

# ¿Cómo cuantificar la diversidad de especies?

Duncan Golicher

La idea que algunas comunidades florísticas o faunísticas son mas "diversas" que otras ha fascinado ecólogos durante varios siglos[4]. La "diversidad" de especies típicamente aumenta en zonas tropicales. Por ejemplo Wallace (1887) escribió "Si un viajero entra la selva amazónica y reconoce una especie de árbol – es probable que tendrá que viajar todo el día hasta que encuentre uno igual". En contraste en los montes del sur de Escocia es posible buscar todo el día sin encontrar más de 40 especies de plantas vasculares. Intuitivamente el bosque tropical es más diverso. Pero.....

¿Qué queremos decir por diversidad de especies?

¿Cómo podemos traducir un concepto intuitivo en algo que podemos medir?

Estas clases se van a introducir conceptos y herramientas fundamentales tanto para la ecología básica como la ecología aplicada a la conservación.

## Metas de las clases.

Las clases son prácticas. El método didáctico empleado intenta proporcionarte una introducción ligera a un problema complejo clásico que no tiene una sola solución. Las clases tienen el proposito de permitirte:

1. Aprender el cálculo y la interpretación de dos índices de diversidad en uso común por tu propia cuenta.
2. Investigar los conceptos detrás del uso de índices de diversidad.
3. Reflexionar sobre como el tamaño de una muestra de individuos puede afectar los valores de los índices usando ejemplos para aclarar el efecto.
4. Aprender como implementar rarefacción (remuestreo).
5. Empezar de desarrollar una actitud crítica hacia el uso de índices de diversidad.
6. Tener contacto con una herramienta de análisis numérica muy popular en Ecología (el idioma R).

Entonces, al final de las dos clases debes haber logrado <sup>1</sup>

1. Aprender de memoria las formulas para calcular los índices de Simpson y Shannon y tener una idea de como interpretarlos.
2. Entender la racionalidad detras del uso de metodos de rarefacción y remuestreo.
3. Entender intuitivamente por qué la cuantificación de "diversidad" y la estimación del número de especies en una "comunidad" no es tan sencilla como parece al principio.

---

<sup>1</sup> Aviso ¡Estas habilidades se pueden comprobar en un examen!

# Una introducción a la diversidad de especies: El calculo de los índices de Shannon y Simpson

Divide en grupos de 3-4. Cada grupo recibe dos bolsas llenas de "especies" que son el resultado conceptual de un muestreo de dos comunidades (figura 1). La meta es sencilla(?). Simplemente hay que decidir cual de las dos comunidades es "más diversa".

Figure 1: Dos muestras de "comunidades"



## Ejercicio 1.

Para cuantificar la diferencia entre las comunidades contesta las siguientes preguntas.

1. ¿Cual muestra contiene mas especies?
2. ¿Cual contiene más individuos?
3. ¿Cual muestra es más equitativa?
4. ¿En cual muestra hay más especies raras?

Ahora imagina que tienes que decidir cual de las dos comunidades debe recibir el estatus de área natural protegida basado en las muestras (solo hay recursos para proteger una de las dos).

1. ¿Cuál vas a proteger y por qué?
2. ¿Cómo puedes justificar tu decisión formalmente?
3. ¿Estas convencido que no has cometido un error?

#### 4. ¿Qué tipo de información adicional debes pedir?

Si estabas pensando bien debe ser claro que aunque el concepto de diversidad de especies parece sencillo, cuando lo ves de cerca y empiezas a pensar no lo es en absoluto. Incluso si puedes decidir en un índice que representa la diversidad en una muestra, ¿Cuáles son los retos en hacer muestreos representativos de estas comunidades? Sería fácil llegar a la misma conclusión como Hurlbert (1971) que diversidad de especies es un "non concepto"[6]. Antes de pensar en una forma tan negativa debemos aprender como se ha intentado medir este concepto tan escurridizo.

Diversidad tiene que estar relacionada con:

1. El numero total de especies (¿?)
2. La equitatividad en sus abundancias (¿?)

¿Qué podemos saber y medir con certidumbre? En una muestra de una comunidad podemos contar el número total de especies y podemos calcular la abundancia relativa de las especies usando un esquema sencillo

S	Número de especies
N	Número total de individuos
$n_i$	Número de individuos de especie i .....i = 1, 2, 3, ... , S
$\Sigma n_i$	=N
$p_i$	$n_i/N$ abundancia relativa de especie i

## Ejercicio 2.

Toma una muestra (aleatoriamente<sup>2</sup>) de 30 individuos de cada uno de las bolsas. Calcula estas cantidades y llena una tabla como el ejemplo (tabla 1) Para automatizar los calculos se puede usar una hoja de cálculo.

Table 1: Calculos de abundancia relativa

Nombre de especie	i	$n_i$	$p_i$	$\ln(p_i)$

La meta fundamental detrás del diseño de la mayoría de los índices de diversidad es unificar estos dos elementos de la diversidad que ya debes haber percibido que son importantes. Uno es equitividad, o sea la falta de variabilidad en las abundancias relativas, y el otro es riqueza, o sea el número total de especies.

Hay dos índices en uso común, aunque ya hay más de 60 publicados en revistas ecológicas. Siendo honesto no se necesita mucho más de estos dos índices de diversidad convencional, dado de que todos miden más o menos lo mismo. Un indice adicional que debe estar considerado es uno de los índices más robustos pero relativamente poco usado (la alpha de Fisher).

Lo más importante es asegurar que tengas un entendimiento general de los propiedades de los índices y especialmente el efecto del tamaño del muestreo sobre su comportamiento.

---

<sup>2</sup>¿Cual es el problema? ¿Puedes realmente tomar una muestra sin sesgo de estas comunidades?

1. El índice de Shannon (Shannon-Weiner)

$$H = -\sum p_i \ln(p_i)$$

2. El índice de Simpson

$$D = \sum p_i \cdot p_i \text{ o sea } D = \sum p^2$$

### Ejercicio 3.

Calcula los índices de Shannon y Simpson para las muestras de comunidades usando la tabla que ya preparaste.

#### El índice de Simpson

Si tienes la buena suerte de poseer intuición matemática debes percibir que el índice de Simpson tiene la tendencia de ser mas pequeño cuando la comunidad es mas "diversa". De hecho la interpretación de D es que es la probabilidad de un encuentro intraspecífico. O sea la probabilidad de que si tomas dos individuos al azar de la comunidad que ambos sean de la misma especie. Lo mas alto que es esta probabilidad lo menos diverso es la comunidad (sensu Wallace).

Para que veas intuitivamente porque esto es cierto piensa en una comunidad completamente equitativa.

Especie	Abun	p	$p^2$
1	5	0.1	0.01
2	5	0.1	0.01
3	5	0.1	0.01
4	5	0.1	0.01
5	5	0.1	0.01
6	5	0.1	0.01
7	5	0.1	0.01
8	5	0.1	0.01
9	5	0.1	0.01
10	5	0.1	0.01

En este caso  $D = 0.01 \cdot 10 = 0.1$  y  $1/D = 10$ . O sea si tomas un individuo aleatoriamente de la comunidad, no importa cual sea su especie, la probabilidad de encontrar otro de la misma especie es 1 en 10.

Dado de que queremos un índice que aumenta con la diversidad en vez de disminuir, seria mejor si podemos interpretar el índice en una forma directa en este sentido. Entonces es común usar el reciproco y el inverso del índice de Simpson

$$\Delta = 1/D$$

$$\text{inv}D = 1/D$$

Para ver mejor como D (la probabilidad de un encuentro intraspecífico) aumenta cuando la comunidad es menos equitativa piensa en el ejemplo de una comunidad con una especie diez veces mas abundante que las demás.

Especie	Abun	p	$p^2$
1	50	0.52632	0.27701
2	5	0.052632	0.0027701
3	5	0.052632	0.0027701
4	5	0.052632	0.0027701
5	5	0.052632	0.0027701
6	5	0.052632	0.0027701
7	5	0.052632	0.0027701
8	5	0.052632	0.0027701
9	5	0.052632	0.0027701
10	5	0.052632	0.0027701

O para simplificarlo aun mas, piensa en una comunidad que consiste en solamente una especie superabundante y algunas especies increíblemente escasas. Casí todos los encuentros estaran entre miembros de la misma especie, haciendo la probabilidad de un encuentro intraespecífico  $\sim 1$ .

## El índice de Shannon

El índice de Shannon,  $H$ , mide más o menos lo mismo que  $D$ , pero su lógica teórica esta más profundamente basada en la teoría informática. Esto hace su interpretación un poco menos intuitiva. Sin ir a más detalle  $H$  normalmente toma valores entre 1 y 4.5. Valores encima de 3 son típicamente interpretados como "diversos". Por razones que no son tan obvias como el caso de Shannon el máximo valor que puede tomar  $H$  es el logaritmo de  $S$ ,  $\ln(S)$ , o sea si la comunidad es completamente equitativa  $\exp(H)=S$ . Para confundir el asunto un poco, la derivación original de Shannon fue con logaritmos al base de dos y algunos autores todavía lo usan así.

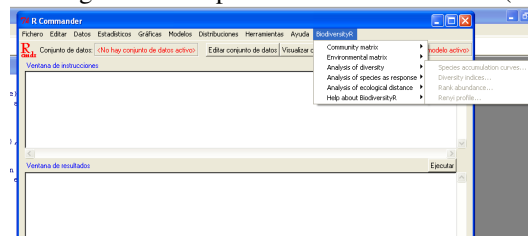
## Clase 2. Usando R para hacer los calculos

El calculo de H y D es fácil usando una hoja de cálculo como Excel. Pero hojas de cálculo no son herramientas ideales para hacer ni cálculos ni graficos científicos. Existen diversos programas comerciales, semi-comerciales y hasta totalmente gratis para calcular índices de diversidad. Al mismo tiempo programas como EstimateS (Colwell et al) o Ecosim (Gotelli et al) hacen remuestreo. También hay programas para hacer análisis multivariado (CANOCO, DECORANA) y análisis espacial (SIG). El problema con el uso de diversos programas es que potenciales pierdes mucho tiempo convirtiendo datos de un formato al otro. Hoy en día el idioma y ambiente analítico R combina todos estos elementos y mucho más.

Desafortunadamente una clase de dos horas no es suficiente para enseñar como usar R. Lo mejor que se puede conseguir es una introducción que muestra la potencia inmensa de esta herramienta. Brevemente R es "Open Source" o sea software libre, que reproduce casi todas las calidades de programas comerciales de alto costo.

Algunos usuarios potenciales lo encuentran incomodo debido a la falta de un interface gráfico. Entonces no todos descubren su potencia. En realidad lo que parece una debilidad es uno de sus mejores ventajas de R siempre y cuando se acostumbre a su estilo y se conozca su sintaxis. Es mucho mas fácil transferir lineas de código que explicar como realizar una operación con un interface. Sin embargo R tiene un interface gráfico llamado "BiodiversityR" que permite al usuario hacer todos los análisis que vamos a hacer sin recurso al interface basado en texto (Fig 2).

Figure 2: Interface gráfico de R para analisis de diversidad (Roeland Kindt)



Para realizar las operaciones vamos a usar un paquete muy completo escrito por y para ecólogos (Oksanen et al) llamado vegan [7]. La dinámica de trabajar en R en estas clases es simplemente copiar y pegar código de un documento al consol de R para lograr resultados exactamente iguales a los del texto. Así que la sustitución de los datos del ejemplo por los tuyos resulta en un análisis. Alternativamente se puede escribir cuidadosamente (cuidado con minúsculas y mayúsculas) los comandos exactamente como aparecen en el documento.

```
install.packages("vegan")
```

Escoge un espejo cercano y sigue los pasos. Luego para tener el paquete disponible escribe

```
library(vegan)
```

Primero vamos a ver algunos datos clásicos de la diversidad de árboles de Barro Colorado. Los datos consisten de 50 parcelas (filas) de 1 hectárea con conteos de los árboles en cada parcela. Se incluyeron todos los árboles con diámetros por encima de 10cm. Hay 225 especies (columnas). Se cargan los datos en la memoria con el comando siguiente:

```
data(BCI)
```

Para ver los datos en una especie de hoja de cálculo escribe:

```
fix(BCI)
```

Si quieres exportar la tabla a Excel pega el código abajo. Este código copia los datos que están en la memoria al portapapeles de Windows. Luego abre una hoja de calculo de Excel en blanco y pegar lo que hay en el portapapeles.

```
write.table (BCI, "clipboard", sep="\t", row.names=F)
```

Guardar el resultado. ¿Cual va a ser el problema obvio de usar Excel para este tipo de datos? Una razón de no usar Excel, entre otras, es que el tamaño físico del matriz de datos hace las operaciones muy torpes. El uso de hojas de cálculo empieza a dificultarse rápidamente una vez que se pasa del punto cuanto no todos los datos caben en una página. De hecho, la fortaleza principal de una hoja de cálculo es que se puede visualizar todos los datos y todos los cálculos en su conjunto. Esta ventaja desaparece en casos de arreglos de datos moderadamente grandes. Si tratas de forzar una hoja de cálculo de rebasar sus límites vas a pasar horas reinventando formas muy enredadas y tediosas de hacer operaciones que son instantáneos en Access, SAS, SPSS, Splus y particularmente en R.

Vamos a calcular el inverso del índice de Shannons para todas las 50 parcelas de BCI.<sup>3</sup>

```
diversity(BCI, "inv")
```

Ahora el índice de Shannon:

```
diversity(BCI, "sh")
```

Una variación es imprimir los sitios en orden.

```
sort(diversity(BCI, "inv"))  
sort(diversity(BCI, "sh"))
```

Prueba

```
boxplot(diversity(BCI, "sh"), col=2, boxwex=0.3)
```

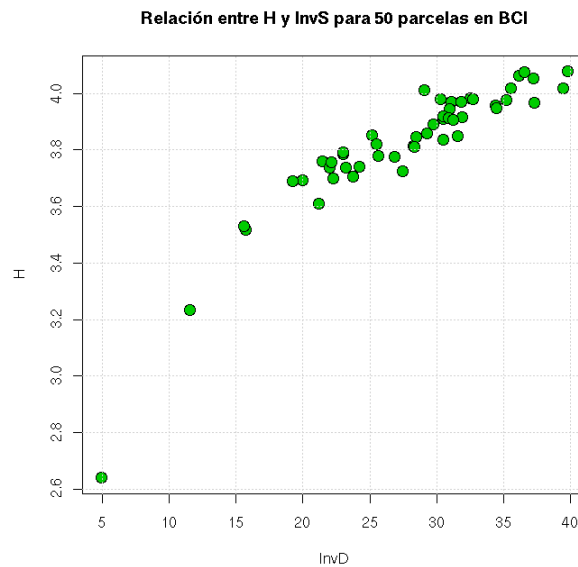
Se puede graficar la relación entre los dos índices (figure 3).

```
H<-diversity(BCI, "sh")  
InvD<-diversity(BCI, "invs")  
plot(InvD, H, pch=21, bg=3, cex=1.5, main="Relación entre H y InvS para 50 parcelas en BCI")  
grid()
```

---

<sup>3</sup>Los comandos tal cual imprimen sus resultados en el consol de R. Este es una forma muy rápida de trabajar. Existen varias formas de exportar los resultados en tablas formateadas para la publicación que se pueden emplear después de una investigación de los datos.

Figure 3:



Incluso se puede identificar las parcelas (haz clic en los puntos del gráfico con el botón izquierdo del ratón y terminar con el botón derecho).

```
identify(InvD, H)
```

### Ejercicio 1.

Se puede calcular el número de especies y el número de individuos en las parcelas en orden usando

```
sort(specnumber(BCI))  
sort(apply(BCI, 1, sum))
```

1. ¿Cual de las parcelas tiene el número mayor de individuos
2. ¿Cual el menor número de especies?
3. ¿Cual de las parcelas es "menos diversa" de acuerdo a este análisis?
4. Investiga las propiedades de los datos originales (puedes escanear el matriz visualmente en Excel por el momento) ¿Se puede proponer una explicación a la diversidad baja en esta parcela?
5. ¿Que quiere decir el siguiente gráfico?

```
plot(H, BCI$Aalseis.blackiana, pch=21, bg=2)
```

Un análisis exploratorio de este tipo no revela mucho de interés teórico. Pero una vez que se haya tomado en cuenta no solamente el valor de un índice, pero también la identidad de una especie clave, tal vez se puede empezar a entender mejor la dinámica de este bosque (ver la caja abajo)



# THE UNUSUAL LIFE HISTORY OF *ALSEIS BLACKIANA*: A SHADE-PERSISTENT PIONEER TREE?

J. W. DALLING,<sup>1,5</sup> K. WINTER,<sup>1</sup> J. D. NASON,<sup>2</sup> S. P. HUBBELL,<sup>1,3</sup> D. A. MURAWSKI,<sup>4</sup> AND J. L. HAMRICK<sup>1,3</sup>

<sup>1</sup>Smithsonian Tropical Research Institute, Washington, D.C. 20560-0580 USA

<sup>2</sup>Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242 USA

<sup>3</sup>Department of Botany, University of Georgia, Athens, Georgia 30602 USA

<sup>4</sup>Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138 USA

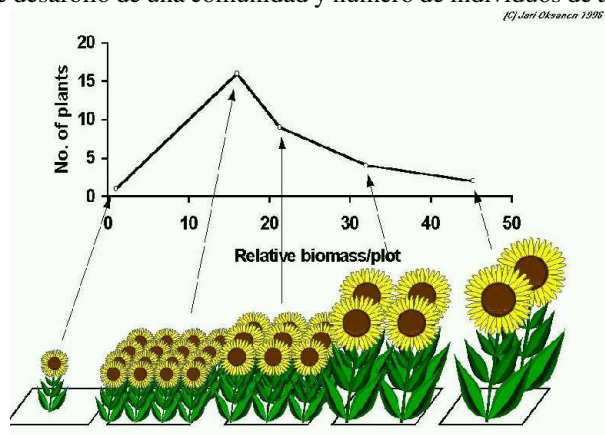
**Abstract.** Classifications of tree species into regeneration guilds or functional groups are often based on a limited number of life history characteristics indicative of requirements for early establishment, with the assumption that these traits are correlated with growth requirements of larger individuals. We tested the validity of this assumption for *Alseis blackiana*, a common canopy tree in central Panama, by examining its seed germination, seedling growth, and sapling distribution, growth, and mortality. We found that the early life history of *Alseis* is characteristic of pioneer species: the seed germination rate was much higher in forest gaps than in the forest understory, small seedlings were absent from the understory, and seedlings showed strong plastic responses to variation in light. Seedling relative growth rate, leaf area production rate, and maximum carbon assimilation rates of *Alseis* are much more similar to those of the pioneer tree *Miconia argentea* than to those of the non-pioneer tree *Tetragastris panamensis*. However, in contrast to *Miconia*, established seedlings of *Alseis* are capable of remarkable persistence in the understory; in a 1-ha plot on Barro Colorado Island, the mortality rate for 722 saplings <1 cm dbh was <4% per year between 1986 and 1998. We suspect that shade tolerance in *Alseis* is achieved by both physiological and morphological adaptations, e.g., efficient light interception by thin, nonoverlapping leaves, tolerance to litterfall damage, and a high resprouting capacity. These results show that existing functional group classifications do not account for the full range of seedling trait combinations found in tropical trees and indicate that a gap requirement for seedling establishment need not limit the abundance of adult individuals in a stand.

**Key words:** *Alseis blackiana*; functional group; *Miconia argentea*; ontogeny; pioneer; regeneration guild; seed germination; seedling growth; shade tolerance; *Tetragastris panamensis*.

## Rarefacción, remuestreo y curvas de acumulación.

Debes haber llegado a la conclusión que ni la diversidad de especies medida por índices convencionales, ni la riqueza de especies (S) dice todo sobre la comunidad. Pero te debes también haber dado cuenta de otro factor incómodo. Una muestra tomada de una comunidad no va a tener las mismas propiedades que la comunidad. En particular es casi imposible que una muestra contenga todas las especies que hay en una comunidad. También debe ser obvio que si la muestra es mas grande en términos de individuos, todos los otros factores siendo iguales, deben contener mas especies. Así que tenemos un problema. Cada muestra de individuos de una comunidad puede tener un tamaño diferente. Por ejemplo una trampa de polillas puede atraer 200 individuos en un lugar, pero solamente 50 en otro. Un cuadrante puede contener 60 árboles juveniles en una área de vegetación secundaria, pero solamente 10 en un bosque maduro. El segundo ejemplo se ha discutido Jari Oksanen (<http://cc.oulu.fi/~jarioksa/>) y en un forum (Journal of Ecology 84, 293-295 (1996)). La idea esta mostrado en fig 4

Figure 4: La relación entre fase de desarrollo de una comunidad y número de individuos de acuerdo del modelo de Jari Oksanen



Tal vez sería conveniente comparar el número de especies en una muestra de (por ejemplo) 20 individuos con remuestras de 20 individuos tomado de una muestra mas grande. Sería mas "justo" para la muestra pequeña. Ahora, regresa a pensar en las bolsas de frijoles. Si tomas una muestra de una bolsa del mismo número de individuos, ¿vas a conseguir el mismo número de especies siempre? Obviamente si haces el remuestreo "a mano" vas a tardar mucho tiempo porque tendrías que repetir el experimento muchas veces. Pero con una computadora es fácil. Abajo hay una función que escribí para hacer remuestreo de una colección de individuos representada como un vector de abundancias  $(n_1, n_2, n_3 \dots n_i)$ .

Hay formulas analíticas de calcular el número esperado de especies en una muestra de un tamaño definido tomado de una lista de especies con abundancias conocidas. Por ejemplo usando la formula de Hurlbert 1971

```
parcela.1<-BCI[1,]  
sum(parcela.1) #Calcular el numero de individuos en parcela 1  
rarefy(parcela.1,448) #Te da el numero de especies en la muestra total  
rarefy(parcela.1,100) #Te da el numero de especies en una muestra de 100
```

Se puede hacer el cálculo para toda la matriz.

```
sort(rarefy(BCI,100))
```

¿Como se compara la conclusión de este análisis con las conclusiones previas sobre cual de las parcelas son mas diversas?

Dado de que hay problemas con el cálculo de intervalos de confianza para rarefacción una táctica común es usar "fuerza bruta" y simular el remuestreo muchas veces. Pega todo el código en la próxima página en R en un solo bloque. Al pegar una función en R no ves resultados. Esto hace que una función nueva estará disponible después para tu análisis, un poco como cargar un paquete.

```

resamp <-
function(X, rep=1,...) {
  require(vegan)
  Samp <- function(size, Indivs) { length(table(sample(Indivs, size,
    replace=F))) }
  N<-X[X>0]
  TotSp <- length(N)
  TotInds<-sum(N)#### Accumulation curve up to the maximum number of
  #individuals
  Sp <- 1:TotSp
  Inds <- rep(1:TotSp, N)
  Size <- rep(floor(TotInds*0.25):TotInds, rep) ###Note its(TotInds*0.25):TotInds
  sp.count <- sapply(Size, Samp, Inds)
  plot(Size,sp.count,xlab="Number of individuals",ylab="Number of species",...)
  Sp.Ac1<-lm(sp.count~log(Size))
  Sp.Ac2<-lm(sp.count~log2(Size))
  fa<-fisher.alpha(N)
  res<-list(Logslope=Sp.Ac1$coefficients[2],Log2slope=Sp.Ac2$coefficients[2],Alpha=fa)
  return(res)
}

```

Ahora puedes probar un remuestreo de todos los datos de BCI en su conjunto, parcela por parcela y visualizar los resultados con

```

par(ask=T)
a<-apply(BCI,1,resamp)

```

Puedes probar la función con este cambio.

```

a<-apply(BCI,1,resamp,log="x")

```

Uno de los elementos de la función es que esta simultaneamente calculando la pendiente de una línea de regresión de S contra el logaritmos del número de individuos y el índice alpha de Fisher. El tema de que representa el siguiente índice es un poco avanzado, pero hay una interpretación intuitiva del pendiente de la regresión.

```

do.call("rbind",a)
  Logslope Log2slope Alpha
1  30.4      21.1      35.7
2  27.7      19.2      31
3  31.5      21.9      33.3
4  29.4      20.4      33.9
5  33.6      23.3      38
6  30.2      20.9      32.5
7  26.6      18.5      30.6

```

8	27.8	19.3	33.4
9	32.8	22.8	35.7
10	32.0	22.2	34.8
11	30.7	21.3	34.2
12	31.9	22.1	34.1
13	31.6	21.9	37.6
14	33.9	23.5	39.2
15	30.2	20.9	35.1
16	31.9	22.1	36.2
17	36.5	25.3	39.2
18	32	22.2	38.7
19	40.9	28.3	46.9
20	35.3	24.4	41
21	37	25.7	41.6
22	34.1	23.6	35.8
23	38.7	26.8	46.9
24	34.4	23.9	39.9
25	37.6	26.0	43.5
26	30.9	21.4	36.4
27	35.7	24.8	41
28	30.5	21.1	33.7
29	32.7	22.6	35.5
30	33.8	23.4	36.9
31	25.0	17.3	27.6
32	28.5	19.7	32.3
33	29.6	20.5	32.1
34	32.3	22.4	35.1
35	33.7	23.4	26.1
36	32.7	22.7	35.9
37	29.5	20.5	33.3
38	29.1	20.2	29.5
39	30.2	20.9	31.4
40	29.1	20.2	27.2
41	37.9	26.3	44.1
42	28.0	19.4	33.6
43	31.4	21.8	33.3
44	28.2	19.6	30.3
45	29.2	20.3	29
46	29.4	20.4	32.3
47	39.9	27.6	42.6
48	31.4	21.7	36

49	33.8	23.4	35.4
50	33.1	22.9	36.4

La interpretación del log2slope puede ser en terminos del número de especies nuevos que esperas encontrar cada vez que tomas el doble del número de individuos. Es claro con este analisis que en un bosque tan diverso como BCI el modelo asimptotico propuesto por algunos autores para comunidades "cerradas" no es muy aplicable.

Incluso se puede hacer un remuestreo de todos los datos tratandolos como una sola parcela. Dado el tamaño este cálculo puede tardar mas de un minuto.

```
b<-apply(BCI, 2, sum)
resamp(b)
```

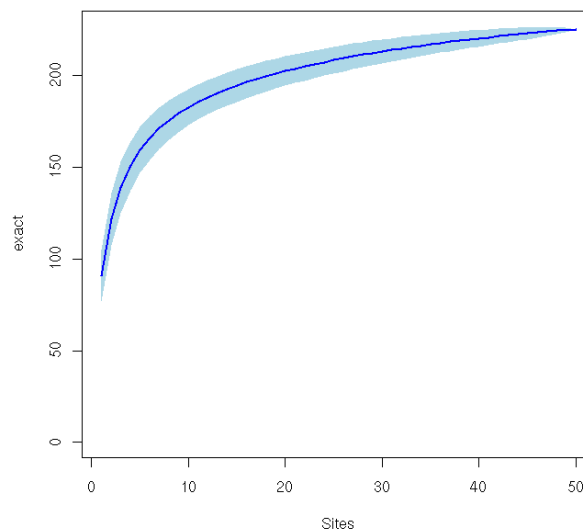
Hay una forma alternativa de hacer un remuestreo. Se puede aleatorizar el orden de los sitios y ver la acumulación de las especies.

```
a<-specaccum(BCI)
plot(a)
```

Se para hacer un gráfico un poco mas bonito como figura 5

```
plot(a, ci.type="poly", col="blue", lwd=2, ci.lty=0, ci.col="lightblue")
```

Figure 5: Relación entre el numero de especies y el numero de sitios (curva de acumulación de especies) para BCI.



## Simulación del muestreo de comunidades con abundancias relativas log normal

Finalmente, en este clase no tenemos tiempo de investigar en profundidad las teorías tras de las abundancias relativas de especies, ni como el efecto del muestreo puede ocultar sus propiedades [3]. Pero para los que tengan interés he escrito una pequeña función para demostrar algunas propiedades de muestras simuladas de comunidades con una distribución subyacente log-normal. Pega la función abajo en R.

```

require(vegan)
div.demo<-function(Gdim=100,n.species=20,sdlog=2){
  coords<-expand.grid(x=1:Gdim,y=1:Gdim)
  species.names<-paste("species",1:n.species,sep=".")
  rel.abun<-sort(rlnorm(n.species,mean=10,sdlog),decreasing=T)
  d<-sample(1:n.species,Gdim^2,prob=rel.abun,replace=T)
  d<-matrix(d,Gdim,Gdim)
  sp.colors<-c("darkred","red","orange",rgb(0,(n.species:0)/n.species,(0:n.species)/n.species))
  #sp.colors=topo.colors(n.species)
  par(mfcol=c(2,2))
  barplot(table(d),col=sp.colors,xlab="Especie",ylab="Numero de individuos",main="Numero de individuos")
  hist(table(d),xlab="Abundancia",ylab="Numero de especies",main="Histograma de abundancia")
  hist(log(table(d)),xlab="Log abundancia",ylab="Numero de especies",main="Histograma de log abundancia")
  image(1:Gdim,1:Gdim,d,col=sp.colors)
  print( diversity(table(d),"shan",2) )
  print( diversity(table(d),"invsim",2) )
  assign("d",d, env = .GlobalEnv)
}
div.demo(100,10,2)

```

## Como importar datos de diversidad a R

El primer detalle importante de mencionar con respecto a la importación de datos es que los datos que hemos visto de BCI estan ya en la forma de un matriz (sitios x especies figura 6).

Este no es una forma convencional de mantener datos de este tipo. Es un paso necesario en su procesamiento despues de su captura en otra formato. Las normas del mantenimiento de bases de datos estipulan estrictamente que nunca debes capturar ni mantener datos en un tipo de formato que puede necesitar un cambio estructural en su diseño (la adición de columnas nuevas) a lo largo de un periodo de captura, uso o analisis.

Figure 6: Parte del matriz de datos sitios x especies

	<i>Abacena.macradenium</i>	<i>Acacia.melanocera</i>	<i>Acalypha.diversifolia</i>	<i>Acalypha.macrostachya</i>	<i>Adelia.triloba</i>	<i>Aegiphila.panamensis</i>	<i>Alchornea.costaricensis</i>
1	0	0	0	0	0	0	2
2	0	0	0	0	0	0	1
3	0	0	0	0	0	0	2
4	0	0	0	0	3	0	18
5	0	0	0	0	1	1	3
6	0	0	0	0	0	0	2
7	0	0	0	0	0	1	0
8	0	0	0	0	0	0	2
9	0	0	0	0	5	0	2
10	1	0	0	0	0	1	2
11	0	0	0	0	0	0	10
12	0	0	0	0	1	1	3
13	0	0	0	0	1	1	1
14	0	0	0	0	0	0	4
15	0	0	0	0	2	0	2
16	0	0	0	0	2	0	2
17	0	0	0	0	0	1	2
18	0	0	0	0	1	1	0
19	0	0	0	0	0	1	1
20	0	0	0	0	0	2	2

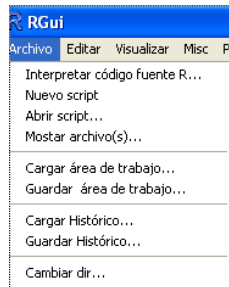
Si los datos nada mas consisten en el conteo de individuos de las especies por sitio la forma extensible de mantener la base de datos es en tres columnas - sitio - especie -conteo (figura 7 )

Figure 7: Forma rectangular (columnar) de mantener los datos

R Editor de datos			
	sitio	especie	abun
1	1	Alchornea.costaricensis	2
2	1	Alseis.blackiana	25
3	1	Annona.spraguei	1
4	1	Apeiba.aspera	13
5	1	Apeiba.tibourbou	2
6	1	Astronium.graveolens	6
7	1	Beilschmiedia.pendula	4
8	1	Brosimum.alicastrum	5
9	1	Casearia.arborea	1
10	1	Casearia.sylvestris	2
11	1	Cassipourea.elliptica	2
12	1	Cecropia.insignis	12
13	1	Chrysophyllum.argenteum	4
14	1	Cordia.alliodora	2
15	1	Cordia.bicolor	12
16	1	Cordia.lasiocalyx	8
17	1	Croton.billbergianus	2
18	1	Cupania.sylvatica	2

Puedes tener los datos en el formato mostrado en figura 7 en una hoja de calculo como Excel, pero luego es aconsejable guardarlos en un formato mas universal, por ejemplo CSV (en ingles significa comma separated variables, o sea variables separado por commas). R puede leer datos directamente desde Excel, pero es considerada una mala practica guardar datos en un formato especifico a un software commercial<sup>4</sup>.

Suponiendo que has guardado los datos como en figura 7 en un archivo de csv puedes leerlos a R una vez que hayas cambiado el directorio de trabajo al directorio donde guardaste los datos. Busca "cambiar dir" en el menu del interface de R y luego selecciona un directorio de trabajo donde has puesto tus archivos.



El siguiente linea lee un archivo llamado BCI.csv a la memoria de R

```
a<-read.csv("BCI.csv")
```

Estos dos lineas reformata los datos como el matriz BCI. Nota que uso tapply y "sum" como operador "dummy" en el sentido que no hay mas de un valor por cada combinaci3n de sitio y especie.

```
BCI<-as.data.frame(tapply(a$abun,list(a$sitio,a$especie),sum))
BCI[] <- lapply(BCI,function(x) replace(x, is.na(x), 0))
```

Nota que para que este funciona tal cual, sin la necesidad de cambios en las lineas, tienes que asegurar que las columnas tengan exactamente los mismos nombres que en el ejemplo. Si los datos son los tuyos entonces el codigo original ya va a funcionar como antes, pero el matriz BCI ahora se referia a los datos nuevos.

<sup>4</sup>¿Que pasa si se hace obsoleto la version?

En la realidad tampoco sería correcto en absoluto capturar los datos en una hoja de calculo en la forma como figura 7. El problema mayor es que habrás escrito el nombre de la especie muchas veces. Entonces hay la posibilidad de errores en la ortografía. Tambien es probable que al momento de captar los datos en el campo no sabias exactamente el nombre de la especie. Tienes que darle un numero de catalogo, pero la identidad taxonomica tal vez tendría que esperar. Tambien si tienes información adicional sobre la especie tendrías que añadir mas campos en la tabla, cada uno lleno de información duplicada. Entonces la forma mucho mas correcta de mantener los datos seria como en figura 8.

Figure 8: Formato para mantener datos de abundancia usando un codigo para la identidad de la especie

	A	B	C
1	sitio	codigo	abun
2	1	7	2
3	1	11	25
4	1	15	1
5	1	16	13
6	1	17	2
7	1	20	6
8	1	23	4
9	1	24	5
10	1	28	1
11	1	31	2
12	1	32	2
13	1	34	12
14	1	44	4

El complemento necesario de esta tabla es obviamente una segunda tabla con la identidad de la especie escrito nada mas una vez. La tabla puede contener campos de información adicional como familia, grupo funcional etc (figura 9 )

Figure 9: Tabla de información sobre especies.

	A	B
1	codigo	especie
2	1	Abarema.macradenium
3	2	Acacia.melanoceras
4	3	Acalypha.diversifolia
5	4	Acalypha.macrostachya
6	5	Adelia.triloba
7	6	Aegiphila.panamensis
8	7	Alchornea.costaricensis
9	8	Alchornea.latifolia
10	9	Alibertia.edulis
11	10	Allophylus.psilospermus
12	11	Alseis.blackiana
13	12	Amaioua.corymbosa

Nota que en este ejemplo es artificial en el sentido de que todo esta en orden alfabetica porque los he derivado de los datos ya procesados. En realidad no importa si los codigos no son en orden, mientras que hay una coincidencia estricta y consistente entre el codigo usado en los dos tablas. Tambien el codigo de identificación de la especie tiene que ser unico. Luego mientras cada tabla contenga un campo con el mismo nombre se juntan dos tablas facilmente en R usando merge.

```
sitios<-read.csv("BCIcodigos.csv")
especies<-read.csv("especies.csv")
a<-merge(especies,sitios)
fix(a)
```

Y ya el codigo de antes va a reformar la matriz original para el analisis.

```
BCI<-as.data.frame(tapply(a$abun,list(a$sitio,a$especie),sum))
BCI[] <- lapply(BCI,function(x) replace(x, is.na(x), 0))
```



Finalmente, que tal si has medido otros aspectos de arboles individuales en las parcelas. Por ejemplo dap. Los datos pueden parecer como figura 10

Figure 10: Ejemplo de datos de dap

	A	B	C	D
1	sitio	individuo	codigo	dap
2		1	1	10
3		1	2	8
4		1	3	7
5		1	4	9
6		1	5	5
7		1	6	3
8		1	7	1
9		1	8	1
10		1	9	1

En este ejemplo muy artificial he tomado solamente los diez primeros especies en la lista anterior y simulado datos de dap. Las abundancias no son los mismos que en los datos originales. Es facil de formar un matriz de nuevo usando la tabla de especies de antes. Nota como habia 225 especies antes y solamente 10 en este ejemplo tengo que usar una pequeña función para quitar los niveles del factor que no se usan.

```
drop.levels <- function(dat){as.data.frame(lapply(dat, function(x) x[,drop=TRUE]))}
dap<-read.csv("dap.csv")
especies<-read.csv("especies.csv")
dap<-drop.levels(merge(dap,especies))
fix(dap)
BCI<-table(dap$sitio,dap$especie)
fix(BCI)
```

## References

- [1] Colwell, R.K., Mao, C.X. & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85: 2717–2727.
- [2] Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J. Foster, R.B., Nunez, P., Aguilar, S., Valencia, R., Villa, G. Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002). Beta-diversity in tropical forest trees. *Science* 295, 666-669.
- [3] Golicher J.D., O'Hara R.B., Ruíz-Montoya L., Cayuela L. 2006. Lifting a veil on diversity: a bayesian approach to fitting relative-abundance models. *Ecological Applications*. 16(1): 202- 212.
- [4] Gotelli, N.J. & Colwell, R.K. (2001). Quantifying biodiversity: procedures and pitfalls in measurement and comparison of species richness. *Ecol. Lett.* 4, 379–391.
- [5] Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–473.
- [6] Hurlbert, S.H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecological Monographs* 54, 187–211.

- [7] Jari Oksanen, Roeland Kindt, Pierre Legendre, Bob O'Hara and M. Henry H. Stevens (2007). *vegan: Community Ecology Package*. R package version 1.8-8. <http://cran.r-project.org/>, <http://r-forge.r-project.org/projects/vegan/>