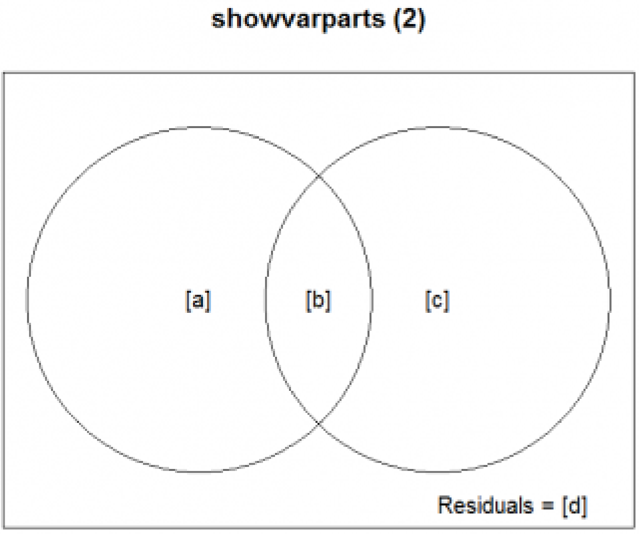
Constrained ordination

Variation partitioning

Note: variation partitioning is sometimes also called ***commonality analysis*** in reference to the *common* (shared) fraction of variation (Kerlinger & Pedhazur 1973). It is also a synonym to ***variance partitioning***[1].

In case we have two or more explanatory variables, one may be interested in variation in species composition explained by each of them. If some of these explanatory variables are correlated, one must expect that variation explained by the first or the other variable cannot be separated - it will be shared.



Venn's diagram, showing variation explained by two environmental variables (or two sets of environmental variables) and coded names of these fractions. Created by function `showvarpart (2)` from library `vegan`.

The way how to approach this problem is variation partitioning, when variation explained by each variable (or set of variables) independently is partitioned into variation attributable purely to given environmental variable, and shared variation attributable to two or more variables.

Variation can be partitioned by individual variables (e.g. variation explained by soil pH vs variation explained by soil Ca) or by groups of variables (e.g. soil variables vs climatic variables).

Results can be visualized using Venn's diagram (see figure on the right). Meaning of the fractions in Venn's diagram is the following[2]:

- [a] - variation explained by variable 1 (conditional (or partial) effect of variable 1, i.e. variation this variable would explain if putting variable 2 as covariable);
- [c] - variation explained by variable 2;
- [b] - shared variation explained by both variables (cannot be decided to which of them should

be attributed, and is a result of correlation between both variables);
- [a+b] - variation explained by variable 1 (independent simple (or marginal) effect of variable 1, i.e. variation this variable would explain if it is as the only explanatory variable in the model);
- [b+c] - variation explained by variable 2;
- [d] - unexplained variation.

If the variation is partitioned among groups with the same number of variables (e.g. two soil and two climatic variables), than variation explained by each group is comparable without adjustement. However, if groups contain different numbers of variables, variation explained by not adjusted R^2 is not comparable, since R2 tends to increase with the number of explanatory variables. Here, use of adjusted R2 is recommended.

The library vegan offers function `varpart`, which can partition variation among up to four variables (or groups of variables). Note that `varpart` is based on redundancy analysis (`rda`) and uses adjusted R^2 to express explained variation. The reason for using only `rda` is that in R, there is still no function available to calculate adjusted R^2 for unimodal ordination methods (like `cca`).

R functions

- **`varpart`** (library vegan) - variation partitioning (using linear constrained ordinatino - `rda`) among up to four matrices of environmental variables. First argument (Y) is dependent (usually species composition) matrix (but could be also only one variable - in that case `varpart` is conductin linear regression). Next arguments (up to four) are (groups of) explanatory variables. Uses either formula interface (with ~) or matrices.
- **`plot`** (library vegan) - draws Venn's diagram with fractions of explained variation. In default setting it doesn't show negative values of explained variation (argument `cutoff = 0`). Consult `?plot.varpart` for more details.
- **`showvarparts`** (library vegan) - draws schema of Venn's diagram with codes of individual fractions.

Examples of use

Use of varpart function (using Barley field weed community dataset)

Use data from Barley field weed community dataset, studying how does weed communities in barley fields changes after application of fertilizer. The problem is that fertilizer may influence the weed community directly, but also indirectly via increased cover of barley (indeed, it grows better if fertilized). The aim of this example is to separate variation in species composition of weed community caused by dose of fertilizer and by estimated cover of barley (both variables are correlated).

First, import the data:

```
fertil.spe <- read.delim
('http://www.davidzeleny.net/anadat-r/data-download/fertil.spe.txt',
row.names = 1)
fertil.env <- read.delim
('http://www.davidzeleny.net/anadat-r/data-download/fertil.env.txt',
row.names = 1)
```

Species data are represented by estimated cover of weed community species in a plot and it is not necessary to transform them. One may first apply DCA ordination on species data to see the length of the first DCA axis and hence seek the recommendation whether linear or unimodal method should be used (the length is 3.8 SD, lying in grey zone). We will use RDA here, since `varpart` method is based only on RDA (if the data are even more heterogeneous, Hellinger transformation would be advisable).

First, let's see how would the variation partitioning looks like if we do it step-by-step manually (without using `varpart` function). We will partition variation explained by variables `dose` and `cover` from `fertil.env`; for this, we need to define the global model (to see how much variation is explained by both of them), and also partial models with one variable as explanatory and the other as covariable:

```
# fractions [a+b+c]:
rda.all <- rda (fertil.spe ~ dose + cover, data = fertil.env)
# fraction [a]:
rda.dose.cover <- rda (fertil.spe ~ dose + Condition (cover), data =
fertil.env)
# fraction [c]:
rda.cover.dose <- rda (fertil.spe ~ cover + Condition (dose), data =
fertil.env)
```

For completeness, let's define also models with simple (marginal) effect of each explanatory variable:

```
# fractions [a+b]:
rda.dose <- rda (fertil.spe ~ dose, data = fertil.env)
# fractions [b+c]:
rda.cover <- rda (fertil.spe ~ cover, data = fertil.env)
```

Now, let's use the variation explained by individual fractions (using `RsquareAdj` function):

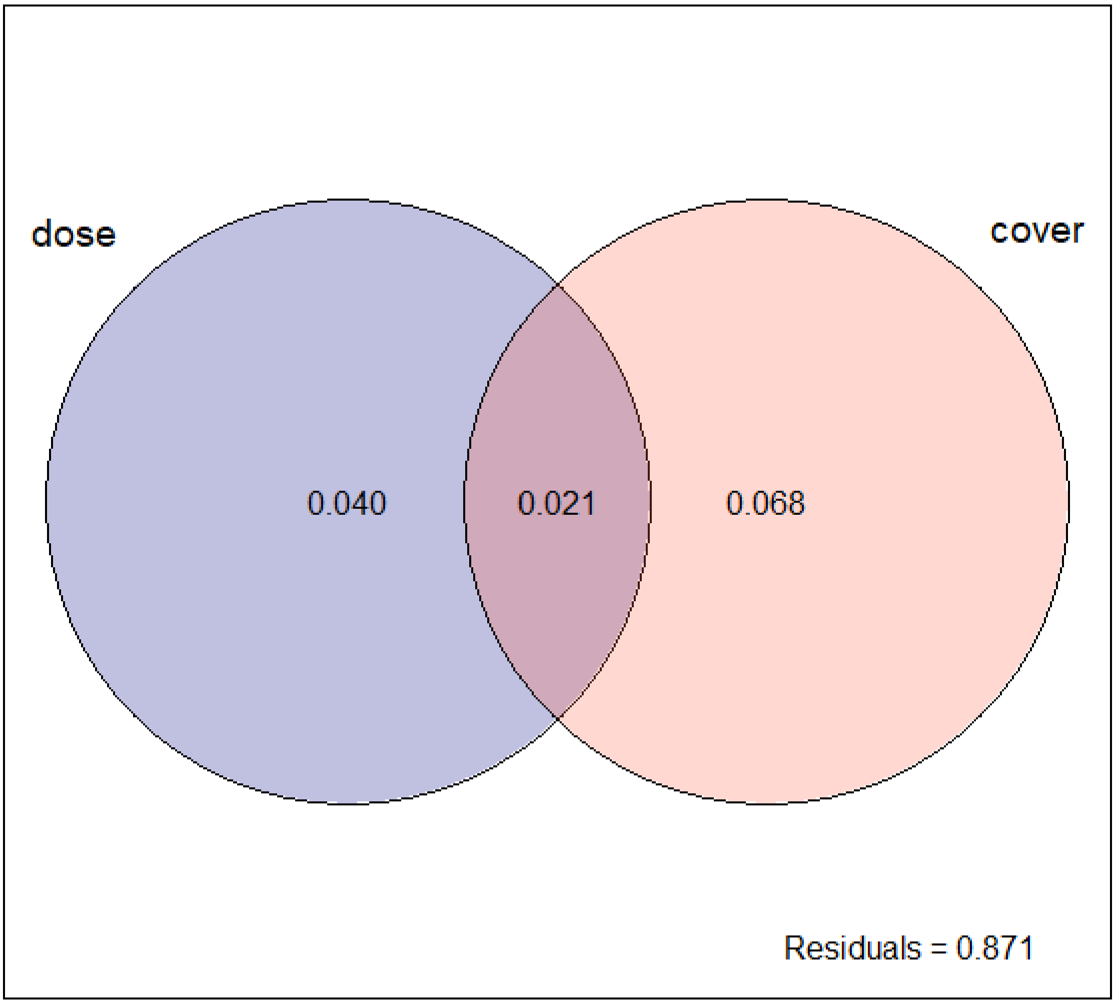
```
## fraction [a+b+c]
RsquareAdj (rda.all)
# $r.squared
# [1] 0.1430381
#
# $adj.r.squared
# [1] 0.1286354

## fraction [a]:
RsquareAdj (rda.dose.cover)
# $r.squared
# [1] 0.04675401
#
# $adj.r.squared
# [1] 0.03988225

## fraction [c]:
RsquareAdj (rda.cover.dose)
# $r.squared
# [1] 0.07414449
#
# $adj.r.squared
# [1] 0.06750099
</file>
For interpretation, we will use the adjusted R<sup>2</sup> stored in
''adj.r.squared'' element of the list returned by ''RsquareAdj'' function:
  * [a+b+c] = 12.9%
  * [a] = 4.0%
  * [c] = 6.8%
  * [b] = [a+b+c]-[a]-[b] = 12.9-4.0-6.8 = 2.1%
  * [d] = 100-[a+b+c] = 100-12.9 = 87.1%

Conditional effect of dose is slightly lower than conditional effect of
cover (4.0% vs 6.9%), and shared variation is not that high (2.1%). Seems
that species composition weeds is more affected by light conditions modified
by increasing cover of barley, than by fertilizer itself.

Now, let's see the same, using function ''varpart''
<code rsplus>
varp <- varpart (fertil.spe, ~ dose, ~ cover, data = fertil.env)
varp
```

```
Partition of variation in RDA

Call: varpart(Y = fertil.spe, X = ~dose, ~cover, data = fertil.env)

Explanatory tables:
X1:  ~dose
X2:  ~cover

No. of explanatory tables: 2
Total variation (SS): 2581.2
            Variance: 21.332
No. of observations: 122

Partition table:
                     Df R.squared Adj.R.squared Testable
[a+b] = X1            1   0.06889       0.06113     TRUE
[b+c] = X2            1   0.09628       0.08875     TRUE
[a+b+c] = X1+X2       2   0.14304       0.12864     TRUE
Individual fractions
[a] = X1|X2           1                 0.03988     TRUE
[b]                   0                 0.02125    FALSE
[c] = X2|X1           1                 0.06750     TRUE
[d] = Residuals                         0.87136    FALSE
---
Use function 'rda' to test significance of fractions of interest
```

The results for individual fractions are identical with our results above. `varpart` reports also simple (marginal) effect (both not-adjusted and adjusted variation) of individual predictors, i.e. fractions [a+b] and [b+c] - manually we would get the same numbers by applying `RsquareAdj` function on `rda.dose` and `rda.cover`.

We may plot the results into Venn's diagram (argument `digits` influences number of decimal digits shown in the diagram):

```
plot (varp, digits = 2)
```



Now, when we know both simple and conditional effect of each variables, we may want to know whether it is significant, and hence worth of interpreting. Results from `varpart` contain the column `testable` with logical values indicating whether given fraction is testable or not. To test each of them, we will need the models defined above, and the function `anova`, which (if applied on single object resulting from `rda` or `cca` method, returns Monte Carlo permutation test of the predictor effect):

```
## fraction [a+b]:
anova (rda.dose)
# Permutation test for rda under reduced model
# Permutation: free
# Number of permutations: 999
#
# Model: rda(formula = fertil.spe ~ dose, data = fertil.env)
# Df Variance      F Pr(>F)
# Model      1   1.4696 8.8789  0.001 ***
#   Residual 120  19.8624
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## fraction [b+c]
anova (rda.cover)
# Permutation test for rda under reduced model
# Permutation: free
# Number of permutations: 999
#
# Model: rda(formula = fertil.spe ~ cover, data = fertil.env)
# Df Variance      F Pr(>F)
# Model      1   2.0539 12.785  0.001 ***
#   Residual 120  19.2781
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## the global model [a+b+c]
anova (rda.all)
# Permutation test for rda under reduced model
# Permutation: free
# Number of permutations: 999
#
# Model: rda(formula = fertil.spe ~ dose + cover, data = fertil.env)
# Df Variance      F Pr(>F)
# Model      2   3.0513 9.9313  0.001 ***
#   Residual 119  18.2808
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## fraction [a]
anova (rda.dose.cover)
# Permutation test for rda under reduced model
# Permutation: free
# Number of permutations: 999
#
# Model: rda(formula = fertil.spe ~ dose + Condition(cover), data =
fertil.env)
# Df Variance      F Pr(>F)
# Model      1   0.9974 6.4924  0.001 ***
#   Residual 119  18.2808
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## fraction [b]
anova (rda.cover.dose)
# Permutation test for rda under reduced model
# Permutation: free
# Number of permutations: 999
#
# Model: rda(formula = fertil.spe ~ cover + Condition(dose), data =
fertil.env)
# Df Variance      F Pr(>F)
# Model      1   1.5817 10.296  0.001 ***
#   Residual 119  18.2808
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From these results you may see that all simple (marginal) and conditional (partial) effects of both predictors are highly significant (P < 0.001 in all cases).

[1] As to the distinction between *variance* and *variation*, Legendre & Legendre (2012) note: “The term variation, a less technical and looser term than variance, is used because one is partitioning the total sum of squared deviations of y from its mean (total SS). In variation partitioning, there is no need to divide the total SS of y by its degrees of freedom to obtain the variance s^2_y.” The first edition of the book *Multivariate Analysis of Ecological Data using CANOCO* (Lepš & Šmilauer 2003) was using the term *variance partitioning*, while in the second edition (Šmilauer & Lepš 2014), authors adopted the term *variation partitioning*, noting: “It was called **variance** partitioning in the original paper, but we prefer, together with Legendre & Legendre 2012, the more appropriate name referring to *variation*, as we include also unimodal ordination methods in our considerations.”

[2] Note that in CANOCO 5, the coding of the fraction follows different logic - [c] is shared variation, and [a] or [b] are partial fractions; meaning of [d] remains the same (unexplained variation).