

Cvičení 6

1. Jednoduchá přímá analýza: Na pakomářích datech z řeky Svratky zjistěte, jak vysoký podíl variability v druhovém složení společenstva larev pakomárů vysvětluje typ biotopu (proměnná *hab*) a otestujte, zda tento podíl je vyšší než náhodný. Použijte lineární model a data standardizujte a transformujte pro ošetření problému dvojitých nul a snížení váhy dominantních druhů. Podíl vysvětlené variability upravte o variabilitu vysvětlitelnou náhodnou proměnnou. Nakreslete příslušný ordinacní diagram.
2. Ještě jednou použijeme již načtená pakomáří data. Použijte opět lineární model se stejnou transformací jako v předchozím bodě, ale nyní se zaměřte na nalezení takové kombinace proměnných prostředí, která vysvětlí co nejvíce variability v pakomářích datech. Jako vysvětlující proměnné použijte pouze numerické proměnné, tedy *bez 'hab' a 'hydr'*. Otázka, na kterou chceme odpovědět tedy je: Jakými proměnnými prostředí lze nejlépe vysvětlit variabilitu pakomářích společenstev? Nebo podél jakých proměnných se jejich společenstva nejvíce mění? Použijte (a) postup plnění prázdného modelu a (b) postup redukování plného modelu. Porovnejte výsledky. Nakreslete příslušný ordinacní diagram.
3. Importujte data obsahující druhové složení a proměnné prostředí na loukách v Bílých Karpatech. Z půdních paramterů (proměnné pHw až C.N) vyberte ty, které mají průkazný vliv na složení vegetace. Proveďte rozklad variability složení vegetace. Jednu skupinu prediktorů budou tvořit vybrané půdní parametry a druhou druh poloparazita, který se vyskytuje na dané lokalitě (proměnná hemiparasite). Otestujte parciální efekty obou skupin proměnných. *Samostatně*: Zkuste nakreslit ordinacní diagram s oběma skupinami prediktorů, tak aby byl kategoriální prediktor zobrazený centroidy a kontinuální prediktor šipkami.
4. *Samostatně*: Importujte data o vlivu pastvy/oplocení na druhové složení vegetace tundry (Saccone et al. 2014 J. Ecol.). Otestujte rozdíly v druhovém složení mezi typy prostředí (proměnná *intertreat*). Zkuste vymyslet jakou metodu použít, zamyslete se nad transformacemi a standardizacemi. Výsledek prezentujte na ordinacním diagramu. Před importováním dat pečlivě prostudujte soubor README.

Trocha teorie k postupnému výběru prediktorů

Pokud jsme postaveni před problémem výběru sady pro studované společenstvo nejdůležitějších prediktorů a nemáme a-priori stanovené hypotézy, obvykle se uchýlíme k postupnému výběru proměnných podle jejich "důležitosti". Kritériem důležitosti je vysvětlená (upravená) variabilita - R^2 a efekt postupně přidávaných proměnných je testován permutačním testem. Klasický postupný výběr se zastaví ve chvíli, kdy již žádný z do modelu ještě nezahrnutých prediktorů není statisticky významný na předem zvolené hladině významnosti (obvykle 5%). Takový postup ale je znám jako příliš liberální. Jednak může vybrat "významný" model i ve chvíli, kdy by žádný významný model neměl být nalezen (a tedy zvýšit chybu prvního řádu), nebo vést k výběru příliš mnoha prediktorů (a tedy nafukující vysvětlenou variabilitu). Jednou z možností, jak tento problém ošetřit je použití korekce p-hodnot pro mnohonásobná testování. Ty ale zase bývají příliš restriktivní.

Alternativní řešení tohoto problému spočívá v použití metody dle Blanchet et al. 2008 neboli „chráněného testu“

- Pro zamezení nafouknutí chyby prvního řádu je nejprve otestován globální model se všemi vysvětlujícími proměnnými zahrnutými. Pouze pokud tento test je významný (významnost tohoto testu je právě ona „ochrana“), pokračujeme postupným výběrem nejdůležitějších prediktorů.
- Abychom omezili riziko, že by do modelu bylo zahrnuto příliš mnoho proměnných, spočítáme upravený koeficient determinace (Adjusted R^2) globálního modelu, který použijeme jako druhé kritérium pro zastavení výběru - postupný výběr zastavujeme buď ve chvíli, kdy už žádný další prediktor není významný, nebo pokud by zahrnutí další proměnné vedlo k překročení tohoto R^2 .

Dva směry výběru prediktorů

Postupný výběr může probíhat dvěma směry: postupným přidáváním proměnných (forward selection), nebo jejich ubíráním (backward selection). První přístup se hodí spíše pro observační data, kde můžeme mít hodně prediktorů, které mohou být i korelované. Když se takové prediktory zahranou šechny do modelu, může backward selection fungovat špatně. Naopak pro experimentální data je výhodná backward selection, protože někdy potřebujeme odhalit třeba efekty interakcí i v případě, že jsou hlavní efekty prediktorů neprůkazné, čímž pádem by neprošly přes forward selection. Přidáváme tu proměnnou, která významně zlepšuje model - která k již vysvětlené variabilitě přidá významný díl vysvětlené variability. Odebíráme pak tu proměnnou, která model významně nezhorší (čili tu, která nepřidává významné množství variability k variabilitě již vysvětlené ostatními proměnnými). Po každé změně modelu se mění i podíly vysvětlené variability a významnosti. Proto výběr probíhá postupně - přidáváme, nebo odebíráme v každém kroku vždy jen jednu proměnnou.

Při postupném výběru je možné a vhodné oba směry kombinovat. V každém kroku otestujeme, zda nám ještě nezahrnuté proměnné významně zlepšují model a zároveň, zda nám vyhození některé z již zahrnutých proměnných model významně nezhorší. To dělá i funkce `ordistep()` s argumentem `direction = "both"`.

R funkce

vegan:

`rda()` - PCA, RDA

`cca()` - CA, CCA

`anova()` - test významnosti modelu

`RsquareAdj()` - výpočet upraveného R^2

`add1()` - otestuje přidává do modelu proměnné a testuje jejich jednoduchý efekt

`drop1()` - postupně odebírá z modelu proměnné a testuje zhoršení modelu

`update()` - přidání nebo odebrání proměnné z již existujícího modelu

`ordistep()` - automatická selekce

`ordiR2step()` - automatická selekce s Blanchetovým kritériem.

`varpart()` - rozklad variability

`Condition()` - specifikace kovariáty v partiální RDA/CCA

Kreslení ordinačního diagramu

`points(ordiplot.objekt, what = "centroids", arrows=F)` přidá centroidy kategoriálních proměnných

`text(ordiplot.objekt, what = "centroids", arrows=F)` přidá popisky kategoriálních proměnných

`text(ordiplot.objekt, what = "biplot", arrows=T)` přidá šipky a popisky kontinuálních prediktorů

```
# RDA a CCA
```

```
# ****
```

```
# se všemi proměnnými
```

```
rda(spe ~ ., data = env)
cca(spe ~ ., data = env)
# s vybranými promennými
rda(spe ~ temp + cond, data = env)
cca(spe ~ temp + cond, data = env)
# parciální PCA (CA), odfiltrování proměnné pH
rda(spe ~ Condition(pH), data = env)
cca(spe ~ Condition(pH), data = env)
# kondicionální, efekt temp po odfiltrování proměnné pH
rda(spe ~ temp + Condition(pH), data = env)
cca(spe ~ temp + Condition(pH), data = env)
# Upravene R2
RsquareAdj(model)
# Forward selection
# -----
# nulový model:
```

```

rda(spe ~ 1, data=env)
# plny model:
rda(spe ~ ., data = env)

# test pridani promenne:
add1(nulovy_model, scope= cilovy_model, test = "permutation",
      permutations=how(nperm=999))
# test ubrani promenne:
drop1(nynejsi_model, test = "permutation",
      permutations=how(nperm=999))

# automaticky vyber:
ordiR2step (nulovy_model, scope= cilovy_model)

# Testovani modelu
# -----
# globalni test modelu
anova()

# test promenych v modelu - sekvenční testování (závisí na pořadí)
anova(..., by= "terms")

# test parciálních efektů (po odečtení vlivů dalších prediktorů)
anova(..., by= "margin")

# test jednoduchých (simple) efektů (ignoruje vlivy dalších
      prediktorů)
add1(nulový model, .~.+pred1+pred2,...)

# test os
anova(..., by= "axis")

```