

## 5. Worksheet: Alpha Diversity

Carter K Stancil; Z620: Quantitative Biodiversity, Indiana University

27 January, 2026

### OVERVIEW

In this exercise, we will explore aspects of local or site-specific diversity, also known as alpha ( $\alpha$ ) diversity. First we will quantify two of the fundamental components of ( $\alpha$ ) diversity: **richness** and **evenness**. From there, we will then discuss ways to integrate richness and evenness, which will include univariate metrics of diversity along with an investigation of the **species abundance distribution (SAD)**.

### Directions:

1. In the Markdown version of this document in your cloned repo, change “Student Name” on line 3 (above) to your name.
2. Complete as much of the worksheet as possible during class.
3. Use the handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the exercise.
4. Answer questions in the worksheet. Space for your answer is provided in this document and indicated by the “>” character. If you need a second paragraph be sure to start the first line with “>”. You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. For the assignment portion of the worksheet, follow the directions at the bottom of this file.
7. When you are done, **Knit** the text and code into a PDF file.
8. After Knitting, submit the completed exercise by creating a **pull request** via GitHub. Your pull request should include this file `AlphaDiversity_Worskheet.Rmd` and the PDF output of Knitr (`AlphaDiversity_Worskheet.pdf`).

### 1) R SETUP

In the R code chunk below, please provide the code to: 1) Clear your R environment, 2) Print your current working directory, 3) Set your working directory to your `Week2-Alpha/` folder folder, and 4) Load the `vegan` R package (be sure to install first if you have not already).

```
rm(list = ls())  
getwd()
```

```
## [1] "C:/github/QB2026_Stancil/QB2026/Week3-Alpha"
```

```
setwd("../Week3-Alpha") # this is running but not actually working... setwd manually via toolbar  
  
#install.packages("vegan")  
require("vegan")
```

```
## Loading required package: vegan

## Warning: package 'vegan' was built under R version 4.5.2

## Loading required package: permute

## Warning: package 'permute' was built under R version 4.5.2
```

## 2) LOADING DATA

In the R code chunk below, do the following: 1) Load the BCI dataset, and 2) Display the structure of the dataset (if the structure is long, use the `max.level = 0` argument to show the basic information).

```
data(BCI)
str(BCI, max.level=0) #structure = str()

## 'data.frame': 50 obs. of 225 variables:
## - attr(*, "original.names")= chr [1:225] "Abarema.macradenium" "Acacia.melanoceras" "Acalypha.diversa"
```

## 3) SPECIES RICHNESS

**Species richness (S)** refers to the number of species in a system or the number of species observed in a sample.

### Observed richness

In the R code chunk below, do the following:

1. Write a function called `S.obs` to calculate observed richness
2. Use your function to determine the number of species in `site1` of the BCI data set, and
3. Compare the output of your function to the output of the `specnumber()` function in `vegan`.

```
S.obs <- function(x = ""){
  rowSums(x>0) * 1
}

# (x = "") is possibly unnecessary, you could probably get away with just (X)
# multiply by 1 to make sure that the final output is NUMERIC not LOGICAL

S.obs(BCI[1,])
```

```
## 1
## 93
```

```
specnumber(BCI[1,])
```

```
## 1
## 93
```

```
S.obs(BCI[1:4,])
```

```
## 1 2 3 4
## 93 84 90 94
```

```
site1 <- BCI[1,]
```

**Question 1:** Does `specnumber()` from `vegan` return the same value for observed richness in `site1` as our function `S.obs`? What is the species richness of the first four sites (i.e., rows) of the BCI matrix?

**Answer 1:** Yes. 93, 84, 90, 94

**Coverage: How well did you sample your site?**

In the R code chunk below, do the following:

1. Write a function to calculate Good's Coverage, and
2. Use that function to calculate coverage for all sites in the BCI matrix.

```
C <- function(x = ""){
  1 - (rowSums(x == 1) / rowSums(x))
}
```

```
# when x == 1, that's a singleton
```

```
C(BCI)
```

```
##      1      2      3      4      5      6      7      8
## 0.9308036 0.9287356 0.9200864 0.9468504 0.9287129 0.9174757 0.9326923 0.9443155
##      9     10     11     12     13     14     15     16
## 0.9095355 0.9275362 0.9152120 0.9071038 0.9242054 0.9132420 0.9350649 0.9267735
##     17     18     19     20     21     22     23     24
## 0.8950131 0.9193084 0.8891455 0.9114219 0.8946078 0.9066986 0.8705882 0.9030612
##     25     26     27     28     29     30     31     32
## 0.9095023 0.9115479 0.9088729 0.9198966 0.8983516 0.9221053 0.9382423 0.9411765
##     33     34     35     36     37     38     39     40
## 0.9220183 0.9239374 0.9267887 0.9186047 0.9379310 0.9306488 0.9268868 0.9386503
##     41     42     43     44     45     46     47     48
## 0.8880597 0.9299517 0.9140049 0.9168704 0.9234234 0.9348837 0.8847059 0.9228916
##     49     50
## 0.9086651 0.9143519
```

**Question 2:** Answer the following questions about coverage:

- a. What is the range of values that can be generated by Good's Coverage?
- b. What would we conclude from Good's Coverage if  $n_i$  equaled  $N$ ?
- c. What portion of taxa in `site1` was represented by singletons?
- d. Make some observations about coverage at the BCI plots.

**Answer 2a:**  $0 < x < 1$

**Answer 2b:** All the species sampled were singletons, so only found once.

**Answer 2c:** 0.93

**Answer 2d:** When Good's Coverage approaches 1, the coverage of our site is higher and we can assume that our sampling effort covered of the natural biodiversity of the population(s) and is therefore more representation of the overall community. All sites had Good's Coverage values of  $> 0.88$ , which indicates generally high coverage of the BCI plots.

## Estimated richness

In the R code chunk below, do the following:

1. Load the microbial dataset (located in the `Week2-Alpha/data` folder),
2. Transform and transpose the data as needed (see handout),
3. Create a new vector (`soilbac1`) by indexing the bacterial OTU abundances of any site in the dataset,
4. Calculate the observed richness at that particular site, and
5. Calculate coverage of that site

```
soilbac <- read.table("soilbac.txt", sep = "\t", header = TRUE, row.names = 1)
soilbac.t <- as.data.frame(t(soilbac))
soilbac1 <- soilbac.t[1,]
S.obs(soilbac1)
```

```
## T1_1
## 1074
```

```
C(soilbac1)
```

```
##      T1_1
## 0.6479471
```

**Question 3:** Answer the following questions about the soil bacterial dataset.

- a. How many sequences did we recover from the sample `soilbac1`, i.e.  $N$ ?
- b. What is the observed richness of `soilbac1`?
- c. How does coverage compare between the BCI sample (`site1`) and the KBS sample (`soilbac1`)?

**Answer 3a:** 1074

**Answer 3b:** 0.6479

**Answer 3c:** Coverage was much higher for the BCI site 1 sampling compared to the KBS site 1 sampling.

## Richness estimators

In the R code chunk below, do the following:

1. Write a function to calculate **Chao1**,
2. Write a function to calculate **Chao2**,
3. Write a function to calculate **ACE**, and
4. Use these functions to estimate richness at **site1** and **soilbac1**.

```
S.chao1 <- function(x = ""){
  S.obs(x) + (sum(x == 1)^2) / (2 * sum(x == 2))
}

# Chao1 cannot use a site-by-species matrix
# singletons in Chao1 are species that appear once at a given site (and doubletons appear twice at a given site)
# Abundance of species at ONE SITE

S.chao2 <- function(site = "", SbyS = ""){
  SbyS = as.data.frame(SbyS)
  x = SbyS[site, ]
  SbyS.pa <- (SbyS > 0) * 1 # convert the SbyS to presence/absence
  Q1 = sum(colSums(SbyS.pa) == 1) # species observed once
  Q2 = sum(colSums(SbyS.pa) == 2) # species observed twice
  S.chao2 = S.obs(x) + (Q1^2)/(2 * Q2)
  return(S.chao2)
}

# singletons in Chao2 are species that appear once across ALL sites (and doubletons only appear twice across ALL sites)
# PRESENCE of species across MULTIPLE SITES

S.ace <- function(x = "", thresh = 10){
  x <- x[x>0] # excludes zero-abundance taxa
  S.abund <- length(which(x > thresh)) # richness of abundant taxa
  S.rare <- length(which(x <= thresh)) # richness of rare taxa
  singlt <- length(which(x == 1)) # number of singleton taxa
  N.rare <- sum(x[which(x <= thresh)]) # abundance of rare individuals
  C.ace <- 1 - (singlt / N.rare) # coverage (prop non-singlt rare inds)
  i <- c(1:thresh) # threshold abundance range
  count <- function(i, y){ # counter to go through i range
    length(y[y == i])
  }
  a.1 <- sapply(i, count, x) # number of individuals in richness i richness classes
  f.1 <- (i * (i - 1)) * a.1 # k(k-1)kf sensu Gotelli
  G.ace <- (S.rare/C.ace)*(sum(f.1)/(N.rare*(N.rare-1)))
  S.ace <- S.abund + (S.rare/C.ace) + (singlt/C.ace) * max(G.ace,0)
  return(S.ace)
}

c(S.chao1(site1), S.chao1(soilbac1))

##          1          T1_1
## 119.6944 2628.5140
```

```
#S.chao2(site1) # error invalid subscript type 'list'
S.chao2(site = 1, SbyS = BCI) # chatGPT suggestion, seems to be an issue with the logic of the function
```

```
##          1
## 104.6053
```

```
S.chao2(site = 1, SbyS = soilbac.t) # same error/fix as above
```

```
##      T1_1
## 21055.39
```

```
c(S.ace(site1), S.ace(soilbac1))
```

```
## [1] 159.3404 4465.9827
```

**Question 4:** What is the difference between ACE and the Chao estimators? Do the estimators give consistent results? Which one would you choose to use and why?

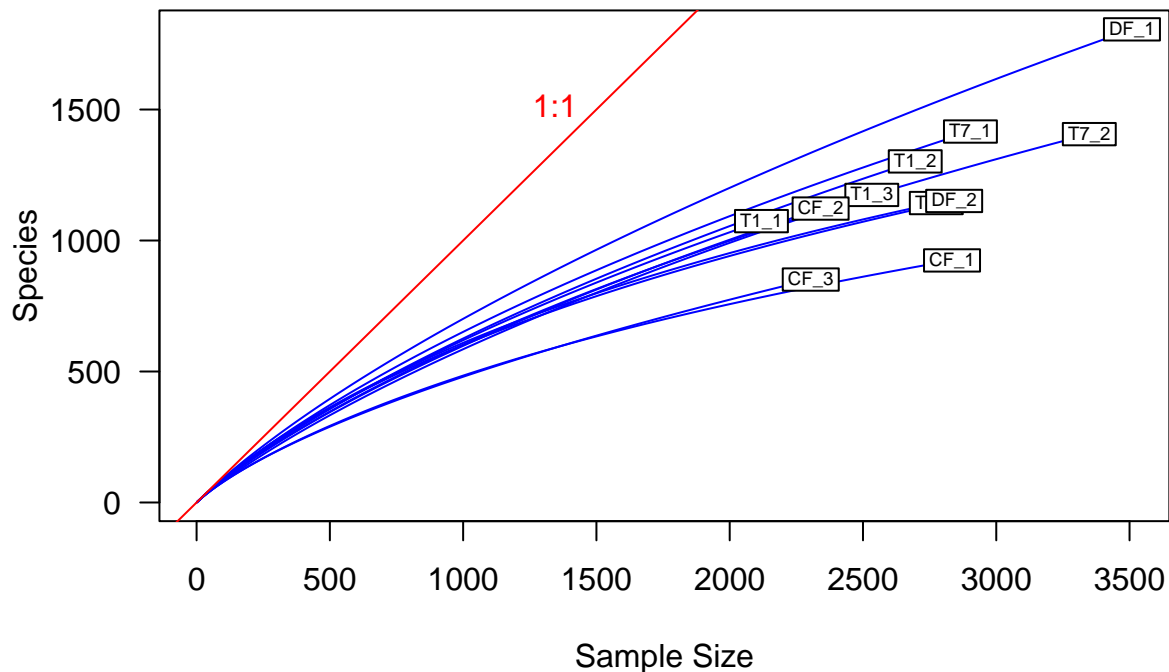
**Answer 4:** site1: Chao1 = 119, Chao2 = 104, ACE = 159 soilbac1: Chao1 = 2628, Chao2 = 21055, ACE = 4465 For site1, the estimators seem relatively consistent, although there is still a pretty large range between 104 (Chao2) and 159 (ACE). For soilbac1, all estimators are different, but Chao2 is significantly higher than the other two estimators, and I don't know why.

## Rarefaction

In the R code chunk below, please do the following:

1. Calculate observed richness for all samples in `soilbac`,
2. Determine the size of the smallest sample,
3. Use the `rarefy()` function to rarefy each sample to this level,
4. Plot the rarefaction results, and
5. Add the 1:1 line and label.

```
soilbac.S <- S.obs(soilbac.t)
min.N <- min(rowSums(soilbac.t))
S.rarefy <- rarefy(x = soilbac.t, sample = min.N, se = TRUE)
rarecurve(x = soilbac.t, step = 20, col = "blue", cex = 0.6, las = 1)
abline(0, 1, col = 'red')
text(1500, 1500, "1:1", pos = 2, col = 'red')
```



#### 4) SPECIES EVNENNESS

Here, we consider how abundance varies among species, that is, **species evenness**.

##### Visualizing evenness: the rank abundance curve (RAC)

One of the most common ways to visualize evenness is in a **rank-abundance curve** (sometime referred to as a rank-abundance distribution or Whittaker plot). An RAC can be constructed by ranking species from the most abundant to the least abundant without respect to species labels (and hence no worries about ‘ties’ in abundance).

In the R code chunk below, do the following:

1. Write a function to construct a RAC,
2. Be sure your function removes species that have zero abundances,
3. Order the vector (RAC) from greatest (most abundant) to least (least abundant), and
4. Return the ranked vector

```
RAC <- function(x = ""){
  x.ab = x[x > 0]
  x.ab.ranked = x.ab[order(x.ab, decreasing = TRUE)]
  as.data.frame(lapply(x.ab.ranked, unlist))
}
```

```
return(x.ab.ranked)
}
```

Now, let us examine the RAC for `site1` of the BCI data set.

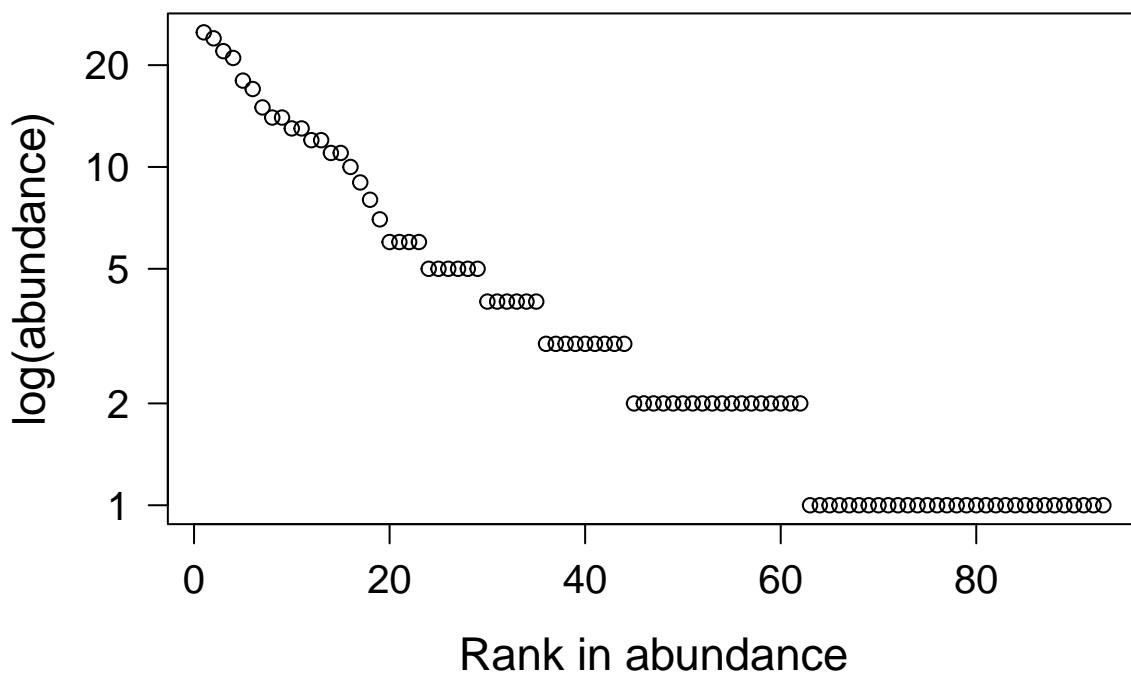
In the R code chunk below, do the following:

1. Create a sequence of ranks and plot the RAC with natural-log-transformed abundances,
2. Label the x-axis “Rank in abundance” and the y-axis “log(abundance)”

```
plot.new()
site1 <- BCI[1, ]

rac <- RAC(x = site1)
ranks <- as.vector(seq(1, length(rac)))
opar <- par(no.readonly = TRUE)           # Saves default plot parameters
par(mar = c(5.1, 5.1, 4.1, 2.1))         # New settings for par
plot(ranks, log(rac), type = 'p', axes = F, # Plots w/o axes
     xlab = "Rank in abundance", ylab = "log(abundance)",
     las = 1, cex.lab = 1.4, cex.axis = 1.25)

box()                                     # Manually adds border
axis(side = 1, labels = T, cex.axis = 1.25) # Manually adds X-axis
axis(side = 2, las = 1, cex.axis = 1.25,    # Manually adds Log-Scaled Y-axis
     labels = c(1, 2, 5, 10, 20), at = log(c(1, 2, 5, 10, 20)))
```





```
par <- opar # Resets plotting parameters
```

**Question 5:** What effect does visualizing species abundance data on a log-scaled axis have on how we interpret evenness in the RAC?

**Answer 5:** Log-scaled axes can lead to misinterpretations of evenness as they downplay the difference between dominant and rare species, making dominant species look less dominant and rare species appear more common.

Now that we have visualized unevenness, it is time to quantify it using Simpson's evenness ( $E_{1/D}$ ) and Smith and Wilson's evenness index ( $E_{var}$ ).

### Simpson's evenness ( $E_{1/D}$ )

In the R code chunk below, do the following:

1. Write the function to calculate  $E_{1/D}$ , and
2. Calculate  $E_{1/D}$  for `site1`.

```
SimpE <- function(x = ""){  
  S <- S.obs(x)  
  x = as.data.frame(x)  
  D <- diversity(x, "inv")  
  E <- (D)/S  
  return(E)  
}  
  
SimpE(site1)
```

```
##           1  
## 0.4238232
```

### Smith and Wilson's evenness index ( $E_{var}$ )

In the R code chunk below, please do the following:

1. Write the function to calculate  $E_{var}$ ,
2. Calculate  $E_{var}$  for `site1`, and
3. Compare  $E_{1/D}$  and  $E_{var}$ .

```
Evar <- function(x){  
  x <- as.vector(x[x > 0])  
  1 - (2/pi) * atan(var(log(x)))  
}  
  
Evar(site1)
```

```
## [1] 0.5067211
```

```
c(SimpE(site1), Evar(site1))
```

```
##           1
## 0.4238232 0.5067211
```

**Question 6:** Compare estimates of evenness for `site1` of BCI using  $E_{1/D}$  and  $E_{var}$ . Do they agree? If so, why? If not, why? What can you infer from the results.

**Answer 6:** # Smith and Wilson's Evenness (Evar) is slightly higher than Simpson's Evenness (SimpE). I am not sure how similar two values need to be to be in agreement, but they appear to be very similar. I can infer that the evenness is neither particularly high or low for this site.

## 5) INTEGRATING RICHNESS AND EVENNESS: DIVERSITY METRICS

So far, we have introduced two primary aspects of diversity, i.e., richness and evenness. Here, we will use popular indices to estimate diversity, which explicitly incorporate richness and evenness. We will write our own diversity functions and compare them against the functions in `vegan`.

### Shannon's diversity (a.k.a., Shannon's entropy)

In the R code chunk below, please do the following:

1. Provide the code for calculating  $H'$  (Shannon's diversity),
2. Compare this estimate with the output of `vegan`'s diversity function using `method = "shannon"`.

```
ShanH <- function(x = ""){
  H = 0
  for (n_i in x){
    if(n_i > 0) {
      p = n_i / sum(x)
      H = H - p*log(p)
    }
  }
  return(H)
}

c(ShanH(site1), diversity(site1, index = "shannon"))
```

```
## [1] 4.018412 4.018412
```

### Simpson's diversity (or dominance)

In the R code chunk below, please do the following:

1. Provide the code for calculating  $D$  (Simpson's diversity),
2. Calculate both the inverse ( $1/D$ ) and  $1 - D$ ,
3. Compare this estimate with the output of `vegan`'s diversity function using `method = "simp"`.

```
SimpD <- function(x = ""){
  D = 0
  N = sum(x)
  for (n_i in x){
    D = D + (n_i^2)/(N^2)
  }
  return(D)
}
D.inv <- 1/SimpD(site1)
D.sub <- 1-SimpD(site1)

diversity(site1, "inv")
```

```
## [1] 39.41555
```

```
D.inv
```

```
## [1] 39.41555
```

```
diversity(site1, "simp")
```

```
## [1] 0.9746293
```

```
D.sub
```

```
## [1] 0.9746293
```

### Fisher's $\alpha$

In the R code chunk below, please do the following:

1. Provide the code for calculating Fisher's  $\alpha$ ,
2. Calculate Fisher's  $\alpha$  for `site1` of BCI.

```
rac <- as.vector(site1[site1 > 0])
invD <- diversity(rac, "inv")
invD
```

```
## [1] 39.41555
```

```
Fisher <- fisher.alpha(rac)
Fisher
```

```
## [1] 35.67297
```

```
ShanH(site1)
```

```
## [1] 4.018412
```

```
Evar(site1)
```

```
## [1] 0.5067211
```

```
E_H <- function(x) {  
  # x = numeric vector of species abundances (e.g., one BCI plot)  
  
  x <- x[x > 0]          # remove absent species  
  p <- x / sum(x)        # relative abundances  
  
  H <- -sum(p * log(p))  # Shannon diversity H'  
  S <- length(x)         # observed species richness  
  
  E <- H / log(S)        # Shannon evenness E_H'  
  return(E)  
}
```

```
E_H(site1)
```

```
## [1] 0.8865579
```

**Question 7:** How is Fisher's  $\alpha$  different from  $E_{H'}$  and  $E_{var}$ ? What does Fisher's  $\alpha$  take into account that  $E_{H'}$  and  $E_{var}$  do not?

**Answer 7:** Fisher's alpha is not an evenness metric (obviously), while the other two are. Fisher's alpha is a parameter of the log-series distribution that attempts to reflect how many rare species there are relative to common species. However, Fisher's alpha is a diversity estimate, not a direct diversity calculation.

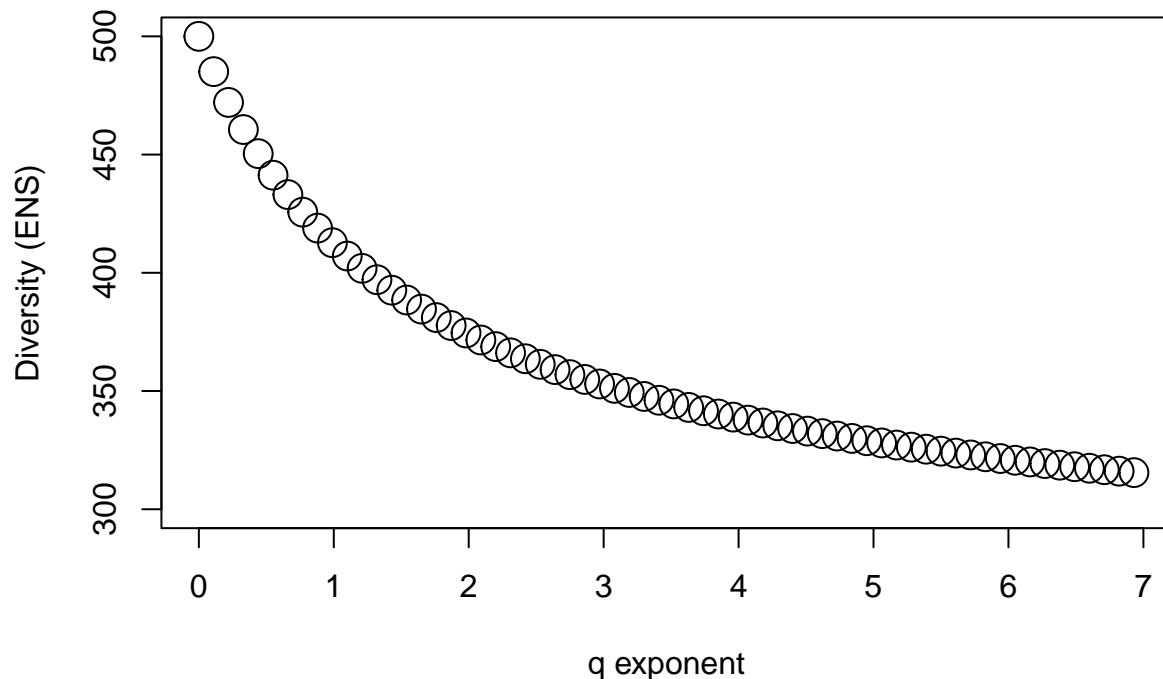
## 6) HILL NUMBERS

Remember that we have learned about the advantages of Hill Numbers to measure and compare diversity among samples. We also learned to explore the effects of rare species in a community by examining diversity for a series of exponents  $q$ .

**Question 8:** Using `site1` of BCI and `vegan` package, a) calculate Hill numbers for  $q$  exponent 0, 1 and 2 (richness, exponential Shannon's entropy, and inverse Simpson's diversity). b) Interpret the effect of rare species in your community based on the response of diversity to increasing exponent  $q$ .

**Answer 8a:**

```
# The following function `profile` calculates the diversity from the equation above along  
# a continuum of q values  
profile <- function(C) {  
  cbind(seq(0, 7, by = 0.11),  
    unlist(lapply(seq(0, 7, by = 0.11), function(q) sum(apply(C, 1, function(x)  
      (x/sum(x))^q)^(1/(1-q)))))) }  
  
set.seed(42)  
C3 <- data.frame(t(sample(1:1000, 500))); colnames(C3) = paste("sp", 1:500)  
C3_profile <- profile(C3)  
plot(C3_profile[,1], C3_profile[,2], ylim=c(300,500), cex = 2,  
xlab = "q exponent", ylab = "Diversity (ENS)")
```



```
# q=0, diversity is species richness
# q=1, diversity is Exponential Shannon diversity
# q=2, diversity is reciprocal of Simpson diversity
```

**Answer 8b:** As  $q$  increases, the weight of rare species is diminished in the diversity metric.

## 7) MOVING BEYOND UNIVARIATE METRICS OF $\alpha$ DIVERSITY

The diversity metrics that we just learned about attempt to integrate richness and evenness into a single, univariate metric. Although useful, information is invariably lost in this process. If we go back to the rank-abundance curve, we can retrieve additional information – and in some cases – make inferences about the processes influencing the structure of an ecological system.

### Species abundance models

The RAC is a simple data structure that is both a vector of abundances. It is also a row in the site-by-species matrix (minus the zeros, i.e., absences).

Predicting the form of the RAC is the first test that any biodiversity theory must pass and there are no less than 20 models that have attempted to explain the uneven form of the RAC across ecological systems.

In the R code chunk below, please do the following:

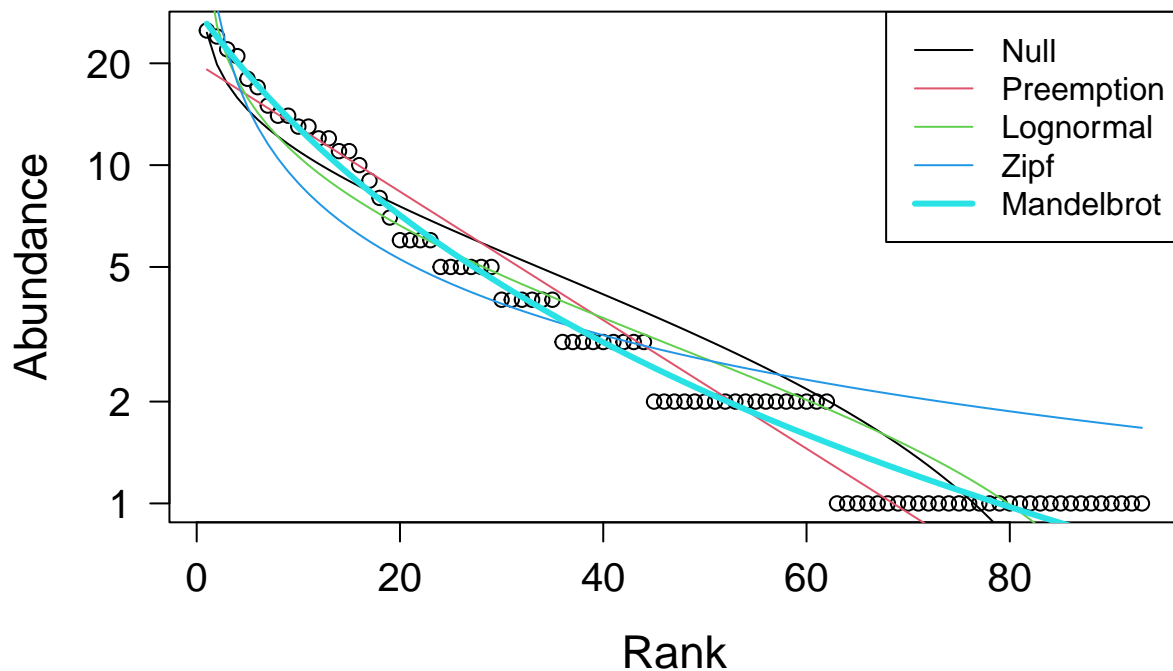
1. Use the `radfit()` function in the `vegan` package to fit the predictions of various species abundance models to the RAC of `site1` in BCI,

2. Display the results of the `radfit()` function, and
3. Plot the results of the `radfit()` function using the code provided in the handout.

```
RACresults <- radfit(site1)
RACresults
```

```
##
## RAD models, family poisson
## No. of species 93, total abundance 448
##
##          par1      par2      par3  Deviance AIC      BIC
## Null                39.5261 315.4362 315.4362
## Preemption 0.042797    21.8939 299.8041 302.3367
## Lognormal  1.0687      1.0186    25.1528 305.0629 310.1281
## Zipf       0.11033 -0.74705    61.0465 340.9567 346.0219
## Mandelbrot 100.52    -2.312     24.084   4.2271 286.1372 293.7350
```

```
plot.new()
plot(RACresults, las = 1, cex.lab = 1.4, cex.axis = 1.25)
```



**Question 9:** Answer the following questions about the rank abundance curves: a) Based on the output of `radfit()` and plotting above, discuss which model best fits our rank-abundance curve for `site1`? b) Can we make any inferences about the forces, processes, and/or mechanisms influencing the structure of our system, e.g., an ecological community?

**Answer 9a:** Mandelbrot has the lowest AIC and BIC values, therefore it should be the best fitting model for our data. **Answer 9b:** There may be competitive hierarchies in the system, but this is not confirmed by simply fitting a model.

**Question 10:** Answer the following questions about the preemption model: a. What does the preemption model assume about the relationship between total abundance ( $N$ ) and total resources that can be preempted? b. Why does the niche preemption model look like a straight line in the RAD plot?

**Answer 10a:** The preemption model assumes that total abundance and total resources available are proportional. Basically, it assumes that there is a fixed amount of resources in the environment. Each individual (or species) preemptively (ahead of each other) consumes a constant fraction of the remaining resources. So if an environment starts out with 100 resources, species A may take 1/10 or 10 resources at the beginning, but if there are 60 left by the time it is species A's turn again, they will take 6 resources. Logically, a species' abundance is proportional to the number of resources captured. Differences in species abundances therefore arise from priority effects or competitive asymmetry because species with higher abundances are somehow getting more resources. **Answer 10b:** The preemption model produces an exponential curve because abundance declines exponentially as resources decline. When plotted on a Log-scale, this curve appears as a straight line.

**Question 10:** Why is it important to account for the number of parameters a model uses when judging how well it explains a given set of data?

**Answer 11:** "A model with more parameters is less constrained and therefore less impressive if it fits well." This is because each parameter reduces how surprising the data looks by fitting more of the data to the curve. You run the risk of overfitting.

## SYNTHESIS

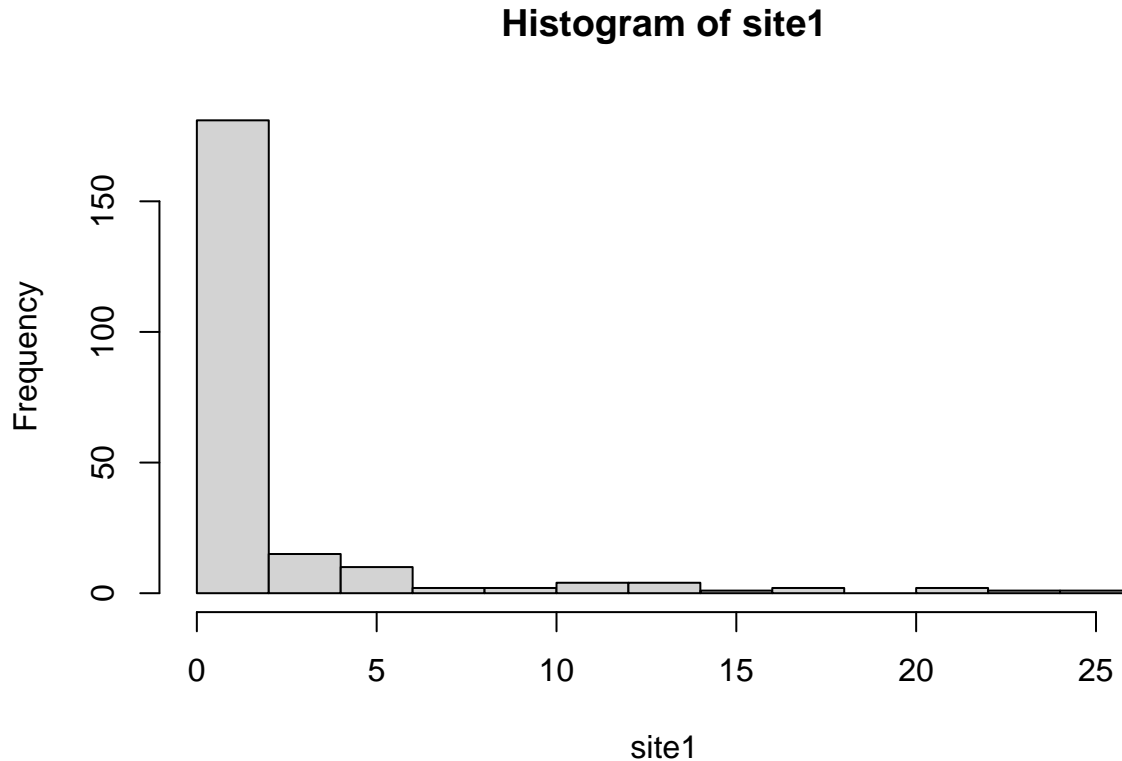
1. As stated by Magurran (2004) the  $D = \sum p_i^2$  derivation of Simpson's Diversity only applies to communities of infinite size. For anything but an infinitely large community, Simpson's Diversity index is calculated as  $D = \sum \frac{n_i(n_i-1)}{N(N-1)}$ . Assuming a finite community, calculate Simpson's D,  $1 - D$ , and Simpson's inverse (i.e.  $1/D$ ) for **site 1** of the BCI site-by-species matrix.

```
D.inv <- 1/SimpD(site1)
D.sub <- 1-SimpD(site1)
print(c(SimpD(site1), D.inv, D.sub))
```

```
## [1] 0.0253707 39.4155538 0.9746293
```

2. Along with the rank-abundance curve (RAC), another way to visualize the distribution of abundance among species is with a histogram (a.k.a., frequency distribution) that shows the frequency of different abundance classes. For example, in a given sample, there may be 10 species represented by a single individual, 8 species with two individuals, 4 species with three individuals, and so on. In fact, the rank-abundance curve and the frequency distribution are the two most common ways to visualize the species-abundance distribution (SAD) and to test species abundance models and biodiversity theories. To address this homework question, use the R function **hist()** to plot the frequency distribution for **site 1** of the BCI site-by-species matrix, and describe the general pattern you see.

```
site1 <- as.numeric(site1)
hist(site1)
```



> Site 1 appears to have many more rare species than common species.

3. We asked you to find a biodiversity dataset with your partner. This data could be one of your own or it could be something that you obtained from the literature. Load that dataset. How many sites are there? How many species are there in the entire site-by-species matrix? Any other interesting observations based on what you learned this week?

```
data(dune)
nrow(dune) # Number of sites
```

```
## [1] 20
```

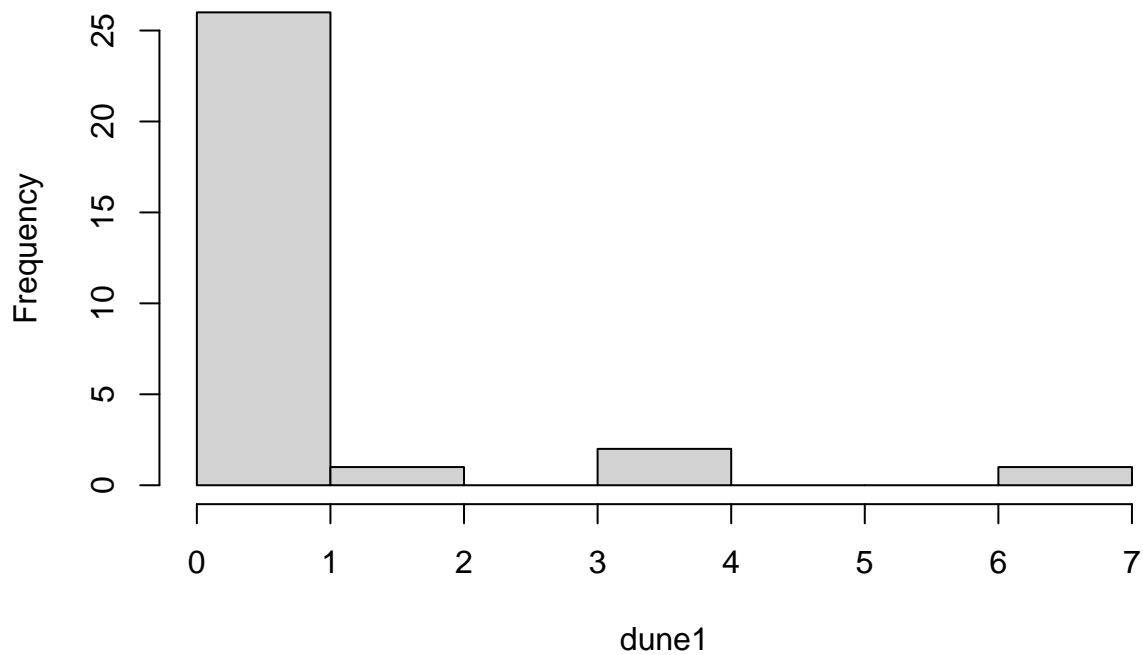
```
ncol(dune) # Number of species
```

```
## [1] 30
```

```
dune1 <- as.numeric(dune[1,])
hist(dune1) # Several rare species
```



## Histogram of dune1



## SUBMITTING YOUR ASSIGNMENT

Use Knitr to create a PDF of your completed 5.AlphaDiversity\_Worksheet.Rmd document, push it to GitHub, and create a pull request. Please make sure your updated repo include both the pdf and RMarkdown files. Unless otherwise noted, this assignment is due on **Wednesday, January 29<sup>th</sup>, 2025 at 12:00 PM (noon)**.