6. Worksheet: Diversity Sampling

Lana Bolin; Z620: Quantitative Biodiversity, Indiana University

29 January, 2019

OVERVIEW

In this worksheet, you will use the jelly bean site-by-species matrix generated from **6. Diversity Sampling**. Along with tools outlined in the **5. Local (alpha) Diversity** and **7. Control Structures** handouts, you will develop questions, visualize data, and test hypotheses related to sampling effects and its effect on estimates of within-sample biodiversity.

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. Refer to previous handout to help with developing of questions and writing of code.
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 `6.DiversitySampling_Worskheet.Rmd` and the PDF output of `Knitr` (`DiversitySampling_Worskheet.pdf`).

1) Group brainstorming

With your team partner and perhaps other students in the class, spend 15 minutes or so brainstorming questions, code, "fantasy figures", and statistical tests that could be used to test questions with the class's data represented in the site-by-species matrix that you have generated.

2) Code

Use the space below for code that is being used to analyze your data and test your hypotheses. Create one (and only one, although it can have multiple panels) *publication quality* figure. Make sure to annotate your code using # symbols so others (including instructors) understand what you have done and why you have done it.

Setup

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

```
## [1] "/Users/lana/GitHub/QB2019_Bolin/2.Worksheets/6.DiversitySampling"
```

```
setwd("~/GitHub/QB2019_Bolin/2.Worksheets/6.DiversitySampling/")
require("vegan")
```

```
## Loading required package: vegan

## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.5-3
```

Load data

```
jelly <- read.table("JellyBeans.txt", sep = "\t", header = TRUE)
jelly$Site <- NULL    # remove Site column
source <- read.table("JellyBeans.Source.txt", sep = "\t", header = TRUE)

# Make a df for each population sampled
jelly.A <- jelly[jelly$Group == "A", ]
jelly.B <- jelly[jelly$Group == "B", ]

# Get sum for each speces
jelly.A.sums <- colSums(jelly.A[, -1])
jelly.B.sums <- colSums(jelly.B[, -1])
```

Find best model fit for RAC curves

Source Community

```
RACresults.source <- radfit(source$Count)
RACresults.source
```

```
##
## RAD models, family poisson
## No. of species 26, total abundance 967
##
##               par1       par2      par3     Deviance AIC      BIC
## Null                                        132.735  265.955 265.955
## Preemption  0.11395                         134.732  269.952 271.210
## Lognormal   3.2358    0.90574                 64.712  201.932 204.448
## Zipf        0.20875  -0.85112                 35.213  172.433 174.949
## Mandelbrot  0.33614  -1.0201    0.7163       30.738  169.958 173.732
```

```
# Mandelbrot has the best fit, so we'll plot that model

RAC.mandelbrot.source <- rad.zipfbrot(source$Count)
```

Samples

```
# Sample A
RACresults.A <- radfit(jelly.A.sums)
RACresults.A
```

```
##
## RAD models, family poisson
## No. of species 28, total abundance 384
##
##               par1       par2      par3     Deviance AIC      BIC
## Null                                         73.055  184.897 184.897
## Preemption   0.13                            82.060  195.901 197.233
## Lognormal   2.0586    1.1033                  37.143  152.985 155.649
## Zipf        0.24844  -0.98266                 25.108  140.949 143.614
```

```
## Mandelbrot  0.41627 -1.1671   0.65773  22.930  140.772 144.769
```

```
  ## Mandelbrot is the best
RAC.mandelbrot.A <- rad.zipfbrot(jelly.A.sums)
```

```
# Sample B
RACresults.B <- radfit(jelly.B.sums)
RACresults.B
```

```
##
## RAD models, family poisson
## No. of species 27, total abundance 286
##
##              par1      par2      par3    Deviance AIC       BIC
## Null                                     11.8569 116.0347 116.0347
## Preemption   0.1114                      16.0307 122.2085 123.5043
## Lognormal    1.9896    0.89268            4.6338 112.8116 115.4032
## Zipf         0.19113 -0.80373            13.6631 121.8408 124.4325
## Mandelbrot   1.5941  -1.4828   3.8234     6.8337 117.0114 120.8989
```

```
  ## Lognormal is the best
RAC.mandelbrot.B <- rad.lognormal(jelly.B.sums)
```

Plot RAC curves

```
plot.new()
par(mfrow = c(1, 3))
plot(RAC.mandelbrot.source, las = 1, cex.lab = 1.4, cex.axis = 1.25, main = "Source Community")
text(23, 155, "(a)", cex = 1.5)
plot(RAC.mandelbrot.A, las = 1, cex.lab = 1.4, cex.axis = 1.25, main = "Sample A")
text(25, 73, "(b)", cex = 1.5)
plot(RAC.mandelbrot.B, las = 1, cex.lab = 1.4, cex.axis = 1.25, main = "Sample B")
text(25, 40, "(c)", cex = 1.5)
```

Source Community **Sample A** **Sample B**
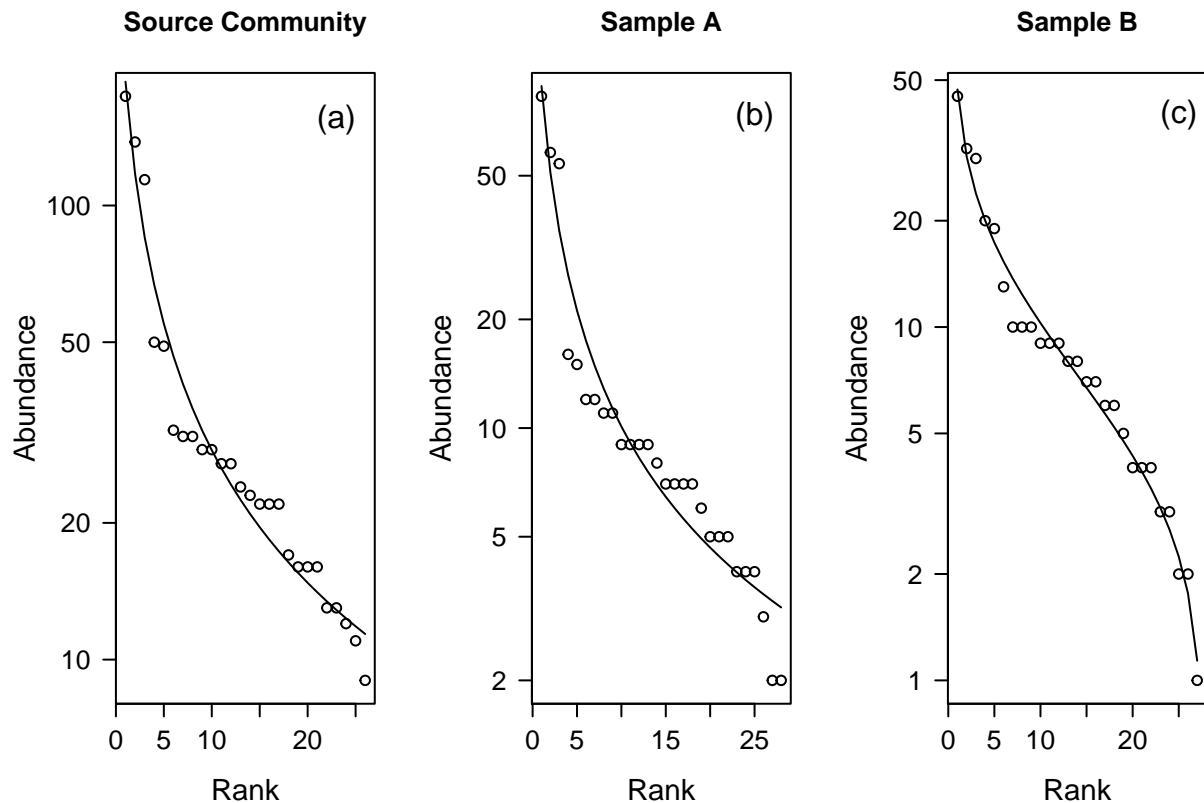
3) Figure caption

Write an informative yet succicnt (~5 sentences) caption that creates a "stand-alone" figure. Take a peek at figures and figure captions in a paper published in your favorite journal for inspiration.

> Rank-Abundance Curves differed between a source community that underwent a vicariance event and samples taken from each subcommunity created by the vicariance event. The best model fit for each curve was selected using the Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC) in the radfit function (vegan package) in R. A Zipf-Mandelbrot model was the best fit for (a) the Source Community, and (b) Sample A (n = 5 sites in subcommunity A), but model coefficients differed (Source Community: c = 0.34, gamma = -1.02, and beta = 0.72. Sample A: c = 0.42, gamma = -1.17, and beta = 0.66). A Lognormal model was the best fit for (c) Sample B (n = 4 sites in subcommunity B). Each point represents a morphospecies.

SUBMITTING YOUR ASSIGNMENT

Use Knitr to create a PDF of your completed 6.DiversisitySampling_Worksheet.Rmd document, push it to GitHub, and create a pull request. Please make sure your updated repo includes both the pdf and RMarkdown files.

Unless otherwise noted, this assignment is due on **Wednesday, January 30th, 2017 at 12:00 PM (noon)**.