Lab 10 Constrained Ordination

The goal of this lab is to apply Constrained Ordination techniques to determine the influence of explanatory variables on patterns of variation in multivariate response variables. Constrained ordination is an extension of unconstrained ordination techniques in which the solution is constrained to be expressed by explanatory variables. The two approaches you will consider are **1) Redundancy Analysis (RDA)**, which assume a linear relationship between response and explanatory variables and builds off of PCA, and **2) Canonical Correspondence Analysis (CCA)**, which assumes a unimodal relationship between response and explanatory variables and builds off of CA.

# Set up R session

## Data

Today you will be using a data set from Northern Finland that includes plant cover data for 44 species (varespec in library *vegan*) and 14 environmental variables (varechem in library *vegan*) across 24 sites.

## Download packages

We will be using the following packages:

library(raster)  
library(vegan)

**You will also be using a function from a Biostats package developed by Kevin McGarigal.** Save this package (i.e., R script) to your working directory and use the source function to call it in:

source("biostats.r")

## Import data

After downloading the vegan library, explore and call in the data sets varespec and varechem:

data(varespec)  
  
data(varechem)  
  
str(varespec)  
summary(varespec)  
  
str(varechem)  
summary(varechem)

To learn more about the data sets:

`?`(varespec)  
  
`?`(varechem)

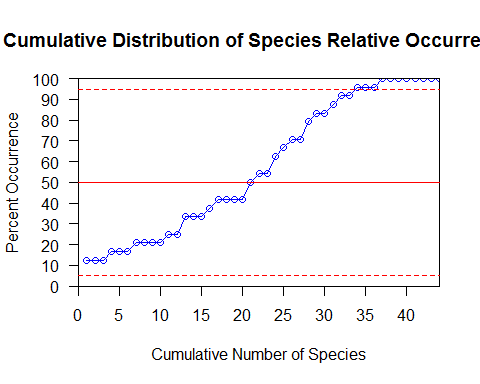
# Data selection, transformation and standardization

Species within community data sets vary greatly in their occurrence, abundance, and habitat specificity. Species that are common, widespread and extremely abundant can obscure patterns in the ordination. Species that are rare and have few occurrences in a data set may not be accurately placed in ecological space. You must decide which species are “rare” and which are super abundant.

## Selecting Species

To explore patterns of rarity and commonness, you will use the foa function from the *Biostats* package. This function will give you a whole series of plots that allow you to explore the occurrence and abundance patterns of the species in your data. The second plot, *Empirical Distribution of Species Relative Occurrence*, will be the one we use to remove common and/or rare species.

occur <- foa.plots(varespec)



rare <- which(occur[, 2] < 5)  
  
common <- which(occur[, 2] > 95)  
  
reduced <- varespec[, -c(rare, common)]

## Species transformations and standardizations

First, check if species abundances are normally distributed across sites:

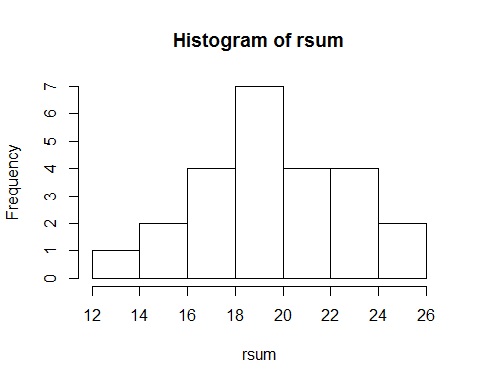
mapply(hist, as.data.frame(varespec[, 1:44]), main = colnames(varespec[, 1:44]),   
 xlab = "abundance")

As you can see, most of the species distributions are right skewed. Use the log transformation (logx+1) to transform the species distributions for both the full and reduced datasets:

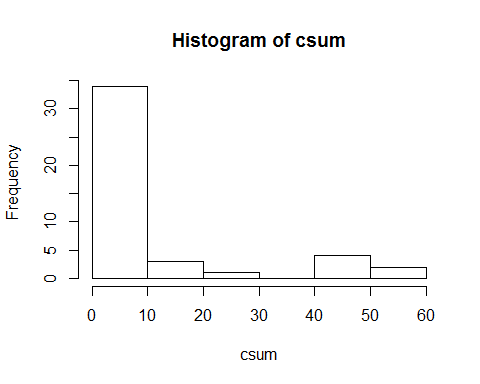
log.full <- log1p(varespec)  
log.red <- log1p(reduced)

Next, check the row and column sum variability using the coefficient of variation (cv) for both data sets:

# Full data set:  
rsum <- rowSums(log.full)  
csum <- colSums(log.full)  
hist(rsum)



hist(csum)



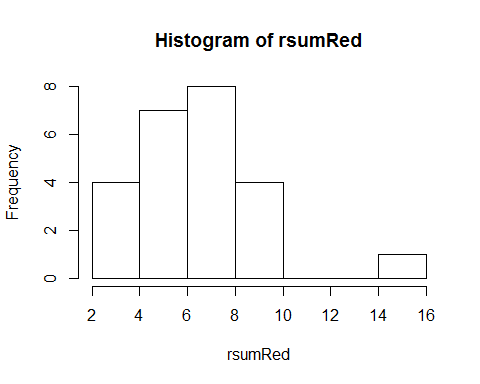
cv(rsum)

## [1] 15.6402

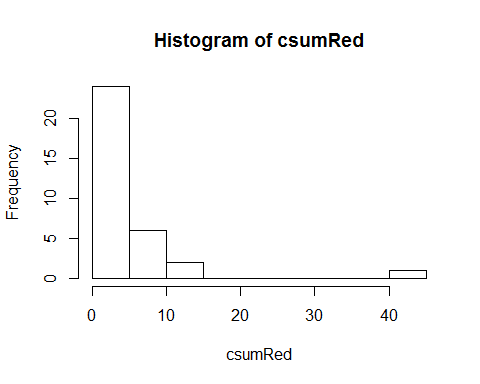
cv(csum)

## [1] 154.4498

# Reduced data set:  
rsumRed <- rowSums(log.red)  
csumRed <- colSums(log.red)  
hist(rsumRed)



hist(csumRed)



cv(rsumRed)

## [1] 41.43926

cv(csumRed)

## [1] 167.6967

If either the row or column sums have cv >50, standardize by the total:

cSpec <- sweep(log.full, 2, csum, "/")  
cSpecRed <- sweep(log.red, 2, csumRed, "/")

##Determine Response Model (RDA vs. CCA)

Now that the data are reduced, transformed and standardized, you need to determine if species abundances show a linear (RDA) or a unimodal (CCA) relationship with the underlying gradient.

First, use Detrended Correspondence Analysis (DCA) to determine the length of the canonical axes. You will use the decorana function in the *vegan* Library. While DCA is a separate analysis with its own assumptions and multifaceted output, you will focus on axis length. An axis length > 3 is evidence of a unimodal relationship. An axis length of <3 is evidence of a linear relationship.

`?`(decorana)

decorana(cSpec)  
decorana(cSpecRed)

Next, plot out each species on the first canonical axis. You need to set the environmental variables first (the next section will get into the details of the explanatory variables). For, now just set them and run the initial CCA to check for linearity.

Set Explanatory Variables:

Vars <- varechem[, c(1, 2, 7)]  
env <- as.data.frame(scale(Vars))

Run CCA:

sp.CCA <- cca(cSpec ~ ., data = env)

Function for plotting species abundances vs. CCA Axis 1:

f2 <- function(x) {  
 plot(x ~ sp.CCA$CC$wa[, 1], xlab = "CCA AXIS 1", ylab = "Abundance ")  
}  
  
# Apply the function across all the species:  
  
mapply(f2, varespec)

# Explanatory Variables

Constrained ordination affords you the ability to include explanatory variables in the ordination. You want to avoid mullitcolinearity among explanatory variables and check if they are measured on the same scale. Based on *a priori* knowledge of this system, use the variables AL, P, and N in the ordination.

First look at all of the pairwise correlations between these variables:

Vars <- varechem[, c(1, 2, 7)]  
Vars  
round(as.dist(cor(Vars)), 2)

Do the variables AL, P, N look like they are measured on different scale? Check the cv to see if you need to z-standardize them:

cv(colSums(Vars))

## [1] 91.17077

You need to make a data frame of the scaled variables to run the Constrained Ordination:

env <- as.data.frame(scale(Vars))

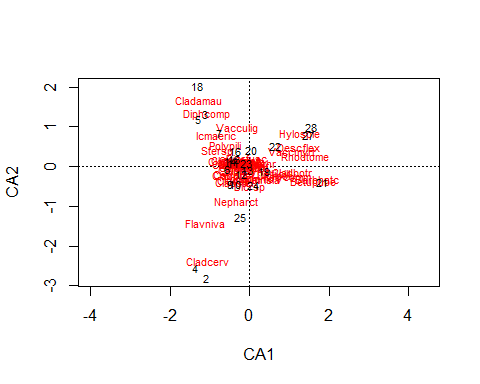
# Running the CCA

You will run the constrained ordination using the cca in the *vegan* library.

`?`(cca)

## Unconstrained Ordination (CA) Before running the constrained model, run an unconstrained ordination (i.e. a regular Correspondence Analysis (CA; See Lab 4). CA will give you a measure of the amount of variation in the site by species matrix that you will try to explain with the explanatory variables (i.e. constraints).

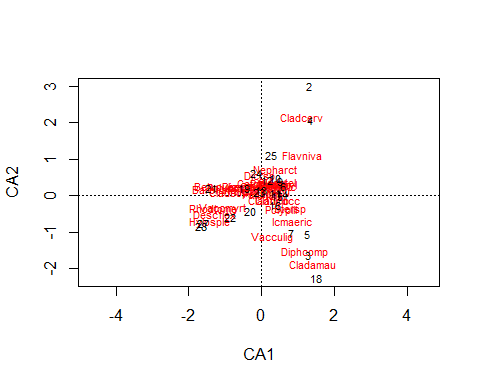
# Full Data  
ca <- cca(cSpec)  
plot(ca)



summary(ca)

##   
## Call:  
## cca(X = cSpec)   
##   
## Partitioning of scaled Chi-square:  
## Inertia Proportion  
## Total 3.342 1  
## Unconstrained 3.342 1  
##   
## Eigenvalues, and their contribution to the scaled Chi-square   
##   
## Importance of components:  
## CA1 CA2 CA3 CA4 CA5 CA6 CA7  
## Eigenvalue 0.5378 0.4014 0.3668 0.3381 0.32691 0.24114 0.1995  
## Proportion Explained 0.1609 0.1201 0.1098 0.1012 0.09782 0.07216 0.0597  
## Cumulative Proportion 0.1609 0.2810 0.3908 0.4920 0.58980 0.66196 0.7217  
## CA8 CA9 CA10 CA11 CA12 CA13  
## Eigenvalue 0.16011 0.14485 0.10309 0.09205 0.07974 0.06703  
## Proportion Explained 0.04791 0.04334 0.03085 0.02754 0.02386 0.02006  
## Cumulative Proportion 0.76957 0.81292 0.84376 0.87131 0.89517 0.91523  
## CA14 CA15 CA16 CA17 CA18 CA19  
## Eigenvalue 0.06280 0.05378 0.04932 0.03876 0.025817 0.020581  
## Proportion Explained 0.01879 0.01609 0.01476 0.01160 0.007725 0.006158  
## Cumulative Proportion 0.93402 0.95011 0.96487 0.97647 0.984192 0.990350  
## CA20 CA21 CA22 CA23  
## Eigenvalue 0.013192 0.008887 0.007156 0.0030136  
## Proportion Explained 0.003947 0.002659 0.002141 0.0009017  
## Cumulative Proportion 0.994298 0.996957 0.999098 1.0000000  
##   
## Scaling 2 for species and site scores  
## \* Species are scaled proportional to eigenvalues  
## \* Sites are unscaled: weighted dispersion equal on all dimensions

# Reduced Data  
ca <- cca(cSpecRed)  
plot(ca)



summary(ca)

##   
## Call:  
## cca(X = cSpecRed)   
##   
## Partitioning of scaled Chi-square:  
## Inertia Proportion  
## Total 4.066 1  
## Unconstrained 4.066 1  
##   
## Eigenvalues, and their contribution to the scaled Chi-square   
##   
## Importance of components:  
## CA1 CA2 CA3 CA4 CA5 CA6 CA7  
## Eigenvalue 0.6236 0.5020 0.4513 0.4136 0.37332 0.29646 0.27791  
## Proportion Explained 0.1534 0.1235 0.1110 0.1017 0.09181 0.07291 0.06835  
## Cumulative Proportion 0.1534 0.2768 0.3878 0.4895 0.58135 0.65425 0.72260  
## CA8 CA9 CA10 CA11 CA12 CA13  
## Eigenvalue 0.18952 0.14892 0.13796 0.12547 0.09429 0.09024  
## Proportion Explained 0.04661 0.03662 0.03393 0.03086 0.02319 0.02219  
## Cumulative Proportion 0.76921 0.80583 0.83975 0.87061 0.89380 0.91599  
## CA14 CA15 CA16 CA17 CA18 CA19  
## Eigenvalue 0.08554 0.07347 0.05977 0.04285 0.03054 0.022178  
## Proportion Explained 0.02103 0.01807 0.01470 0.01054 0.00751 0.005454  
## Cumulative Proportion 0.93703 0.95509 0.96979 0.98033 0.98784 0.993295  
## CA20 CA21 CA22 CA23  
## Eigenvalue 0.012906 0.007665 0.005027 0.001667  
## Proportion Explained 0.003174 0.001885 0.001236 0.000410  
## Cumulative Proportion 0.996469 0.998354 0.999590 1.000000  
##   
## Scaling 2 for species and site scores  
## \* Species are scaled proportional to eigenvalues  
## \* Sites are unscaled: weighted dispersion equal on all dimensions  
##   
##   
  
## Biplot scores for constraining variables  
##   
## CCA1 CCA2 CCA3 CA1 CA2 CA3  
## N -0.5069 0.02840 -0.8615 0 0 0  
## P 0.2819 -0.94801 0.1476 0 0 0  
## Al 0.9184 0.02965 -0.3944 0 0 0

The first thing you should focus on in the summary is the proportion of “inertia” (i.e. variance) explained by the Constrained Ordination. Notice that the total amount of inertia is the same as the Unconstrained Ordination you just ran.

Now look at the eigenvalue and proportion and cumulative amount of variation.

## Monte Carlo testing of the significance of the constrained axis.

The permutation allows you to test if you constrained axes explain more variation than would be expected randomly. You will use the anova.cca function in vegan to conduct the permutation. It is “anova-like” but not an anova. Global Test (i.e. all variables together):

anova(sp.CCA)

## Permutation test for cca under reduced model  
## Permutation: free  
## Number of permutations: 999  
##   
## Model: cca(formula = cSpec ~ N + P + Al, data = env)  
## Df ChiSquare F Pr(>F)   
## Model 3 0.70046 1.7678 0.001 \*\*\*  
## Residual 20 2.64149   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Axes Tests (i.e. each axis individually):

anova(sp.CCA, by = "axis")

## Permutation test for cca under reduced model  
## Forward tests for axes  
## Permutation: free  
## Number of permutations: 999  
##   
## Model: cca(formula = cSpec ~ N + P + Al, data = env)  
## Df ChiSquare F Pr(>F)   
## CCA1 1 0.32219 2.4395 0.005 \*\*  
## CCA2 1 0.23057 1.7457 0.065 .   
## CCA3 1 0.14770 1.1183 0.317   
## Residual 20 2.64149   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Variable Tests (i.e. each variable individually):

anova(sp.CCA, by = "terms")

## Permutation test for cca under reduced model  
## Terms added sequentially (first to last)  
## Permutation: free  
## Number of permutations: 999  
##   
## Model: cca(formula = cSpec ~ N + P + Al, data = env)  
## Df ChiSquare F Pr(>F)   
## N 1 0.19261 1.4583 0.054 .   
## P 1 0.23123 1.7508 0.022 \*   
## Al 1 0.27661 2.0944 0.002 \*\*  
## Residual 20 2.64149   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

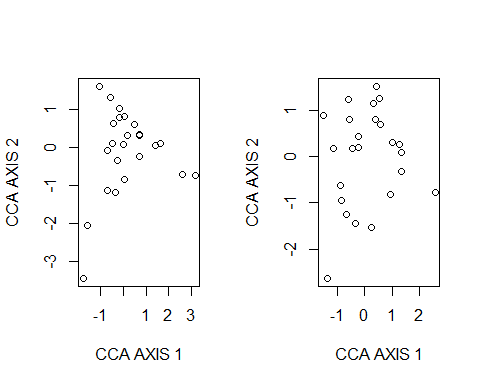
## Observed (F matrix) and Predicted (Z Matrix) Site Scores

Now look back at you cca summary again:

summary(sp.CCA)

The matrix labeled “Site scores (weighted averages of species scores)” is the F matrix and the matrix labeled “Site constraints (linear combinations of constraining variables)”is the Z matrix. Look at these two sets of site scores projected in ordination space:

par(mfrow = c(1, 2))  
plot(sp.CCA$CC$wa[, 1], sp.CCA$CC$wa[, 2], xlab = "CCA AXIS 1", ylab = "CCA AXIS 2")  
plot(sp.CCA$CC$u[, 1], sp.CCA$CC$u[, 2], xlab = "CCA AXIS 1", ylab = "CCA AXIS 2")



Look at the correlation between these two matrices. These correlations can lend insight as to how well the predicted site locations match the observed ones. However, they are not to be trusted as the only line of evidence.

spenvcor(sp.CCA)

## CCA1 CCA2 CCA3   
## 0.8870958 0.8806445 0.6690905

## Intra-set correlations and biplot scores for the constraining variables.

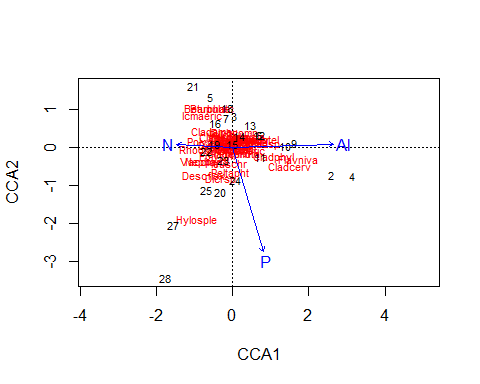
Correlations between the Z matrix (predicted site scores) and the environmental variables provide information on which variables have the largest influence on the constrained ordination. These also denote the placement of the environmental variables as vectors on the CCA tri-plot.

sp.CCA$CCA$biplot

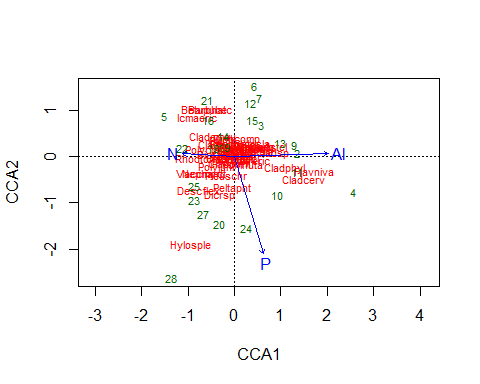
## CCA1 CCA2 CCA3  
## N -0.5069393 0.02839513 -0.861514  
## P 0.2819078 -0.94801285 0.147647  
## Al 0.9184416 0.02964719 -0.394444

##The Tri-Plot (using the site scores from the F matrix)

plot(sp.CCA, choices = c(1, 2), display = c("wa", "sp", "bp"), scaling = 2)

 and using the site scores from the Z matrix:

plot(sp.CCA, choices = c(1, 2), display = c("lc", "sp", "bp"), scaling = 2)



## 

## \*\*Now run the constrained ordination with the reduced data set. Does excluding the common species improve the effectiveness of the constrained ordination and or change your interpretation?