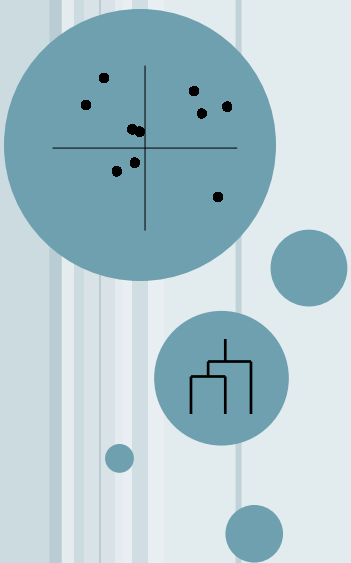
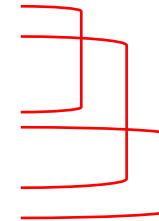


ĚKOLOGICKÁ PODOBNOST (*ECOLOGICAL RESEMBLANCE*)

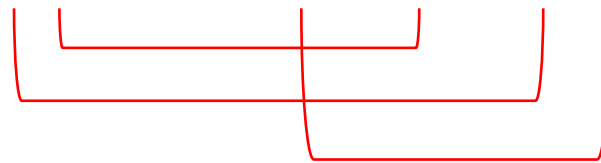


EKOLOGICKÁ PODOBNOST Q VS R ANALÝZA

Vzorky	Druhy		
	druh 1	druh 2	druh 3
vzorek 1	0	1	1
vzorek 2	1	0	0
vzorek 3	0	4	4



vztahy mezi vzorky
Q analýza



vztahy mezi druhy
(nebo obecně mezi deskriptory)
R analýza

PODOBNOSTI X VZDÁLENOSTI (Q ANALÝZA)

Indexy podobnosti

- slouží k vyjádření podobnosti mezi vzorky, ne k jejich umístění do mnohorozměrného prostoru (například ordinace)
- nejnižší hodnota 0 – vzorky nesdílejí žádný druh
- nejvyšší hodnota (1 nebo jiná) – vzorky jsou identické

Vzdálenosti mezi vzorky

- slouží k umístění vzorků v mnohorozměrném prostoru
- nejnižší hodnota 0 – vzorky jsou identické (ve stejné lokaci)
- hodnota se zvyšuje se zvyšující se nepodobností mezi vzorky

INDEXY PODOBNOSTI

kvalitativní vs kvantitativní

- kvalitativní – pro presenčně-absenční data
- kvantitativní – pro data vyjadřující abundance, počty aj.

symetrické vs asymetrické

- dvojité nepřítomnosti („*double-zero*“) – počet druhů, které **chybí** zároveň v obou vzorcích, v kontrastu s počtem druhů které se vyskytují zároveň v obou vzorcích
- symetrické – dvojité nepřítomnosti hodnotí stejně jako dvojité přítomnosti (totiž že vyjadřují podobnost mezi vzorky); v ekologii se prakticky nepoužívají
- asymetrické – dvojité nepřítomnosti ignorují; nejčastější typ indexů podobnosti v ekologii

PROBLÉM DVOJITÝCH NEPŘÍTOMNOSTÍ (*DOUBLE-ZEROS*)

Fakt, že druh chybí zároveň v obou snímcích, může mít několik významů:

- vzorky leží mimo ekologickou niku druhu
 - nemůžeme ale říct, jestli oba vzorky leží na stejné straně ekologického gradientu mimo druhovou niku (a jsou si tedy docela podobné) nebo na stranách opačných (a jsou pak úplně odlišné)
- vzorky leží uvnitř ekologické niky druhu, ale druh se ve vzorku nevyskytuje, protože
 - se tam nedostal (*dispersal limitation*)
 - jsme ho přehlédli a nezaznamenali (*sampling bias*)
 - nachází se právě v dormantním stadiu a není proto vidět (jednoletky, geofyty)

PROBLÉM DVOJITÝCH NEPŘÍTOMNOSTÍ (DOUBLE-ZEROS)

	vlhkostilný druh 1	vlhkostilný druh 2	mezický druh 1	mezický druh 2	suchostilný druh 1	suchostilný druh 2
snímek 1	1	1	0	0	0	0
snímek 2	0	1	1	1	1	0
snímek 3	0	0	0	0	1	1

- snímky 1 až 3 jsou seřazeny podle vlhkosti stanoviště – snímek 1 je nejvlhčí, snímek 3 nejsušší
- snímek 1 a 3 neobsahují ani jeden mezický druh – snímek 1 je pro tyto druhy příliš vlhký, snímek 3 příliš suchý
- **symetrické indexy** podobnosti: dvojitá nepřítomnost mezických druhů bude zvyšovat podobnost snímků 1 a 3
- **asymetrické indexy**: dvojitě nepřítomnosti jsou ignorovány

INDEXY PODOBNOSTI PRO KVALITATIVNÍ DATA

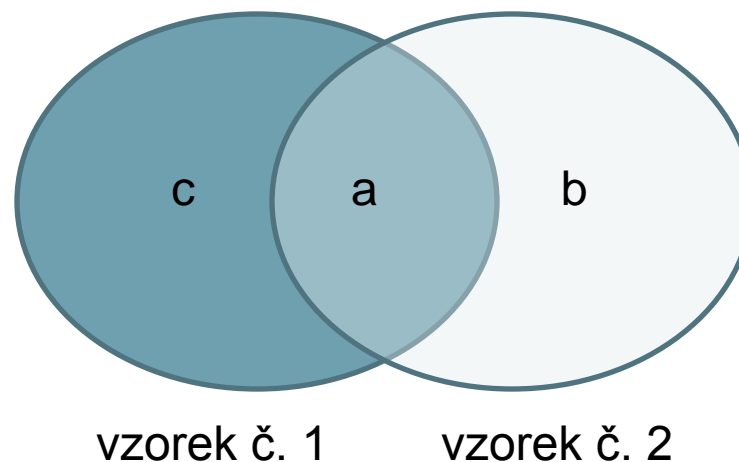
druh je		ve vzorku č. 1	
		přítomen	nepřítomen
ve vzorku č. 2	přítomen	a	b
	nepřítomen	c	d

a – počet druhů přítomných v obou vzorcích

b, c – počet druhů přítomných jen v jednom vzorku

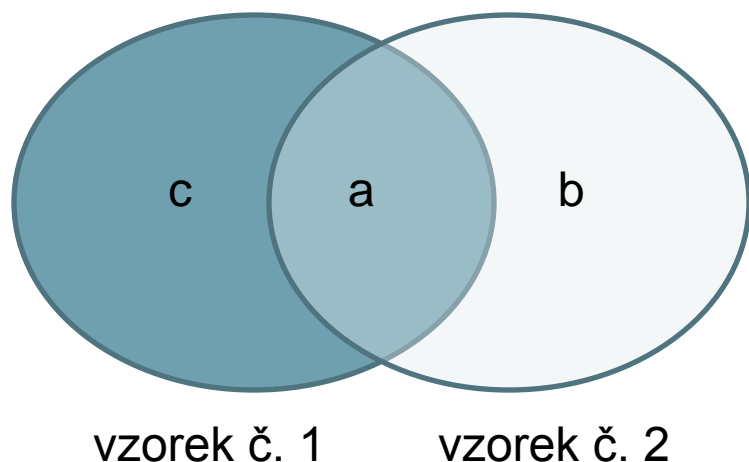
d – počet druhů, které chybí v obou vzorcích („double zeros“)

Pokud nebereme v úvahu druhy nepřítomné v obou vzorcích (d), lze zobrazit i pomocí Vennova diagramu →



INDEXY PODOBNOSTI PRO KVALITATIVNÍ DATA

- Jaccardův koeficient $J = a / (a + b + c)$
- Sørensenův koeficient $S = 2a / (2a + b + c)$
 - přítomnosti druhu v obou vzorcích (a) přisuzuje dvojnásobnou váhu
- Simpsonův koeficient $Si = a / [a + \min(b,c)]$
 - vhodný pro vzorky velmi odlišné počtem druhů



INDEXY PODOBNOSTI PRO KVANTITATIVNÍ DATA

- např. zobecněný Sørensenův koeficient (procentická podobnost, *percentage similarity*)

$$PS = [2 \sum \min (x_i, y_i)] / \sum (x_i + y_i)$$

- x_i, y_i ... kvantita i-tého druhu ve srovnávaných vzorcích
- má rozsah od 0 do 1
- pro presenčně absenční data přechází v $2a / (2a + b + c)$
- velmi vhodný pro ekologická data
- *percentage dissimilarity (PD, Bray-Curtis index)* = $1 - PS$

VZDÁLENOSTI MEZI VZORKY (*DISTANCE MEASURES*)

- všechny indexy podobnosti (kvalitativní i kvantitativní) lze převést na distance

$$D = 1 - S, \text{ nebo } D = \sqrt{1 - S}$$

- kde D je vzdálenost (*distance*) a S je podobnost (*similarity*)
- odmocninový převod se používá například pro Sørensenův koeficient
- neplatí obráceně (ne všechny vzdálenosti se dají převést na podobnosti – např. Euklidovská vzdálenost)

VZDÁLENOSTI MEZI VZORKY (*DISTANCE MEASURES*)

- Euklidovská vzdálenost (*Euclidean distance*)

$$ED = \sqrt{\sum (x_i - y_i)^2}$$

- rozsah: od 0 (identické vzorky), horní mez není dána
- rozsah hodnot výrazně záleží na použitých jednotkách
- míra citlivá na odlehlé body - nevhodná pro ekologická data

- Tětivová vzdálenost (*chord distance, relativized Euclidean distance*)

- Euklidovská vzdálenost použitá na datech standardizovaných přes vzorky (*by sample norm*)
- rozsah: od 0 (identické vzorky) do $\sqrt{2}$ (vzorky nesdílí žádný druh)

- Chi-kvadrát vzdálenost (*chi-square distance*)

- málokdy se používá přímo na výpočet vzdálenosti mezi vzorky
- vyjadřuje vzdálenost mezi vzorky v unimodálních ordinačních metodách (např. v korespondenční analýze, CA)

EUKLIDOVSKÁ VZDÁLENOST

PARADOX

- o může se stát, že dva vzorky, které sdílí některé druhy (vzorky 1 a 3), budou mít větší vzdálenost než dva vzorky, které nesdílí ani jeden druh (vzorky 1 a 2)

Vzorky	Druhy		
	druh 1	druh 2	druh 3
vzorek 1	0	1	1
vzorek 2	1	0	0
vzorek 3	0	4	4

1,732
4,243

$$\text{Eucl (vzorek 1, vzorek 2)} = \sqrt{(0-1)^2 + (1-0)^2 + (1-0)^2} = 1,732$$

$$\text{Eucl (vzorek 1, vzorek 3)} = \sqrt{(0-0)^2 + (1-4)^2 + (1-4)^2} = 4,243$$

INDEXY PODOBNOSTI MEZI DRUHY (R ANALÝZA)

V kolika vzorcích je ...		druh č. 1	
		přítomen	nepřítomen
druh č. 2	přítomen	a	b
	nepřítomen	c	d

- Diceův index $\text{Dice} = 2a / (2a + b + c)$
 - stejný jako Sørensenův index pro podobnost mezi vzorky
 - uveden dříve než Sørensen (Dice 1945 vs Sørensen 1948)
- Pearsonův korelační koeficient r
 - není vhodný pro data s velkým počtem nul, ani po transformaci

MATICE PODOBNOSTÍ (VZDÁLENOSTÍ) MEZI VZORKY (NEBO DRUHY)

- je symetrická (podobnost mezi 2. a 3. snímkem = podobnost mezi 3. a 2. snímkem)
- diagonála obsahuje pouze nuly (matice vzdáleností) nebo pouze jedničky (matice podobností)

	1	2	3	4	5	6	7	8	9	10
1	0	12.37	11.70	17.92	13.86	10.58	11.92	10.54	13.82	15.59
2	12.37	0	11.14	13.34	16.58	13.96	9.64	13.56	13.64	13.42
3	11.70	11.14	0	14.42	16.16	11.53	10.34	13.71	14.90	13.78
4	17.92	13.34	14.42	0	18.36	15.78	9.64	17.03	14.42	7.48
5	13.86	16.58	16.16	18.36	0	13.71	14.49	9.00	14.04	15.46
6	10.58	13.96	11.53	15.78	13.71	0	11.31	11.87	10.54	12.85
7	11.92	9.64	10.34	9.64	14.49	11.31	0	13.82	12.77	9.43
8	10.54	13.56	13.71	17.03	9.00	11.87	13.82	0	10.95	14.35
9	13.82	13.64	14.90	14.42	14.04	10.54	12.77	10.95	0	10.39
10	15.59	13.42	13.78	7.48	15.46	12.85	9.43	14.35	10.39	0

matice Euklidovských vzdáleností mezi 10 vzorky

MATICE „VZORKY × DRUHY“ V EKOLOGII SPOLEČENSTEV (*SPARSE MATRIX*, ŘÍDKÁ MATICE)

více než 90% hodnot tvoří nuly, u
velkých souborů až 99%

