Local (alpha) Diversity

Student Name; Z620: Quantitative Biodiversity, Indiana University

16 December, 2016

OVERVIEW

In this exercise, we will explore aspects of local or site-specific diversity, also known as alpha (α) diversity. First we will quantify two of the fundamental components of (α) 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. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the exercise as possible during class; what you do not complete in class will need to be done on your own outside of 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. Be sure to **answer the questions** in this exercise document; they also correspond to the handout. 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 ">".
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. For homework, 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, please submit the completed exercise by creating a **pull request** via GitHub. Your pull request should include this file *alpha_exercise.Rmd* and the PDF output of `Knitr` (*alpha_exercise.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 */Alpha* folder, and 4) load the `vegan` R package (be sure to install if needed).

2) LOADING DATA

In the R code chunk below, do the following: 1) load the BCI dataset from the `vegan` package, and 2) display the structure of the dataset (if the structure is long, use `max.level=0` to show just basic information).

3) SPECIES RICHNESS

Species richness (S) is simply 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`, and 3. Compare the output of your function to the output of the `specnumber()` function in vegan.

```
# S.obs <- function(   ){
#   rowSums(     ) *
#   }
```

Question 1: Does `specnumber()` from `vegan` return the same value for observed richness of `site1` as our function `S.obs`? What is the species richness of the first 4 sites (i.e., rows) of the BCI matrix?

Answer 1:

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.

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` were represented as singletons?
d. Make some observations about coverage at the BCI plots.

Answer 2a:

Answer 2b:

Answer 2c:

Answer 2d:

Estimated Richness

In the R code chunk below, do the following: 1. load the microbial dataset (located in the */Alpha/data* folder), 2. transform and transpose the data as needed (see handout), 3. create a vector (soilbac1) with the bacterial OTU abundances at any site in the dataset, 4. calculate the observed richness at that particular site, and 5. calculate the coverage at that particular site

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:

Answer 3b:

Answer 3c:

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**, and 3. use these functions to estimate richness at both `site1` and `soilbac1`.

Rarefaction

In the R code chunk below, please do the following: 1. calculate observed richness for all samples in `soilbac` (our function will work with the site-by-species matrix), 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.

4) SPECIES EVENNESS

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, please 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

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

In the R code chunk below, please 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)"

Question 4: What effect does visualizing species abundance data on a log-scaled axis have on how we interpret evenness in the RAC?

Answer 4:

Now that we have visualized unevennes, it is time to quantify it. Here, we will introduce two metrics of evenness that meet the above criteria: 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`.

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}.

Question 5: Compare estimates of evenness for `site1` of BCI using $E_{1/D}$ and E_{var}. Do they agree? If so, why? If not, why not? What can you infer from the results.

Answer 5:

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 agains 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".

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, 2. compare this estimate with the output of `vegan's` diversity functionusing method = "simp".

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

Answer 6:

6) MOVING BEYOND UNIVARIATE METRICS OF α 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 and 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 2. plot the results of the `radfit()` function using the code provided in the handout.

Question 7: Based on the output of `radfit()` and plotting above, discuss which model best fits our rank-abundance curve for `site1`? Can we make any inferences about the forces, processes, and/or mechanisms influencing the structure of our system, e.g., an ecological community?

Answer 7a: ***Answer 7b***:

Question 8: 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 8a:
Answer 8b:

Question 9: 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 9:

HOMEWORK

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.

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.

3. Use Knitr to create a pdf of your completed alpha_exercise.Rmd document, push it to GitHub, and create a pull request. The due date for this assignment will be announced in class and/or canvas.