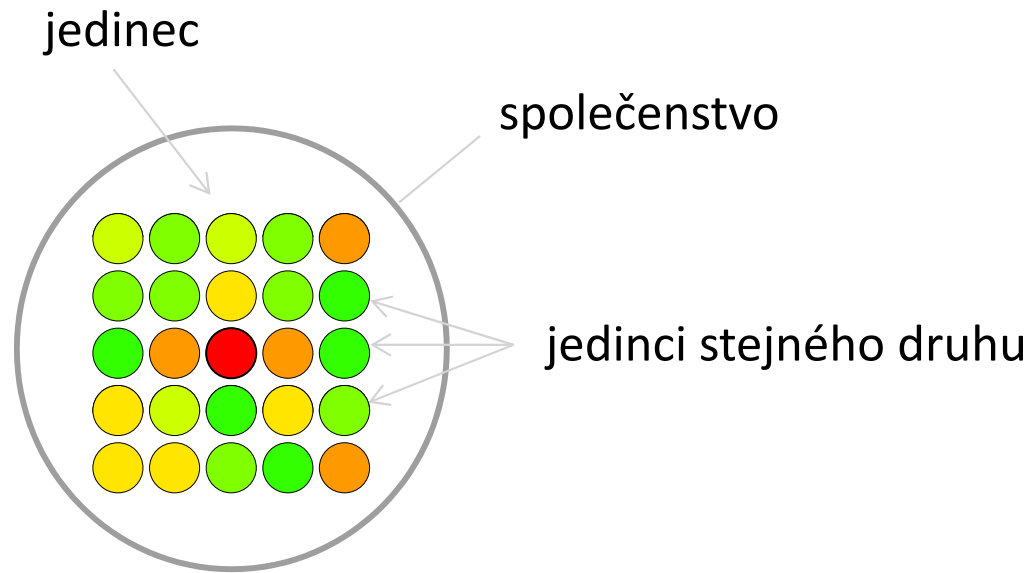


EKOLOGICKÁ NEPODOBNOST
(ECOLOGICAL DISSIMILARITY)

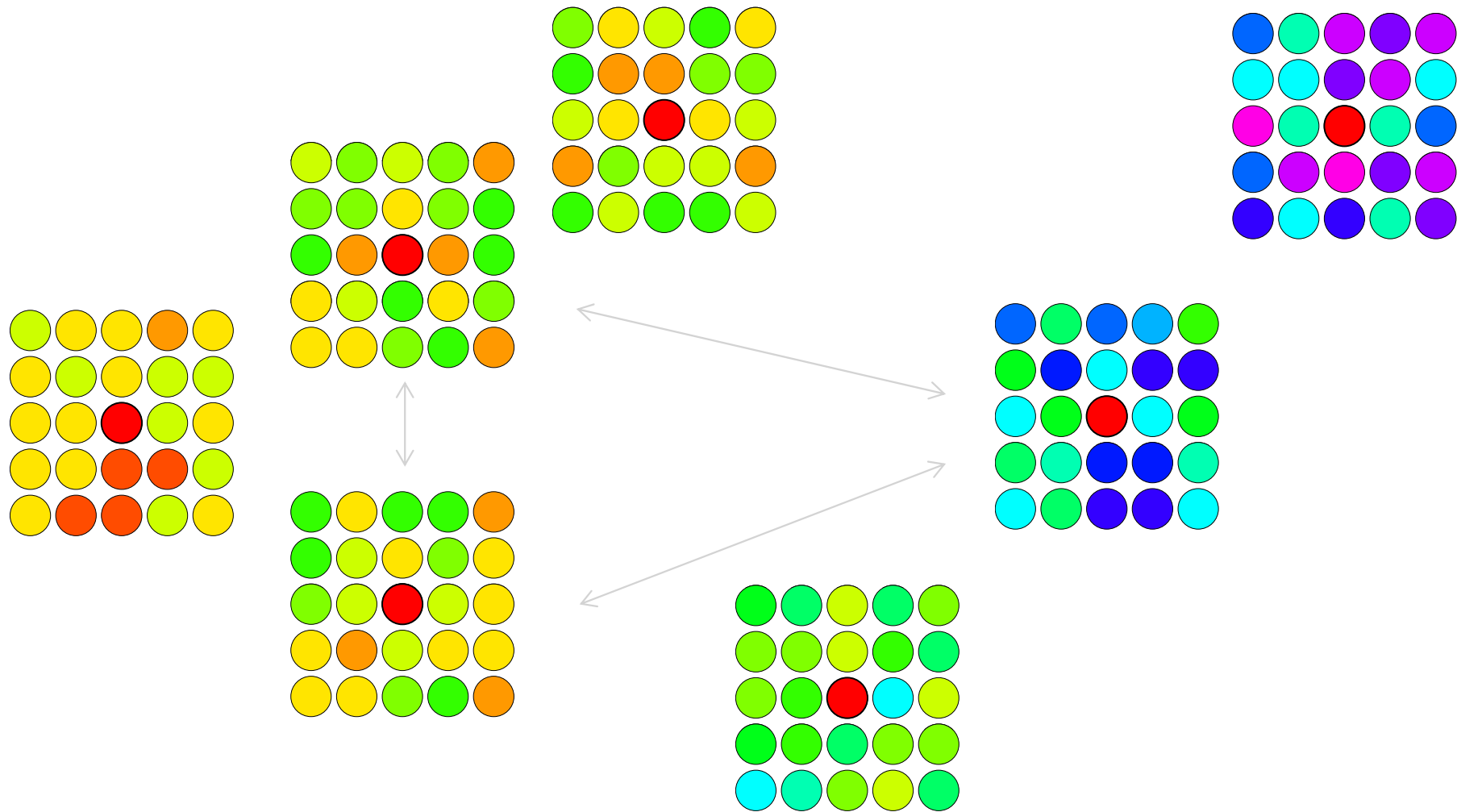
VÍCEROZMĚRNÁ NEPODOBNOST (*MULTIVARIATE DISSIMILARITY*)

- Výpočet nepodobností (typicky mezi pozorováními) je prvním krokem mnohorozměrné analýzy
 - Výsledkem je matice nepodobností
- Nepodobnost (*dissimilarity*) je opak podobnosti (*similarity*)
 - $D = 1 - S$
- Vzdálenost = nepodobnost založená na **metrické** vzdálenosti
 - Lze zobrazit v prostoru (n-rozměrném)

EKOLOGICKÁ NEPODOBNOST



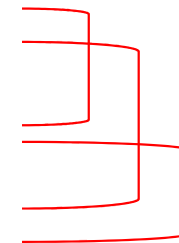
EKOLOGICKÁ NEPODOBNOST



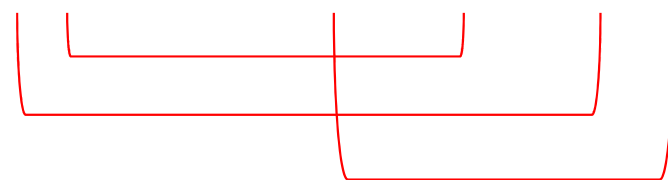
EKOLOGICKÁ NEPODOBNOST

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 NEPODOBNOSTI (Q ANALÝZA)

Nepodobnosti a jejich koeficienty (*dissimilarity coefficients*)

- slouží k umístění vzorků v mnohorozměrném prostoru
- nejnižší hodnota 0 – vzorky mají shodné vlastnosti
- hodnota se zvyšuje se zvyšující se nepodobností mezi vzorky

Indexy podobnosti (*similarity coefficients*)

- 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

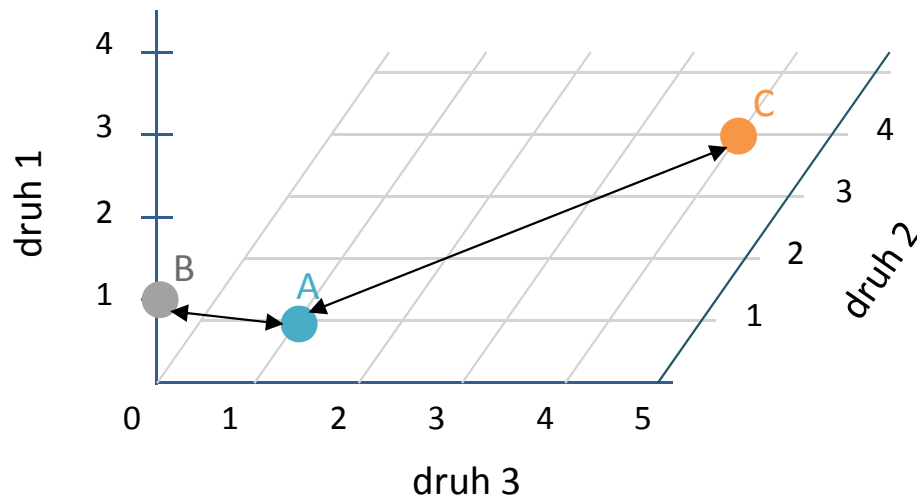
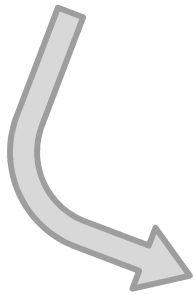
VZDÁLENOST – CO TO JE?

	druh 1	druh 2	druh 3
vzorek A	0	1	1
vzorek B	1	0	0
vzorek C	0	4	4

Párové Euklidovské vzdálenosti
(jako bychom změřili pravítkem)

$$D_{(1,2)} = \sqrt{(0-1)^2 + (1-0)^2 + (1-0)^2} = 1.732$$

$$D_{(1,3)} = \sqrt{(0-0)^2 + (1-4)^2 + (1-4)^2} = 4.243$$



EUKLIDOVSKÁ VZDÁLENOST

PARADOX PŘI POUŽITÍ ABUNDANČNÍCH DAT

- při použití abundančních dat se může 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)

Community matrix

	druh 1	druh 2	druh 3
vzorek A	0	1	1
vzorek B	1	0	0
vzorek C	0	4	4

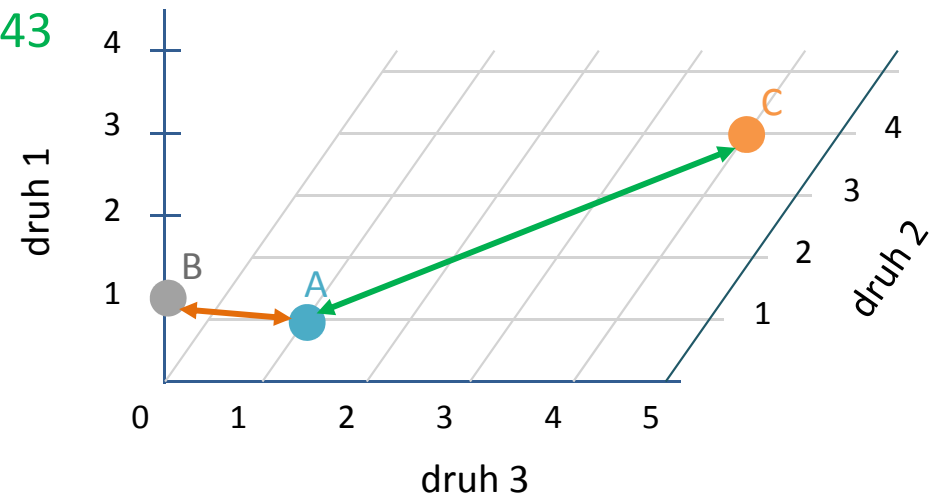
1,732

4,243



Dissimilarity matrix

	A	B	C
A	0	1.73	4.24
B	1.73	0	5.74
C	4.24	5.74	0



INDEXY (NE)PODOBNOSTI (DIS-SIMILARITY COEFFICIENTS)

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*)

Skutečnost, že druh je přítomen zároveň v obou snímcích znamená, že:

- vzorky leží uvnitř jeho ekologické niky
 - lokality jsou si **podobné**

Naproti tomu skutečnost, že druh chybí zároveň v obou snímcích, může znamenat, že:

- vzorky leží mimo ekologickou niku druhu
 - nevíme ale, zda oba vzorky leží na stejné straně ekologického gradientu mimo niku druhu (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*)

PROBLÉM DVOJITÝCH NEPŘÍTOMNOSTÍ (DOUBLE-ZERO PROBLEM)

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

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

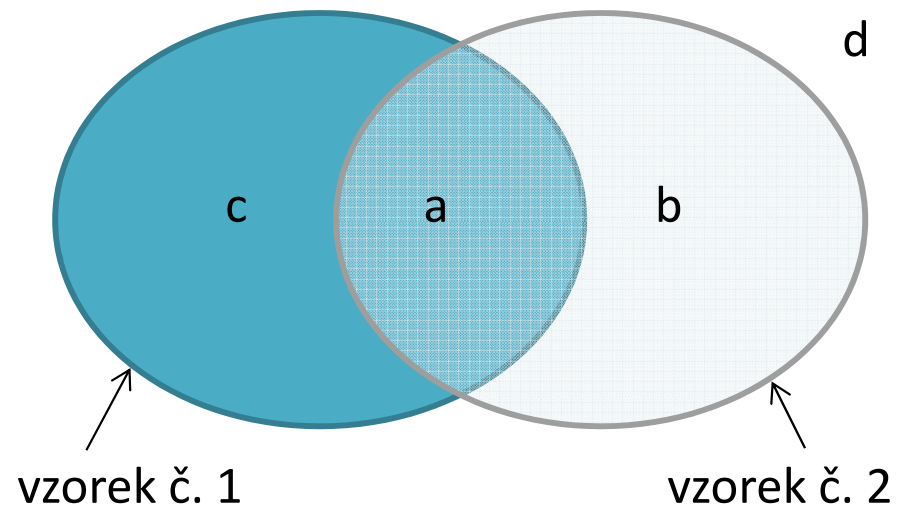
INDEXY PODOBNOSTI PRO KVALITATIVNÍ DATA

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“)



INDEXY PODOBNOSTI PRO KVALITATIVNÍ DATA

○ Jaccardův koeficient podobnosti

- podíl shodných druhů ku počtu všech druhů zaznamenaných na páru lokalit
- -> metrická distance

$$S_{Jac} = \frac{a}{a+b+c}$$

$$\left(D_{Jac} = 1 - S_{Jac} = \frac{b+c}{a+b+c} \right)$$

○ Sørensenův koeficient

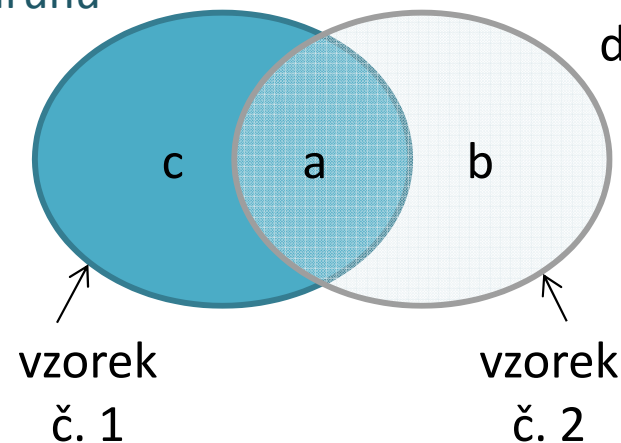
- oproti Jaccardovi má společná přítomnost druhů (a) dvojnásobnou váhu
- -> semimetrická distance
- binární forma Steinhausova, tzv. Bray-Curtis indexu

$$S_{Sor} = \frac{2a}{2a+b+c}$$

○ Simpsonův koeficient

- vhodný pro vzorky s velmi rozdílnými počty druhů

$$S_{Simp} = \frac{a}{a + \min(b, c)}$$



INDEXY (NE)PODOBNOTI PRO KVANTITATIVNÍ DATA

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

$$S_{Sor} = \frac{2W}{A+B} = \frac{2 \sum_{i=1}^p \min(x_i, y_i)}{\sum_{i=1}^p (x_i + y_i)}$$

- W – množství „shodných“ jedinců, A – počet jedinců v jednom vzorku, B – počet jedinců ve druhém vzorku
- x_i, y_i ... kvantita i -tého druhu ve srovnávaných vzorcích x a y
- 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)* = $1 - S_{Sor}$

VZDÁLENOSTI MEZI VZORKY (*DISTANCE COEFFICIENTS*)

- 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 – zajistí, že vzniklá distance je metrická
- naopak to ale vždy neplatí
 - např. převod Euklidovské vzdálenosti na podobnost lze provést jen na relativní škále

VZDÁLENOSTI MEZI VZORKY (*DISTANCE MEASURES*)

○ Euklidovská vzdálenost (*Euclidean distance*)

- rozsah: od 0 (identické vzorky), horní mez není dána
- rozsah hodnot výrazně závisí na použitých jednotkách
- míra citlivá na odlehlé body – nevhodná pro ekologická data
- symetrická míra vzdálenosti – trpí problémem dvojitých nul

$$D_{(x,y)} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

○ tětiová vzdálenost (*chord distance, relativized Euclidean distance*)

- Euklidovská vzdálenost použitá na datech standardizovaných přes vzorky (method = „normalize“, MARGIN = 1 ve funkci decostand)
- rozsah: od 0 (identické vzorky) do $2^{1/2}$ (vzorky nesdílí žádný druh)
- netrpí problémem dvojitých nul

$$y'_{ij} = \frac{y_{ij}}{\sqrt{\sum_{j=1}^p y_{ij}^2}}$$

○ Hellingerova vzdálenost (*Hellinger distance*)

- Euklidovská vzdálenost aplikovaná na data po aplikaci Hellingerovy standardizace
- netrpí problémem dvojitých nul
- vlastně tětiová vzdálenost vypočítaná na odmocninách abundancí

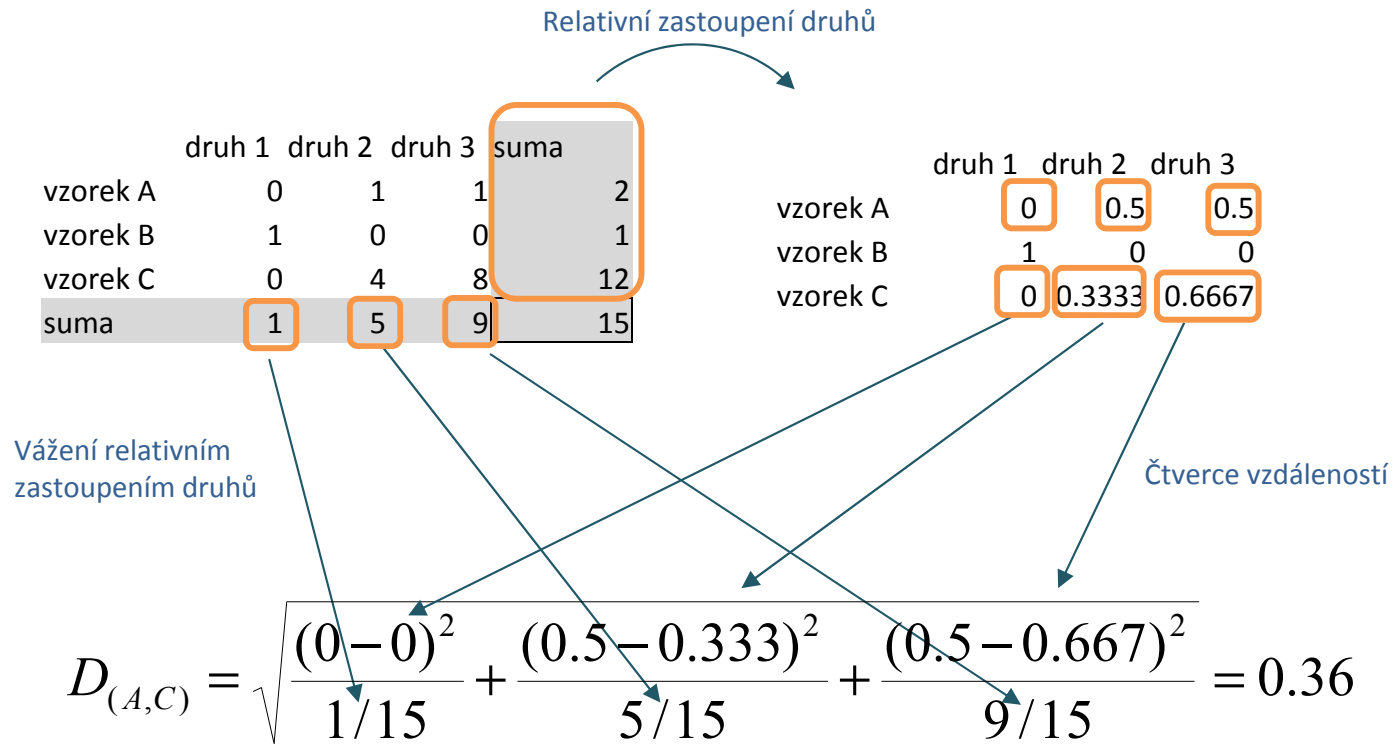
$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}}$$

○ 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)

$$D_{(x1,x2)} = \sqrt{\sum_{j=1}^p \frac{1}{y_{+j} / y_{++}} \left(\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

CHI-KVADRÁT VZDÁLENOST



	vzorek A	vzorek B	vzorek C
vzorek A	0	4.0092	0.36
vzorek B	4.0092	0	4.0208
vzorek C	0.36	4.0208	0

INDEXY (NE)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

$$Dice = \frac{2a}{2a + b + c}$$

- 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

- Lze použít i chi-kvadrát distanci

MATICE (NE)PODOBNOSTÍ MEZI VZORKY (NEBO DRUHY)

- je symetrická (podobnost mezi 2. a 3. snímkem = podobnost mezi 3. a 2. snímkem)
- diagonála obsahuje pouze nuly (matice nepodobností) nebo pouze jedničky (matice podobností)
- V R – *dist* objekt: pouze oblast pod diagonálou

	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 NEPODOBNOSTÍ (BRAY-CURTIS) V R (DATASET BÍLÉ KARPATY)

	CMA	CME	CMI	HUP	HUS	JAZ	KAZ	KOR	LES1
CME	0.3990148								
CMI	0.4188563	0.5626911							
HUP	0.8827586	0.8534923	0.8080000						
HUS	0.8284672	0.8126126	0.8077572	0.4752852					
JAZ	0.7668919	0.6594324	0.5102041	0.7719298	0.8252788				
KAZ	0.9366667	0.9242175	0.8852713	0.7439446	0.7582418	0.8881356			
KOR	0.8747764	0.8162544	0.8178808	0.7355680	0.7386139	0.8324226	0.6804309		
LES1	0.8945455	0.8563734	0.8218487	0.8143939	0.7983871	0.8000000	0.8284672	0.6765286	

MANTELŮV TEST

- Testuje korelaci mezi dvěma (stejně velkými) maticemi (ne)podobností
 - Spočte Pearson r na základě hodnot nepodobností matic
 - Testuje pomocí permutačního testu
 - Permutuje se pořadí vzorků v jedné z matic a spočte se permutované r
 - To se provede mnohokrát -> nulová distribuce r
 - děleno celkovým počtem
 - $P = k/(n+1)$
 - k je počet permutací, kdy permutované $r \geq$ skutečné r
 - n je celkový počet permutací (typicky 999 nebo 9999)
- Poměrně „primitivní“ metoda
- Lze použít třeba na testování korelace mezi prostorovou vzdáleností a nepodobností společenstev