Lab 11: RDA and Variance partitioning

The goal of this lab is to apply Redundancy Analysis (RDA) 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. You will also explore variable selection and variance partitioning to further interpret the results of RDA. These methods can also be applied to CCA.

# Set up R session

## Download packages

We will be using the following packages:

library(vegan)

## Import Data

Today you will be using a Doubs fish data set (Verneaux 1973) from the Doubs River in France and Switzerland. This data set includes a *site by species* matrix (“DoubsSpe.csv”) that includes fish abundances for 27 species across 30 sites and a site by environment (“DoubsEnv.csv”) matrix containing measurements of 11 environmental variables across 30 sites.

spe <- read.csv("../Data/lab\_11/DoubsSpe.csv", row.names=1)  
env <- read.csv("../Data/lab\_11/DoubsEnv.csv", row.names=1)  
  
  
  
#########Modify Data  
#remove site 8  
spe <- spe[-8, ]  
env <- env[-8, ]  
  
env$pen

## [1] 48.0 3.0 3.7 3.2 2.3 3.2 6.6 1.2 9.9 4.1 1.6 2.1 1.2 0.5 2.0  
## [16] 0.5 0.8 0.5 0.8 1.0 1.4 1.2 0.3 0.5 0.5 1.2 0.3 0.6 0.2

# Remove the 'das' variable from the env dataset  
env <- env[, -1]

Explore the data set and modify it in the following way: 1. Notice that one row (i.e., site) has no species. Remove that row before proceeding. 2. Also, remove the variable “das” from the data set. 3. Recode the slope variable (pen) into a factor (qualitative) variable (to explore how these are handled in the ordinations). Here you divide the data into quantiles and assign a factor level to each quantile (very steep, steep, moderate, low):

pen2 <- rep("very\_steep", nrow(env))  
pen2[env$pen <= quantile(env$pen)[4]] <- "steep"  
pen2[env$pen <= quantile(env$pen)[3]] <- "moderate"  
pen2[env$pen <= quantile(env$pen)[2]] <- "low"  
pen2 <- factor(pen2, levels=c("low", "moderate", "steep", "very\_steep"))  
table(pen2)

## pen2  
## low moderate steep very\_steep   
## 8 8 6 7

1. Incorporate the new qualitative slope variable “pen2” into the new environmental data frame:

env2 <- env  
env2$pen <- pen2

# Data transformation

A common transformation for species data in RDA is the Hellinger transformation. This transformation is simply the square root of row normalized data (i.e., relative abundance). This transformation has been shown to have good statistical properties for constrained ordination. You will use the function deconstand in the *vegan* package:

spe.hel <- decostand(spe, "hellinger")

# Run RDA using function rda in vegan using all the environmental variables.

?rda

spe.rda <- rda(spe.hel ~ ., env2)

Explore the results:

summary(spe.rda)

Unadjusted R^2 retrieved from the rda object

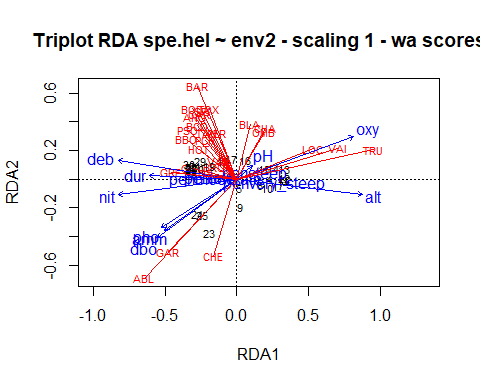
R2 <- RsquareAdj(spe.rda)$r.squared

Adjusted R^2 retrieved from the rda object

R2adj <- RsquareAdj(spe.rda)$adj.r.squared

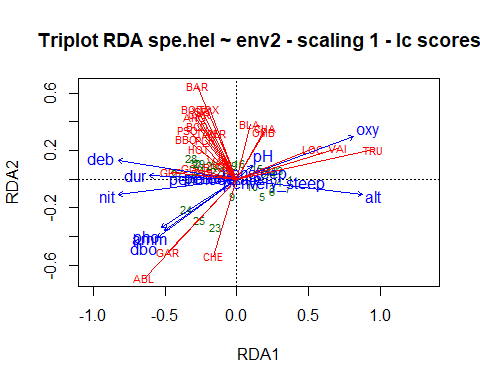
Plot using the F-scores:

plot(spe.rda, scaling=1,   
main="Triplot RDA spe.hel ~ env2 - scaling 1 - wa scores")  
spe.sc <- scores(spe.rda, choices=1:2, scaling=1, display="sp")  
arrows(0, 0, spe.sc[, 1], spe.sc[, 2], length=0, lty=1, col="red")



Plot using the Z-scores:

plot(spe.rda, scaling=1, display=c("sp", "lc", "cn"),   
main="Triplot RDA spe.hel ~ env2 - scaling 1 - lc scores")  
arrows(0, 0, spe.sc[, 1], spe.sc[, 2], length=0, lty=1, col="red")



**??Raise your hand and give me an interpretation of the tri-plot??**

Conduct a permutation test using anova function in vegan to test the significance of the model and individual axes:

anova(spe.rda, step=1000)

## Permutation test for rda under reduced model  
## Permutation: free  
## Number of permutations: 999  
##   
## Model: rda(formula = spe.hel ~ alt + pen + deb + pH + dur + pho + nit + amm + oxy + dbo, data = env2)  
## Df Variance F Pr(>F)   
## Model 12 0.36537 3.5522 0.001 \*\*\*  
## Residual 16 0.13714   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Tests of all canonical axes:

anova(spe.rda, by="axis", step=1000)

## Permutation test for rda under reduced model  
## Forward tests for axes  
## Permutation: free  
## Number of permutations: 999  
##   
## Model: rda(formula = spe.hel ~ alt + pen + deb + pH + dur + pho + nit + amm + oxy + dbo, data = env2)  
## Df Variance F Pr(>F)   
## RDA1 1 0.228081 26.6098 0.001 \*\*\*  
## RDA2 1 0.053696 6.2646 0.004 \*\*   
## RDA3 1 0.032124 3.7478 0.338   
## RDA4 1 0.023207 2.7075 0.774   
## RDA5 1 0.008707 1.0159 1.000   
## RDA6 1 0.007218 0.8421 1.000   
## RDA7 1 0.004862 0.5673 1.000   
## RDA8 1 0.002919 0.3406 1.000   
## RDA9 1 0.002141 0.2497 1.000   
## RDA10 1 0.001160 0.1353 1.000   
## RDA11 1 0.000913 0.1066 1.000   
## RDA12 1 0.000341 0.0397 1.000   
## Residual 16 0.137141   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Variable Selection: forward selection using ordiR2step in *vegan*.

We realize that we have used too many variables and that many variables are closely related in our data set. Here we want to reduce the number of variables while maintaining a model that describes as much variance as possible. Here selection is based on increasing and R2 and p-value of the permutation test. See settings in ordiR2step for more options.

??ordiR2step  
step.forward <- ordiR2step(rda(spe.hel ~ 1, data=env2),   
scope=formula(spe.rda), R2scope = F, direction="forward", pstep=1000)

**??What is the most parsimonious model??**

Conduct an RDA with this model.

# Partial RDA

Partial RDA allows us to quantify the unique contribution of an explanatory variable in describing variation in a response data set. It allows us to hold other explanatory variables constant while assessing the role of the focal variable. Here, I show a partial RDA for the altitude variable “alt”. You can conduct a partial RDA for any or all explanatory variables.

partial.alt <- rda(spe.hel ~ alt + Condition(oxy + dbo + dur) , data=env2)

Explore the results of this model and conduct a permutation test to assess the significance of this model. Feel free to explore partial RDA for other variables in the data set.

# Variance partitioning

Variance partitioning allows us quantify the amount of variation attributed to each variable and the amount that is shared between variables. In some ways, it is an extension of partial RDA. We can partition variance between up to four variables (or matrices) with the varpart function in *vegan*. You can use more variables but have to partition by hand.

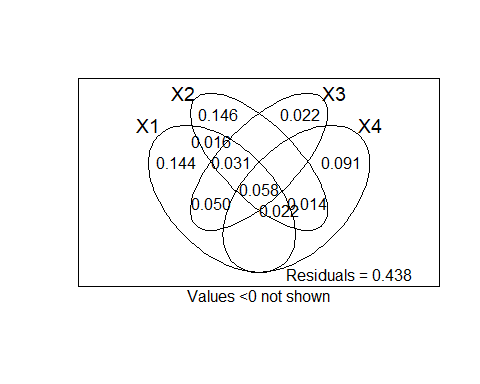
??varpart

Here we partition the variance for the model we constructed trough forward selection above:

spe.part <- varpart(spe.hel,~ alt, ~oxy,~ dur ,~dbo ,data=env2)  
spe.part

##   
## Partition of variance in RDA   
##   
## Call: varpart(Y = spe.hel, X = ~alt, ~oxy, ~dur, ~dbo, data = env2)  
##   
## Explanatory tables:  
## X1: ~alt  
## X2: ~oxy  
## X3: ~dur  
## X4: ~dbo   
##   
## No. of explanatory tables: 4   
## Total variation (SS): 14.07   
## Variance: 0.50251   
## No. of observations: 29   
##   
## Partition table:  
## Df R.square Adj.R.square Testable  
## [aeghklno] = X1 1 0.32807 0.30318 TRUE  
## [befiklmo] = X2 1 0.30248 0.27665 TRUE  
## [cfgjlmno] = X3 1 0.17349 0.14287 TRUE  
## [dhijkmno] = X4 1 0.19244 0.16253 TRUE  
## [abefghiklmno] = X1+X2 2 0.49209 0.45302 TRUE  
## [acefghjklmno] = X1+X3 2 0.36084 0.31167 TRUE  
## [adeghijklmno] = X1+X4 2 0.44604 0.40343 TRUE  
## [bcefgijklmno] = X2+X3 2 0.38730 0.34017 TRUE  
## [bdefhijklmno] = X2+X4 2 0.39263 0.34591 TRUE  
## [cdfghijklmno] = X3+X4 2 0.31007 0.25699 TRUE  
## [abcefghijklmno] = X1+X2+X3 3 0.52791 0.47125 TRUE  
## [abdefghijklmno] = X1+X2+X4 3 0.58942 0.54016 TRUE  
## [acdefghijklmno] = X1+X3+X4 3 0.47904 0.41653 TRUE  
## [bcdefghijklmno] = X2+X3+X4 3 0.48052 0.41818 TRUE  
## [abcdefghijklmno] = All 4 0.62469 0.56214 TRUE  
## Individual fractions   
## [a] = X1 | X2+X3+X4 1 0.14396 TRUE  
## [b] = X2 | X1+X3+X4 1 0.14561 TRUE  
## [c] = X3 | X1+X2+X4 1 0.02199 TRUE  
## [d] = X4 | X1+X2+X3 1 0.09089 TRUE  
## [e] 0 0.01557 FALSE  
## [f] 0 -0.00889 FALSE  
## [g] 0 0.05028 FALSE  
## [h] 0 -0.01287 FALSE  
## [i] 0 0.01397 FALSE  
## [j] 0 -0.00376 FALSE  
## [k] 0 0.02214 FALSE  
## [l] 0 0.03108 FALSE  
## [m] 0 -0.00084 FALSE  
## [n] 0 -0.00499 FALSE  
## [o] 0 0.05800 FALSE  
## [p] = Residuals 0 0.43786 FALSE  
## Controlling 2 tables X   
## [ae] = X1 | X3+X4 1 0.15954 TRUE  
## [ag] = X1 | X2+X4 1 0.19425 TRUE  
## [ah] = X1 | X2+X3 1 0.13109 TRUE  
## [be] = X2 | X3+X4 1 0.16119 TRUE  
## [bf] = X2 | X1+X4 1 0.13672 TRUE  
## [bi] = X2 | X1+X3 1 0.15958 TRUE  
## [cf] = X3 | X1+X4 1 0.01310 TRUE  
## [cg] = X3 | X2+X4 1 0.07227 TRUE  
## [cj] = X3 | X1+X2 1 0.01823 TRUE  
## [dh] = X4 | X2+X3 1 0.07802 TRUE  
## [di] = X4 | X1+X3 1 0.10486 TRUE  
## [dj] = X4 | X1+X2 1 0.08713 TRUE  
## Controlling 1 table X   
## [aghn] = X1 | X2 1 0.17638 TRUE  
## [aehk] = X1 | X3 1 0.16880 TRUE  
## [aegl] = X1 | X4 1 0.24090 TRUE  
## [bfim] = X2 | X1 1 0.14985 TRUE  
## [beik] = X2 | X3 1 0.19729 TRUE  
## [befl] = X2 | X4 1 0.18338 TRUE  
## [cfjm] = X3 | X1 1 0.00850 TRUE  
## [cgjn] = X3 | X2 1 0.06352 TRUE  
## [cfgl] = X3 | X4 1 0.09446 TRUE  
## [dijm] = X4 | X1 1 0.10025 TRUE  
## [dhjn] = X4 | X2 1 0.06926 TRUE  
## [dhik] = X4 | X3 1 0.11412 TRUE  
## ---  
## Use function 'rda' to test significance of fractions of interest

plot(spe.part, digits=2)

 **You can also test the significance of the variance fractions using a permutation test.**

# Handmade RDA adapted from Borcard et al. (2011)

Yc <- as.matrix(spe.hel) #your Hellinger transformed site by species matrix  
  
 Xcr <- as.matrix(env)# here we are using the slope variable “pen” as continuous.  
  
 # 2. Computation of the multivariate linear regression  
 # \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
  
 # Matrix of regression coefficients (eq. 11.4)  
 B <- solve(t(Xcr) %\*% Xcr) %\*% t(Xcr) %\*% Yc  
  
 # Matrix of fitted values   
 Yhat <- Xcr %\*% B  
  
 # Matrix of residuals  
 Yres <- Yc - Yhat  
  
 # Dimensions  
 n <- nrow(Yc)  
 p <- ncol(Yc)  
 m <- ncol(Xcr)  
  
  
 # 3. PCA on fitted values  
 # \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
 # Covariance matrix   
 S <- cov(Yhat)  
  
 # Eigenvalue decomposition-gives eigenvalues and eigenvectors (factor loadings)  
 eigenS <- eigen(S)  
  
 # How many canonical axes to keep/explore?  
 kc <- length(which(eigenS$values > 0.00000001))  
  
 # Eigenvalues of canonical axes  
 ev <- eigenS$values[1:kc]  
  
 # Total variance (inertia) of the response matrix  
 trace = sum(diag(cov(Yc)))  
   
 # Orthonormal eigenvectors (species score/FACTOR LOADINGS)  
 U <- eigenS$vectors[,1:kc]  
 row.names(U) <- colnames(Yc)  
  
 # Site scores (F-SCORES)  
 F <- Yc %\*% U  
 row.names(F) <- row.names(Yc)  
  
 # Site constraints (Z-SCORES)  
 Z <- Yhat %\*% U  
 row.names(Z) <- row.names(Yc)  
  
# Canonical coefficients (regression coefficients for each explanatory variable on #each rda axis)  
 CC <- B %\*% U  
 row.names(CC) <- colnames(Xcr)  
  
 # Explanatory variables  
 # Species-environment correlations  
 corXZ <- cor(Xcr,Z)  
  
 # Diagonal matrix of weights  
 D <- diag(sqrt(ev/trace))  
  
 # Biplot scores of explanatory variables  
 coordX <- corXZ %\*% D # Scaled  
   
  
 # Unadjusted R2  
 R2 <- sum(ev/trace)  
 # Adjusted R2  
 R2a <- 1-((n-1)/(n-m-1))\*(1-R2)  
  
  
  
result <- list(trace, R2, R2a, ev, CC, U, F, Z, coordX)  
  
names(result) <- c("Total\_variance", "R2", "R2adj", "Can\_ev", "Can\_coeff",   
 "Species\_sc1", "wa\_sc1", "lc\_sc1", "Biplot\_sc1")