

Análisis de comunidades ecológicas

Octubre 21, 2024

FECHA DE ENTREGA	Noviembre 4, 2024
MODALIDAD	Individual
FORMATO	Digital (Apellidos_Comunidades.pdf)

Objetivos

1. Reconocer los principales índices utilizados en ecología para representar la diversidad de una comunidad.
2. Aprender a manejar el paquete de funciones *vegan*.
3. Explorar repositorios de datos ecológicos para identificar posibles fuentes de material para análisis.

Preparación

Trabajo en la nube Todo el trabajo que realizaremos en el curso con relación a R puede llevarse a cabo en la plataforma <https://rstudio.cloud>. Los servicios ofrecidos a través de la cuenta gratuita son suficientes para los talleres que realizaremos. En caso de no tener espacio en el computador personal para instalar R/RStudio, esta es una buena alternativa.

Instalación local R es de código abierto y de acceso libre. Se puede descargar para instalar del enlace <https://cran.r-project.org/> de acuerdo al sistema operativo propio. RStudio, un *Interactive Development Environment (IDE)* para R, se ofrece en una versión de código abierto y de libre acceso, la cual se puede obtener de <https://rstudio.com/products/rstudio/>. Es importante haber instalado R antes de intentar instalar RStudio.

Motivación

Conocer la diversidad de una comunidad biótica es uno de los primeros pasos para explorar los procesos que llevan a esta, y el cómo la diversidad manifiesta refleja procesos ecosistémicos.

La principal razón por la que utilizamos R en este curso es la amplia disponibilidad de paquetes de funciones enfocados en análisis ecológicos. En CRAN (el repositorio de paquetes de R) podemos consultar por tareas para explorar los paquetes disponibles dentro de un universo en continua expansión. En <https://cran.r-project.org/web/views/>

hagan click en *Environmetrics*. Este recurso muestra todos los paquetes de R que están enfocados en análisis ecológicos y ambientales, y sus posibles aplicaciones. Trabajaremos principalmente con *vegan*.

```
1 if (!require("vegan")) {
2   install.packages("vegan")
3   library(vegan)
4 }
```

El estimado más sencillo de diversidad empieza por contar el número de distintas especies. Este nos indica cuántas categorías distintas (de acuerdo a como hayamos definido las categorías) existen en nuestra muestra, pero ignora tanto la distribución de las abundancias relativas de cada categoría, como la identidad de las mismas. A esta variable la vamos a denominar **Riqueza**, y a asignar el símbolo S (pueden recordarlo pensando que S simboliza *Species*). En la tabla vemos cómo para dos comunidades de estructura muy distinta, el valor de S es idéntico.

Especie	Com. 1	Com. 2
A	1	16
B	10	0
C	21	18
D	59	21
E	4	25
F	0	15

Aunque con los datos de la tabla es fácil contar el número de especies presentes, con datos más masivos resulta engorroso (y propenso a error) realizar el cálculo a ojo. Afortunadamente en *vegan* podemos encontrar una función que realiza dicho cálculo partiendo de una tabla de abundancias o de frecuencias. Para explorar estos comandos, vamos primero a hacer uso de otro gran recurso en *vegan*, los conjuntos de datos que este paquete proporciona.

El primer conjunto de datos que utilizaremos lleva el nombre de *BCI*, las iniciales de *Barro Colorado Island*. Esta es una estación biológica experimental manejada por el *Smithsonian Institute* en una isla ubicada a lo largo del Canal de Panamá.

```
1 ?BCI # para ver la descripción del conjunto de datos
2 data("BCI") # para cargar el conjunto de datos a una variable
```

Al cargar el conjunto de datos *BCI*, obtenemos una variable del mismo nombre que contiene una estructura *data frame*. Esta se compone de especies en las filas, y parcelas en las columnas. Los números en cada una de las celdas indican el conteo de individuos presentes de cada especie en la parcela correspondiente.

La función para contar el número de especies es *specnumber()*, a la cual se le puede entregar el *data frame* que generamos anteriormente, así:

```
1 S <- specnumber(BCI)
```

Con una sola línea de código, hemos obtenido el conteo de el número de especies en cada una de las 50 parcelas de *BCI*.

Índices de diversidad

El número de especies, hemos declarado, es un indicador incompleto de la diversidad de cada parcela. La **Diversidad** formalmente definida tiene en cuenta también la distribución de las abundancias de cada especie. Para expresar diversidad existen varios índices. El primero que exploraremos es el índice de **Shannon-Wiener**, definido como:

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

en donde p_i es la abundancia relativa de especie i ; y S , cómo ya lo habíamos mencionado, corresponde a la riqueza (número total de especies) de la comunidad a evaluar. Por convención, este índice suele simbolizarse H' .

Tomemos las comunidades de la tabla para calcular el índice de **Shannon-Wiener**.

```
1 com1 <- c(1, 10, 21, 59, 4, 0) # vector de abundancias de la
  comunidad 1
2 com2 <- c(16, 0, 18, 21, 25, 15) # vector de abundancias de la
  comunidad 2
3 # para calcular la abundancia relativa de cada una de las especies ...
4 p1 <- com1/sum(com1)
5 p2 <- com2/sum(com2)
6 # y ahora calculamos H'
7 H1 <- -sum(p1*log(p1))
8 H1
```

Algo no anda bien. El resultado de H1 nos da *Not a Number* (NaN). El problema aquí es que en el vector de ambas comunidades tenemos especies con abundancia 0, y el logaritmo de 0 no está definido. Entonces, para calcular H' , primero debemos eliminar los elementos de valor 0, así:

```
1 p1 <- p1[p1 != 0]
2 H1 <- -sum(p1*log(p1))
3 #
4 p2 <- p2[p2 != 0]
5 H2 <- -sum(p2*log(p2))
```

El procedimiento parece sencillo. Pero, nuevamente si tuviéramos un alto número de comunidades por evaluar, el proceso sería tedioso. Podríamos escribir una función personalizada, o utilizar la función `diversity()` de *vegan*. Para calcular la diversidad de ambas comunidades, podríamos escribir:

```
1 comunidades <- rbind(com1, com2)
2 diversity(comunidades)
```

Pregunta : ¿Concuerdan los estimados con los que obtuvimos anteriormente de manera manual?

La función `diversity()` es la principal herramienta para calcular índices de diversidad en *vegan*. Al consultar la documentación de la función, encontrarán que uno de los parámetros (*index*) permite indicar cual de los índices se busca aplicar. Entre estos se incluye el índice de **Simpson**, que en su sentido original se expresa como:

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

Esta expresión indica la probabilidad de que dos individuos seleccionados al azar de la misma comunidad correspondan a la especie i , cuyos individuos representan el p_i de la comunidad.

La formulación original del índice de *Simpson* resulta en una expresión poco intuitiva, pues para la mayor diversidad posible (cuando todas las poblaciones se encuentren uniformemente representadas), el índice alcanzaría su valor mínimo de $D = 1/S$. Una formulación un poco más intuitiva se obtiene con la transformación $1 - D$, denominada el índice *Gini-Simpson*. Si el índice de *Simpson* indicaba la probabilidad de encuentros *intra*-específicos, el índice *Gini-Simpson* representa la probabilidad de encuentros *inter*-específicos. El índice *Gini-Simpson* se ha convertido en la versión más comúnmente utilizada, y frecuentemente se le encuentra como el índice de *Simpson*, aunque su origen sea más complejo. Una expresión alternativa del índice de *Gini-Simpson* que encontrarán, parte del número de individuos y no de su abundancia relativa, de tal modo que:

$$D = 1 - \frac{\sum_{i=1}^S n_i(n_i - 1)}{N(N - 1)} \quad (3)$$

donde n_i representa el número de individuos de la especie i ; y N el número total de individuos en la muestra. Expresar de este modo el índice facilita su interpretación, pues un valor de 0 indica una comunidad compuesta de una sola especie (0 diversidad), mientras que un valor de 1 correspondería a diversidad infinita (lo cual realmente es una asíntota del índice).

Un índice adicional que no se encuentra codificado en la función *diversity()* es el índice de *equitatividad* o índice de *Pielou*, el cual se expresa como:

$$J' = \frac{H'}{H'_{max}} \quad (4)$$

donde H' corresponde al índice de *Shannon-Wiener* y H'_{max} , el máximo valor del mismo índice para la comunidad en cuestión, $H'_{max} = \ln S$. Este índice también alcanza un valor máximo $J' = 1$ cuando todas las poblaciones de la comunidad se encuentran representadas por el mismo número de individuos.

Diversidad vs. Entropía - Números de Hill

En la primera parte de *Análisis de Comunidades* exploramos algunos índices que se utilizan frecuentemente para la descripción de la diversidad de comunidades. Estos índices buscan integrar aspectos de *riqueza* de poblaciones (grupos o especies) y de la *uniformidad* con que estas se encuentran representadas dentro de la comunidad. Entre la multiplicidad de índices que existen, exploramos el índice de *Shannon-Wiener* (H' ; Shannon, 1948), el índice de *equitatividad de Pielou* (J), y el índice de *Gini-Simpson* (D ; Simpson, 1949).

Índices como el de Shannon-Wiener o el de Gini-Simpson en realidad son índices de entropía, no de diversidad. En parte por esto, los índices presentan dificultades en lo poco intuitivo de sus valores, su deficiente comportamiento estadístico, o su falta de cambio proporcional a los cambios en composición de especies de dos comunidades.

Una expresión unificada de diversidad que supera todos estos problemas es el conjunto de **Números de Hill**, introducidos por Robert MacArthur en 1965 (en concepto, pues su nombre se deriva del trabajo de 1973 de O. M. Hill), pero popularizados por Lou Jost en la primera década del siglo XXI (Jost, 2006). La formulación general es:

$${}^qD = \left(\sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad (5)$$

Como en todas las ocasiones anteriores, S representa el número de especies en la muestra (*riqueza*), y p_i la abundancia relativa de la especie i ($p_i = n_i/N$). El exponente q es un entero positivo, el cual define el número de Hill del que se esté tratando (p. ej., cuando $q = 1$, hablamos del número 1 de Hill). A medida que q aumenta, la expresión de diversidad resultante asigna creciente importancia a las especies comunes y descuenta la influencia de las especies raras.

A diferencia de los demás índices, cuyos valores carecen de unidades o de formas claras de expresarlos, los valores de los número de Hill se expresan como el **número efectivo de especies**. Este concepto se entiende como *el número equivalente de especies, presentes todas en abundancias relativas iguales*, es decir con una distribución uniforme de las abundancias relativas.

Aunque técnicamente no hay límite para los números de Hill (ni tampoco prohibición en asignar valores negativos a q), los valores más frecuentemente utilizados son $q = \{0, 1, 2\}$. Para $q = 0$,

$${}^0D = \left(\sum_{i=1}^S p_i^0 \right)^{1/(1-0)} = S \quad (6)$$

Es decir, 0D equivale a la riqueza (número de especies) de la comunidad. Cuando $q = 1$, la ecuación 5 no se puede resolver de manera analítica (el exponente $1/(1-q)$ no está definido), pero se puede calcular de manera numérica, y en el límite se acerca a:

$${}^1D = e^{(-\sum_{i=1}^S p_i \ln p_i)} = e^{H'} \quad (7)$$

que corresponde al exponencial del índice de Shannon. Cuando $q = 2$,

$${}^2D = \left(\sum_{i=1}^S p_i^2 \right)^{-1} = \frac{1}{\left(\sum_{i=1}^S p_i^2 \right)} \quad (8)$$

es decir, el número 2 de Hill (2D) es equivalente al inverso del índice de *Simpson*. Aunque los números de Hill son indicadores robustos de diversidad, no obstante padecen de la influencia del esfuerzo de muestreo en sus estimaciones. Es a ese aspecto al que ahora tornamos.

Rarefacción

La *Relación Especies Área* (*Species Area Relationship, SAR*) nos dice que dentro de un hábitat contiguo, la riqueza estimada dependerá del área (o esfuerzo) de muestreo, de modo que:

$$S = cA^z \quad (9)$$

en donde S (nuevamente) representa la riqueza o número de especies, y A es equivalente al área muestreada; c es una constante determinada por la unidad de muestreo, y z es la pendiente en la relación. Siendo esto así, ¿cómo podemos comparar la diversidad o la riqueza entre comunidades, en las que condiciones logísticas y ambientales pueden traducirse en diferencias en el esfuerzo de muestreo?

Para estandarizar las estimaciones de riqueza empleamos un método estadístico de submuestreo llamado **rarefacción**. A partir de este construimos *curvas de rarefacción*, las cuales representan simulaciones de los estimados de especies basado en submuestreos de distintas profundidades.

Tomemos por ejemplo el conjunto de datos BCI del paquete *vegan*. Si consideramos las 50 hectáreas como la comunidad total, tendríamos una riqueza de 225 especies, y una abundancia total de cada especie que obtenemos sumando las abundancias de cada especie a lo largo de las distintas submuestras.

```
1 library(vegan)           # Ya el paquete debe estar instalado.
2 data(BCI)
3 Nspe <- colSums(BCI)      # Abundancia total de cada especie.
4 Ntot <- sum(Nspe)         # Número total de individuos.
5 #
6 # Recuerden que pueden indagar sobre cada función precediéndola de ?
```

Podríamos generar entonces una serie de submuestreos, obteniendo un subconjunto de individuos registrados, así:

```
1 subsamples <- c(1, seq(from=100, to=21400, by=100), Ntot)
```

En esta serie de submuestreos, incluimos un muestreo de un sólo individuo, y un submuestreo que equivale a la totalidad de todos los individuos registrados. Ahora utilizamos la función *rarefy()*, la cual va a realizar replicados de submuestreos de los tamaños que le indicamos en el objeto que creamos anteriormente (*subsamples*), para generar el estimado de especies a cada nivel y el intervalo de error.

```
1 BCI_rar <- rarefy(Nspe, sample=subsamples, se=T, MARGIN=2)
2 #
3 plot(subsamples, BCI_rar[1,], type="n", las=1,
4       xlab="Número de individuos", ylab="Número de especies (S)",
5       main="Curva de Rarefacción para BCI")
6 lines(subsamples, BCI_rar[1, ], type = "b", pch=20, cex=0.5)
7 lines(subsamples, BCI_rar[1, ] + 2 * BCI_rar[2, ], lty = 3)
8 lines(subsamples, BCI_rar[1, ] - 2 * BCI_rar[2, ], lty = 3)
```

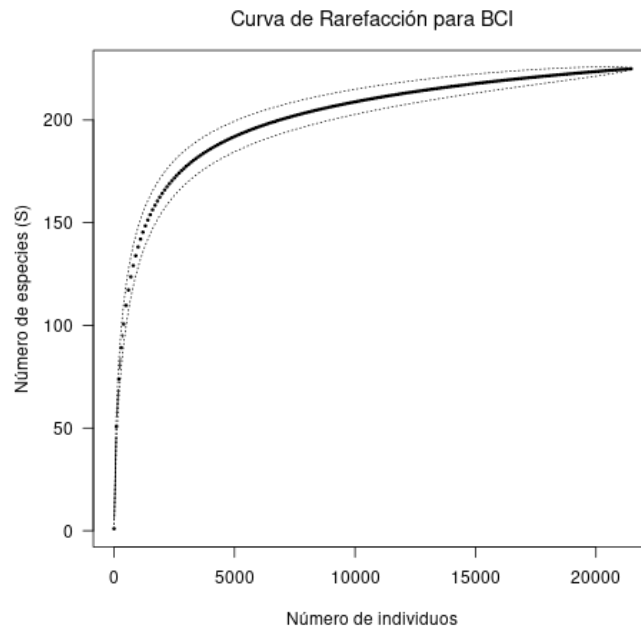


Figura 1: Curva de rarefacción para la comunidad completa del conjunto de datos BCI (Estación Biológica de Barro Colorado, paquete *vegan*).

En la figura 1 podemos apreciar el influjo que tiene el tamaño de muestreo en el estimado total de especies S (0D). Esta representación constituye una **interpolación** de los valores de S , pues llena los estimados de distintos submuestreos. En la figura 2 podemos ver que los estimados de los números de Hill 1 (1D) y 2 (2D) también son influenciados por la profundidad de muestreo, aunque en menor grado. Esto se debe a que a medida que q (el número de Hill) aumenta, el cálculo de diversidad le presta menor importancia a las especies raras. También es posible a partir de estas observaciones **extrapolar** la curva para estimar el número total de especies presentes en la comunidad, como veremos en la siguiente sección.

Estimadores no paramétricos (asintóticos) de S

En pasadas discusiones nos hemos discutido cómo los modelos de *Ranked Abundance Distributions* (RAD) nos permiten estimar el posible número total de especies en la comunidad a partir de las observaciones empíricas. Vimos cómo en el trabajo de Steege *et al.* (2013) utilizaron la *serie logarítmica* para a partir del α de Fisher estimar el número total de especies en Amazonia. También vimos cómo a partir de las *octavas de Preston* podíamos aproximar las abundancias relativas a una *serie log-normal*, y a partir de las propiedades de esta distribución, estimar también el número total de especies. El gran problema de estos métodos de estimación de riqueza es que dependen de que la distribución de las abundancias observadas se ajuste a la distribución real. La distribución en muchas comu-

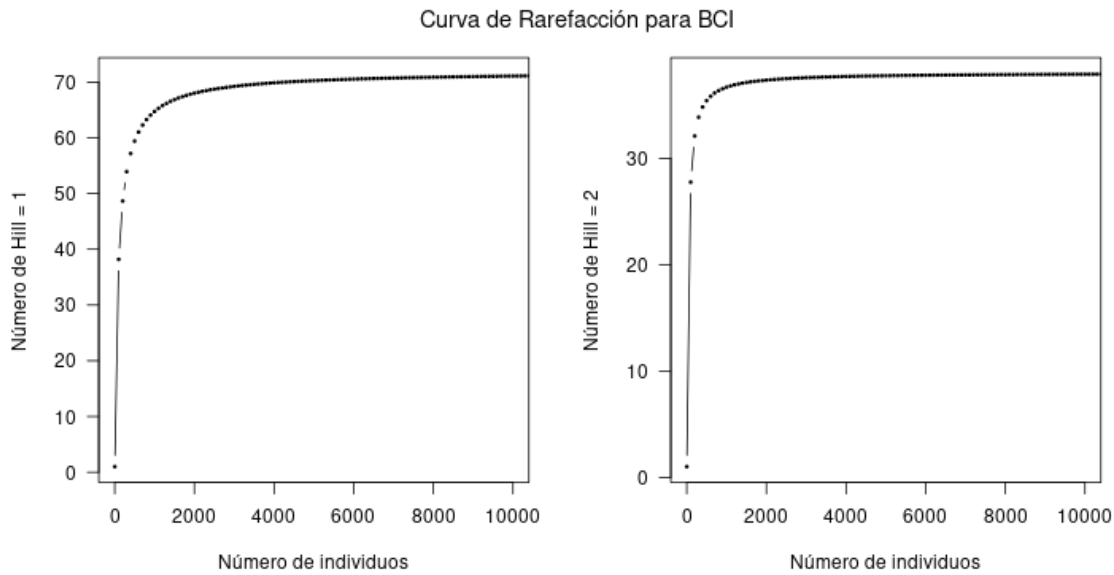


Figura 2: Curva de rarefacción para la comunidad completa del conjunto de datos BCI (Estación Biológica de Barro Colorado, paquete *vegan*), estimando los números de Hill 1 y 2 (1D y 2D).

nidades no se ajusta de manera satisfactoria a ninguno de los modelos propuestos, o se ajusta a múltiple de ellos.

Una alternativa es el uso de estimadores no paramétricos, es decir, que no asumen nada acerca de la distribución de las abundancias de las poblaciones dentro de una comunidad. En su lugar, se basan en modelos estadísticos que hacen uso de la frecuencia del número de especies *raras* en la muestra, específicamente del número de especies representadas por 1 (f_1) o 2 (f_2) individuos. El siguiente estimador:

$$Chao1 = S_{obs} + \frac{f_1^2}{2f_2} \text{ si } f_2 > 0 \quad (10)$$

$$Chao1 = S_{obs} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)} \text{ si } f_2 = 0 \quad (11)$$

es un ejemplo de estimadores a los cuales se les denomina también como *estimadores asintóticos*, pues se basan en las especies raras que son las únicas que aparecen en la muestra a medida que la curva de acumulación de especies se acerca a la asíntota¹. El

¹Aunque la pregunta de ¿Cuántas especies habitan un ecosistema? ha sido central en el estudio de comunidades, los métodos estadísticos para su estimación continúan siendo un tema de gran interés. Los últimos 15 años han visto un gran avance en este aspecto de la teoría ecológica con el trabajo de Anne Chao (Taiwan) y Robert Colwell (EE.UU.). El estimador *Chao1*, nombrado así por Colwell en honor a Chao, es uno de dichos avances.

estimador *Chao1* utiliza datos de abundancia de cada una de las especies dentro de una comunidad. Para calcularlo en R, podemos utilizar la función *estimaR()* (de *vegan*), así:

```
1 estimaR(Nspe) # Para estimar la riqueza en las 50 hectáreas.
2 estimaR(BCI) # Para estimar la riqueza de CADA UNA de las parcelas.
```

Al utilizar esta función notarán que los resultados incluyen no sólo el estimado basado en *Chao1*, sino también algo llamado *ACE*. Este último se refiere al *Abundance-based Coverage Estimator*, otro estimador asintótico cuya fórmula se puede encontrar en la documentación de *estimaR()*. De igual modo notarán el estimado de error estándar (se) para el estimado de riqueza (*S*). En la figura 3 podemos observar los estimados de *S* basados en la curva presentada en la figura 1.

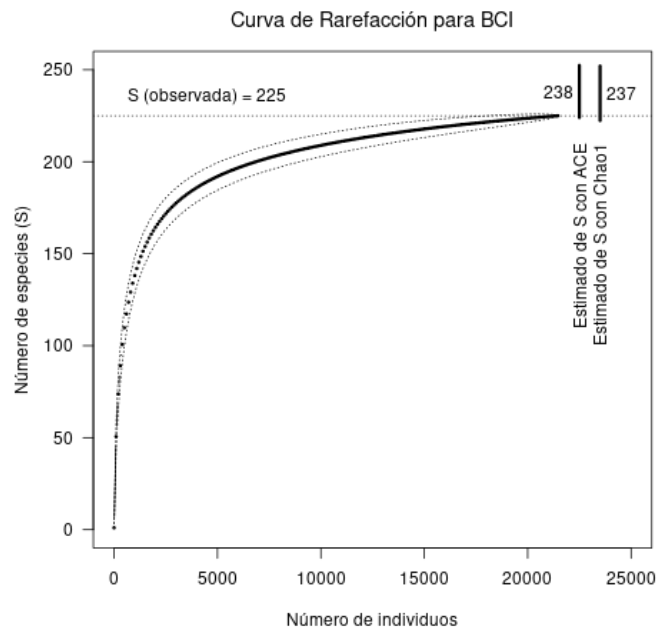


Figura 3: Como en la figura 1, con la inclusión de estimados de riqueza utilizando ACE y *Chao1*. Para el código utilizado para generar esta gráfica, ver el Apendix al final del documento.

En el paso anterior, utilizamos datos de abundancia para realizar la estimación de *S*. El estimador *Chao2* es el equivalente de *Chao1*, pero utilizando frecuencias de ocurrencias en distintas subunidades. Es decir, *Chao2* es un estimador para datos a partir de submuestras. En este caso, podríamos estimar la riqueza en las 50 hectáreas de los datos de BCI, pero utilizando cada una de las parcelas como una submuestra de toda la comunidad. Para utilizar el estimador *Chao2* podemos hacer uso de la función *specpool()*, de *vegan*, así:

```
1 specpool(BCI)
```

Noten que en este caso, el conjunto de datos que le entregamos a la función no es el acumulado de abundancia de cada especie, si no el conjunto de muestras.

Estimados asintóticos de diversidad

Dada la importancia de los métodos para estimación de número de especies, herramientas aún más poderosas han sido desarrolladas. El programa *EstimateS*, desarrollado por Robert Colwell (<http://viceroy.colorado.edu/estimates/EstimateSPages/AboutEstimateS.htm>) es un ejemplo de ellas. Para propósitos similares tenemos el paquete de funciones *iNEXT* en R, desarrollado por el grupo de Anne Chao.

Para aprender a manejar las herramientas para estimación de diversidad, vamos a partir del tutorial contenido en <https://cran.r-project.org/web/packages/iNEXT/vignettes/Introduction.html>. Para realizar los ejercicios necesitarán ejecutar las siguientes líneas:

```
1 install.packages("iNEXT")
2 install.packages("ggplot2")
3 library(iNEXT)
4 library(ggplot2)
```

Luego, deberán estudiar la viñeta para familiarizarse con el paquete. Como entrega, generarán las curvas de interpolación y extrapolación para al menos tres de las parcelas o unidades de muestreo del conjunto de datos que hayan utilizado, y entregarán las gráficas correspondientes. Para alimentar estos datos a *iNEXT*, deberán utilizar la siguiente expresión:

```
1 t(BCI[c(X,Y,Z)])
```

En el ejemplo que se les da, X, Y y Z representan los índices de tres parcelas de BCI. Este comando ilustra cómo entregarle datos a *iNEXT*. Para el caso de los datos de Haghkerdar et al. (2019; ver abajo en la descripción del entregable), esta expresión la reemplazarían por los vectores correspondientes a cada una de sus unidades de muestreo. La función *t()* en esa expresión es para cambiar columnas por filas en el conjunto de datos BCI para ajustarse a cómo los pide *iNEXT*. El resultado de esa expresión se puede utilizar directamente en donde *iNEXT* pide *x* (el objeto de abundancia de especies) o puede ser guardado como un objeto el cual se le entregue a *iNEXT*. Dependiendo de cómo hayan organizado los datos utilizados para el primer taller, es posible que esta función no sea necesaria.

Comparando comunidades - diversidad β

Recordemos que en ecología, la diversidad se puede describir en tres dimensiones distintas: la diversidad local (α), la diversidad regional (γ), y la diversidad de diferenciación o de recambio (β). La definición de las dos primeras es clara; la de la diversidad β es un poco más ambigua. En principio, y según fue definida por Robert Whittaker, la diversidad β busca entender que tan marcados son los gradientes de diversidad a lo largo de gradientes ambientales o espaciales; entre mayor sea la diversidad β , menor será el componente de la diversidad regional γ representado por cada una de las diversidades locales α .

Podemos empezar a concebir entonces la diversidad β como diferencias entre las estructuras de comunidades locales. Las diferencias la podemos conceptualizar como *dis-*

tancias: entre más distintas sean dos comunidades, más distante se encontrarán en el espacio ecológico.

Todos nos encontramos con conceptos de distancia temprano en las clases de geometría, cuando aprendimos la distancia entre dos puntos en el plano, lo que también conocemos como distancia *Euclidiana*, en el que:

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (12)$$

Pregunta: ¿Cómo aplicamos este concepto a la comparación entre comunidades?

Nuevamente trabajando con *vegan*, esta vez hacemos uso de las funciones *vegdist*, *decostand*, y *metaMDS*.

Lo primero que debemos hacer es cargar los datos con los que vamos a trabajar.

```
1 library(vegan)
2 data(BCI)
```

Cómo nuestros datos son de conteos, y cada una de las parcelas tiene números distintos de individuos, es importante normalizar nuestros datos, utilizando la función *decostand()*.

```
1 BCI.wisconsin <- wisconsin(BCI)
```

Para calcular las distancias, utilizamos la función *vegdist()*,

```
1 BCI.braycurtis <- vegdist(BCI.wisconsin, method="bray")
2 # y graficamos los resultados
3 plot(BCI.braycurtis)
4 #
5 # oh oh ...
```

La función *vegdist()* entrega un objeto de distancias, el cual es más fácil de representar a través de herramientas de agrupamiento. En este caso, utilizamos un método llamado *agrupamiento jerárquico*:

```
1 BCI.bc.ward <- hclust(BCI.braycurtis, method="ward.D2")
2 plot(BCI.bc.ward, las=1)
```

Aunque la gráfica anterior nos entrega un acercamiento intuitivo, las distancias multidimensionales entre las comunidades también pueden representarse a través de técnicas de escalamiento multidimensional (*Multi Dimensional Scaling*), en el que representamos las distancias entre las distintas comunidades proyectadas en dos dimensiones.

```
1 BCI.mds <- metaMDS(BCI.braycurtis)
2 plot(BCI.mds, display="sites", type="text")
```

Ya que este conjunto de datos está acompañado de datos sobre variables ambientales de cada sitio de muestreo, podemos intentar ajustar estas variables sobre la ordenación anterior para explorar la influencia que ejercen dichas variables.

```
1 BCI.ENV.FIT <- envfit(BCI.mds, BCI.env, permu=999)
```

Y para visualizarlas variables significativas sobre el gráfico anterior,

```
1 plot(BCI.ENV.FIT, p.max=0.05)
```

Entrega

Para fortalecer los conceptos tratados en este taller, vamos a explorar y analizar los datos de Haghkerdar et al. (2019). Los datos correspondientes los pueden descargar del repositorio **DRYAD**, a través del enlace <https://datadryad.org/stash/dataset/doi:10.5061%2Fdryad.bv1d63h>. Allí encontrarán información acerca de los datos, y un enlace para acceder al artículo original. Para entender los datos, será preciso leer este último y tener así el contexto metodológico del estudio.

1. Descargar los datos asociados indicados, y descomprimir el archivo obtenido; los datos de abundancia de Familias se encuentran en el archivo “*haghkerdar-community_EandE2019.csv*”.
2. Ingresar este archivo en R (pueden explorar las funciones *read.csv()* o *read.txt()*).
3. Calcular a nivel de “channel” (para entender qué significa, deberás consultar la información sobre la generación de los datos) los índices discutidos durante la clase, específicamente:
 - (a) Riqueza de especies (noten que aunque los datos expresen la abundancia a nivel del ranking de Familia, aún podemos utilizar la función *specnumber*; sólo es importante recordar que los resultados se expresen en términos de número de Familias presentes, y no de Especies)
 - (b) Índice de Shannon-Wiener
 - (c) Índice de Simpson (Gini-Simpson)
 - (d) Índice de Pielou
4. Construir gráficas de barras (función *barplot()*) que representen para cada una de las comunidades los índices calculados (una gráfica por variable) para los estimados de diversidad α . Construir un gráfico de agrupación jerárquica para ilustrar la disimilitud de las comunidades (diversidad β). Recordar que la función para exportar los gráficos generados es *png()* (ó *jpg()*). Por favor utilizar esta para aprender a manejarla, en lugar de tomar el atajo de adquirir pantallazos de RStudio.
5. Proporcionar gráficas de iNEXT que ilustren los estimadores asintóticos para números de Hill 0 y 1.
6. Generar un breve reporte que incluya información sobre el conjunto de datos seleccionado, cualquier preprocesamiento que se le haya hecho a los datos, y una breve discusión acerca de las observaciones realizadas.

Referencias

- Haghkerdar, J. M., McLachlan, J. R., Ireland, A., & Greig, H. S. (2019). Repeat disturbances have cumulative impacts on stream communities. *Ecol. Evol.*, 9(5), 2898-2906. <https://doi.org/10.1002/ece3.4968>
- Hill, M. O. (1973). Diversity and Evenness : A Unifying Notation and Its Consequences. *Ecology*, 54(2), 427-432. <https://doi.org/10.2307/1934352>

- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363-375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27, 379-423. <https://doi.org/10.1145/584091.584093>
- Simpson, E. H. (1949). Measurement of Diversity. *Nature*, 688(1943), 688. <https://doi.org/10.1038/163688a0>