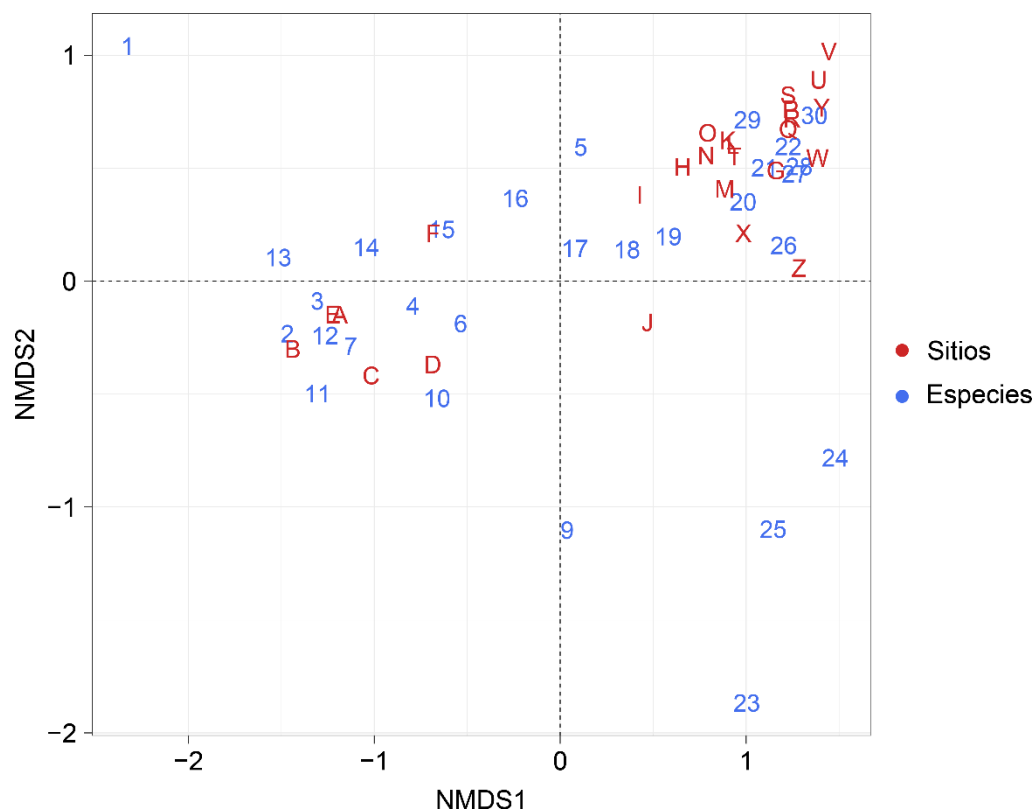
Es importante tener presente que los PCoA son métodos paramétricos, es decir, parten de ciertos supuestos sobre la estructura de los datos, en particular que la matriz de distancias sea métrica o se pueda aproximar a una matriz métrica (p.e., basada en distancias euclidianas). Además, tratan de maximizar la varianza, por lo que cada componente explica un porcentaje de la variación original. En contraste, el NMDS<sup>3</sup> es un método no paramétrico, lo que lo hace más flexible y robusto cuando se usan índices no métricos como los utilizados en ecología de comunidades. A diferencia del PCoA, el NMDS no maximiza la varianza explicada, sino que trata de ubicar los puntos en el espacio respetando el orden de similitud entre ellos, por lo que sus componentes no son necesariamente los que más varianza tienen, ni tiene sentido entenderlos en términos de variación.

---

<sup>3</sup> Para que un NMDS se considere confiable, su valor de estrés (*stress*) debe ser idealmente menor a 0.1 o en su defecto, no mayor a 0.2.

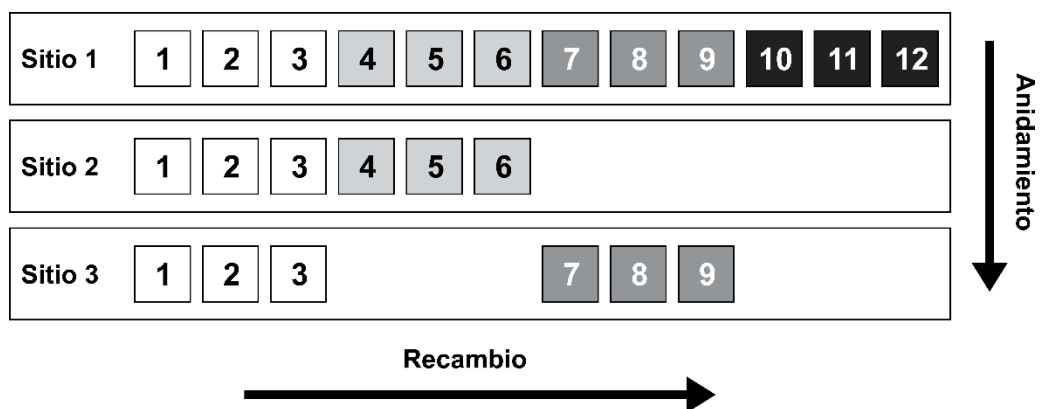


En una ordenación NMDS (Fig. 11), la posición relativa de los sitios indica su similitud en composición, es decir, entre más cerca se encuentren dos sitios más similares son en términos de especies presentes o abundantes. Por otro lado, la posición de las especies también es interpretativa, especies que aparecen próximas entre sí tienden a compartir patrones de abundancia o presencia en los mismos sitios. Además, la orientación de las especies respecto al centro del gráfico permite identificar asociaciones, las especies que se ubican en la dirección de ciertos sitios están generalmente asociadas a ellos, ya sea por su presencia exclusiva o por una abundancia especialmente alta en esos lugares.



**Figura 11.** Ejemplo de una ordenación NMDS usando el índice de Jaccard.

Ahora bien, aunque la diversidad beta ( $\beta$ ) suele entenderse tradicionalmente como recambio, este es solo uno de sus componentes. La visión más actual de la diversidad beta ( $\beta$ ) implica la suma de dos componentes (Baselga 2010b): el recambio (*turnover*) y el anidamiento (*nestedness*). El recambio es la proporción de reemplazo de especies entre comunidades, mientras que el anidamiento, la proporción de especies presentes en una comunidad que son un subconjunto de otra (Fig. 12). Estos componentes son mutuamente informativos, de modo que, una comunidad con alto recambio tiende a tener bajo anidamiento, y viceversa.



**Figura 12.** Componentes de la diversidad beta ( $\beta$ ). Adaptado de Baselga (2010b).

## 11. Aplicando la teoría usando R

A continuación, utilizaremos un conjunto de datos real para aprender a calcular y graficar distintas métricas de diversidad utilizando el lenguaje de programación R. Es importante aclarar que esta no es una guía de programación, por lo tanto, si el lector no está familiarizado con los conceptos básicos del lenguaje, se recomienda tomarse el tiempo para aprenderlos primero. Para este propósito, he preparado un pequeño tutorial disponible en mi GitHub (<https://github.com/VR-Daniel/Tutoriales-R>). Si bien existen otros programas con interfaces gráficas más amigables para el usuario, he optado por un enfoque programático debido a la escalabilidad que ofrece. También cabe mencionar que los códigos presentados no están optimizados deliberadamente, con el fin de facilitar el seguimiento paso a paso para usuarios menos experimentados. Además, cada línea está comentada para explicar su función.

### 11.1. Primeros pasos: librerías y datos

Lo primero que debemos hacer es descargar y cargar las librerías que vamos a utilizar. En este caso, las seis primeras incluyen funciones enfocadas en los análisis de diversidad biológica, mientras que las siguientes, en el manejo, análisis y visualización de datos. Adicionalmente, vamos a crear un tema base para definir algunos de los elementos estéticos de nuestras gráficas y enlazar unas funciones personalizadas que usaremos más adelante.

```
# Cargamos las librerías
library(entropart)
```

```

library(iNEXT) # Debe ser la v.2.0.20
library(vegan)
library(tabula)
library(betapart)
library(ade4)
library(tidyverse)
library(stats)
library(car)
library(FSA)
library(ggplot2)
library(dendextend)
library(parallel)

# Enlazamos las funciones auxiliares (las usaremos luego)
URL <- "https://raw.githubusercontent.com/VR-Daniel/Tutoriales-R/"
sub <- "refs/heads/main/Diversidad%20en%20R/auxfuns/"
auxfun_1 <- paste0(URL, sub, "estimadores_SpadeR.R")
auxfun_2 <- paste0(URL, sub, "estimadores_acum_SpadeR.R")
Janzen73 <- paste0(URL, sub, "Janzen73.R")

# Creamos un tema base para ggplot
theme_base <- function() {
  theme_bw() +
  theme(
    axis.text = element_text(size = 28,
                              color = "black"),
    axis.text.x = element_text(margin = margin(t = 8)),
    axis.text.y = element_text(margin = margin(r = 8)),
    axis.title = element_text(size = 28),
    axis.title.x = element_text(margin = margin(t = 15)),
    axis.title.y = element_text(margin = margin(r = 15)),
    axis.ticks.length = unit(0.3, "cm"),
    panel.spacing = unit(1, "cm"),
    strip.text = element_text(size = 28,
                              margin = margin(t = 20, b = 20)),
    legend.text = element_text(size = 24),
    legend.title = element_blank(),
    legend.key.size = unit(1.2, "cm")
  )
}

```

Para la primera parte de este ejercicio, utilizaremos el conjunto de datos BCI, que viene precargado en el paquete ‘vegan’ e incluye datos de abundancia de 225 especies de árboles en la isla de Barro Colorado, Panamá, muestreadas en 50 parcelas.

```

# Cargamos el conjunto de datos BCI
data(BCI, package = "vegan")

```

En el caso de tener nuestros propios datos podemos importarlos desde un archivo de texto delimitado, usando las funciones `read.table()`, `read.csv()` o `read.delim()`. Es

importante que nuestros datos estén organizados de modo que las especies estén en las columnas y las unidades de muestreo o comunidades en las filas.

## 11.2. Índices de riqueza

Comenzaremos por calcular los índices de Margalef y Menhinick, para los cuales no disponemos de una función predefinida, pero que podemos crear fácilmente, ya que sus fórmulas son bastante sencillas. Para calcular estos índices necesitamos dos datos básicos: el número de especies y el total de individuos. Dado que contamos con 50 unidades de muestreo, es necesario obtener estos datos para cada una de ellas.

```
# Número de especies con abundancia > 0 por unidad de muestreo
S <- rowSums(BCI > 0)

# Suma total de individuos por unidad de muestreo
N <- rowSums(BCI)

# Índices
Margalef <- (S - 1) / log(N)
Menhinick <- S / sqrt(N)

# Tabla de índices
indices_riqueza <- data.frame("Margalef" = Margalef,
                              "Menhinick" = Menhinick)

# Imprimimos el resultado
head(indices_riqueza)
```

	Margalef	Menhinick
1	15.07013	4.393837
2	13.66177	4.027492
3	14.50048	4.182655
4	14.92662	4.170576
5	16.06540	4.494441
6	13.95112	4.187649

Como vemos, el resultado es una matriz con los valores de cada índice para los 50 sitios muestreados. Aquí, para ahorrar espacio, usamos `head()` para echar un vistazo solo a los primeros seis.

## 11.3. Índices de dominancia y equitatividad

A continuación, calcularemos los índices de dominancia y equitatividad. Para ello, utilizaremos la librería ‘tabula’. Si bien el paquete ‘vegan’ es mucho más popular entre ecólogos, resulta más limitado en cuanto a las opciones que ofrece para calcular las variantes

de algunos índices, siendo incluso necesario en algunos casos escribir funciones personalizadas que pueden llegar a ser complejas.

```
# Índice de Gini-Simpson
# --- Usa por defecto la versión no corregida, pero
# --- se puede modificar con unbiased == TRUE
Gini <- 1 - apply(BCI, 1, tabula::index_simpson, unbiased = FALSE)

# Índice de Shannon
# --- Usa por defecto base = exp(1) [= ln], pero se puede modificar
Shannon <- apply(BCI, 1, tabula::index_shannon, base = exp(1))

# Índice de Brillouin
# --- Usa por defecto base = exp(1), no se puede modificar
Brillouin <- apply(BCI, 1, tabula::index_brillouin)

# Combinamos los índices en una tabla
indices_dominancia <- cbind("Simpson" = Gini,
                           "Shannon" = Shannon,
                           "Brillouin" = Brillouin)

#Imprimimos el resultado
head(indices_dominancia)
```

	Simpson	Shannon	Brillouin
1	0.9746293	4.018412	3.719341
2	0.9683393	3.848471	3.570265
3	0.9646078	3.814060	3.539980
4	0.9716117	3.976563	3.703080
5	0.9678267	3.969940	3.684281
6	0.9627557	3.776575	3.489278

```
# Equitatividad de Simpson
eqSimpson <- apply(BCI, 1, index_simpson, evenness = TRUE)

# Equitatividad de Shannon (= índice de Pielou)
eqShannon <- apply(BCI, 1, index_shannon, evenness = TRUE)

# Equitatividad de Brillouin
eqBrillouin <- apply(BCI, 1, index_brillouin, evenness = TRUE)

# Combinamos los índices en una tabla
indices_equitatividad <- cbind("eq.Simpson" = eqSimpson,
                              "eq.Shannon" = eqShannon,
                              "eq.Brillouin" = eqBrillouin)

# Imprimimos el resultado
head(indices_equitatividad)
```

	eq.Simpson	eq.Shannon	eq.Brillouin
1	0.4238232	0.8865579	0.8896033
2	0.3760104	0.8685692	0.8711838
3	0.3139420	0.8476046	0.8498499

4	0.3747422	0.8752597	0.8778150
5	0.3077392	0.8602030	0.8618114
6	0.3158792	0.8500724	0.8523669

#### 11.4. Índices de diversidad verdadera

Siguiendo el orden del capítulo, ahora vamos a calcular las diversidades verdaderas. Para esto tenemos dos opciones: transformar los índices que ya calculamos (Ecs. 13–15), o usar la función `renyi()` del paquete ‘vegan’. Para este ejemplo utilizaremos ambos métodos, con el objetivo de mostrar que son equivalentes.

```
# Transformamos los índices tradicionales
indices_manual <- data.frame("S" = S,
                             "Shannon" = indices_dominancia[, 2],
                             "Simpson" = indices_dominancia[, 1],
                             "q0" = S,
                             "q1" = exp(Shannon),
                             "q2" = 1 / (1 - Gini))

# Calculamos las diversidades verdaderas usando 'vegan'
indices_vegan <- renyi(BCI,
                      scales = c(0, 1, 2), # Ordenes de diversidad
                      hill = TRUE)

# Imprimimos el resultado
head(cbind(indices_manual, indices_vegan))
```

	S	Shannon	Simpson	q0	q1	q2	0	1	2
1	93	4.01841	0.97463	93	55.61270	39.41555	93	55.61270	39.41555
2	84	3.84847	0.96834	84	46.92127	31.58488	84	46.92127	31.58488
3	90	3.81406	0.96461	90	45.33410	28.25478	90	45.33410	28.25478
4	94	3.97656	0.97161	94	53.33341	35.22577	94	53.33341	35.22577
5	101	3.96994	0.96783	101	52.98137	31.08166	101	52.98137	31.08166
6	85	3.77658	0.96276	85	43.66624	26.84973	85	43.66624	26.84973

#### 11.5. Réplicas y unidades de muestreo

Es importante recordar que, en nuestra matriz de datos, cada fila corresponde a un sitio o unidad de muestreo que se asume como «independiente». Sin embargo, en ocasiones es común contar con réplicas, por ejemplo, según el tipo de cobertura vegetal o hábitat. Así, para continuar con el ejemplo, vamos a suponer que las 50 parcelas de nuestro conjunto de

datos corresponden a cinco grupos diferentes<sup>4</sup>. En este caso, utilizaremos una partición que definí previamente con base en otros análisis (los cuales no detallaré aquí), tratando de maximizar la similitud en composición entre sitios y así simular particiones ecológicas más realistas. En la vida real, la naturaleza ya hizo ese trabajo por nosotros, y si existen reglas de ensamblaje de las comunidades, nuestras réplicas deberían parecerse más entre sí que con los sitios de otros grupos.

```
# Generamos un vector de posiciones para los grupos ficticios
# --- Aquí son números, pero podrían ser categorías (p.e. cobertura)
k <- c(4, 4, 2, 2, 2, 5, 4, 2, 2, 2,
      4, 4, 3, 4, 2, 5, 4, 3, 4, 2,
      5, 1, 3, 4, 4, 5, 1, 1, 4, 4,
      5, 5, 5, 1, 1, 5, 5, 1, 1, 1,
      5, 5, 5, 5, 5, 3, 1, 5, 5, 5)

# Generamos una tabla de riqueza y abundancia por replica
replicas <- data.frame("Grupo" = factor(k),
                      "Riqueza" = S,
                      "Abundancia" = N)

# Imprimimos el resultado
head(replicas)
```

	Grupo	Riqueza	Abundancia
1	4	93	448
2	4	84	435
3	2	90	463
4	2	94	508
5	2	101	505
6	5	85	412

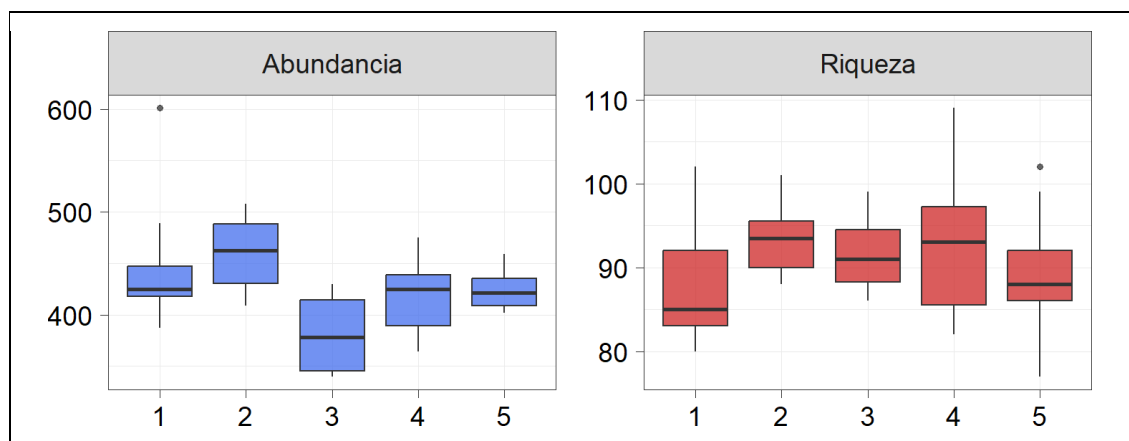
Lo siguiente que deberíamos preguntarnos es si nuestras réplicas son realmente réplicas. Por definición, una réplica debería presentar poca variación respecto a las demás, es decir, el conjunto de réplicas debe ser preciso, ya que de lo contrario estaríamos introduciendo mucho ruido en los análisis. Una forma sencilla de validarlo es mediante los diagramas de cajas y bigotes, con los que podemos evaluar visualmente qué tan dispersos están nuestros datos

<sup>4</sup> Para el propósito de este ejercicio, y por facilidad en la clasificación de los datos, estamos utilizando un diseño desbalanceado, en el que cada grupo tiene un número desigual de réplicas. Los análisis que vamos a emplear pueden manejar esta situación, pero es importante tener presente que no todos los análisis son adecuados para este tipo de datos. Para un muestreo balanceado, lo ideal es planificarlo durante la etapa de diseño, aunque si los datos ya fueron tomados, podríamos «corregirlo», por ejemplo, mediante un proceso de rarefacción.

dentro de cada grupo. Como en este ejemplo forzamos las particiones tratando de imitar un criterio biológico, los datos pueden presentar más variación de la que esperaríamos.

```
# Convertimos la tabla resumen de las réplicas a formato largo
# --- Este formato es necesario para que algunos paquetes
# --- puedan entender la estructura de los datos
replicas_long <- replicas %>%
  pivot_longer(cols = c(Riqueza, Abundancia),
               names_to = "Variable",
               values_to = "Valor")

# Graficamos los boxplots para la abundancia y la riqueza
ggplot(replicas_long, aes(x = Grupo, y = Valor, fill = Variable)) +
  geom_boxplot(size = 1, outlier.size = 3, alpha = 0.75) +
  scale_fill_manual(values = c("royalblue2", "firebrick3")) +
  facet_wrap(~ Variable, scales = "free") +
  xlab(element_blank()) +
  ylab(element_blank()) +
  theme_base() +
  theme(legend.position = "none")
```



Ahora, con el fin de aplicar ciertas pruebas estadísticas, primero debemos validar dos supuestos importantes en nuestros datos: la normalidad y la homogeneidad de varianzas (homocedasticidad). Según los resultados de estas pruebas, podremos decidir si utilizar pruebas paramétricas o no paramétricas.

```
# Creamos una lista con los grupos únicos
grupos <- sort(unique(replicas$Grupo))

# Creamos vectores vacíos para guardar los resultados
p.riqueza <- numeric(length(grupos))
p.abundancia <- numeric(length(grupos))
```



```
# Calculamos los p-valores de Shapiro por grupo
for(i in 1:length(grupos)){
  grupo <- replicas[replicas$Grupo == grupos[i], ]
  p.riqueza[i] <- shapiro.test(grupo$Riqueza)$p.value
  p.abundancia[i] <- shapiro.test(grupo$Abundancia)$p.value
}

# Generamos una tabla con los p-valores para ambas variables
shapiro <- data.frame(
  "Grupo" = grupos,
  "p.riqueza" = round(p.riqueza, 3),
  "p.abundancia" = round(p.abundancia, 3)
)

# Imprimimos el resultado
# --- Se rechaza la normalidad si  $p < \alpha$  ( $= 0.05$ )
shapiro
```

	Grupo	p.riqueza	p.abundancia
1	1	0.231	0.012
2	2	0.320	0.571
3	3	0.858	0.278
4	4	0.418	0.606
5	5	0.848	0.252

```
# Probamos homogeneidad de varianzas con la prueba de Bartlett
# --- Si se acepta normalidad, usar el test de Bartlett
bar.riqueza <- bartlett.test(Riqueza ~ Grupo, data = replicas)
bar.abundancia <- bartlett.test(Abundancia ~ Grupo, data = replicas)

# Probamos homogeneidad de varianzas con la prueba de Levene
# --- Si se rechaza normalidad, usar el test de Levene
lev.riqueza <- leveneTest(Riqueza ~ Grupo, data = replicas)
lev.abundancia <- leveneTest(Abundancia ~ Grupo, data = replicas)

# Generamos una tabla con los p-valores para ambas variables
homocedasticidad <- data.frame(
  row.names = c("Riqueza", "Abundancia"),
  "Bartlett" = c(bar.riqueza$p.value, bar.abundancia$p.value),
  "Levene" = c(lev.riqueza[1, 3], lev.abundancia[1, 3])
)

# Imprimimos el resultado (redondeado a tres decimales)
# --- Se rechaza la homocedasticidad si  $p < \alpha$  ( $= 0.05$ )
round(homocedasticidad, 3)
```

	Bartlett	Levene
Riqueza	0.491	0.504
Abundancia	0.001	0.193

Según los resultados, casi todas nuestras comunidades cumplen con los supuestos de las pruebas paramétricas. No obstante, debemos tener presente que, para la abundancia, usamos

la abundancia total de cada réplica, pero si visualizáramos la distribución de las abundancias por especie, observaríamos algo diferente.

```
# Creamos una tabla para la abundancia total por grupo
# --- Aunque el diseño es desbalanceado, la matriz es
# --- simétrica respecto al tamaño del grupo más grande
# --- Para los datos faltantes, asignamos NA
ab_replicas <- matrix(NA, nrow = max(table(k)), ncol = length(grupos))

# Asignamos nombres a las filas y columnas
colnames(ab_replicas) <- paste0("G", grupos)

# Llenamos la matriz con los valores
for(i in 1:length(grupos)){
  suma <- rowSums(BCI[k == grupos[i], ])
  ab_replicas[1:length(suma), i] <- suma
}

# Creamos una tabla para la abundancia total por especie
ab_especies <- matrix(nrow = ncol(BCI), ncol = length(grupos))

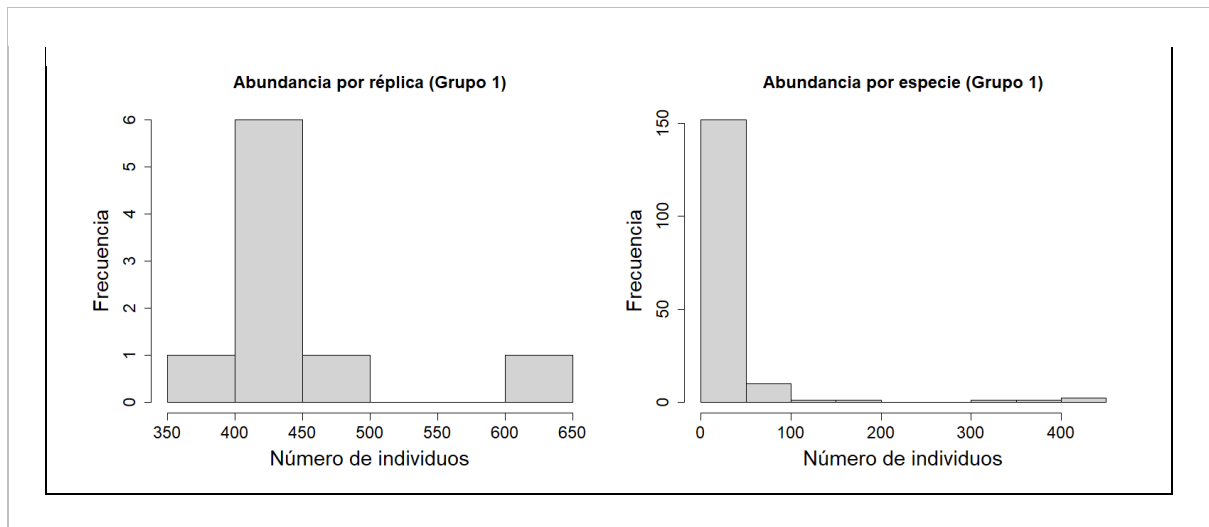
# Asignamos nombres a las filas y columnas
colnames(ab_especies) <- grupos
rownames(ab_especies) <- colnames(BCI)

# Llenamos la matriz con los valores
for(i in grupos){
  suma <- colSums(BCI[k == i, ])
  ab_especies[, as.character(i)] <- suma
}

# Reemplazamos los ceros por NA (para que no cuenten en el histograma)
ab_especies[ab_especies == 0] <- NA

# Histograma para abundancia por réplica del Grupo 1
hist(ab_replicas[, 1],
     breaks = 7,
     main = "Abundancia por réplica (Grupo 1)",
     xlab = "Número de individuos",
     ylab = "Frecuencia",
     cex.main = 1.5,
     cex.lab = 1.8,
     cex.axis = 1.5)

# Histograma para abundancia por especie del Grupo 1
hist(ab_especies[, 1],
     breaks = 7,
     main = "Abundancia por especie (Grupo 1)",
     xlab = "Número de individuos",
     ylab = "Frecuencia",
     cex.main = 1.5,
     cex.lab = 1.8,
     cex.axis = 1.5)
```



Como vemos, a diferencia de la abundancia total por réplica, donde el número de individuos varía alrededor de un valor central, la abundancia por especie está altamente sesgada a la derecha, lo cual es esperable en la mayoría de las comunidades biológicas. Es decir, la mayoría de las especies tienden a tener abundancias bajas, y solo unas pocas presentan abundancias muy altas. Por eso, es fundamental entender qué estamos haciendo con nuestros datos, para poder limitar nuestras inferencias a aquello que realmente pueden decirnos.

Ahora bien, aunque las pruebas paramétricas como la ANOVA son relativamente robustas a la falta de normalidad (más que a la falta de homocedasticidad), debido a desviación de la normalidad en al menos una de las variables por grupo y a la asimetría de los datos por especie, para el análisis de comunidades suelo preferir pruebas no paramétricas como el Kruskal-Wallis. A continuación, calcularemos si existen diferencias en las medias de la abundancia total y número de especies entre réplicas.

```
# Probamos si existen diferencias en las medias de la abundancia
# --- Se rechaza que las medias sean iguales si p < α (= 0.05)
kruskal.test(Abundancia ~ Grupo, data = replicas)
```

Kruskal-Wallis rank sum test

data: Abundancia by Grupo

Kruskal-Wallis chi-squared = 11.066, df = 4, p-value = 0.02583

```
# Probamos si existen diferencias en las medias del número de especies
# --- Se rechaza que las medias sean iguales si p < α (= 0.05)
kruskal.test(Riqueza ~ Grupo, data = replicas)
```

Kruskal-Wallis rank sum test				
data: Riqueza by Grupo				
Kruskal-Wallis chi-squared = 5.1878, df = 4, p-value = 0.2686				
# Prueba post-hoc para las abundancias				
# --- Ya que el Kruskal para las abundancias resultó significativo,				
# --- necesitamos una prueba pareada para ver diferencias específicas				
dunnTest(Abundancia ~ Grupo, data = replicas, method = "holm")\$res				
	Comparison	Z	P.unadj	P.adj
1	1 - 2	-1.0413765	0.297700838	0.59540168
2	1 - 3	2.1551153	0.031152816	0.21806971
3	2 - 3	2.9411532	0.003269928	0.03269928
4	1 - 4	1.1864407	0.235448297	0.94179319
5	2 - 4	2.2548406	0.024143336	0.19314669
6	3 - 4	-1.3369560	0.181236998	0.90618499
7	1 - 5	1.1455128	0.251996794	0.75599038
8	2 - 5	2.2816148	0.022512095	0.20260885
9	3 - 5	-1.4806889	0.138689486	0.83213692
10	4 - 5	-0.1351458	0.892496608	0.89249661

Con base en lo anterior, podemos decir que nuestros grupos tienen un número similar de especies, pero al menos un par de ellos (Grupo 2 y 3) difiere en cuanto a sus abundancias (lo cual también es evidente en el histograma que realizamos previamente).

## 11.6. Estimadores de riqueza

Pasamos ahora a los estimadores de riqueza. Para esto, continuaremos trabajando con el conjunto de datos resumido en cinco grupos, en lugar de las 50 parcelas originales. No obstante, la misma lógica se puede aplicar en caso de que quisiéramos conservar los datos sin agrupar. Si bien existen diversos paquetes con funciones que permiten calcular estimadores de riqueza, la mayoría se quedan cortos al ofrecernos opciones para elegir variantes o modificar parámetros. De todos ellos, ‘SpadeR’ es el único que permite seleccionar entre una combinación de estimadores basados en abundancia o en distintos tipos de incidencia, utilizar versiones corregidas o no corregidas de los índices, modificar el número de especies raras para estimadores como ACE o ICE, y ajustar el nivel de confianza para el cálculo de intervalos.

Para facilitar el análisis, he creado una función personalizada que devuelve los estimadores presentados en la Sección 5.2 de este capítulo para datos basados en abundancia. Por ahora no abordaremos los estimadores basados en incidencia, ya que estos requieren un formato de entrada distinto y no siempre pueden calcularse si los datos no presentan cierta variación. Si el lector quiere abordar el cálculo basado en datos de incidencia, puede referirse a paquetes como ‘iNEXT’ o ‘fossil’.

```
# Generamos la matriz de datos resumida por grupo (opcional)
BCI_sum <- rowsum(BCI, group = k)

# Función personalizada para obtener los índices de abundancia
# --- Carga una función llamada estimadores_SpadeR()
source(auxfun_1)

# Imprimimos el resultado (para el Grupo 1)
# --- k controla el número de especies raras para ACE/ICE
# --- conf controla el nivel de confianza para los intervalos
estimadores_SpadeR(BCI_sum, k = 10, conf = 0.95)[[1]]
```

	Estimado	s.error	Inferior	Superior
Chao1	177.305	5.327	171.277	194.420
Chao1-bc	176.553	5.011	170.950	192.802
ACE	182.740	4.160	176.568	193.360
ACE-bc	185.265	4.160	178.838	195.503
Jack1	189.995	8.820	178.319	214.882
Jack2	186.003	11.225	173.855	223.353

```
# Calculamos el número de singletons
sing <- rowSums(BCI_sum == 1)

# Calculamos el número de doubletons
doub <- rowSums(BCI_sum == 2)

# Imprimimos la tabla concatenada
head(cbind(sing, doub))
```

	sing	doub
1	22	26
2	21	10
3	35	17
4	25	17
5	28	9

En este ejemplo, todas nuestras unidades de muestreo tienen al menos una especie con una abundancia de dos (*doubletons*), por lo que no es necesario utilizar los estimadores corregidos.

## 11.7. Curvas de acumulación

Para poder graficar la curva de acumulación de especies de nuestra comunidad, primero debemos calcular los valores de diversidad promedio acumulada. Hasta ahora, hemos obtenido valores totales para toda la comunidad, pero las curvas de acumulación, como su nombre lo indica, son acumulativas, por lo que necesitamos calcular la diversidad a medida que se agregan nuevas unidades de muestreo o individuos. Además, como vimos en la Sección 6, las curvas de acumulación suavizadas requieren aleatorizar la matriz.

Desafortunadamente, los paquetes especializados en el análisis de diversidad se quedan cortos de nuevo en cuanto a opciones. Si bien ‘vegan’ nos permite obtener datos permutados para las curvas de acumulación de forma sencilla, nos limita a un par de estimadores y no entrega intervalos de confianza directamente (pero se pueden calcular a partir del error estándar). Por ende, vamos a reciclar y modificar la función personalizada que habíamos creado antes para ‘SpadeR’, pero esta vez, con los 50 sitios de los datos originales.

```
# Función personalizada para obtener los estimadores acumulados
# --- Carga una función llamada estimadores_acum_SpadeR()
source(auxfun_2)

# Calculamos los estimadores
# --- k controla el número de especies raras para ACE/ICE
# --- conf controla el nivel de confianza para los intervalos
# --- perm controla el número de permutaciones
curva_ac <- estimadores_acum_SpadeR(BCI, perm = 100)

# Imprimimos en pantalla (solo las columnas 1, 25 y 50)
lapply(curva_ac, function(df) df[, c(1, 25, 50)])
```

\$promedio	1	25	50
Sobs	91.180	208.530	225.000
Chao1	129.200	232.928	238.884
Chao1-bc	125.888	229.878	237.214
ACE	130.562	222.776	238.218
ACE-bc	146.953	225.855	241.184
Jack1	125.949	231.598	243.999
Jack2	143.508	242.677	249.999
Singletons	34.850	23.070	19.000
Doubletons	17.250	11.990	13.000
Abundancia	433.330	10737.320	21457.000

\$s.error	1	25	50
Sobs	6.905	3.560	0

Chao1	19.383	11.859	0
Chao1-bc	17.882	10.460	0
ACE	14.959	6.310	0
ACE-bc	22.068	7.487	0
Jack1	10.808	6.319	0
Jack2	15.774	10.106	0
Singletons	5.169	3.508	0
Doubletons	3.534	2.928	0
Abundancia	41.543	161.950	0

\$inferior			
	1	25	50
Sobs	89.827	207.832	225.000
Chao1	125.401	230.604	238.884
Chao1-bc	122.383	227.828	237.214
ACE	127.630	221.539	238.218
ACE-bc	142.628	224.387	241.184
Jack1	123.831	230.359	243.999
Jack2	140.416	240.696	249.999
Singletons	33.837	22.382	19.000
Doubletons	16.557	11.416	13.000
Abundancia	425.188	10705.578	21457.000

\$superior			
	1	25	50
Sobs	92.533	209.228	225.000
Chao1	132.999	235.252	238.884
Chao1-bc	129.392	231.928	237.214
ACE	133.494	224.013	238.218
ACE-bc	151.279	227.322	241.184
Jack1	128.067	232.837	243.999
Jack2	146.599	244.658	249.999
Singletons	35.863	23.758	19.000
Doubletons	17.943	12.564	13.000
Abundancia	441.472	10769.062	21457.000

```
# Convertimos los estimadores promedio a data.frame
curva_ac_p <- as.data.frame(t(curva_ac$promedio))
curva_ac_p$Unidades <- seq(1, nrow(curva_ac_p))
curva_ac_p <- curva_ac_p %>%
  pivot_longer(
    cols = -c(Unidades, Abundancia),
    names_to = "Estimador",
    values_to = "Promedio")

# Convertimos los intervalos de confianza a data.frame
curva_ac_i <- as.data.frame(t(curva_ac$inferior))
curva_ac_i$Unidades <- seq(1, nrow(curva_ac_i))
curva_ac_i <- curva_ac_i %>%
  pivot_longer(
    cols = -c(Unidades, Abundancia),
    names_to = "Estimador",
    values_to = "Inferior")

curva_ac_s <- as.data.frame(t(curva_ac$superior))
curva_ac_s$Unidades <- seq(1, nrow(curva_ac_s))
curva_ac_s <- curva_ac_s %>%
```

```

pivot_longer(
  cols = -c(Unidades, Abundancia),
  names_to = "Estimador",
  values_to = "Superior")

# Unimos las tres tablas en un único data.frame formato largo
curva_ac_long <- cbind(curva_ac_p,
  "Inferior" = curva_ac_i$Inferior,
  "Superior" = curva_ac_s$Superior)

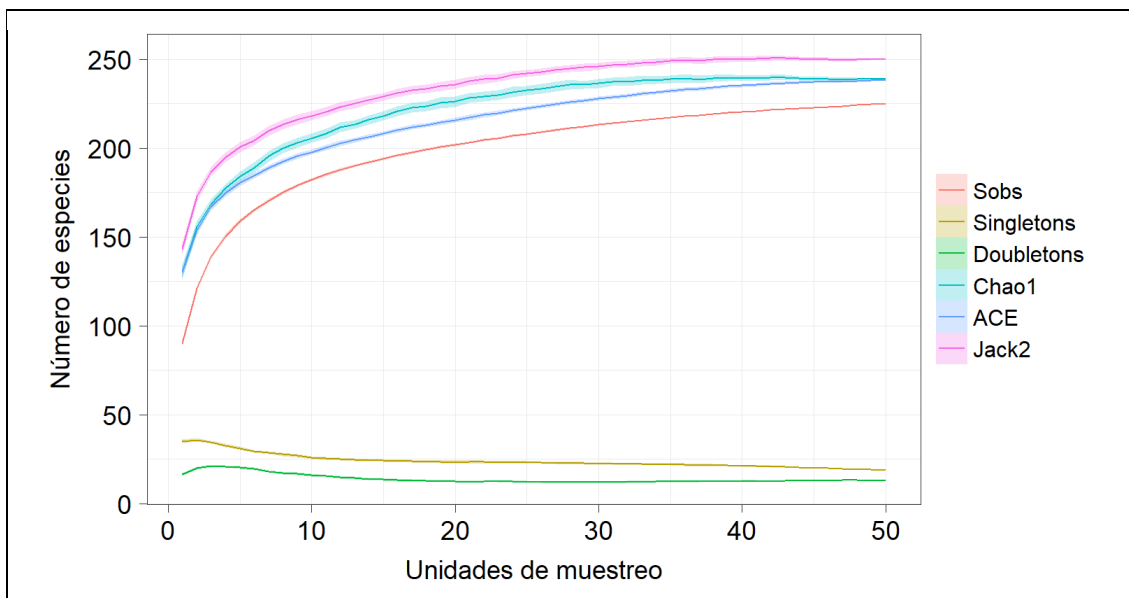
# Filtramos los estimadores a graficar
estimadores <- c("Sobs",
  "Singletons",
  "Doubletons",
  "Chao1",
  "ACE",
  "Jack2")

curva_ac_long <- curva_ac_long %>%
  filter(Estimador %in% estimadores)

# Reordenamos los niveles del factor de los estimadores
curva_ac_long$Estimador <- factor(curva_ac_long$Estimador,
  levels = estimadores)

# Graficamos la curva usando unidades de muestreo acumuladas
# --- Si quisiéramos usar abundancia acumulada: x = Abundancia
ggplot(curva_ac_long, aes(x = Unidades,
  color = Estimador,
  fill = Estimador)) +
  geom_line(aes(y = Promedio), linewidth = 1) +
  geom_ribbon(aes(ymin = Inferior, ymax = Superior),
    alpha = 0.25, linewidth = 0, color = NA) +
  xlab("Unidades de muestreo") +
  ylab("Número de especies") +
  theme_base() +
  theme(legend.position = "right")

```





## 11.8. Curvas de rarefacción

Para el ejercicio de rarefacción y los perfiles de diversidad, cambiaremos de conjunto de datos. Como ya hemos discutido, este método busca comparar comunidades que han sido muestreadas con diferente intensidad, algo que no se cumple con los datos de BCI que hemos venido usando hasta ahora. Por eso, a partir de este punto, emplearemos un subconjunto de los datos presentados por Janzen (1973), correspondiente a dos comunidades de escarabajos tropicales recolectados en dos tipos de bosque: uno maduro (112 especies y 237 individuos) y otro en regeneración (140 especies y 976 individuos).

```
# Cargamos los datos
source(Janzen73)

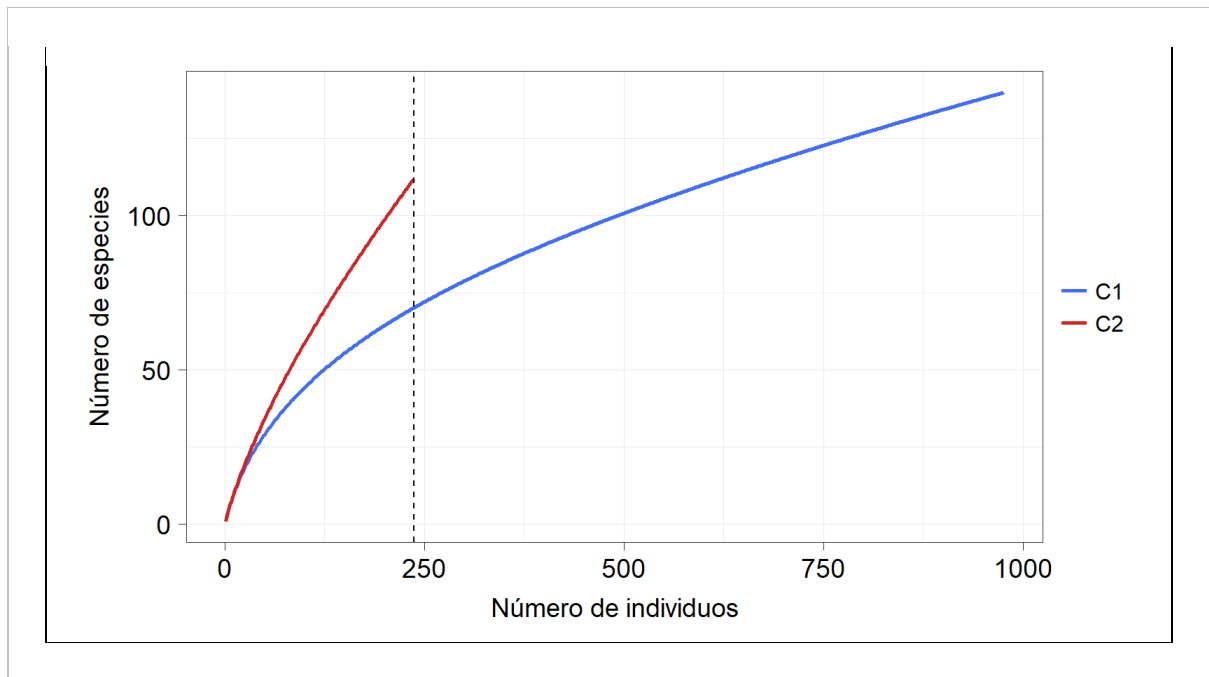
# Imprimimos en pantalla (solo las primeras 15 columnas)
Janzen73[, 1:15]
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
C1	1	1	2	12	77	20	1	1	2	5	3	7	1	1	8
C2	3	1	2	6	1	1	0	3	1	1	0	1	1	0	0

Una vez que tenemos nuestros datos, podemos aplicar la rarefacción fácilmente utilizando el paquete ‘vegan’, el cual genera una curva de acumulación basada en el tamaño de muestra (número de individuos) para cada una de las filas de nuestra matriz de datos.

```
# Generamos la rarefacción (basada en tamaño de muestra)
# --- tidy = TRUE nos da una tabla de datos en lugar de un gráfico
rare_muestra <- rarecurve(Janzen73, tidy = TRUE)

# Graficamos la curva de rarefacción
ggplot(rare_muestra, aes(x = Sample, y = Species, color = Site)) +
  geom_line(linewidth = 2) +
  geom_vline(xintercept = min(rowSums(Janzen73)),
             linewidth = 1,
             linetype = "dashed") +
  scale_color_manual(values = c("royalblue2", "firebrick3")) +
  xlab("Número de individuos") +
  ylab("Número de especies") +
  theme_base() +
  theme(legend.position = "right")
```



Ya sea a partir de la gráfica o de los datos, podemos obtener los valores de riqueza rarificada suponiendo que ambas comunidades tuvieran el mismo número de individuos. No obstante, recordemos que este tipo de rarefacción basada en el tamaño de muestra no es la opción más adecuada, ya que la cobertura de muestra puede diferir entre comunidades.

```
# Obtenemos las abundancias totales para cada comunidad
n <- rowSums(Janzen73)

# Obtenemos el número de especies al mismo tamaño de muestra
rare_muestra[which(rare_muestra$Sample == min(n)), ]

# Calculamos el número de singletons (f1) y doubletons (f2)
f1 <- rowSums(Janzen73 == 1)
f2 <- rowSums(Janzen73 == 2)

# Calculamos la cobertura de muestra para cada comunidad
cobertura <- 1 - (f1/n) * ((n-1) * f1) / (((n-1) * f1) + (2*f2))

# Imprimimos en pantalla
cobertura
```

	C1	C2
	0.9283144	0.6459268

Ya que sabemos que nuestras coberturas de muestra difieren considerablemente, vamos a realizar la rarefacción utilizando el valor mínimo como referencia. Para ello, necesitaremos la librería ‘iNEXT’. Es importante tener en cuenta que, a diferencia de otros paquetes que hemos utilizado, esta librería espera que los datos estén organizados con las especies en las filas y las unidades de muestreo en las columnas, por lo que primero tendremos que transponer nuestra matriz.

```
# Transponemos nuestra matriz de datos
t.Janzen73 <- as.data.frame(t(Janzen73))

# Comando general de iNEXT (calcula muchas cosas)
# --- q es el orden de diversidad, en este caso riqueza (q = 0)
estimadores_iNEXT <- iNEXT(t.Janzen73, q = 0, datatype = "abundance")

# Imprimimos en pantalla
estimadores_iNEXT
```

```
$DataInfo: basic data information
  site  n S.obs   SC f1 f2 f3 f4 f5 f6 f7 f8 f9 f10
1  C1 976  140 0.928 70 17  4  5  5  5  5  3  1  2
2  C2 237  112 0.646 84 10  4  3  5  1  2  1  0  0

$iNextEst: diversity estimates [...]
$C1
      m method order      qD  qD.LCL  qD.UCL    SC SC.LCL SC.UCL
1     1  inter.     0   1.000   1.000   1.000 0.034  0.031  0.037
10    488  inter.     0  99.658  93.377 105.940 0.902  0.889  0.915
20    976   obs.     0 140.000 128.627 151.373 0.928  0.916  0.941
30   1438  extra.     0 169.592 153.601 185.583 0.943  0.929  0.957
40   1952  extra.     0 195.425 173.687 217.162 0.956  0.941  0.971

$C2
      m method order      qD  qD.LCL  qD.UCL    SC SC.LCL SC.UCL
1     1  inter.     0   1.000   1.000   1.000 0.040  0.026  0.054
10   118  inter.     0  66.440  60.968  71.912 0.576  0.527  0.626
20   237   obs.     0 112.000 101.071 122.929 0.646  0.594  0.698
30   349  extra.     0 149.518 133.163 165.872 0.684  0.625  0.743
40   474  extra.     0 186.679 163.456 209.902 0.721  0.654  0.788

$AsyEst: asymptotic diversity estimates [...]
  Site Diversity Observed Estimator  s.e.    LCL    UCL
1  C1  Richness  140.000   283.970  50.474 213.868 420.600
2  C1  Shannon   51.194    59.202   2.939  53.442  64.962
3  C1  Simpson   28.876    29.726   1.632  28.876  32.924
4  C2  Richness  112.000   463.311 136.273 280.660 843.766
5  C2  Shannon   58.277   123.467  20.392  83.499 163.435
6  C2  Simpson   22.549    24.815   4.976  22.549  34.568
```

La salida de esta función es una lista compuesta por tres objetos. El primero (`$DataInfo`) contiene información general sobre la comunidad, como el número de especies, la abundancia total, y el número de *singletons* y *doubletons*. El segundo (`$iNextEst`) presenta los estimadores de riqueza a diferentes niveles de cobertura (columna SC), incluyendo tanto el valor observado como los valores interpolados y extrapolados. Finalmente, el tercer objeto (`$AsyEst`) incluye los índices de diversidad verdadera asintóticos para nuestras comunidades.

Para realizar la rarefacción basada en la cobertura de muestra, utilizaremos los estimadores que calculamos con ‘iNEXT’. Por defecto, esta función genera estimaciones para 40 valores de abundancia, desde uno hasta el doble del número total de individuos por comunidad, usando intervalos igualmente espaciados. Sin embargo, al dividir esos 40 puntos entre el rango total de abundancias, es posible que no obtengamos exactamente la cobertura deseada para rarificar. Por ello, podemos incluir el parámetro adicional `knots` en la función, el cual permite especificar el número de puntos de estimación. Por ejemplo, si una comunidad tiene 976 individuos, la función estimará por defecto hasta 1952 individuos (el doble), usando 40 pasos. En cambio, si establecemos `knots = 1952`, obtendremos estimaciones de uno en uno, es decir, con intervalos de una unidad, lo que nos da mayor precisión para acercarnos a la cobertura deseada.

```
# Obtenemos el valor máximo a extrapolar por comunidad
maxExtrap <- n * 2

# Calculamos los estimadores para cada comunidad con knots por unidad
estimadores_C1 <- iNEXT(t.Janzen73[, 1],
                        q = 0,
                        datatype = "abundance",
                        knots = maxExtrap[1])

estimadores_C2 <- iNEXT(t.Janzen73[, 2],
                        q = 0,
                        datatype = "abundance",
                        knots = maxExtrap[2])
```

Una vez hecho esto, podemos extraer el valor de riqueza estimado para nuestras dos comunidades suponiendo que ambas hubiesen sido muestreadas al mismo nivel de cobertura, que en este caso es del 65%. Debido a cuestiones de redondeo decimal, es posible que

obtenemos más de un valor de riqueza asociado a una misma cobertura dentro de una comunidad, en ese caso, simplemente tomamos el promedio.

```
# Filtramos los valores de ambas comunidades al 65% de cobertura
cobertura_C1 <- dplyr::filter(estimadores_C1$iNextEst,
                             round(SC, 2) == 0.65)

cobertura_C2 <- dplyr::filter(estimadores_C2$iNextEst,
                             round(SC, 2) == 0.65)

# Imprimimos en pantalla
head(cobertura_C1)
head(cobertura_C2)
```

	m	method	order	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
1	55	inter.	0	30.901	29.703	32.099	0.646	0.624	0.668
2	56	inter.	0	31.255	30.036	32.473	0.650	0.628	0.672
3	57	inter.	0	31.605	30.366	32.844	0.653	0.631	0.675

	m	method	order	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
1	236	inter.	0	111.646	98.021	125.270	0.646	0.578	0.713
2	237	obs.	0	112.000	98.312	125.688	0.646	0.578	0.714
3	238	extra.	0	112.354	98.602	126.106	0.646	0.578	0.714
4	239	extra.	0	112.708	98.892	126.524	0.647	0.579	0.715
5	240	extra.	0	113.061	99.181	126.941	0.647	0.579	0.715
6	241	extra.	0	113.414	99.470	127.358	0.647	0.579	0.716

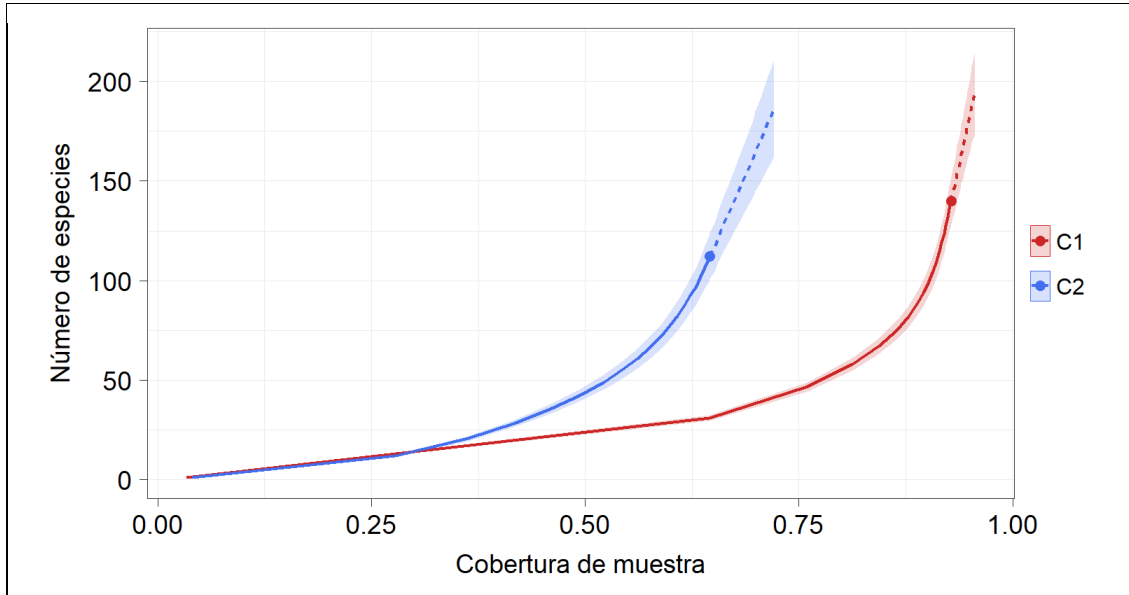
```
# Calculamos los valores promedio de la riqueza al 65%
cbind(C1 = mean(cobertura_C1$qD), C2 = mean(cobertura_C2$qD))
```

	C1	C2
[1,]	31.25367	116.0405

Posteriormente, podemos representar gráficamente el resultado de la rarefacción basada en cobertura, incluyendo los valores interpolados, extrapolados y sus respectivos intervalos de confianza.

```
# Generamos la rarefacción (basada en cobertura de muestra)
# --- Curva basada en cobertura: type 3
# --- Curva basada en tamaño: type 1
ggiNEXT(estimadores_iNEXT, type = 3) +
  scale_color_manual(values = c("C1" = "firebrick3",
                                "C2" = "royalblue2")) +
  scale_fill_manual(values = c("C1" = "firebrick3",
                                "C2" = "royalblue2")) +
  scale_shape_manual(values = c("C1" = 16, "C2" = 16)) +
```

```
xlab("Cobertura de muestra") +
ylab("Número de especies") +
guides(linetype = "none") +
theme_base() +
theme(
  legend.position = "right",
  legend.key.width = unit(0.8, "cm"),
  legend.key.spacing.y = unit(0.5, "cm"))
```



### 11.9. Perfiles de diversidad

Para terminar esta primera parte, lo único que nos resta es calcular los perfiles de diversidad, para lo cual necesitamos los tres órdenes de diversidad verdadera con significado biológico. No obstante, para darle a la gráfica un acabado más suave, calcularemos algunos órdenes adicionales en intervalos de 0.2, desde el orden cero hasta el orden dos. Para los perfiles de diversidad es importante utilizar el mismo esfuerzo de muestreo, en este caso, como cobertura de la muestra. De lo contrario, como ya vimos, las conclusiones podrían ser erradas.

```
# Calculamos el número de ind. de cada comunidad a la misma cobertura
ind_C1 <- Coverage2Size(t.Janzen73[, 1],
                        SampleCoverage = min(cobertura))
ind_C2 <- Coverage2Size(t.Janzen73[, 2],
                        SampleCoverage = min(cobertura))

# Calculamos los estimadores para cada comunidad a la misma cobertura
estimadores_C1 <- iNEXT(t.Janzen73[, 1],
```

```

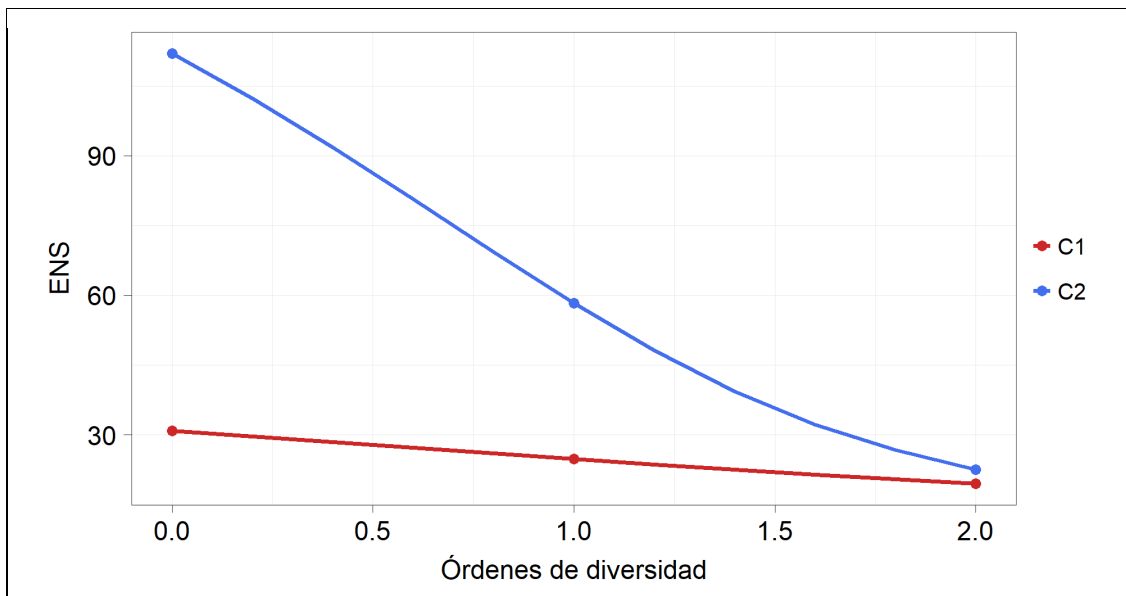
      q = seq(0, 2, by = 0.2),
      datatype = "abundance",
      size = c(0, ind_C1))$iNextEst %>%
      filter(m != 0) %>% # Eliminamos los falsos 0
      mutate(Comunidad = "C1")

estimadores_C2 <- iNEXT(t.Janzen73[, 2],
      q = seq(0, 2, by = 0.2),
      datatype = "abundance",
      size = c(0, ind_C2))$iNextEst %>%
      filter(m != 0) %>% # Eliminamos los falsos 0
      mutate(Comunidad = "C2")

# Combinamos los dos data frames
estimadores_rare <- rbind(estimadores_C1, estimadores_C2)

# Generamos los perfiles de diversidad
ggplot(estimadores_rare, aes(x = order, y = qD, color = Comunidad)) +
  geom_line(linewidth = 2) +
  geom_point(data = estimadores_rare
    %>% filter(order %in% c(0, 1, 2)),
    aes(x = order, y = qD, color = Comunidad),
    size = 5) +
  scale_color_manual(values = c("C1" = "firebrick3",
                                "C2" = "royalblue2")) +
  xlab("Órdenes de diversidad") +
  ylab("ENS") +
  theme_base() +
  theme(
    legend.position = "right",
    legend.key.width = unit(0.8, "cm"),
    legend.key.spacing.y = unit(0.5, "cm"))

```



### 11.10. Recambio y anidamiento

Lo último que nos queda por hacer en este tutorial es generar algunas métricas y gráficos relacionados con la diversidad beta ( $\beta$ ), con el fin de explorar cómo varía la composición de especies entre comunidades. Para esta sección, utilizaremos un conjunto de datos diferente, correspondiente a una serie de comunidades de peces muestreadas a lo largo del río Doubs, en Francia. A diferencia de los conjuntos de datos que utilizamos para calcular la diversidad alfa ( $\alpha$ ), esta matriz incluye al menos un sitio en el que no se registró ninguna especie, lo cual puede generar problemas al momento de realizar los análisis. Por ello, es fundamental revisar y depurar nuestras matrices antes de proceder, eliminando aquellos sitios que contienen únicamente ceros.

```
# Cargamos el conjunto de datos de peces del rio Doubs
data(doubs, package = "ade4")

# Eliminamos los sitios sin datos
doubs <- doubs$fish[rowSums(doubs$fish) > 0,]
```

Para calcular los índices de diversidad beta ( $\beta$ ) podemos usar la función `vegdist()` del paquete ‘vegan’, que permite generar matrices pareadas de forma rápida y sencilla para una gran variedad de índices de similitud o medidas de distancia.

```
# Calculamos la matriz de distancias
# --- Usando ?vegdist podemos ver todos los métodos disponibles
# --- Por defecto, vegdist entrega índices de disimilitud
beta <- vegdist(doubs, method = "jaccard")

# Imprimimos en pantalla la matriz (primeras 5 filas y columnas)
as.matrix(beta)[1:5, 1:5]
```

	1	2	3	4	5
1	0.0000000	0.7500000	0.8125000	0.8571429	0.9428571
2	0.7500000	0.0000000	0.2500000	0.5000000	0.8205128
3	0.8125000	0.2500000	0.0000000	0.3181818	0.8095238
4	0.8571429	0.5000000	0.3181818	0.0000000	0.6585366
5	0.9428571	0.8205128	0.8095238	0.6585366	0.0000000

Si por el contrario quisiéramos calcular la diversidad beta ( $\beta$ ) particionada, podemos usar las funciones `beta.multi()` y `beta.pair()` del paquete ‘betapart’. La diferencia entre ambas



funciones es que la primera nos da los valores totales de diversidad, mientras que la segunda, genera matrices de disimilitud para cada par de sitios.

```
# Convertimos nuestra matriz a binaria (solo si tenemos abundancias)
doubts_bin <- decostand(doubts, method = "pa")

# Calculamos la diversidad beta particionada
# --- index.family = "jaccard" o index.family = "sorensen"
beta_part <- beta.multi(doubts_bin, index.family = "jaccard")

# Imprimimos en pantalla el resultado
# --- La letra J (Jaccard) o S (Sørensen) indica el índice
# --- TU = turnover, NE = nestedness, AC = overall dissimilarity
beta_part
```

```
$beta.JTU
[1] 0.7885784

$beta.JNE
[1] 0.1470249

$beta.JAC
[1] 0.9356033
```

### 11.11. Dendrogramas de similitud

A continuación, podemos generar un dendrograma de similitud utilizando alguno de los índices que ya hemos visto, con el fin de representar gráficamente la similitud en la composición de nuestras comunidades de estudio. En los dendrogramas, la cercanía entre ramas indica similitud en el número de especies y/o en sus abundancias, según el índice utilizado. Esta clasificación jerárquica permite identificar visualmente ciertos grupos de sitios que se parecen más entre sí que a otros.

```
# Calculamos la matriz de distancias
# --- binary = TRUE convierte los datos de abundancia a binarios
jaccard_dist <- vegdist(doubts, method = "jaccard", binary = TRUE)

# Generamos el cluster jerárquico
jaccard_clust <- hclust(jaccard_dist, method = "average")

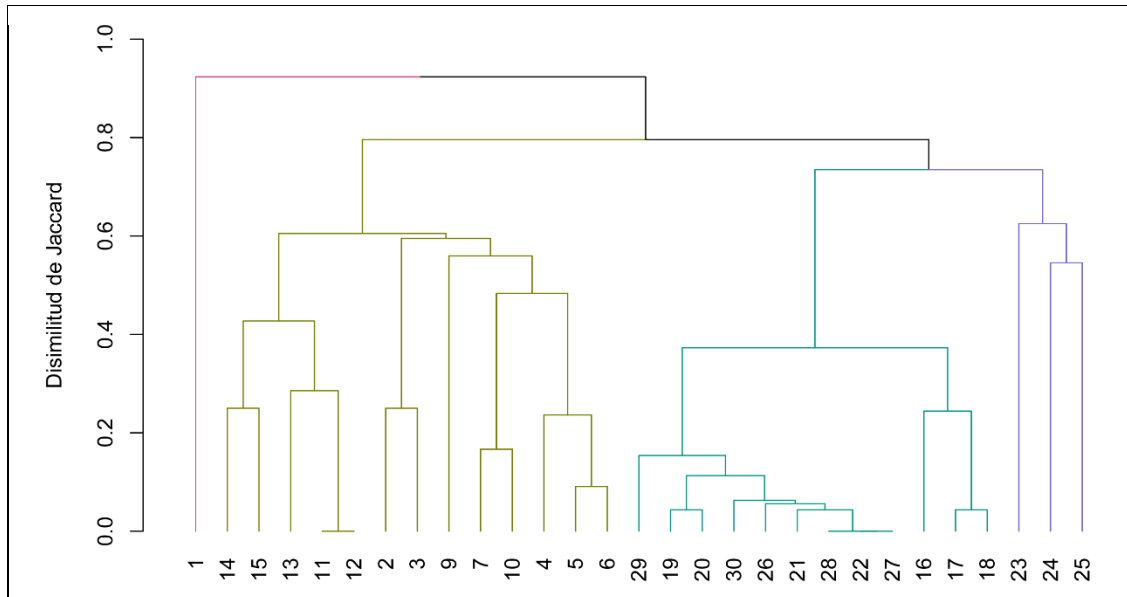
# Convertimos el cluster a dendrograma
dend <- as.dendrogram(jaccard_clust)

# Definimos el número de grupos que deseamos (opcional)
# --- El valor de k es subjetivo y genera un número k de grupos
```

```
# --- según la estructura del árbol
k <- 4

# Coloreamos el árbol según el número de grupos
dend_col <- color_branches(dend, k = k)

# Graficamos el dendrograma
plot(dend_col, ylim = c(0, 1), ylab = "Disimilitud de Jaccard")
```



## 11.12. NMDS

Finalmente, pasamos al terreno de los análisis multivariados, donde calcularemos un NMDS que nos permitirá visualizar de forma complementaria al dendrograma que ya hemos generado, la similitud en composición y/o abundancia de nuestras comunidades y hacer inferencias al respecto.

```
# Definimos una semilla (para la reproducibilidad del código)
set.seed(123)

# Realizamos el escalamiento multidimensional (NMDS)
# --- Si queremos considerar abundancias distance = "bray"
# --- Si no queremos considerar abundancias distance = "jaccard"
nmds <- metaMDS(doubs, distance = "bray", k = 2, trymax = 100)

# Validamos el valor de estres
nmds$stress
```

```
[1] 0.07376227
```

```

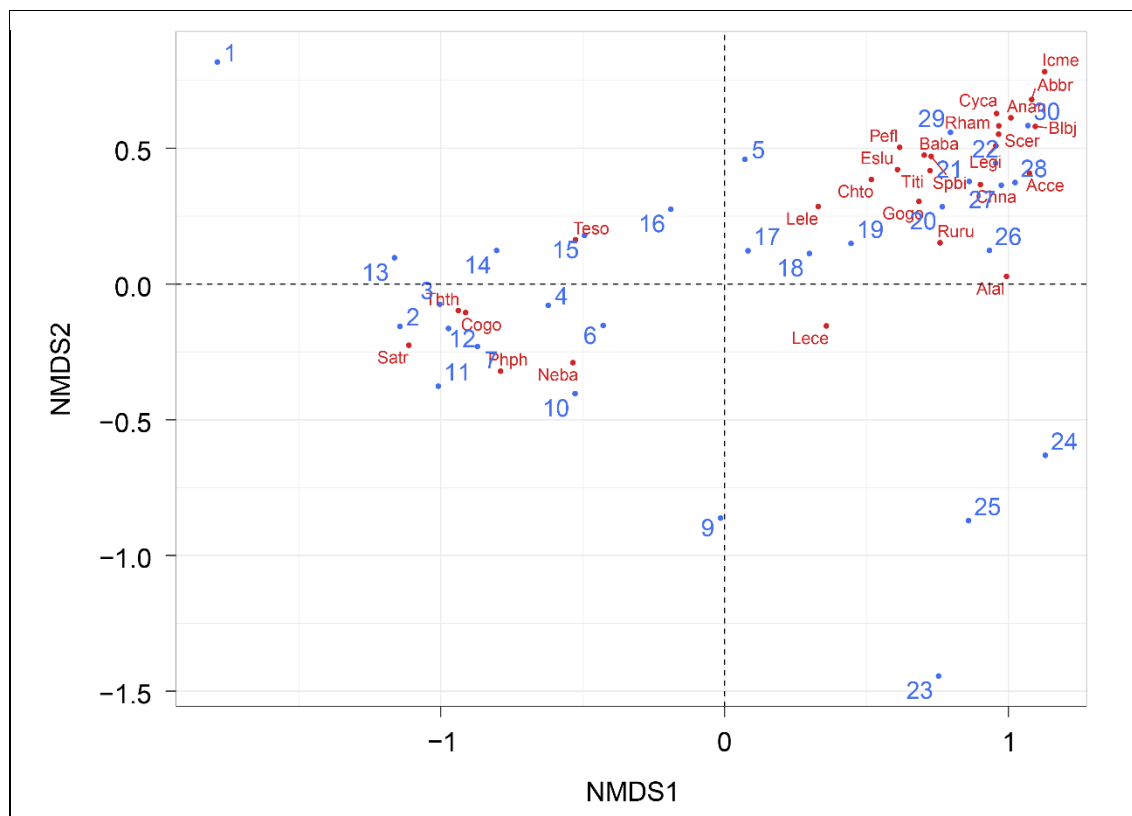
# Extraemos las coordenadas del NMDS para los sitios y las especies
sitios <- as.data.frame(scores(nmds, display = "sites"))
especies <- as.data.frame(scores(nmds, display = "species"))

# Agregamos a las tablas los grupos y las etiquetas de los datos
sitios$Grupo <- "Sitios"
especies$Grupo <- "Especies"
sitios$Etiqueta <- rownames(sitios)
especies$Etiqueta <- rownames(especies)

# Convertimos la tabla a formato largo
doubts_long <- rbind(sitios, especies)

# Graficamos el NMDS
ggplot(doubts_long, aes(x = NMDS1, y = NMDS2, color = Grupo)) +
  geom_point() +
  geom_text_repel(data = subset(doubts_long, Grupo == "Especies"),
                  aes(label = Etiqueta),
                  size = 5,
                  max.overlaps = Inf) +
  geom_text_repel(data = subset(doubts_long, Grupo == "Sitios"),
                  aes(label = Etiqueta),
                  size = 7,
                  max.overlaps = Inf) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  scale_color_manual(values = c("firebrick3", "royalblue2")) +
  theme_base() +
  theme(legend.position = "none")

```



## Bibliografía

- Baselga, A. (2010a). Multiplicative partition of true diversity yields independent alpha and beta components; additive partition does not. *Ecology*, 91(7), 1974–1981.  
<https://doi.org/10.1890/09-0320.1>
- Baselga, A. (2010b). Partitioning the turnover and nestedness components of  $\beta$  diversity. *Global Ecology and Biogeography*, 19(1), 134–143. <https://doi.org/10.1111/j.1466-8238.2009.00490.x>
- Chao, A. (1984). Non-parametric estimation of the classes in a population. *Scandinavian Journal of Statistics*, 11(4), 265–270.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43(4), 783. <https://doi.org/10.2307/2531532>
- Chao, A. (2005). *Species estimation and applications*. In: Kotz, S., Read, C. B., Balakrishnan, N. & Vidakovic, B. (Eds.). *Encyclopedia of statistical sciences*. 2nd. Vol. 12. pp. 7907–7916.
- Chao, A., Chiu, C. H. & Jost, L. (2010). Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3599–3609. <https://doi.org/10.1098/rstb.2010.0272>
- Chao, A., Chiu, C. H. & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 297–324.  
<https://doi.org/10.1146/annurev-ecolsys-120213-091540>
- Chao, A. & Jost, L. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, 93(12), 2533–2547.  
<https://doi.org/10.1890/11-1952.1>
- Chao, A. & Lee, S. M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, 87(417), 210–217.  
<https://doi.org/10.1080/01621459.1992.10475194>
- Colwell, R. K. (2019). EstimateS 9.1 User's Guide. EstimateS. Fecha de acceso: 2 de junio, 2025. Disponible en: <https://www.robertkcolwell.org>
- Colwell, R. K. & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 345(1311), 101–118. <https://doi.org/10.1098/rstb.1994.0091>

- Daly, A., Baetens, J. & Baets, B. D. (2018). Ecological diversity: Measuring the unmeasurable. *Mathematics*, 6(7), 119. <https://doi.org/10.3390/math6070119>
- Hammer, Ø., Harper, D. A. T. & Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1), 1–9.
- Hortal, J., Borges, P. A. V. & Gaspar, C. (2006). Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*, 75(1), 274–287. <https://doi.org/10.1111/j.1365-2656.2006.01048.x>
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432. <https://doi.org/10.2307/1934352>
- Hunter, P. R. & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpsons index of diversity. *Journal of Clinical Microbiology*, 26(11), 2465–2466. <https://doi.org/10.1128/jcm.26.11.2465-2466.1988>
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x>
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), 2427–2439. <https://doi.org/10.1890/06-1736.1>
- Jost, L. (2010). The relation between evenness and diversity. *Diversity*, 2(2), 207–232. <https://doi.org/10.3390/d2020207>
- Koleff, P., Gaston, K. J. & Lennon, J. J. (2003). Measuring  $\beta$  diversity for presence–absence data. *Journal of Animal Ecology*, 72(3), 367–382. <https://doi.org/10.1046/j.1365-2656.2003.00710.x>
- Magurran, A. E. (2004). *Measuring biological diversity*. Blackwell: Oxford, United Kingdom.
- Margalef, R. (1958). Information theory in ecology. *General Systems*, 3, 36–71.
- Menhinick, E. F. (1964). A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology*, 45(4), 859–861. <https://doi.org/10.2307/1934933>
- Oksanen, J., Simpson, G. L., Blanchet, F. G. *et al.* (2024). vegan: Community Ecology Package. R package version 2.6-8. Disponible en: <https://CRAN.R-project.org/package=vegan>
- Palacio, F. X., Apodaca, M. J. & Crisci, J. V. (2020). *Análisis multivariado para datos biológicos: teoría y su aplicación utilizando el lenguaje R*. Fundación de Historia Natural Félix de Azara: Buenos Aires, Argentina.

- Peet, R. K. (1974). The Measurement of Species Diversity. *Annual Review of Ecology and Systematics*, 5, 285–307.
- Pielou, E. C. (1975). *Ecological Diversity*. Wiley: New York.
- Pla, L., Casanoves, F. & Di Rienzo, J. (2012). *Quantifying Functional Biodiversity*. Springer: New York.
- Rajaram, R., Castellani, B. & Wilson, A. N. (2017). Advancing Shannon entropy for measuring diversity in systems. *Complexity*, 2017, 1–10.  
<https://doi.org/10.1155/2017/8715605>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(4148), 688–688.  
<https://doi.org/10.1038/163688a0>
- Smith, W. & Grassle, J. F. (1977). Sampling properties of a family of diversity measures. *Biometrics*, 33(2), 283. <https://doi.org/10.2307/2529778>
- Tucker, C., Cadotte, M W., Carvalho, S. B. *et al.* (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92, 698–715. <https://doi.org/10.1111/brv.12252>
- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279–338. <https://doi.org/10.2307/1943563>
- Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, 21(2/3), 213–251. <https://doi.org/10.2307/1218190>