

Česká verze stránek není od roku 2013 aktualizována. Aktuální verzi této stránky najdete v její anglické verzi zde.

Přímá ordinace

Přímá ordinace (*constrained ordination*, *direct gradient analysis*) se používá k testování hypotéz o vlivu vysvětlujících proměnných na druhové složení společenstev. Jde o to popsat variabilitu v druhových datech, která se vztahuje ke známým proměnným prostředí (které například změřím). Tím se liší přímá ordinace od ordinace nepřímé, která má za úkol najít hlavní směry variability v druhových datech, tedy skryté gradienty, které se můžeme následně pokusit interpretovat na základě jejich korelací s měřenými faktory prostředí, případně na základě vlastních zkušeností. V případě přímých ordinací se budeme zabývat pouze metodami založenými na výpočtu eigenvalues (eigenvalue based)[1)]

RDA (Redundancy Analysis)

Metoda přímé ordinace založené na PCA (Principal Component Analysis). Pro RDA použijeme data z třebíčských trávníků, které mají krátký ekologických gradient (trávníky jsou to veskrze suché acidofilní, občas mírně bazičtější, ale žádné dramatické rozdíly). Vlastní druhová data rovnou odmocníme, protože jde o procentické abundance:

```
grasslands.spe <- read.csv2
('http://www.davidzeleny.net/anadat-r/data-download/grasslands-spe.csv', row.names
= 1)
grasslands.env <- read.csv2
('http://www.davidzeleny.net/anadat-r/data-download/grasslands-env.csv', dec = '.')

grasslands.spe <- sqrt (grasslands.spe)
```

rda (library vegan)

Zajímalo by mě například, jaký vliv na vegetaci má pH půdy. Využiji přitom výhody zápisu ve formě tzv. formule: `Y ~ x1 + x2 + x3, data = X`, kde X je datový rámec s vysvětlujícími proměnnými, x1 ad. jsou názvy jednotlivých proměnných a Y je datový rámec s vysvětlovannými proměnnými (druhovým složením).

```
library (vegan)
rda (grasslands.spe ~ pH.H, data = grasslands.env)
```

```
Call: rda(formula = grasslands.spe ~ pH.H, data = grasslands.env)

               Inertia Proportion Rank
Total         67.48476    1.00000
Constrained    3.44274    0.05102    1
Unconstrained 64.04202    0.94898   46
Inertia is variance

Eigenvalues for constrained axes:
 RDA1
```

```
3.443

Eigenvalues for unconstrained axes:
  PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
8.862 6.587 5.764 4.346 2.929 2.758 2.528 2.409
(Showed only 8 of all 46 unconstrained eigenvalues)
```

Ve výstupu mě zajímá především rozklad celkové variance (`Total inertia`) na část vysvětlenou použitými faktory prostředí (`Constrained`, součet eigenvalues os s omezením - v tomto případě pouze jedné osy) a nevysvětlenou variabilitu (`Unconstrained`, součet eigenvalues zbývajících 46 os bez omezení). Součet eigenvaues pro omezené osy (*constrained axes*) je v našem případě 3.44, celková variabilita (*Total inertia*) je 67.48. Vysvětlená variabilita je tedy 3.44/67.48 = 0.05097, tedy 5.1%. To nakonec vyčtu i z hodnot ve sloupečku `Proportion`, které udávají rozklad variability v proporcích (celková varibilita je rovna jedné, omezené osy vysvětlí 0.051).

Jak vypadá výstup RDA, pokud zahrnu dvě vysvětlující proměnné? Vedle pH použiji ještě proměnnou `soildepth`, tedy hloubku půdy:

```
rda (grasslands.spe ~ pH.H + soildepth, data = grasslands.env)
```

```
Call: rda(formula = grasslands.spe ~ pH.H + soildepth, data =
grasslands.env)

              Inertia Proportion Rank
Total         67.4848     1.0000
Constrained    5.2236     0.0774    2
Unconstrained 62.2612     0.9226   45
Inertia is variance

Eigenvalues for constrained axes:
 RDA1  RDA2
3.721 1.503

Eigenvalues for unconstrained axes:
  PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
8.813 6.095 5.729 4.221 2.920 2.729 2.523 2.383
(Showed only 8 of all 45 unconstrained eigenvalues)
```

Total inertia je stejné jako v předchozím případě, protože používám stejná data. Zvýšila se variabilita vysvětlená proměnnými prostředí, ale nijak dramaticky (7.7%). Omezené osy jsou nyní dvě - platí, že počet omezených os je roven počtu vysvětlujích proměnných (může jich být ale méně, v případě že některé vysvětlující proměnné jsou spolu silně korelované).

Z výstupní tabulky funkce `rda` vyčtu ještě jednu zajímavou věc - první neomezená osa (PC1) má hodnotu eigenvalue podstatně vyšší (8.813) než je součet eigenvalues obou omezených os. To znamená, že mimo gradienty prostředí, které jsem použil jako vysvětlující proměnné (pH a hloubka půdy) je datový soubor strukturovaný podle jiného gradientu, a to poměrně výrazně.

CCA (Canonical Correspondence Analysis)

Použití je stejné jako v případě RDA. CCA se používá na heterogenní data (podle kuchařky Lepš-Šmilauer jde o data, která mají první DCA osu delší než 4 jednotky).

cca (library vegan)

Výstup příkazu `cca` vypadá stejně jako v případe `rda`. Ukážeme si to na příkladu odmocněných dat z údolí Vltavy, kde jako vysvětlující proměnné použijeme pH ve vodě (pH.H) a orientaci (ASPSSW):

```
vltava.spe <- read.csv2
('http://www.davidzeleny.net/anadat-r/data-download/vltava-spe.csv', row.names = 1)
vltava.env <- read.csv2
('http://www.davidzeleny.net/anadat-r/data-download/vltava-env.csv', dec = '.')
vltava.spe <- sqrt (vltava.spe)
```

```
cca.1 <- cca (vltava.spe ~ pH.H + ASPSSW, data = vltava.env)
cca.1
```

Tentokrát jsem výsledky funkce `cca` přiřadil do nové proměnné, se kterou můžu dále pracovat.

```
Call: cca(formula = vltava.spe ~ pH.H + ASPSSW, data = vltava.env)

              Inertia Proportion Rank
Total         8.65579    1.00000
Constrained   0.55075    0.06363    2
Unconstrained 8.10505    0.93637   94
Inertia is mean squared contingency coefficient

Eigenvalues for constrained axes:
  CCA1   CCA2
0.3044 0.2464

Eigenvalues for unconstrained axes:
   CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8
0.4964 0.3979 0.3099 0.3028 0.2707 0.2512 0.2265 0.2171
(Showed only 8 of all 94 unconstrained eigenvalues)
```

Vysvětlující proměnné vysvětlí 6.4% variability, poměrně dost variability zůstává i na první ordinační ose bez omezení.

add1 (library vegan)

step (library vegan)

Cvičení

Použijte data z Bryce Canyonu a použijte je v CCA analýze. Data jsou na stupnici 0-5, není je tedy třeba transformovat.

1. použijte proměnné `elev`, `slope` a `av` (viz tabulka s vysvětlivkami k datům)
2. kolik celkové variability v druhových datech vysvětlí dohromady tyto tři proměnné?
3. jaké procento z této vysvětlené variability je zachyceno na první, druhé resp. třetí ose?
4. kolik z celkové variability zachycuje první osa bez omezení (*unconstrained*)?

Řešení

[1)]

Přímé ordinace založené na výpočtu distancí (*distance based*) také existují, jen nejsou příliš používané. Viz například dbRDA (*distance based* RDA), která funguje shruba následovně: z původní matice druhových dat (samples x species) se vypočte matice distancí mezi vzorky (samples x samples), která se podrobí analýze hlavních koordinát (PCOA) a skóre jednotlivých vzorků na osách PCoA se použijí coby vysvětlovaná proměnná v klasické RDA spolu s jednou nebo více vysvětlujícími proměnnými. Výhoda je v tom, že pro výpočet distanční matice můžu použít libovolnou míru distance (*distance measure*), nejsem tedy vázán pouze na Euklidovskou vzdálenost se kterou pracuje PCA (potažmo RDA).

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

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

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