Nejdříve načtěme data a upravíme je (druhová data zlogaritmujeme a z dat o proměnných prostředí odstraníme první dvě proměnné):

```
library (vegan)
data (BCI)

BCI.soil <- read.delim
('http://www.davidzeleny.net/anadat-r/lib/exe/fetch.php?media=data:bci.soil.txt')

log.BCI <- log1p (BCI)

BCI.soil2 <- BCI.soil[,-1:-2]
```

Vypočteme globální RDA na zlogaritmovaných datech se všemi vysvětlujícími proměnnými a zjistíme, jestli je tento model signifikantní:

```
rda.all <- rda (log.BCI ~ ., BCI.soil2)
anova (rda.all)
```

```
Permutation test for rda under reduced model

Model: rda(formula = log.BCI ~ Al + B + Ca + Cu + Fe + K + Mg + Mn + P + Zn + N +
N.min. + pH, data = BCI.soil2)
         Df    Var      F N.Perm Pr(>F)
Model    13 24.838 2.3429    199  0.005 **
Residual 36 29.358
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model se všemi proměnnými je průkazný, můžeme tedy přistoupit k postupnému výběru proměnných. Nejdříve zjistíme adjustovaný R^2 celého modelu a uložíme ho do proměnné:

```
adj.r <- RsquareAdj (rda.all)$adj.r.squared
adj.r
```

```
[1] 0.2626876
```

A nyní vlastní postupný výběr s adjustovaným R^2 jako kritériem pro zastavení výběru:

```
fw <- forward.sel (log.BCI, BCI.soil2, adjR2thresh = adj.r)
```

```
Testing variable 1
Testing variable 2
Testing variable 3
Testing variable 4
Testing variable 5
Testing variable 6
Testing variable 7
Procedure stopped (alpha criteria): pvalue for variable 7 is 0.081000 (superior to
0.050000)
```

Sedmá proměnná už nebyla do modelu zahrnuta, protože nevyšla signifikantně (v tomto případě se tedy výběr zastavil kvůli signifikanci a ne kvůli tomu, že by R^2_{adj} modelu s vybranými proměnnými přesáhla R^2_{adj} globálního modelu.

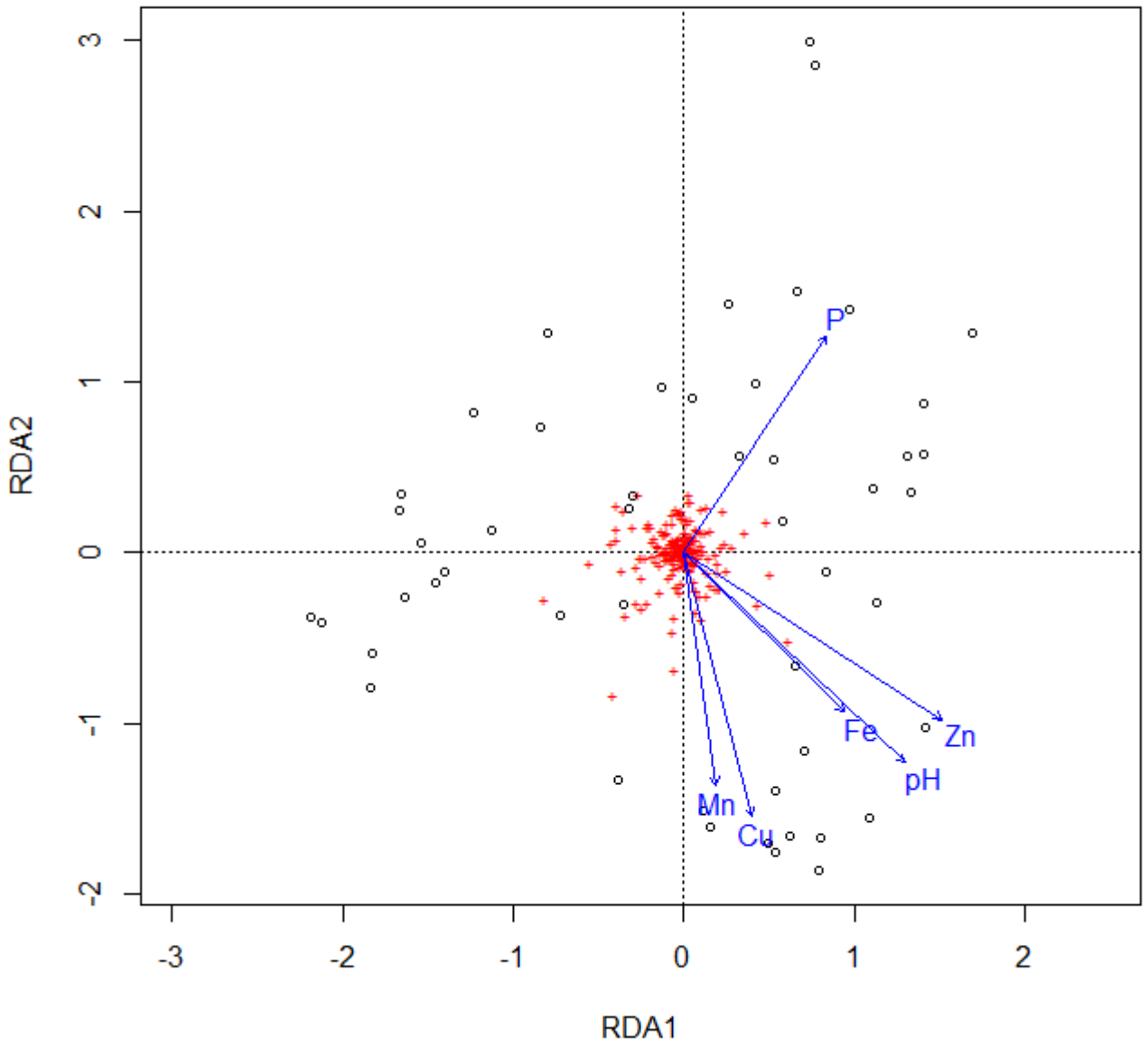
```
fw
```

```
  variables order         R2      R2Cum   AdjR2Cum        F  pval
1        pH    13 0.09279191 0.09279191 0.07389174 4.909581 0.001
2         P     9 0.07492951 0.16772141 0.13230530 4.231380 0.001
3        Zn    10 0.05916907 0.22689049 0.17647030 3.520559 0.001
4        Cu     4 0.04827906 0.27516955 0.21074017 2.997332 0.001
5        Fe     5 0.03314920 0.30831874 0.22971860 2.108724 0.001
6        Mn     8 0.02597042 0.33428916 0.24139927 1.677497 0.003
```

Je vidět, že jak pH, tak P (fosfor) každá vysvětlí poměrně slušné procento varibility. Ostatní proměnné přidají další procenta. Vidíme, že model s 6 proměnnými vysvětlil 24% variability, zatímco globální model se všemi 13 proměnnými vysvětlil jen o dvě procenta více (26%, hodnota uložená v proměnné `adj.r`).

Na závěr nakreslíme ordinační diagram RDA s vybranými chemickými proměnnými:

```
rda.fw <- rda (log.BCI, BCI.soil2[, fw$var])
ordiplot (rda.fw)
```



Z ordinačního diagramu je patrné, že hlavní dvě proměnné, pH a P, spolu prakticky nejsou korelované, zatímco všechny ostatní vybrané proměnné jsou do více či méně korelované s pH.

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/cs:forward_sel:tropical_forest_exercise_solution

Last update: **2017/10/11 20:36**