Example 2 - RDA

Maxime Hervé, Florence Nicolè and Kim-Anh Lê Cao

16/10/2017

Contents

Packages to load

```
library(Hotelling)      # Function needed: clr
library(vegan)          # Function needed: rda
library(RVAideMemoire) # Functions needed: MVA.synt, MVA.anova, pairwise.factorfit,
                        #                   MVA.plot
```

Data loading

The dataset is composed of 40 rows (40 samples, 5 per species/organ/time) and 73 columns (the plant species, the organ, the time of the day and the relative proportion of 70 volatile compounds).

```
tab <- read.table("Example 2.txt",header=TRUE)
```

As the data are compositional, the sum of all compounds for a given sample is always equal to 100 % (or almost because of rounding errors). We check this is the case:

```
rowSums(tab[,4:73])
```

```
 sample1  sample2  sample3  sample4  sample5  sample6  sample7  sample8
  99.999  100.001  100.001  100.000  100.001   99.999  100.001   99.999
 sample9 sample10 sample11 sample12 sample13 sample14 sample15 sample16
  99.999  100.000  100.001  100.001  100.000  100.000  100.000   99.999
sample17 sample18 sample19 sample20 sample21 sample22 sample23 sample24
 100.000   99.999  100.000  100.000  100.000  100.000  100.001  100.000
sample25 sample26 sample27 sample28 sample29 sample30 sample31 sample32
  99.999  100.002   99.998  100.000  100.002   99.997  100.000   99.999
sample33 sample34 sample35 sample36 sample37 sample38 sample39 sample40
 100.000   99.999  100.000   99.998   99.999  100.000   99.999  100.000
```

Pre-treatment

We transform compositional data using the Centered LogRatio method. Since zeroes are present, we add a small constant value to the whole data, that is much lower than the minimal value of the whole data (one order of magnitude smaller).

The minimal non-zero value is:

```
min(tab[,4:73][tab[,4:73] != 0])
```

```
[1] 0.002
```

Thus we decide to add an offset of 0.0001 to all values:

```
Chemistry <- clr(tab[,4:73] + 0.0001)
```

The data are then autoscaled:

```
Chemistry.scaled <- scale(Chemistry)
```

Analysis

We fit the RDA. Since all interactions between species, organ and time are relevant, they are all included in the model:

```
RDA <- rda(Chemistry.scaled~Species*Organ*Time,data=tab)
```

How much total variance does the experimental design explain?

```
MVA.synt(RDA)
```

```
Criterion: total variance (%)
               Proportion Cumulative
   Constrained      38.64      38.64
 Unconstrained      61.36     100.00

Criterion: constrained variance (%)
 Axis Proportion Cumulative
    1      44.65      44.65
    2      29.28      73.94
    3       9.20      83.13
    4       8.59      91.72
    5       4.13      95.85

Criterion: unconstrained variance (%)
 Axis Proportion Cumulative
    1      13.89      13.89
    2      10.01      23.89
    3       9.44      33.33
    4       7.20      40.53
    5       6.82      47.35
```

The experimental design (the three factors and their interactions), taken together, explain 39 % of the total variance in the chemical data.

We test for the significance of this explained variance, *i.e.* that it is higher than under the null hypothesis of no effect of the experimental design:

```
anova(RDA)
```

```
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 999

Model: rda(formula = Chemistry.scaled ~ Species * Organ * Time, data = tab)
         Df Variance      F Pr(>F)
Model     7   27.051 2.8793  0.001 ***
Residual 32   42.949
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significant global effect of the experimental design, then we test for the individual effects of the factors (and interactions):

```
MVA.anova(RDA)
```

```
Permutation test for rda under reduced model
Type II tests
Permutation: free
Number of permutations: 999

Model: rda(formula = Chemistry.scaled ~ Species * Organ * Time, data = tab)
                  Df Variance      F Pr(>F)
Species            1    5.678 4.2303  0.001 ***
Organ              1    4.826 3.5960  0.001 ***
Time               1    6.341 4.7246  0.001 ***
Species:Organ      1    4.778 3.5600  0.001 ***
Species:Time       1    1.605 1.1959  0.237
Organ:Time         1    2.472 1.8417  0.020 *
Species:Organ:Time 1    1.351 1.0070  0.417
Residual          32   42.949
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two interactions are significant: 1/ species x organ and 2/ organ x time. The effects of species and time are independent. See here for an explanation of what interactions mean.

We perform pairwise comparisons using r^2-based permutation tests to study the first significant interaction:

```
pairwise.factorfit(RDA,tab$Species:tab$Organ)
```

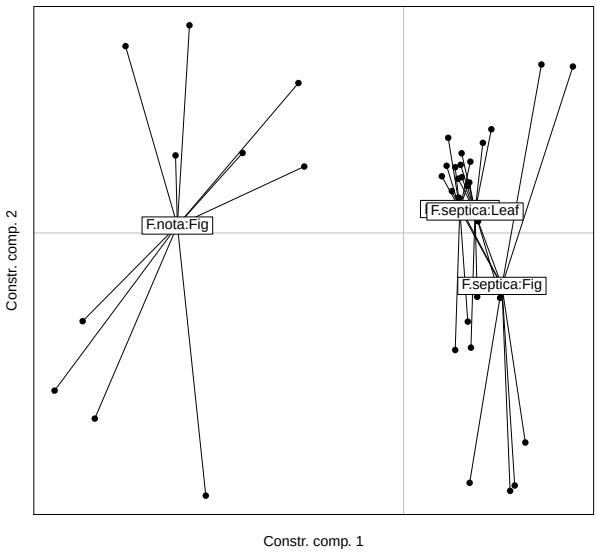
```
    Pairwise comparisons using factor fitting to an ordination

data:  RDA by tab$Species:tab$Organ
999 permutations

               F.nota:Fig F.nota:Leaf F.septica:Fig
F.nota:Leaf    0.003      -           -
F.septica:Fig  0.004      0.264       -
F.septica:Leaf 0.003      0.750       0.264

P value adjustment method: fdr
```

Figs of *Ficus nota* differ from all other combinations (leaves of *F. nota*, leaves and figs of *F. septica*), which

are not different from each other.

We draw a score plot to illustrate this interaction:

```
MVA.plot(RDA,fac=tab$Species:tab$Organ,drawextaxes=FALSE)
```



We use the same procedure to study the second significant interaction:

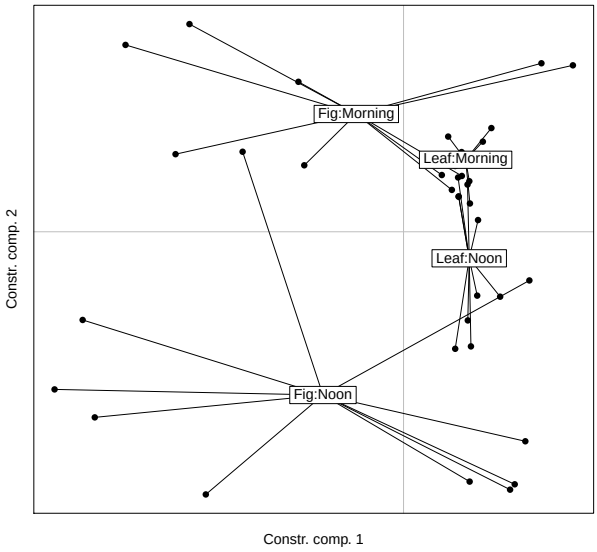
```
pairwise.factorfit(RDA,tab$Organ:tab$Time)
```

```
    Pairwise comparisons using factor fitting to an ordination

data:  RDA by tab$Organ:tab$Time
999 permutations

             Fig:Morning Fig:Noon Leaf:Morning
Fig:Noon     0.0015      -        -
Leaf:Morning 0.0280      0.0015   -
Leaf:Noon    0.0024      0.0015   0.0015

P value adjustment method: fdr
```

All combinations are significantly different from each other.

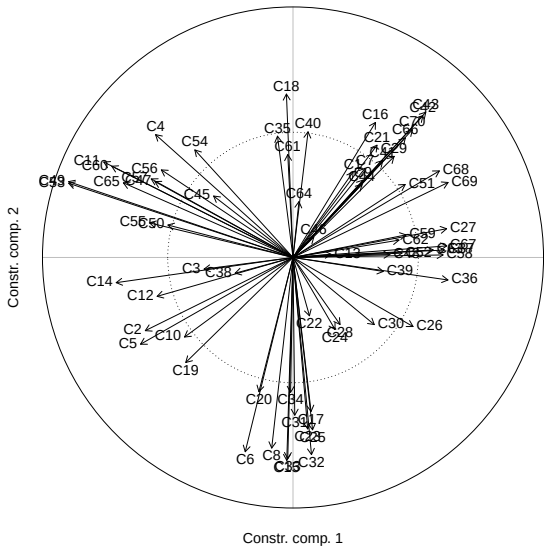
We draw a score plot to illustrate this interaction:

```
MVA.plot(RDA,fac=tab$Organ:tab$Time,drawextaxes=FALSE)
```



We draw the correlation circle plot to visualize the discriminant compounds, and how they contribute to each component:

```
MVA.plot(RDA,"corr")
```



Information on the current R session

```
sessionInfo()
```

```
R version 3.4.0 (2017-04-21)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1
```

```
Matrix products: default

locale:
[1] LC_COLLATE=French_France.1252  LC_CTYPE=French_France.1252
[3] LC_MONETARY=French_France.1252 LC_NUMERIC=C
[5] LC_TIME=French_France.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] RVAideMemoire_0.9-68 vegan_2.4-4          lattice_0.20-35
[4] permute_0.9-4        Hotelling_1.0-4      corpcor_1.6.9
[7] knitr_1.17

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.13       cluster_2.0.6      magrittr_1.5
 [4] splines_3.4.0      MASS_7.3-47        minqa_1.2.4
 [7] car_2.1-5          stringr_1.2.0      tools_3.4.0
[10] pbkrtest_0.4-7     nnet_7.3-12        parallel_3.4.0
[13] grid_3.4.0         nlme_3.1-131       mgcv_1.8-22
[16] quantreg_5.33      MatrixModels_0.4-1 htmltools_0.3.6
[19] lme4_1.1-14        yaml_2.1.14        ade4_1.7-8
[22] rprojroot_1.2      digest_0.6.12      Matrix_1.2-11
[25] nloptr_1.0.4       evaluate_0.10.1    rmarkdown_1.6
[28] stringi_1.1.5      compiler_3.4.0     backports_1.1.1
[31] SparseM_1.77
```