Advanced Ecology 462 R tutorials

4-Mar-2016

Rank-abundance revisited

1. Rank-abundance curves of trophic groups

Last week, we used a loop to produce rank-abundance curves for each Pacific island in our diversity dataset. Let's consider the rank-abundance curves for different trophic groups - understanding species dominance in reef fishes may be most relevant when we consider species that have similar functional roles.

We have been using the vegan package and the table() function to create a community abundance matrix. Instead of presence/absence data, we're going to use the count variable to calculate true species abundances.

Remember to load in your clean data frame before we start examining the data.

```
## Sum counts across species, by island with aggregate()
## save piscivores and herbivores as different data frames
ranks<-aggregate(count ~ species + trophic + island + state, cred, sum)
pisc<-ranks[ranks$trophic=="Pisc",]
herb<-ranks[ranks$trophic=="Herb",]

### piscivore rank abundance for each island, in one plot, with a loop
island.vec<-unique(cred$island)
plot(sort(pisc$count[pisc$island==island.vec[1]], decreasing=TRUE),bg=1,type="b",
ylab="Abundance", xlab="Rank", pch=21, ylim=c(0, max(pisc$count)), main="Piscivore
 abundance")
for (i in 2:8){
    points(sort(pisc$count[pisc$island==island.vec[i]], decreasing=TRUE), bg=i, ty
pe="b", pch=21)
    }
legend(20, 400, legend=levels(cred$island), pt.bg=1:8, pch=21)
```

Notice how we plot the first island, then use a loop that goes from 2 to 8 to add lines for each island.

```
## Repeat for herbivores
plot(sort(herb$count[herb$island==island.vec[1]], decreasing=TRUE),bg=1,type="b",
ylab="Abundance", xlab="Rank", pch=21, ylim=c(0, max(herb$count)), main="Herbivore
 abundance")
for (i in 2:8){
    points(sort(herb$count[herb$island==island.vec[i]], decreasing=TRUE), bg=i, ty
pe="b", pch=21)
    }
legend(20, 400, legend=levels(cred$island), pt.bg=1:8, pch=21)
```

The biggest change in community structure appears to happen to the piscivore trophic group. The rank-abundance plot shows us that piscivores on populated islands (Kauai, Oahu, Hawaii) are less abundant, there are fewer species, and have higher dominance.

How does this match up with your analyses of piscivore biomass from earlier tutorials?

What is the pattern at Johnston atoll?

2. Statistical models - multiple regression

We've run some basic statistical tests to examine the drivers of reef fish diversity, but we haven't used multiple linear regression yet. In observational studies, it's highly unlikely that we'll be able to say that changes in our response variable are due to a single explanatory variable.

In the CRED dataset, published studies and ecological theory tell us that productivity, SST and anthropogenic disturbances can influence community structure (e.g. biomass distributions) and species diversity (e.g. richness and dominance?). We can use multiple regression to examine how each of these variables contributes to changes in diversity.

Multiple regression is a form of **statistical modelling** where we try to explain the variance in a response variable in terms of explanatory variables. Statistical modelling is used for **hypothesis testing and prediction** - be clear on the distinction with conceptual, simulation or mathematical modelling (simulating ecological processes with equations, e.g. Lotka-Volterra models, stock-recruitment models).

The dataset we're currently using only covers 8 islands, and we realised last week that we had quite a low sample size. These research cruises collected data in 2010, but the CRED programme has been running for over ten years. Let's add in some data from other cruises.

```
cred$year<-2010     ### add variable to old dataset
cred11<-read.csv("CRED_reef_diversity_2011_cruises.csv")
cred11$year<-2011
dim(cred); dim(cred11)   # compare sizes of datasets
```

The 2011 data **are** clean. So we can simply bind these datasets together (but you should always check for errors when you get new datasets!).

```
## use rbind to 'row bind' the two datasets
cred_full<-rbind(cred, cred11)
head(cred_full)
dim(cred_full)
## are there any new islands in the dataset?
unique(cred_full$island)
```

Now let's calculate the species richness (not rarefied - how would you do this?) for each island.

```
richness<-aggregate(species ~ island + state + region + productivity + log_populat
ion + SST, cred_full, function(x) length(unique(x)))
richness

## multiple linear regression for species richness
mod<-lm(species ~ region   + productivity + SST, richness)
summary(mod)
```

So how do we interpret these estimates, standard errors and P values? the visreg() package can help guide our understanding here..

```
require(visreg)
par(mfrow=c(2,2)) ## set up the plot space for all predictor variables (n=3)
visreg(mod)
```

The (Intercept) estimate (-14.659) tells us species diversity **when all predictor variables are set to zero**. This often isn't ecologically relevant. Why not?

We are more interested in the estimates for our parameters. Let's take the continuous variables first - productivity and SST. The coefficient estimates are positive, indicating that species richness increases with productivity and SST. The slope of the relationships are 189.133 (for productivity) and 1.537 (for SST). The Std. Error can be used to estimate 95% confidence intervals, and the P value indicates the significance of each coefficient - in a linear model, the null hypothesis ($P > 0.05$) is that the slope is 0.

Therefore, the productivity slope has a significant P value (0.0337), and so that slope is significantly different from 0. What about SST?

So what do we conclude? Productivity is a significant predictor of species richness of Pacific coral reef fishes (piscivores and herbivores).

Don't forget to check your model assumptions! See last weeks' tutorial for residual plots in R.

3. A toolkit for ecological analysis in R

Over the past 6 tutorials you have learned how to:

1. Read data into R and to clean that data

2. Manipulate data frames in order to calculate summary statistics
3. Create basic plots with colours and legends
4. Run t-tests, anova, linear regression
5. Use ecological functions in the vegan package
6. Create logical loops
7. Write your own functions

These skills set you up to start your own ecological analyses. Remember some key guidelines when you're working with your own datasets:

1. **Clean your data** - look for NAs, spelling mistakes, outliers in all numeric variables
2. **R is more diverse and more powerful than you can imagine** - there are many different ways to manipulate datasets in R - talk to your classmates and colleagues, use online tutorials and R forums, look out an R textbook, ask your TA for help (!). No one way is better than another, but some are much much faster...
3. **Plots are powerful** - the best papers often have relevant and informative plots. Become familiar with different plot types and always use **colour**!
4. **Be exhaustive in your exploratory analyses** - plot everything and anything that is relevant to your hypotheses. This is the only way to become familiar with your dataset and to ensure that your results are robust.

Obviously, we've only scratched the surface of R but you're ready to teach yourself from here. I'd suggest exploring the mapping functions, data aggregation packages ("plyr" is very useful), other plotting functions ("ggplot2" is popular), and become familiar with different statistical methods (almost every ecologist uses mixed effects models these days).

And if you feel like breaking out of R studio, there are lots of other text editors to explore (Sublime Text, Atom).

Good luck!