Richness Analyses

Jeff Oliver

August 15, 2018

Testing for a significant difference in species richness (i.e. number of species) between the two collection methods, Pollard walks and Malaise traps. I used a paired t-test for the analysis.

Methods

Setup

There is some data wrangling that needed to happen, and I had to use the `tidyr` data for some of that (specifically, the `spread` function). For a nice figure, I used the `ggplot2` package.

```r
library("tidyr")
library("ggplot2")
bioscan <- read.csv(file = "data/BioScanData.csv")

# Drop any rows missing data
bioscan <- na.omit(bioscan)
```

Data wrangling

For the paired t-test, we only want to use data for those sites where we have information for both collection types. So we need to identify which sites have Pollard walk data, which sites have Malaise trap data, and which sites have **both** (the intersection of Pollard and Malaise).

```r
# Identify sites with data for each of the two collection methods
pollard.sites <- bioscan$Site.Number[bioscan$Collection.Method == "Pollard Walk"]
malaise.sites <- bioscan$Site.Number[bioscan$Collection.Method == "Malaise"]

# Identify sites with data for *both* collection methods
sites.with.both <- intersect(x = pollard.sites, y = malaise.sites)

# Reduce dataset to only those sites with both types of data
bioscan <- bioscan[bioscan$Site.Number %in% sites.with.both, ]
rownames(bioscan) <- NULL
```

Richness calculation

To calculate species richness, we count, for each site/collection method combination, the number of species for which at least one individual was observed. Our data are currently organized so that a single row represents a single site/collection method combination, so we can perform this operation once for each row and store the data in a new column called `richness`.

```r
# Identify those columns with species data
species.cols <- c(5:33)

# Calculate richness for each row (total number of species with at least one
```

```
# individual observed)
bioscan$richness <- apply(X = bioscan[, species.cols],
                          MARGIN = 1,
                          FUN = function(x) {
                            sum(x > 0)
                          })
```

More(!) data wrangling

Now we need to transform the data into a data frame with each site in a single row and richness values for the two collection methods in two columns. E.g.:

Site Number	Pollard walk richness	Malaise trap richness
1	3	9
2	2	4
. . .	. . .	. . .
n	10	6

Here we use `tidyr`'s `spread` function to transform our `bioscan` data frame.

```
# Create data frame for t-test. One column for Pollard, one for Malaise
richness.df <- bioscan[, c("Site.Number", "Collection.Method", "richness")]
richness.df <- richness.df %>%
  spread(Collection.Method, richness)
```

Statistical analyses

Finally, we can run a paired t-test, comparing richness of Malaise trap samples to richness of Pollard walks at the same site.

```
richness.t <- t.test(x = richness.df$Malaise,
                     y = richness.df$`Pollard Walk`,
                     paired = TRUE)
```

Results

The t-test shows a significant difference between the two collection methods (t = -4.657, $p < 0.001$, 95%CI of the mean difference between Maliase traps and Pollard walks [-7.835, -2.915]). On average, Pollard walks reported 5.375 more species than were collected in Malaise traps.

We can compare the two types visually with a boxplot:

```
richness.boxplot <- ggplot(data = bioscan, mapping = aes(x = Collection.Method, y = richness)) +
  geom_boxplot() +
  xlab(label = "Collection Method") +
  ylab(label = "Species Richness")
print(richness.boxplot)
```