

Table of Contents

Ekologická podobnost	1
<i>Podobnost, nepodobnost a vzdálenost</i>	1
<i>R funkce</i>	1
<i>Příklad</i>	2
<i>Rozšiřující informace</i>	2



Česká verze stránek není od roku 2013 aktualizována. Aktuální verzi této stránky najdete v její anglické verzi [zde](#).

Ekologická podobnost

Výpočet ekologické podobnosti mezi vzorky je základní nástroj, kterým přistupujeme k mnohorozměrným ekologickým datům. Pod pojmem **ekologická podobnost** v tomto případě myslím rozdíly v druhovém složení dvou vzorků (např. vegetačních snímků). Na ekologické podobnosti je založena jak numerická klasifikace, tak i ordinační analýza. U mnohých ordinačních technik není vztah k ekologické podobnosti na první pohled patrný, ale je tam; například algoritmus CA (korespondenční analýza) rozmístí vzorky v ordinačním prostoru tak, že jejich vzdálenosti jsou vyjádřeny pomocí chí-kvadrát vzdáleností¹⁾.

Podobnost, nepodobnost a vzdálenost

Intuitivní přístup je uvažovat o **podobnosti** (*similarity*) mezi vzorky - čím jsou si vzorky podobnější svým druhovým složením, tím je jejich podobnost vyšší. Indexy podobnosti nabývají hodnot 0 (vzorky nesdílejí žádný druh) až 1 (vzorky jsou identické, sdílejí všechny druhy, a pokud uvažují i o abundancích, abundance všech druhů jsou také stejné). Ordinační i klasifikační techniky nepracují s podobnostmi, ale s tzv. vzdálenostmi vzorků, protože potřebují umístit vzorky do (mnohorozměrného) prostoru. **Vzdálenost** (*distance*) mezi vzorky prakticky odpovídá jejich ekologické **nepodobnosti** (*dissimilarity*) (i když teoreticky ne každá nepodobnost se dá vyjádřit jako vzdálenost bodů v euklidovském prostoru), a dá se odvodit buď z jednotlivých indexů podobnosti jejich odečtením od jedničky, případně existují vlastní míry vzdáleností (např. euklidovská). Platí přitom, že zatímco všechny míry podobnosti se dají převést na nepodobnost (vzdálenost), ne všechny míry nepodobnosti (vzdálenosti) se dají převést na podobnost (to platí opět např. o euklidovské vzdálenosti).

Indexů podobnosti, resp. měr vyjadřujících vzdálenost mezi vzorky je celá řada ([Legendre & Legendre 2012](#) jich uvádějí kolem 30) a vyznat se v nich není úplně jednoduchá záležitost. Při výběru se musíme nejdříve rozhodnout, jestli je naším cílem R nebo Q analýza²⁾. [Legendre & Legendre 2012](#) nabízí jakýsi klíč k výběru vhodné míry v případě daného typu dat a studovaného problému (tabulky 7.4-7.6). Obecně (a silně zjednodušeně) se dá říci, že Bray-Curtis, Sørensen nebo tětiovová vzdálenost (*Chord distance*)) fungují na ekologických datech lépe než euklidovská nebo chí-kvadrát vzdálenost.

V názvech jednotlivých indexů je docela chaos, a stejný název se někdy používá pro index podobnosti i nepodobnosti. Aby se to nepletlo, budeme v celé kapitole uvažovat jen o indexech nepodobnosti, tedy o **vzdálenostech** (*distances*).

R funkce

- **dist** - základní funkce, která ale nabízí jen omezený výběr distancí - z těch používaných častěji v ekologii jsou to `euclidean`, `canberra` a `manhattan`. Výsledkem této i ostatních distančních funkcí je distanční matice (*distance matrix*), v tomto případě objekt třídy `dist`.

- **vegdist (library vegan)** - výchozí míra nepodobnosti v případě této funkce je Bray-Curtis distance, která je (narozdíl od euklidovské distance) považována za velmi vhodná na ekologická data (jedná se o kvantitativní míru odvozenou od Sørensenova indexu podobnosti, která ignoruje dvojité nuly).
- **dsvdis (library labdsv)** - podobná `vegdist`, ale v jiné knihovně.
- **designdist (library vegan)** - pokud znáte vzoreček pro některou z měr nepodobnosti, které nejsou zahrnuty v žádné z funkcí, pomocí této funkce si ji můžete na základě tohoto vzorečku vypočítat.
- **daisy (library cluster)** - nabízí mimo jiné také `manhattan` a `gower` distance.
- **distance (library ecodist)** - nabízí sedm vzdáleností, ale funkce spíše než pro praktické použití je určena k demonstraci toho, jak se dají jednotlivé vzdálenosti spočítat (pro větší matice je funkce pomalá).

Příklad

(`veg.data` obsahuje matici vzorky x druhy):

```
dist (x = veg.data, method = 'euclidean')
```

výsledkem je distanční matice s euklidovskými vzdálenostmi mezi vzorky.

Rozšiřující informace

Další inspiraci hledejte například v přehledu metod pro analýzu ekologických dat [CRAN Task View: Analysis of Ecological and Environmental Data](#), kterou spravuje Gavin Simpson, v sekci *Dissimilarity coefficients*.

1)

Podobně v případě PCA (analýzy hlavních komponent) jsou vzorky rozmístěny na základě euklidovské vzdálenosti.

2)

R analýza se zaměřuje na podobnost mezi druhy, Q analýza na podobnost mezi vzorky). Pokud nám jde o vztahy mezi vzorky (Q analýza), nejlepší volbou pro ekologická data jsou indexy ignorující dvojité nepřítomnosti (nebo dvojitě nuly, *double zeros*). Dvojitá nepřítomnost ve zkratce znamená, že některý druh se nevyskytuje zároveň v obou vzorcích; pokud z matice vzorky x druhy vyseknu pouze ony dva řádky se vzorky, které chci porovnat, řada druhů bude chybět zároveň v obou vzorcích. Některé indexy se k těmto dvojitým nepřítomnostem druhů (0-0) chovají stejně jako k dvojitým přítomnostem (1-1) a předpokládají, že druh chybějící v obou vzorcích svědčí o podobnosti těchto vzorků (toto je případ i často používané euklidovské vzdálenosti). To ale v ekologii neplatí - druh může v jednom snímku chybět z několika různých důvodů, a samotná absence druhu by proto neměla být považována za znak podobnosti (Příkladem mohou být dva vzorky, jeden na vlhkém a druhý na suchém stanovišti. V obou vzorcích zároveň chybí mezické druhy, ale v každém z jiného důvodu - ve vlhkém proto, že je tam na něj moc vlhko, v suchém zase proto, že je na něj příliš sucho. O podobnosti na základě absence druhu zde tedy nemůže být řeč

From:

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

Permanent link:

<https://anadat-r.davidzeleny.net/doku.php/cs:similarity>

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