



Statistické hodnocení biodiverzity

Jiří Jarkovský

Institut biostatistiky a analýz, Masarykova univerzita, 2009



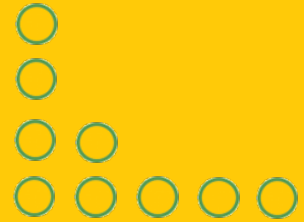
Program kurzu

- | | |
|--------|---|
| 22.9. | <ol style="list-style-type: none"> 1. Biodiverzita jako pojem. 2. Biodiverzita a biostatistika |
| 29.9. | <ol style="list-style-type: none"> 3. Vizualizace biodiverzity 4. Indexy diverzity a jejich statistická spolehlivost |
| 6.10. | <ol style="list-style-type: none"> 5. Species-abundance křivky a stochastické modely 6. Niche-oriented species - abundance modely |
| 13.10. | <ol style="list-style-type: none"> 7. Aplikovatelnost parametrických a neparametrických statistických technik při hodnocení biodiverzity 8. Parametrické hodnocení biodiverzity ve vícerozměrných analýzách |
| 20.10. | <ol style="list-style-type: none"> 9. Možnosti frakcionace biologických společenstev a následná analýza biodiverzity získaných podjednotek 10. Markovovy řetězy jako technika využitelná pro analýzu dat týkajících se biodiverzity |
| 27.10 | <ol style="list-style-type: none"> 11. Případové studie: Parazitární společenstva 12. Případové studie: Lišejníky a znečištění ovzduší |





I. Biodiverzita jako pojem



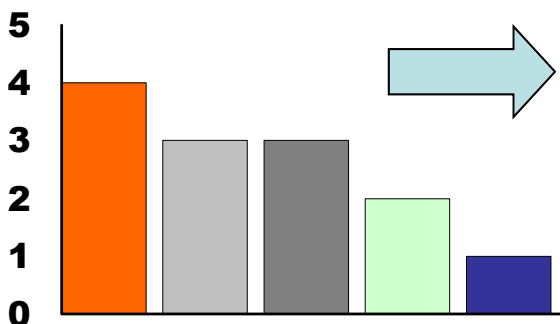
Biodiverzita ?

- ◆ Místo na Zemi → žijí zde organismy, tj. je zde biodiverzita → jak ji popsát, vysvětlit a co to znamená ?

- Dvě složky biodiverzity:
 - I. Různorodost – počet různých organismů (kvalita)
 - II. Relativní abundance – poměr výskytu organismů (kvantita)



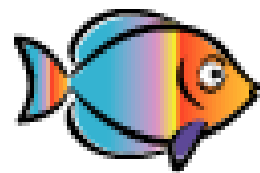
5



Proč ?



Definice biodiversity I

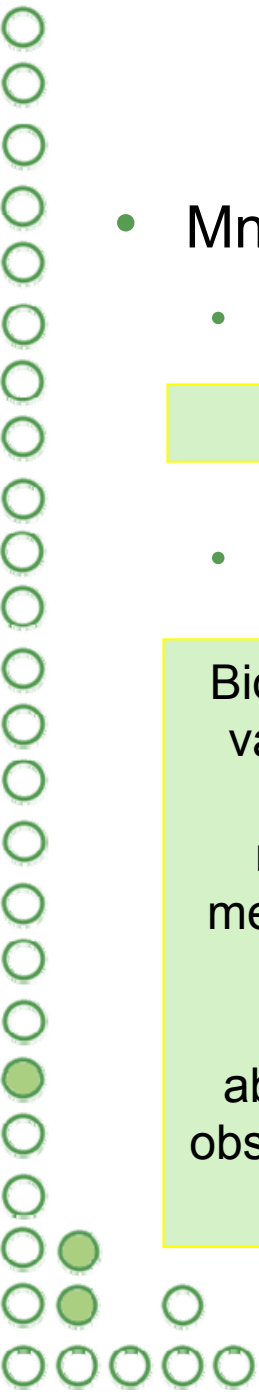
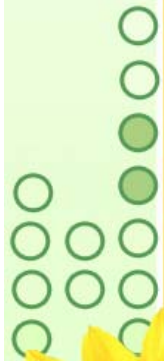


- Mnoho definic biodiversity 😊
 - Nejjednodušší:

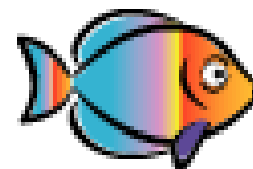
Počet druhů (druhová bohatost) (Fiedler and Jain, 1992)

- Komplexní definice:

Biodiversity is an attribute of an area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes, whether naturally occurring or modified by humans. Biodiversity can be measured in terms of genetic diversity and the identity and number of different types of species, assemblages of species, biotic communities and biotic processes, and the amount (e.g., abundance, biomass, cover, rate) and structure of each. It can be observed and measured at any spatial scale ranging from microsites and habitat patches to the entire biosphere (DeLong, 1996)



Definice biodiverzity II

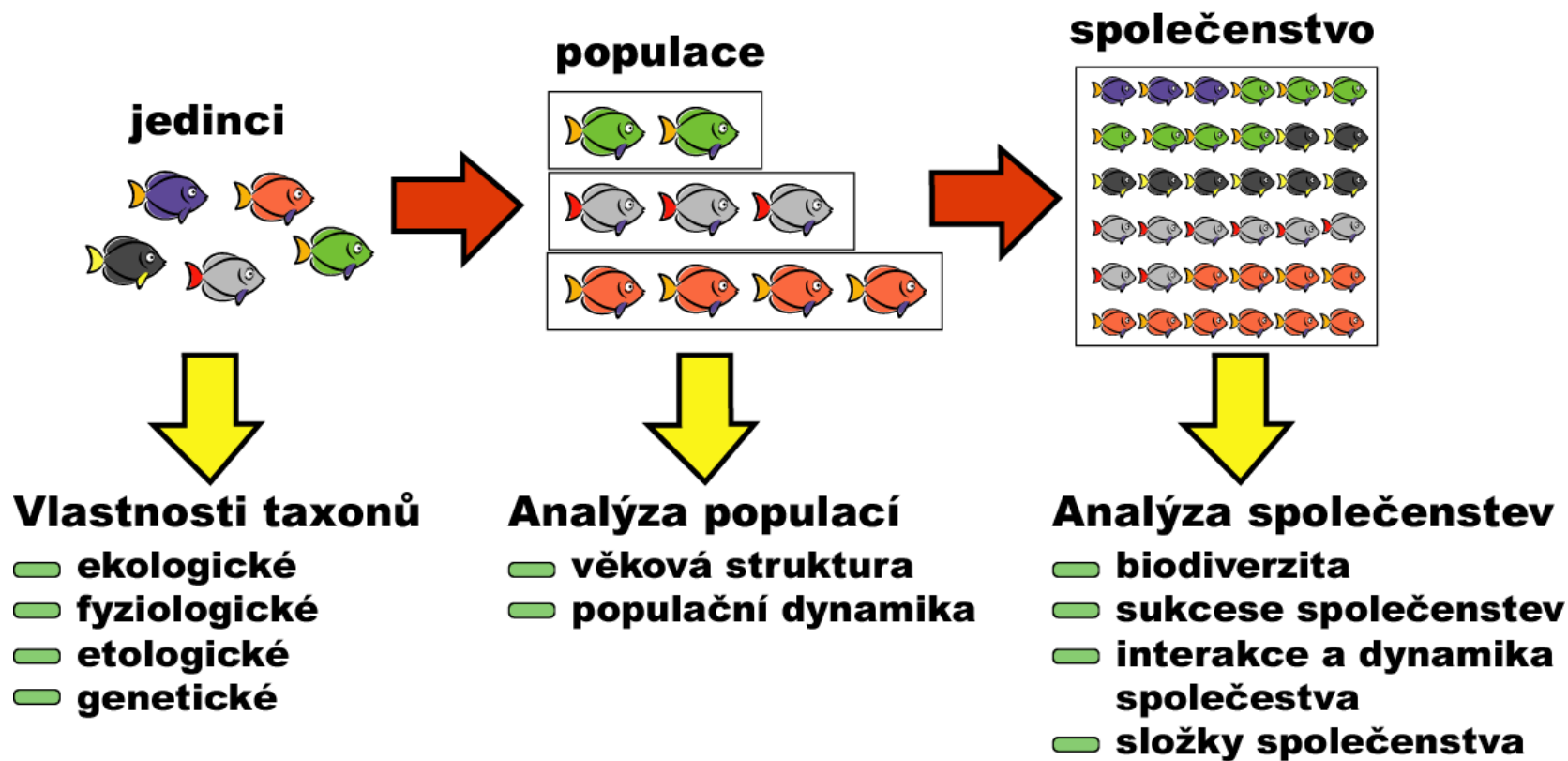


- Biodiverzita je odrazem pestrosti organismů
- V zcela nejjednodušším významu odpovídá biodiverzita počtu druhů
- V složitějším významu odráží také relativní abundance taxonů, tj. strukturu společenstva
- Kromě druhového složení je možné popisovat diverzitu i na nižší nebo vyšší úrovni než je organismus
 - Příkladem nižší úrovně může být např. genetická diverzita druhu
 - Příkladem vyšší úrovně jsou např. ekologické skupiny organismů
- V celkovém důsledku ji lze definovat jako komplexní pohled (taxonomický, ekologický, genetický) na složení společenstva organismů



Jedinci, populace a společenstva

- Teoretickou minimální jednotkou v ekologii je jedinec druhu



Typy biodiverzity

- Na biodiverzitu můžeme nahlížet z různých pohledů:
 - Taxonomická diverzita – výskyt a četnost jedinců druhů nebo jiných taxonomických jednotek



- Genetická – výskyt různých kombinací alel v populacích organismů

aa Aa AA



- Ekologická/funkční – funkce, kterou organismy vykonávají v rámci společenstva (predátor, parazit, dekompozitor, sesilní mobilní organismy atd.)



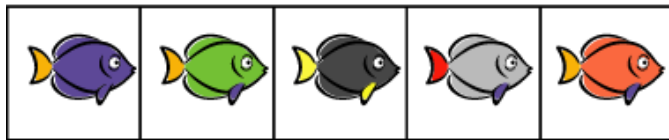
- Fyziologická/biochemická diverzita – způsoby a biochemické dráhy používané organismy k zpracování substrátu



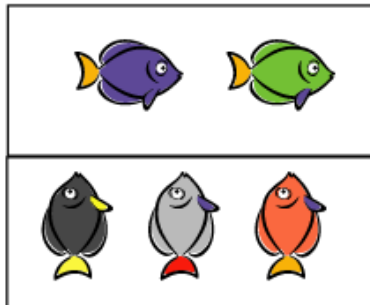
Různorodost organismů (kvalita)

- Určení minimální kvalitativní jednotky odpovídá typu biodiverzity, kterou chceme postihnout

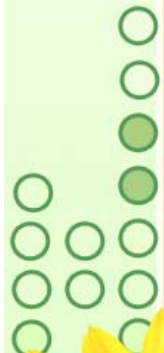
Taxonomie



Skupiny - ekologie, etologie, fyziologie



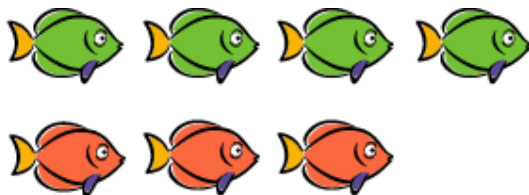
- ◆ Použití různých typů rozdělení organismů poskytuje různé pohledy na složení a typ jejich společenstva, proto je vhodné použít při hodnocení více možností kvalitativního přístupu
- ◆ V případě některých organismů lze použít pouze určité přístupy k jejich biodiverzitě (např. z důvodu nedostatku informací nebo nemožnosti některých postupů)



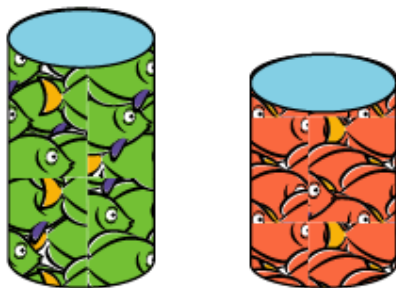
Kvantita organismů

- Kvantita organismů může být měřena různými způsoby

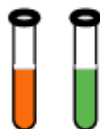
Počty



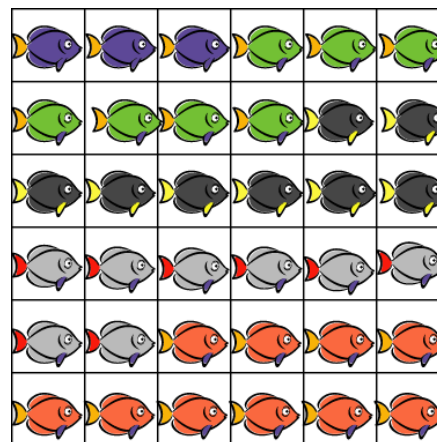
Biomasa



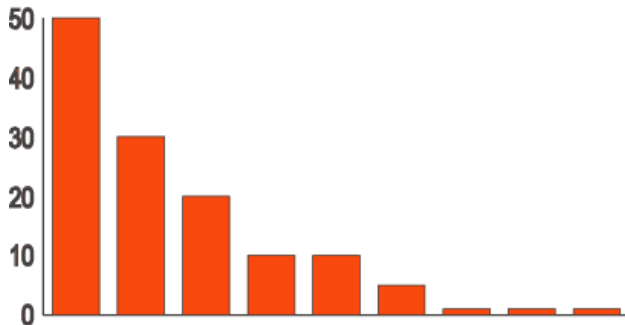
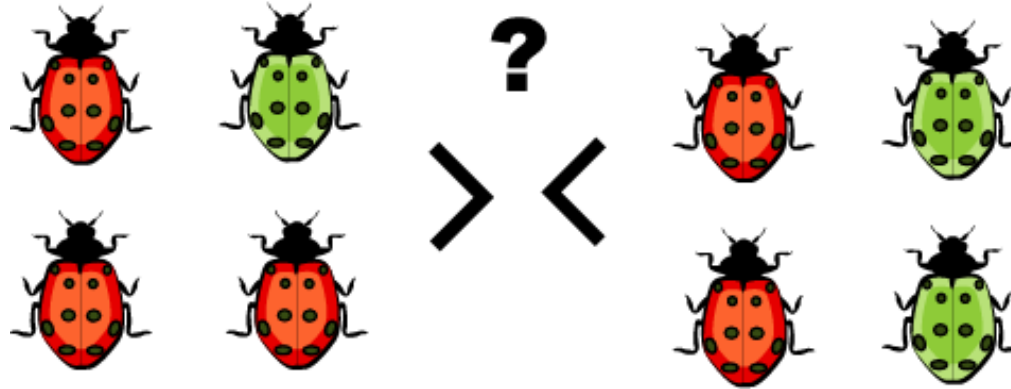
Aktivita



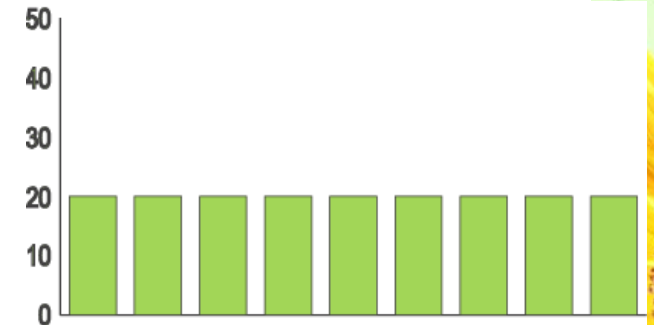
- ◆ Různé způsoby měření mají použití pro určité typy organismů nebo za určitých situací
- ◆ Kvantita organismů odráží obsazení prostředí organismy – lze analyzovat vzhledem k parametrům a historii daného prostředí



Problémy s biodiverzitou – co znamená větší diverzitu ?



?



Biodiversity in contemporary biology and environmental sciences

Biodiversity is one of the leading concepts in biology, that is defined at many levels ...

Genes

Individuals

Populations

Species

Communities

Ecosystems

... and very suitable for bioindication

Effects made in past

Effects of chronic exposure

Impact of low dose exposures

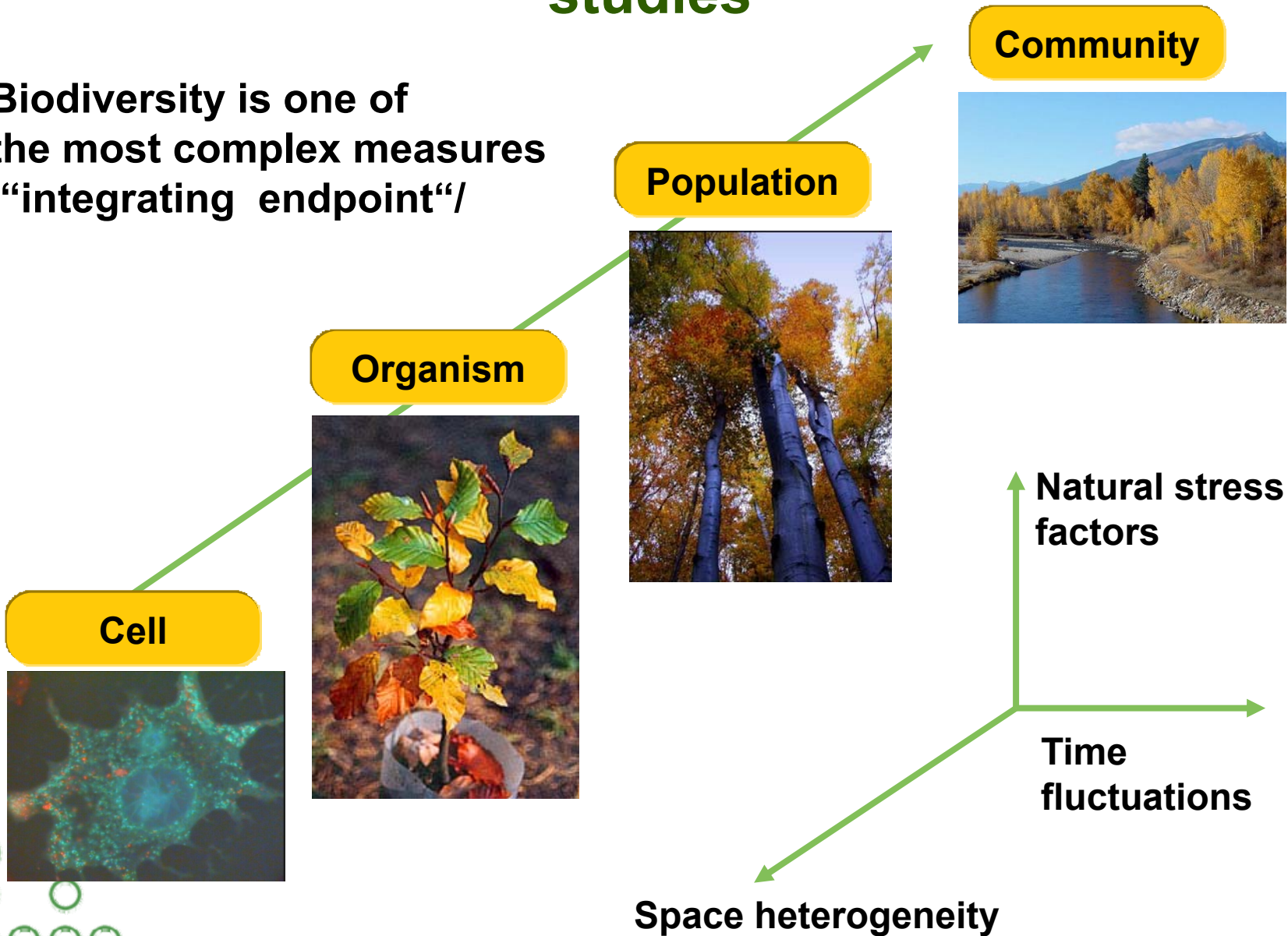
Impact on relationships in communities

Biodiversity bears very useful and lasting information, even when „short-term“ methods fail (biotests, chemical analyses)



Biodiversity as end-point in environmental studies

Biodiversity is one of the most complex measures /“integrating endpoint“/



Biological communities as typical assessment endpoint at ecosystem level

VALUES & BENEFITS



DRAWBACKS

Long-term „memory“

Interpretation at ecosystem level

Relation to functionality and stability of ecosystems

End-point with evolutionary background

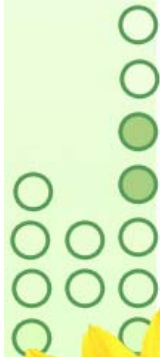
Time-consuming measurement

Problem with representative sampling

Dependent on experts (taxonomists, ...)

Long-term representative data are required for benchmarking

Diversity = variability: numerous masking effects



Biological communities as typical assessment endpoint at ecosystem level

DRAWBACKS

Time-consuming measurement

Problem with representative sampling

Dependent on experts (taxonomists, ...)

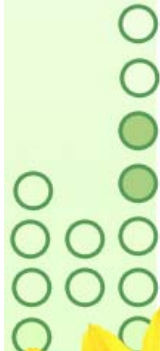
Long-term representative data are required for benchmarking

Diversity = variability:
numerous masking effects

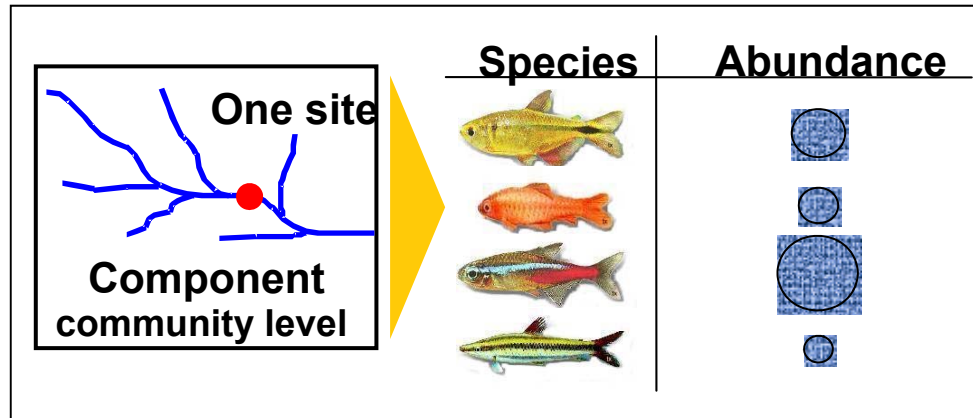
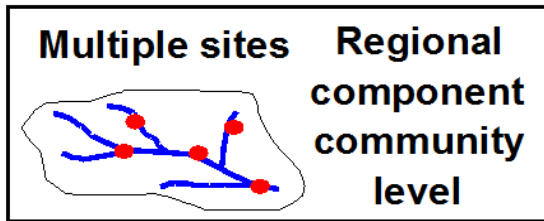
CONSEQUENCES IN BIOMONITORING

Trend to simplify design of studies or to reduce number of examined species

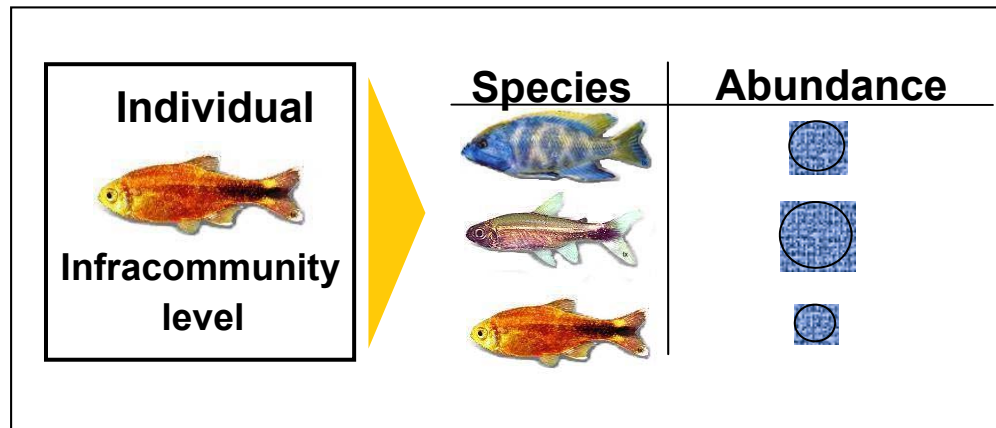
Searching for indicator species of components of communities, susceptible to stress



Biodiversity as end-point in environmental studies



Also objective stratification of communities strongly decreases sample size !!



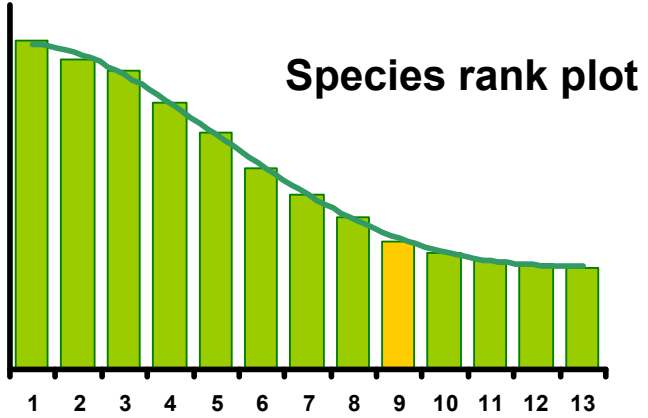
How to quantify and communicate complex assessment endpoint?

As aesthetic nature

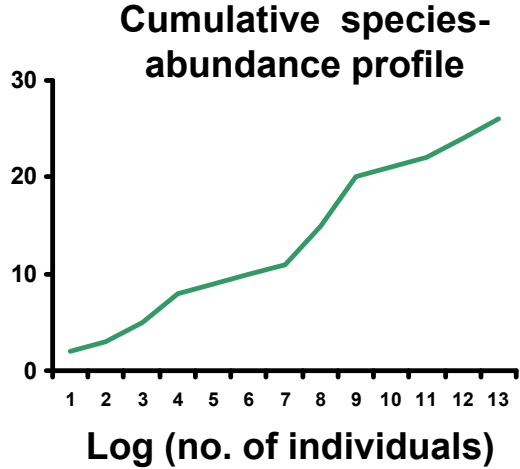
As parametrically standardized end-point



No. of individuals



Species rank



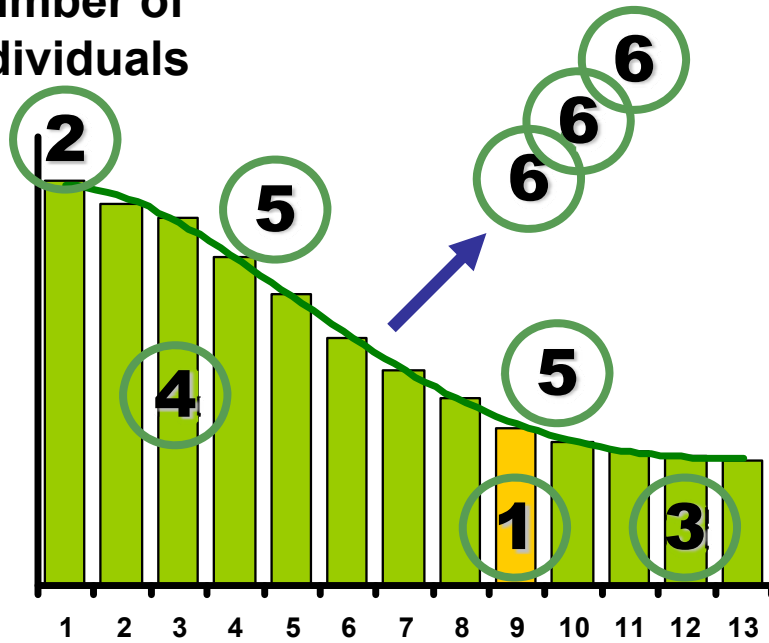
Log (no. of individuals)



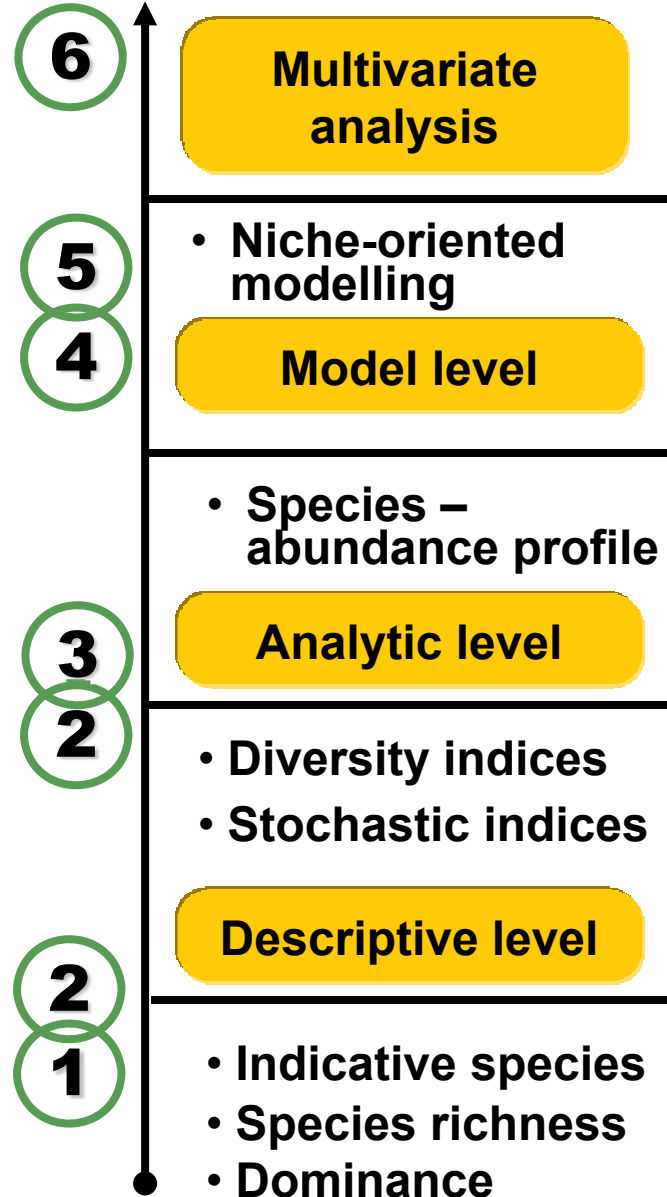
Complex endpoints typically offers redundant list of possible parameters

Species-abundance profile as standard output

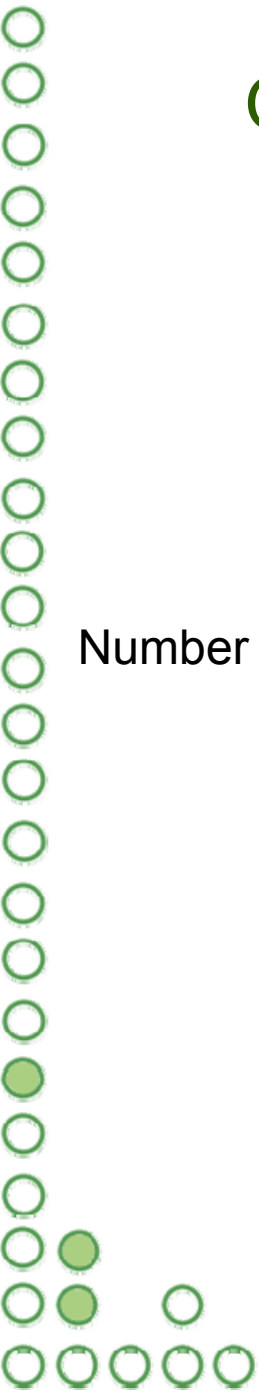
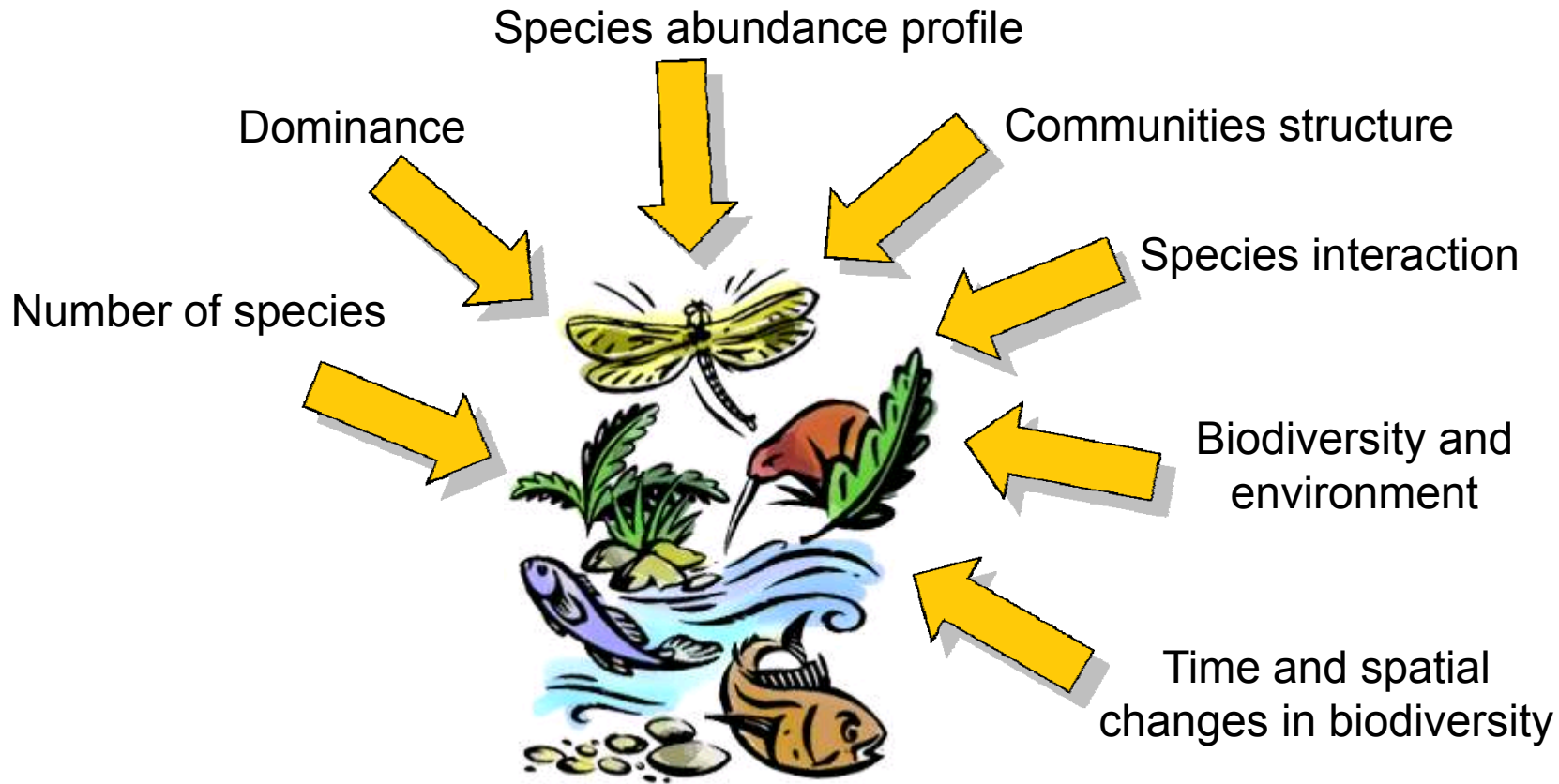
Number of individuals



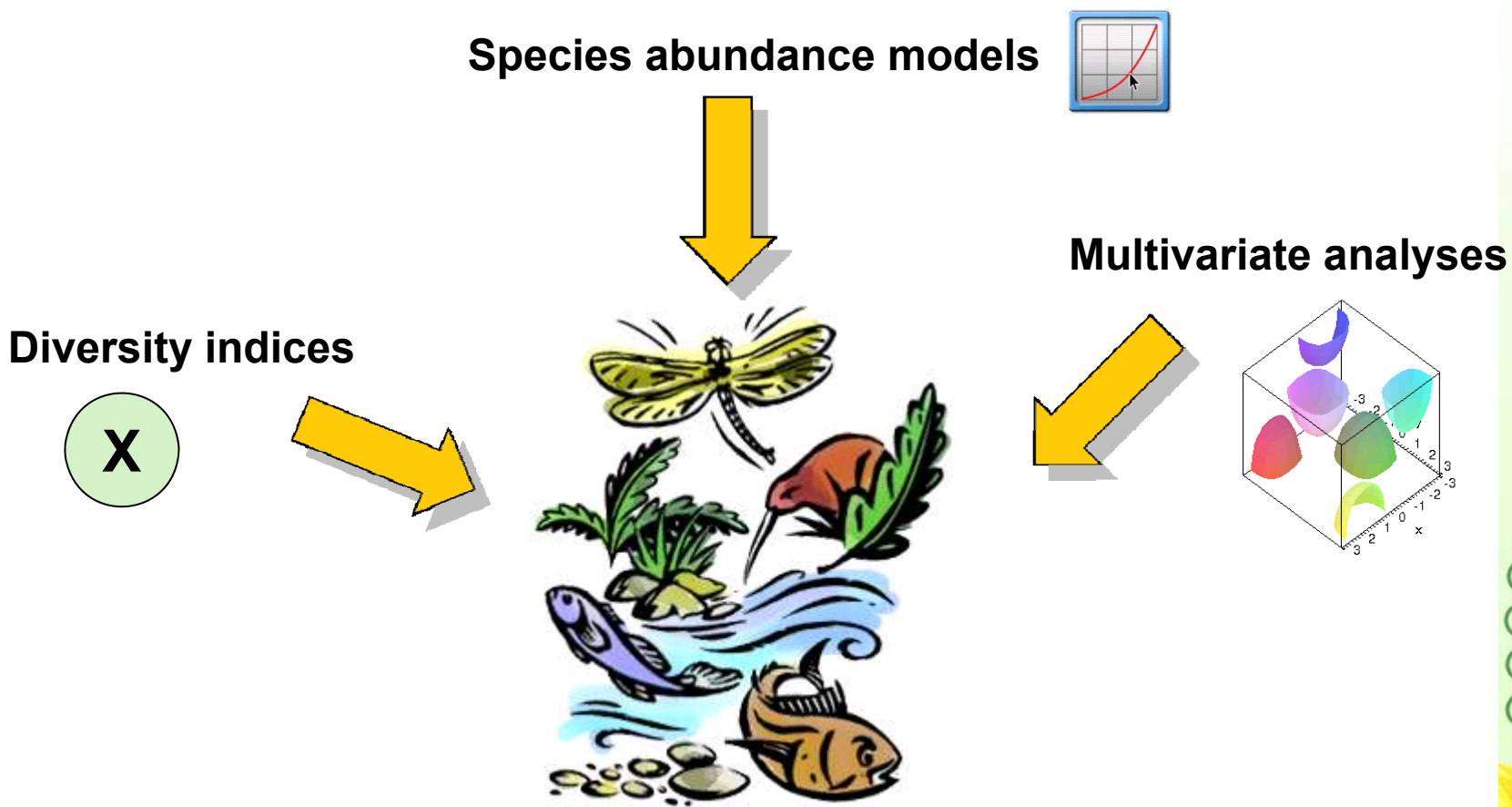
Species rank



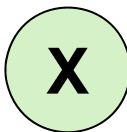
Questions in analysis of biodiversity data



Methods of biodiversity analysis



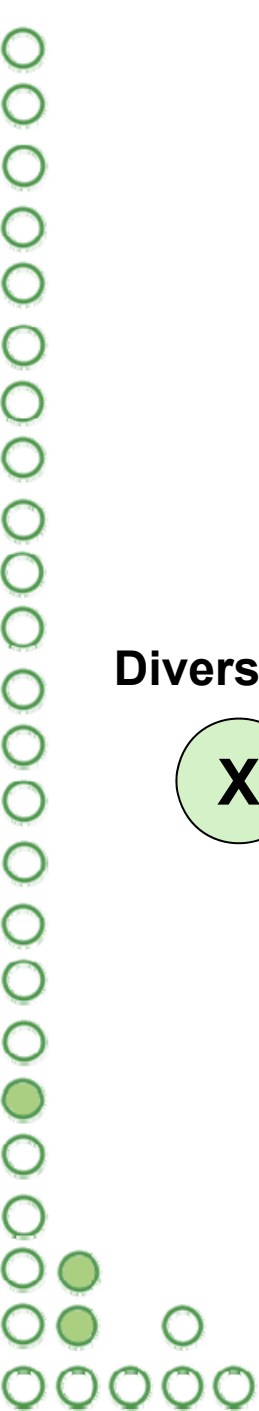
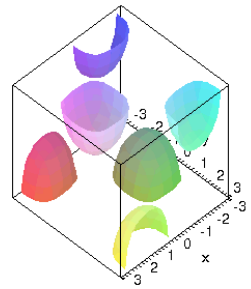
Diversity indices

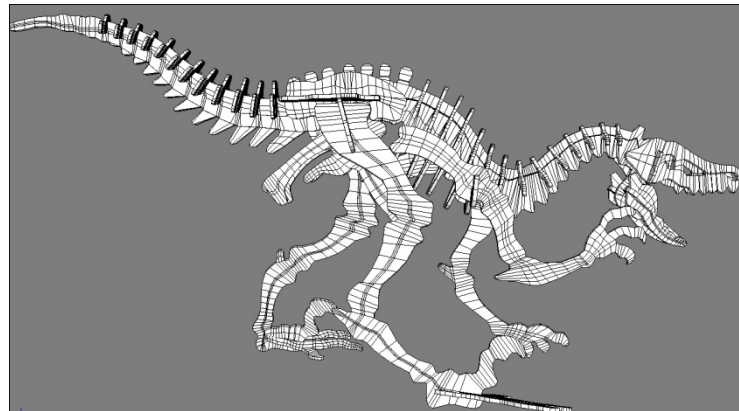


Species abundance models

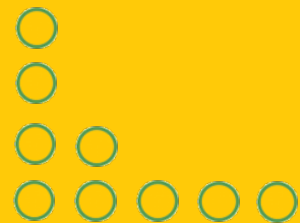


Multivariate analyses



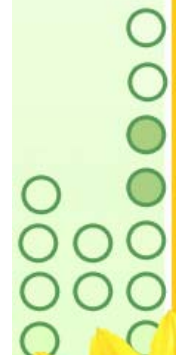


II. Biodiverzita a biostatistika



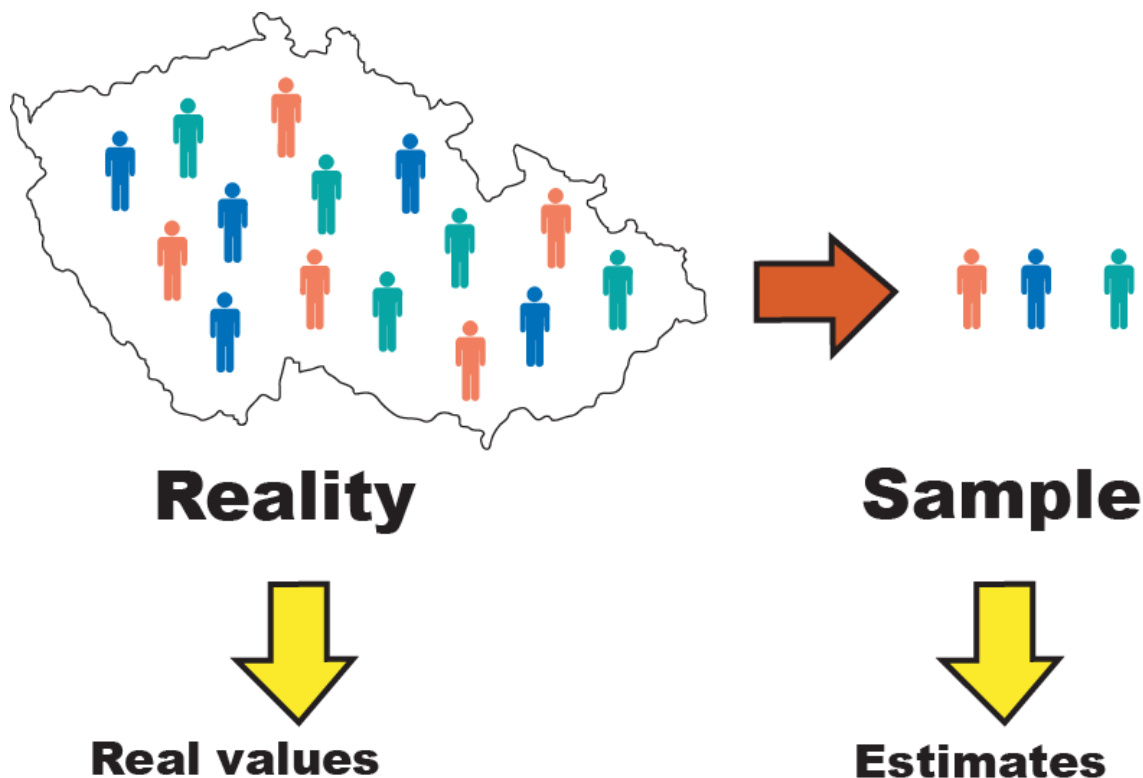
Vazba mezi biodiverzitou a biostatistikou?

- Co má společného analýza biodiverzity s klasickou biostatistikou?
 - Stejné základní principy
 - Vzorkování
 - Odhady
 - Testování významnosti
 - Analogie mezi „klasickou“ biostatistikou a analýzou biodiverzity
 - Popisná statistika
 - Modelová rozložení
 - Vícerozměrná analýza



Vzorkování biodiverzity

- Stejné teoretické předpoklady jako jakékoliv jiné vzorkování
 - Náhodné a reprezentativní vzorkování
 - Korektní design experimentu a opakovatelnost vzorkování

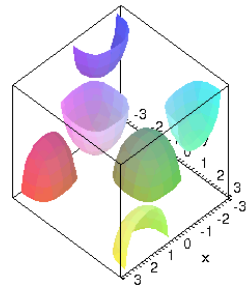
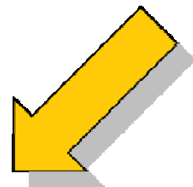


Metody analýzy biodiverzity

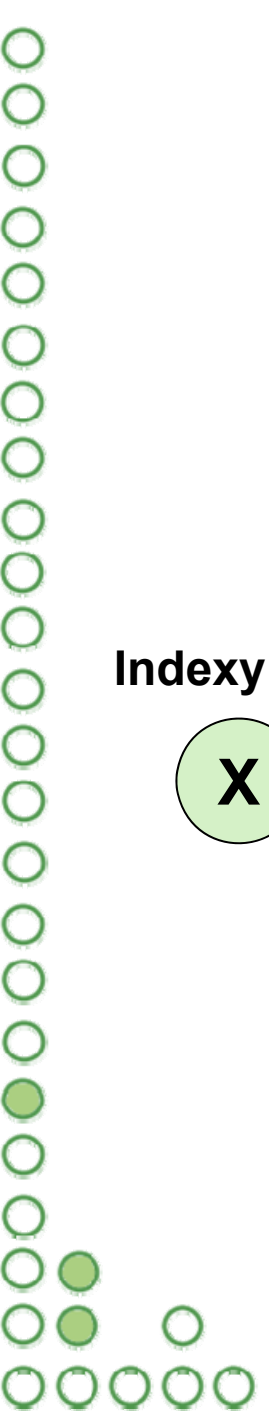
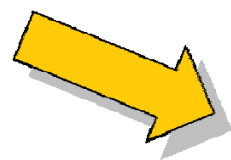
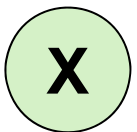
Species abundance modely



Vícerozměrná analýza



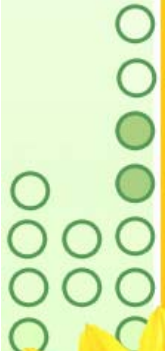
Indexy diverzity



Indexy diverzity: výhody a nevýhody

X

- Indexy diverzity je možné brát jako analogii k **popisné statistice**
- Celé společenstvo je agregováno **jediným číslem**, které reprezentuje **počet druhů a/nebo jejich dominanci ve společenstvu**
- Pro popisnou statistiku diverzity je možné získat intervaly spolehlivosti a dostupné jsou i statistické testy
- **Výhody:**
 - Měření diverzity v jediném čísle
- **Nevýhody:**
 - Redukce individuality taxonů
 - V některých případech nejasná interpretace (stejná hodnota indexu může být spočítána z velmi odlišných společenstev)



Species abundance models: výhody a nevýhody



- Mohou být pokládány za **analogii proložení statistického rozložení** v klasické biostatistice
- **Kvantitativní struktura** biologických společenstev (profil abundance taxonů) je **důsledkem ekologických procesů**
 - Modelový profil společenstva s ekologickou interpretací může být proložen s reálným profilem společenstva
- **Výhody:**
 - Analýza zahrnuje celý profil abundance ve společenstvu
 - Druhy jsou zastoupeny svým pořadím abundance (jde tedy o testování hypotézy vlivu ekologických procesů na kvantitativní strukturu společenstva)
- **Nevýhody:**
 - Ne zcela ujasněná metodika prokládání pozorovaných a očekávaných profilů abundance
 - Nedostatek ověřených informací o ekologických hypotézách spjatých s modely profilů abundance a jejich testování



Vícerozměrná analýza společenstev: výhody a nevýhody

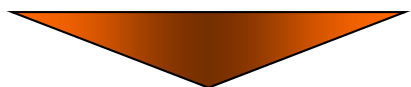


- Na data biodiverzity může být aplikována řada shlukovacích, ordinačních, regresních a klasifikačních vícerozměrných technik.
- Tyto metody hledají v rozsáhlých datech **vícerozměrné vzory společenstev** umožňující odpovědět na následující otázky:
 - Vztah druhů k prostředí
 - Prostorové vztahy
 - Interakce taxonů
- **Výhody:**
 - Shrnující výsledky postihující všechny aspekty dat
 - Identifikace skrytých interakcí a vztahů mezi proměnnými
- **Nevýhody:**
 - Náročné na data a metodiku
 - Vyžadují expertní znalosti jak v oblasti statistické metodiky, tak biologických společenstev, v opačném případě mohou vést k nesprávným závěrům a interpretacím

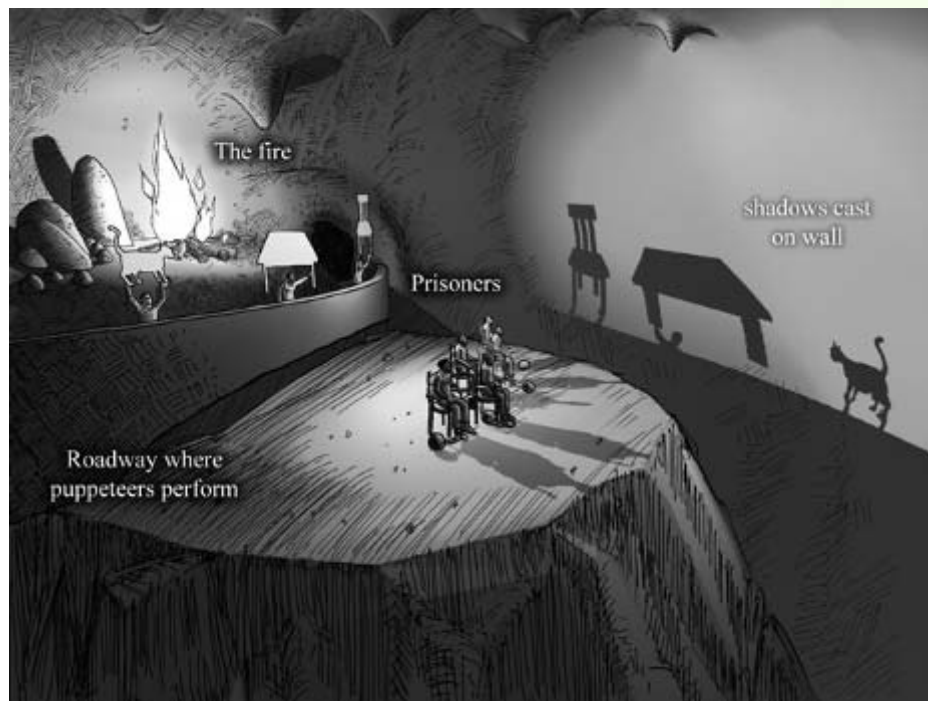


Výzkum, realita, statistika

- Výzkum je naším způsobem porozumění realitě
- Ale jak přesné a pravdivé je naše porozumění?

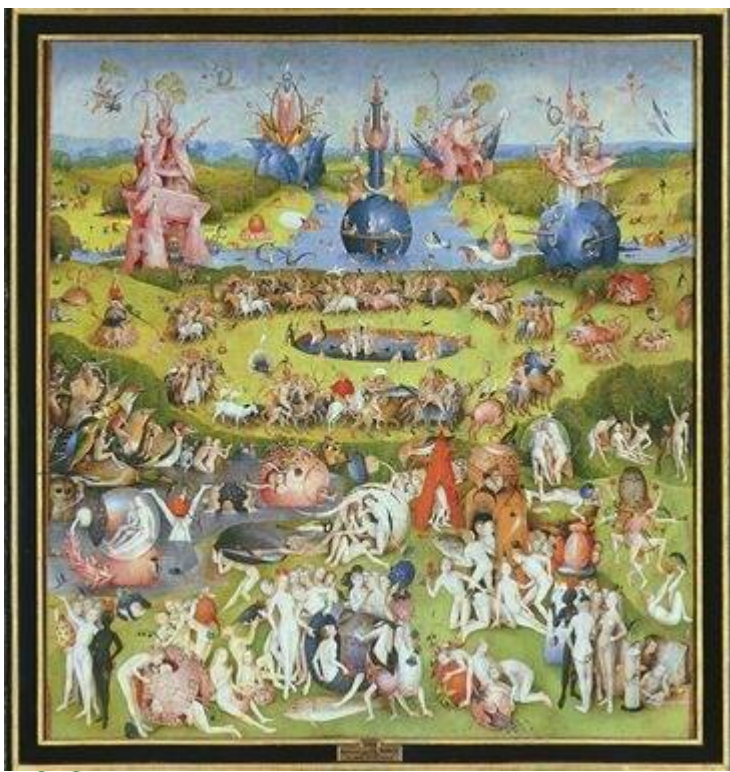


Statistika je jedním z nástrojů vnášejších do našich výsledků určitou spolehlivost.



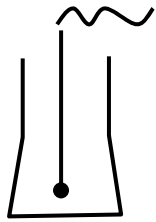
Význam variability

- Naše realita je variabilní a statistika je vědou zabývající se variabilitou
- Korektní analýza variabilita a její pochopení přináší užitečné informace o naší realitě
- V případě deterministického světa by statistická analýza nebyla potřebná



Biostatistika - různé přístupy k variabilitě

Variabilita opakovaných měření



Data

2,1
2,8
3,2
1,2
5,2
2,9

chyba

Variabilita znaku v populaci



165 cm



140 cm



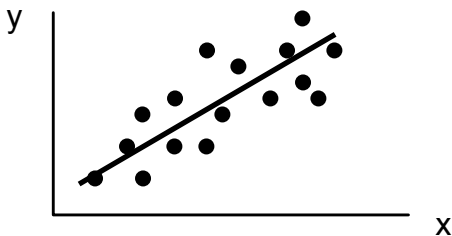
182 cm



163 cm

rozptyl znaku, přirozená variabilita

Variabilita modelovaných dat



chyba = nepřesnost modelu

Variabilita časových řad



čas

fluktuace, časová proměnlivost

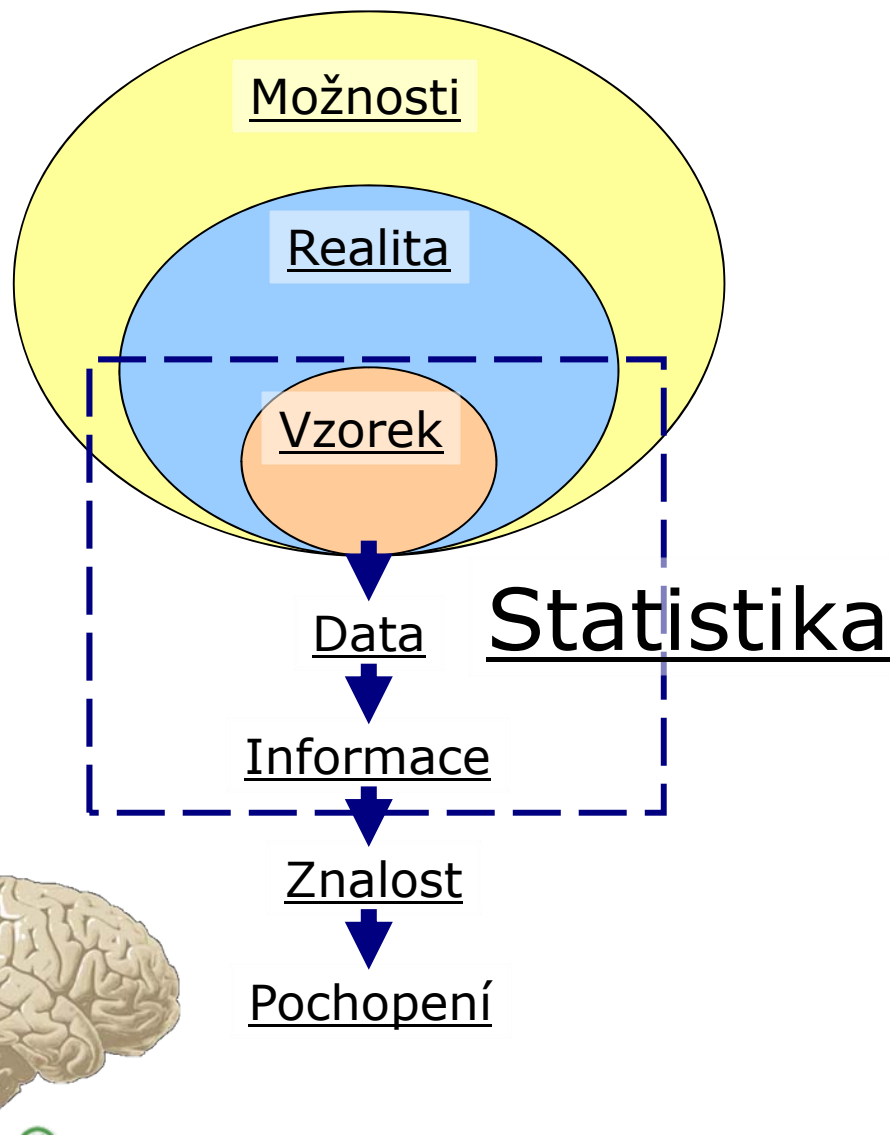
Variabilita ve skladbě biologických společenstev

DRUH 1	15
DRUH 2	30
DRUH 3	40
DRUH 4	14



biodiverzita

Co může statistika říci o naší realitě?



Statistika není schopna činit závěry o jevech neobsažených v našem vzorku.

Statistika je nasazena v procesu získání informací z vzorkovaných dat a je podporou v získání naší znalosti a pochopení problému.

Statistika není náhradou naší inteligence !!!

Cílová populace

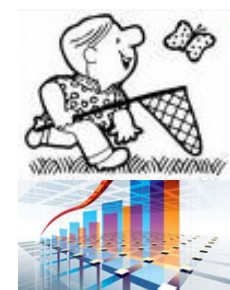
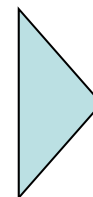
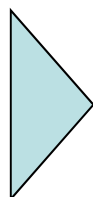
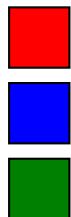
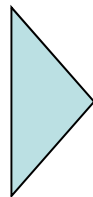
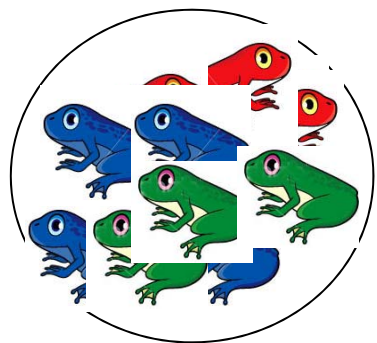
- **Cílová populace** – klíčový pojem statistického zpracování
 - Skupina objektů o nichž se chceme něco dozvědět (např. pacienti s danou diagnózou, všichni lidé nad 60 let, měření hemoglobinu v dané laboratoři)
 - Musí být definována ještě před zahájením sběru dat
 - Na cílové populaci probíhá vzorkování dat, které musí cílovou populaci dobře (reprezentativně) charakterizovat

Cílová populace

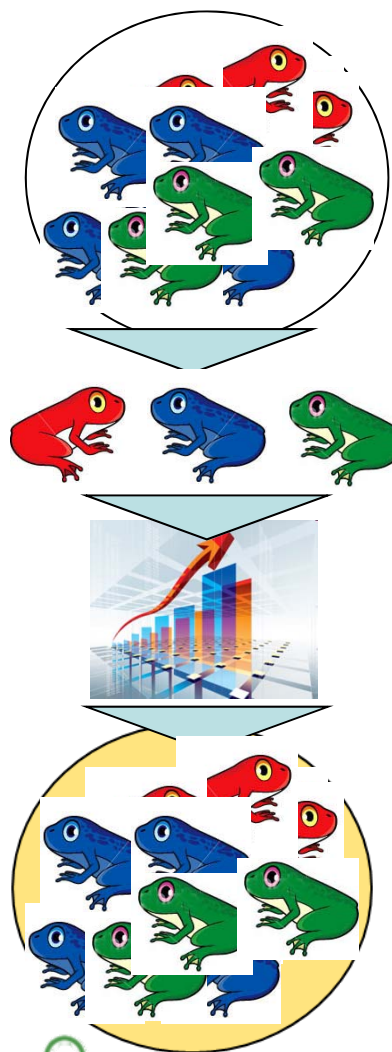
Klíčové faktory
cílové populace

Design
experimentu a
vzorkovací plán

Vzorkování a
analýza dat



Statistika a zobecnění výsledků



Neznámá cílová populace

Vzorek

Analýza

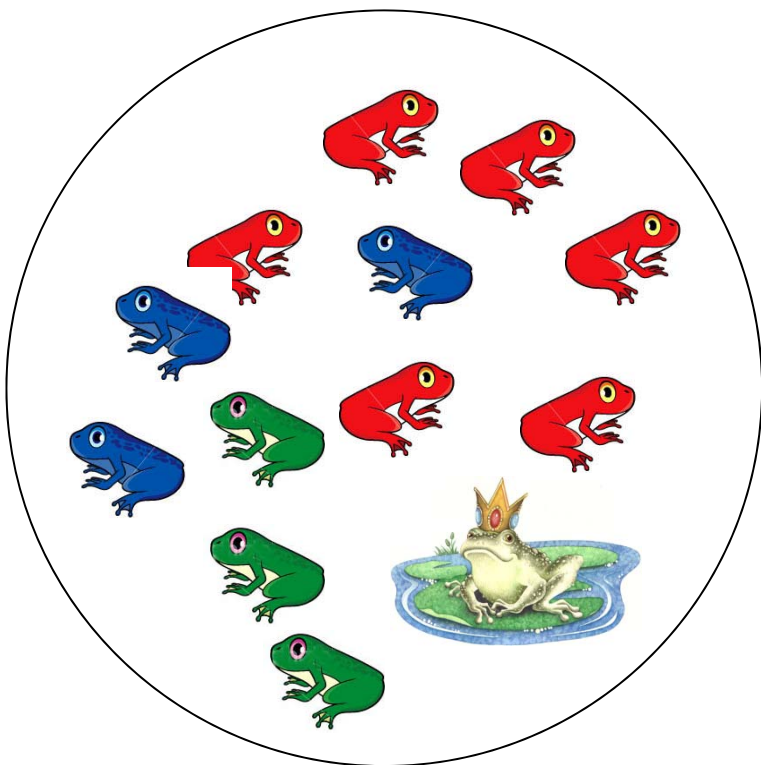
Díky zobecnění výsledků známe vlastnosti cílové populace

- Cílem analýzy není pouhý popis a analýza vzorku, ale zobecnění výsledků ze vzorku na jeho cílovou populaci
- Pokud vzorek nereprezentuje cílovou populaci, vede zobecnění k chybným závěrům



Vzorkování a jeho význam ve statistice

- Statistika hovoří o realitě prostřednictvím vzorku!!!
 - Statistické předpoklady korektního vzorkování



Representativnost: struktura vzorku musí maximálně reflektovat realitu

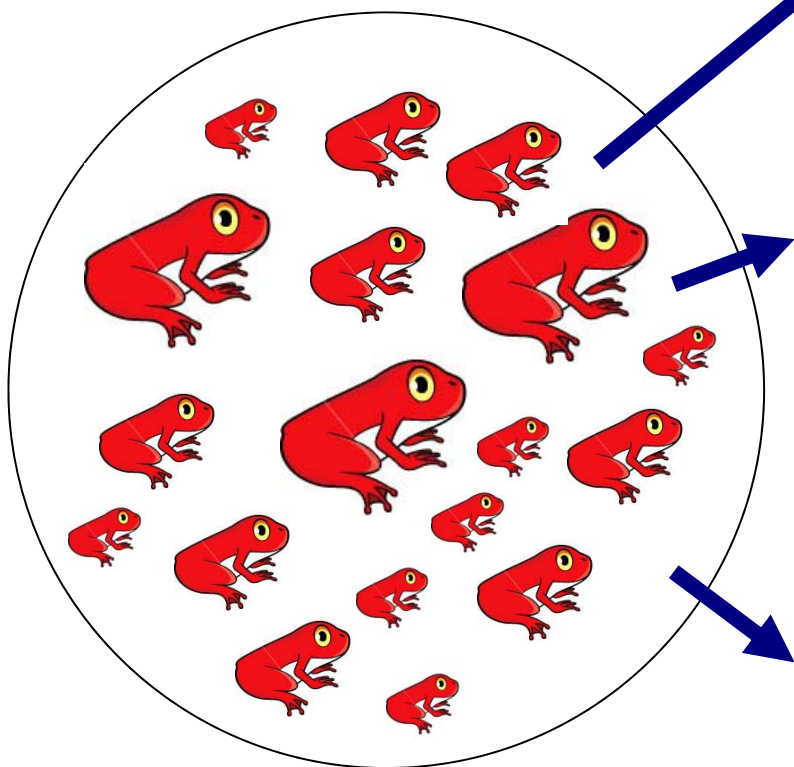


Nezávislost: několikanásobné vzorkování téhož objektu nepřináší ze statistického hlediska žádnou novou informaci



Velikost vzorku a přesnost statistických výstupů


Existuje skutečné rozložení
a skutečný průměr měřené
proměnné



Z jednoho měření nezjistíme nic

Vzorek:  → ?????

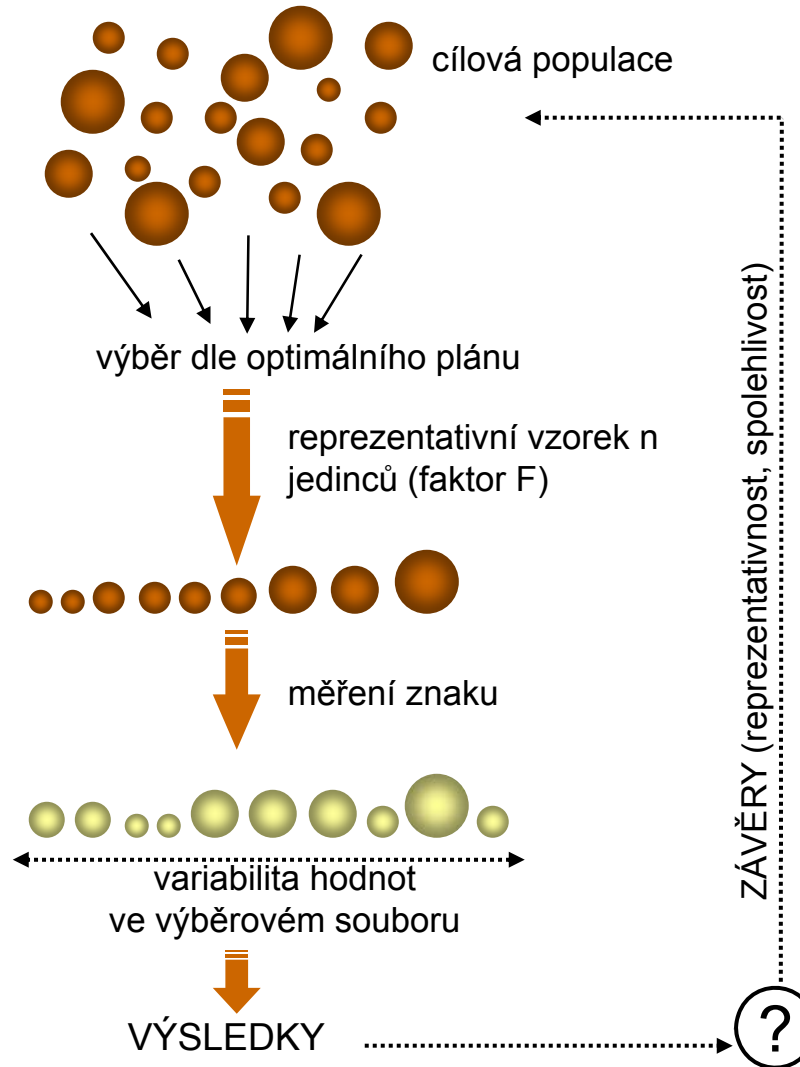
Vzorek určité velikosti poskytuje
odhad reálné hodnoty s definovanou
spolehlivostí

Vzorek:  → Odhad
průměru
atd.

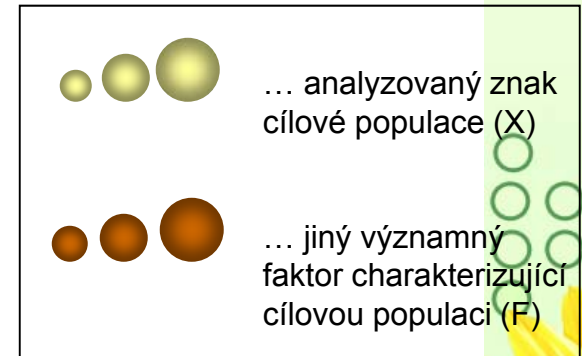
Vzorkování všech existujících
objektů poskytne skutečnou
hodnotu dané popisné statistiky,
nicméně tento přístup je ve většině
případech nereálný.

Experimentální design: nezbytná výbava biologa

Účel analýzy:
Popisný

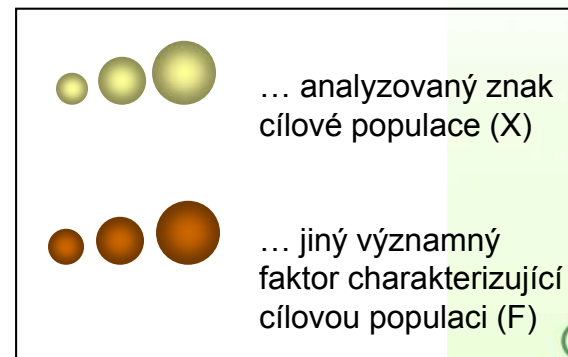
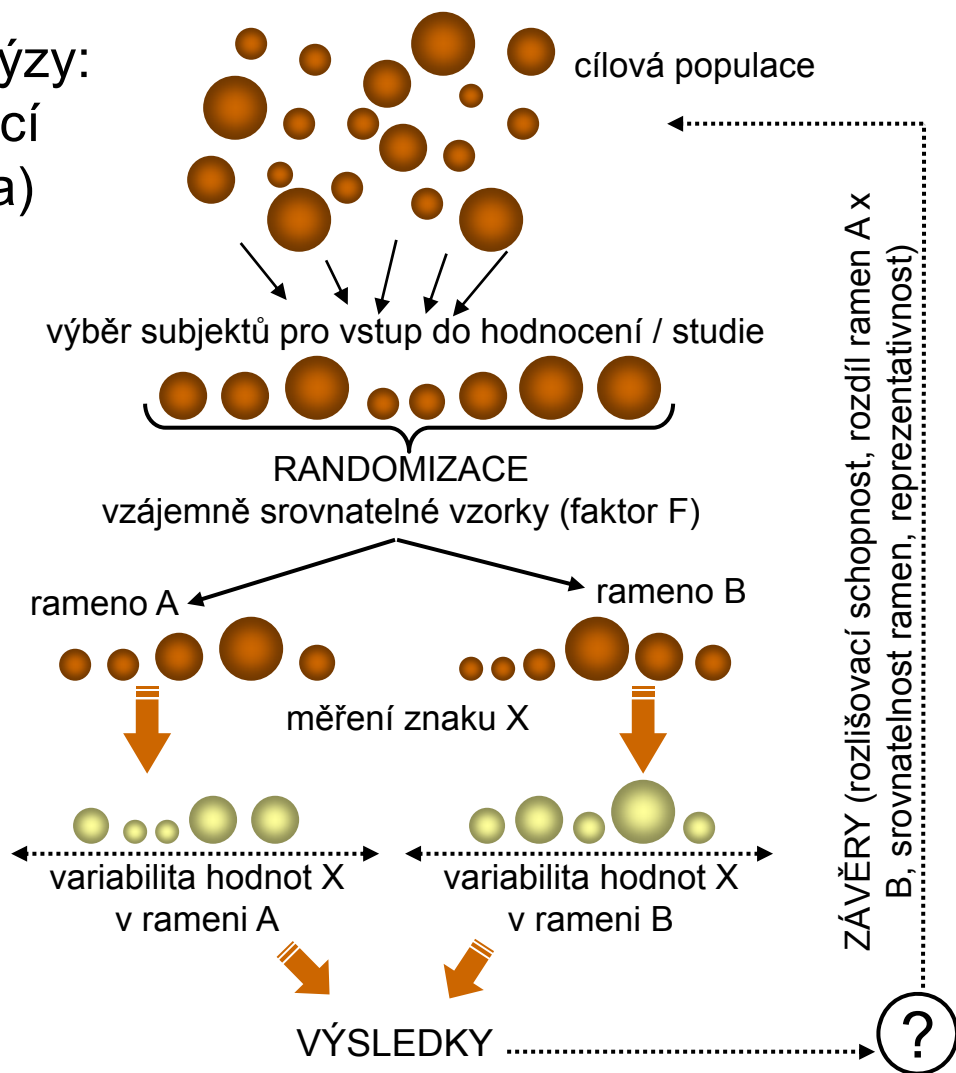


?
Reprezentativnost
Spolehlivost
Přesnost



Experimentální design: nezbytná výbava biologa

Účel analýzy:
Srovnávací
(2 ramena)



?
Srovnatelnost
Spolehlivost
Přesnost



Praktická a statistická významnost

- Samotná statistická významnost nemá žádný reálný význam, je pouze měřítkem náhodnosti hodnoceného jevu
- Pro vyhodnocení reálné významnosti je nezbytné znát i reálně významné hodnoty

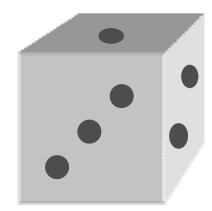
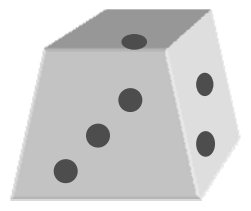
		Praktická významnost	
		ANO	NE
Statistická významnost	ANO	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr	Významný výsledek je statistický artefakt velkého vzorku, prakticky nevyužitelné
	NE	Výsledek může být pouhá náhoda, neprůkazný výsledek	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr



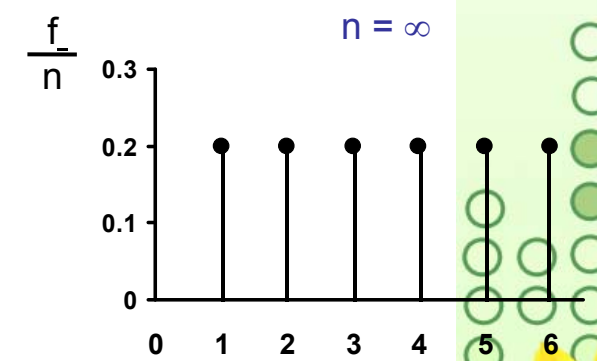
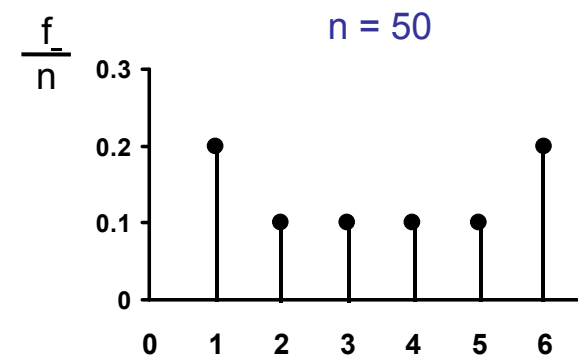
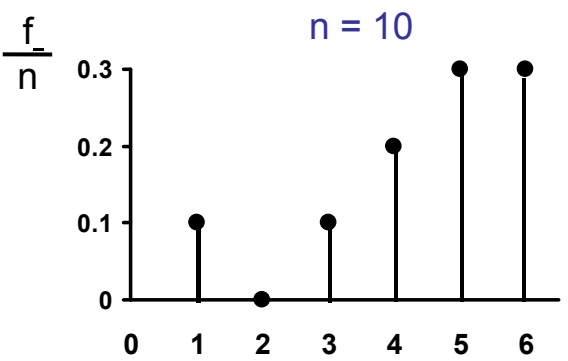
JAK vznikají informace ?

„Empirical approach“

„Classical approach“



Empirický postup



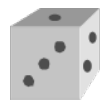
možné jevy: čísla 1 – 6

n – počet hodů (opakování)

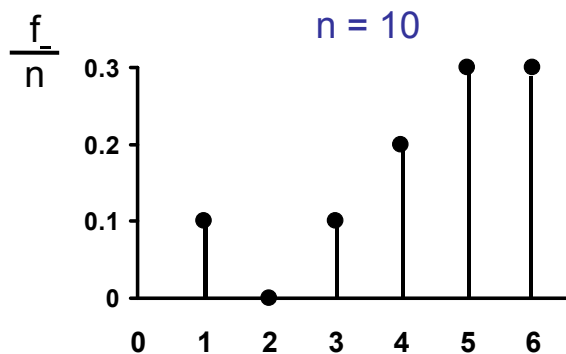
U složitých stochastických systémů se pravda získá až po odvedení značného množství experimentální práce: musíme dát systému šanci se projevit



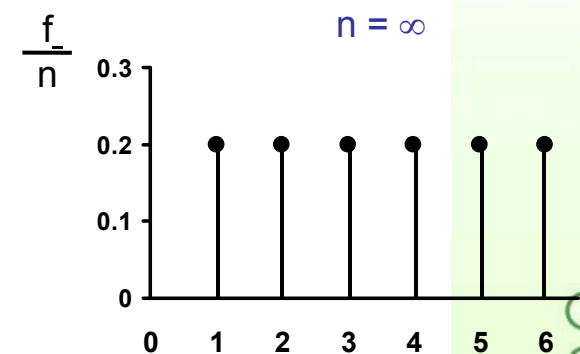
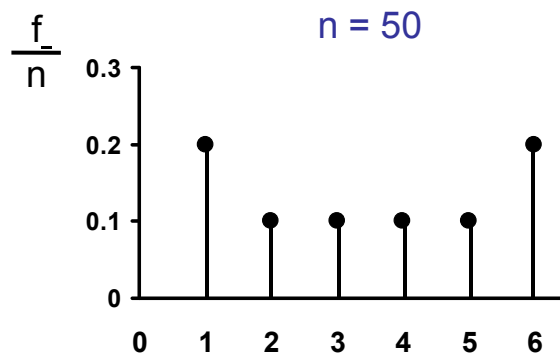
JAK vznikají informace ?



Empirický postup



možné jevy: čísla 1 – 6



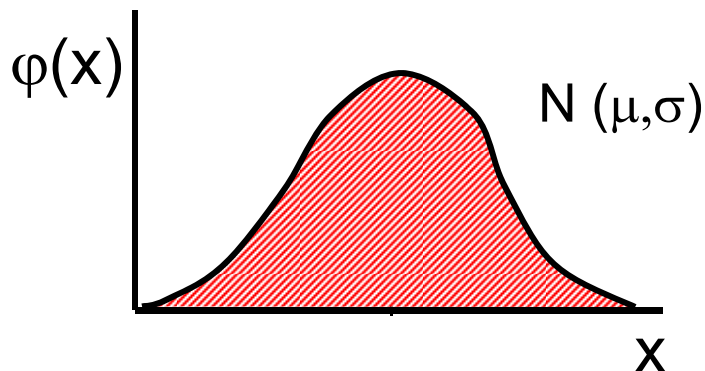
n – počet hodů (opakování)



Při realizaci náhodného experimentu roste se zvyšujícím se počtem opakování pravdivá znalost systému (výsledky se stávají stabilnější) diskutabilní je ale ovšem míra zobecnění konkrétního experimentu



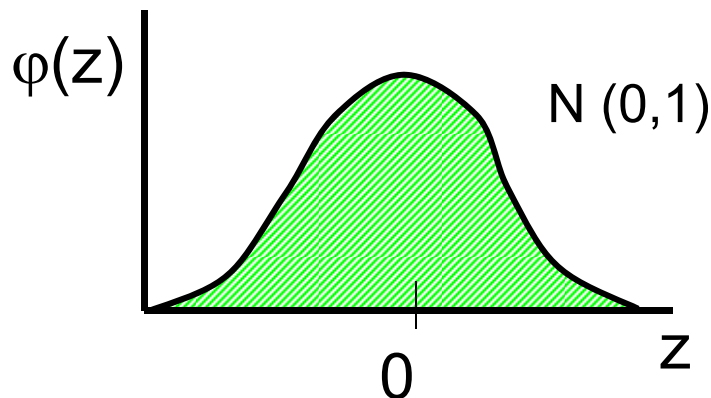
Rozložení hodnot jako model: Normální rozložení



$$\varphi(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$z = \frac{x - \mu}{\sigma}$$

Standardizovaná forma



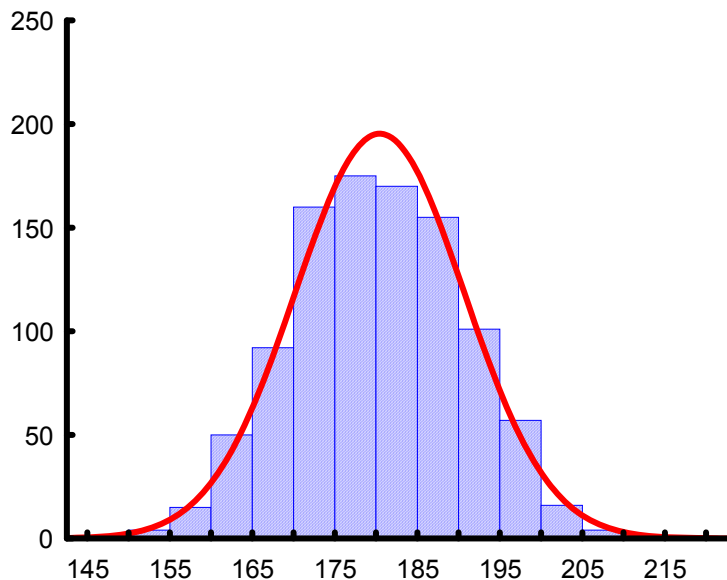
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

Tabelovaná podoba



Testy normality

- Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.



- **Test dobré shody**

V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí χ^2 testu dobré shody. Test dává dobré výsledky, ale je náročný na n , tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

- **Kolmogorov Smirnov test**

Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložení. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Měl by být počítán pouze v případě, že známe průměr a směrodatnou odchylku hypotetického rozložení, pokud tyto hodnoty neznáme, měla by být použita jeho modifikace – Lilieforsův test.

- **Shapiro-Wilk's test**

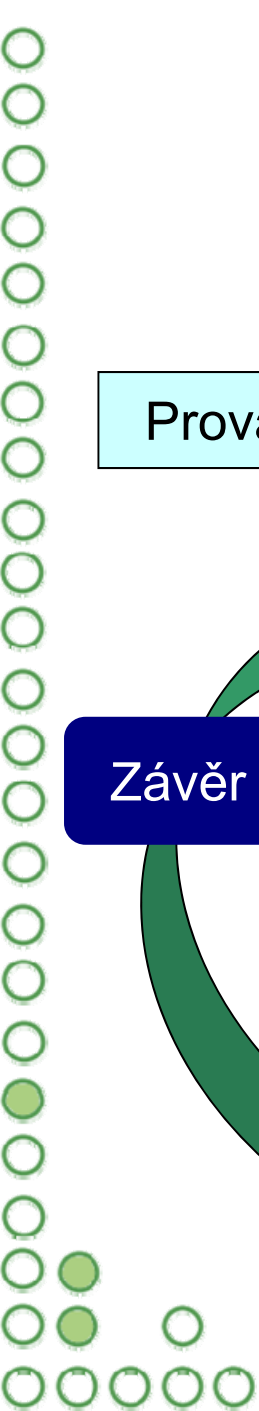
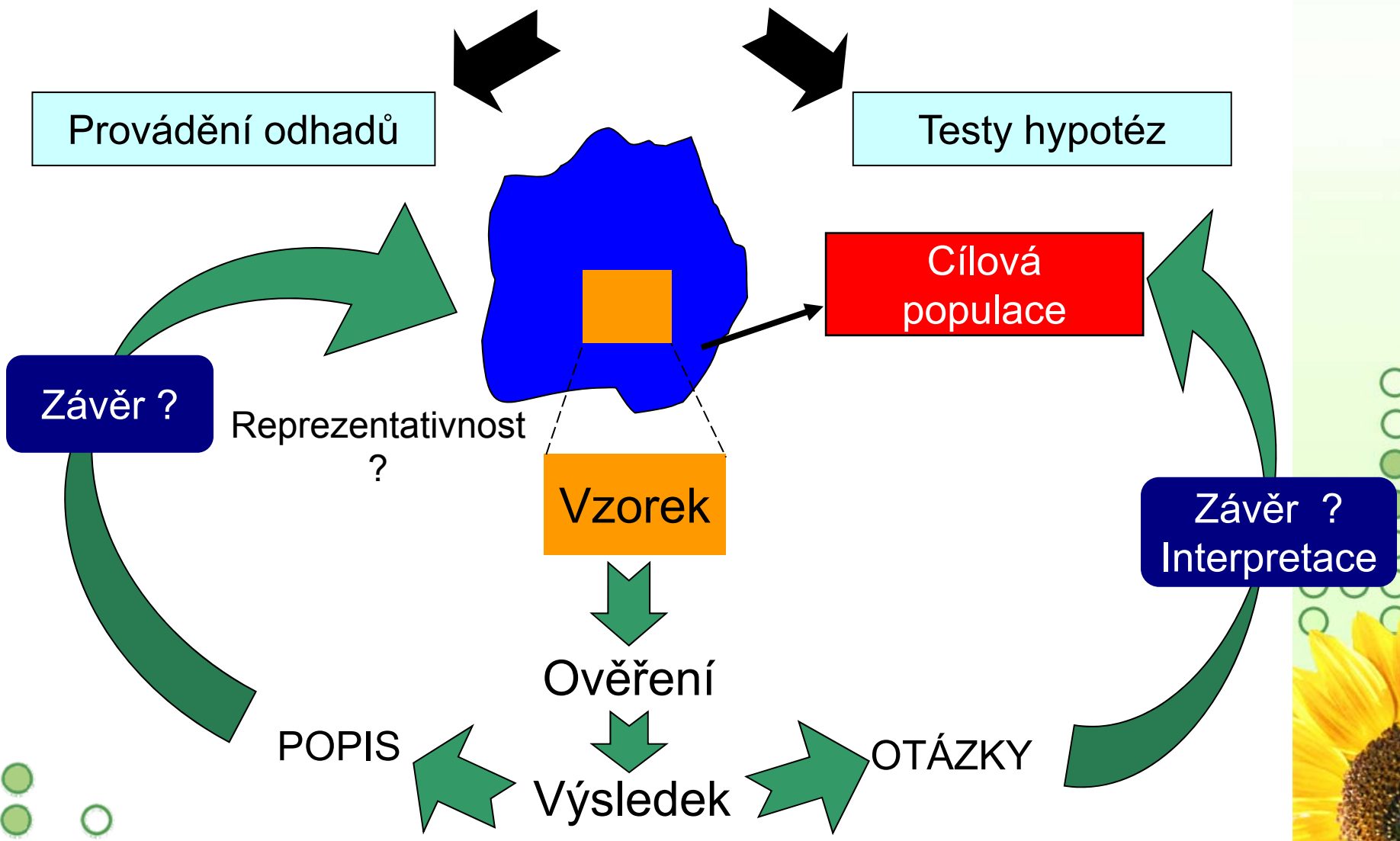
Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou silou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.



Stručný přehled modelových rozložení I.

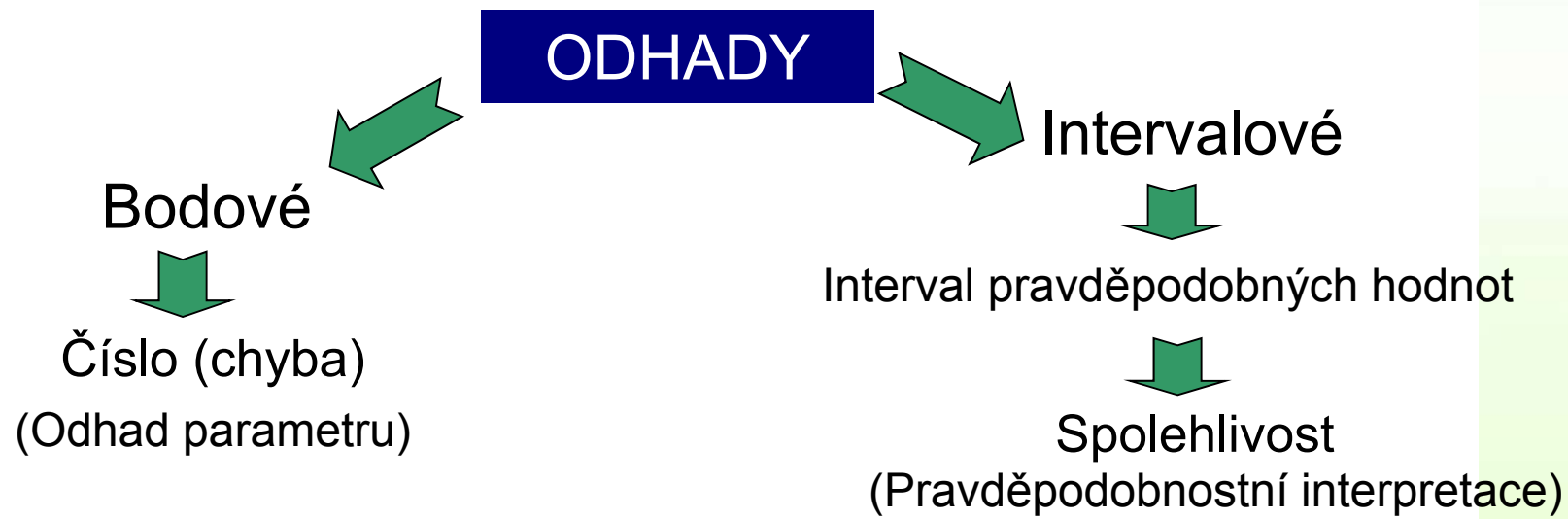
Rozložení	Parametry	Stručný popis
Normální	Průměr (μ) Rozptyl (σ^2)	Symetrická funkce popisující intervalovou hustotu četnosti; nejpravděpodobnější jsou průměrné hodnoty znaku v populaci.
Log-normální	Medián Geometrický průměr Rozptyl (σ^2)	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
Weibullovo	α - parametr tvaru β - parametr rozsahu hodnot	Změnou parametru a lze modelovat distribuci doby přežití, např. stresovaného organismu. Rozložení využívané i jako model k odhadu LC_{50} nebo EC_{50} u testů toxicity.
Rovnoměrné	Medián Geometrický průměr Rozptyl (σ^2)	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
Triangulární	$f(x) = [b - \text{ABS}(x - a)] / b^2$ $a - b < x < a + b$	Pravděpodobnostní funkce pro typ rozložení, kdy jsou střední hodnoty výrazně pravděpodobnější než hodnoty okrajové.
Gamma	Parametry distribuční funkce: α - parametr tvaru β - parametr rozsahu hodnot	Umožňuje flexibilně modelování distribučních funkcí nejrůznějších tvarů. Např. χ^2 rozložení je rozložení typu Gamma. Gamma rozložení s $a = 1$ je známo jako exponenciální rozložení.

Statistika v průzkumném studiu



INTERVAL SPOLEHLIVOSTI

velmi užitečná míra věrohodnosti odhadů



Obecný tvar:

$$P (L_1 < \text{Odhad} < L_2) \geq 1 - \alpha/2$$

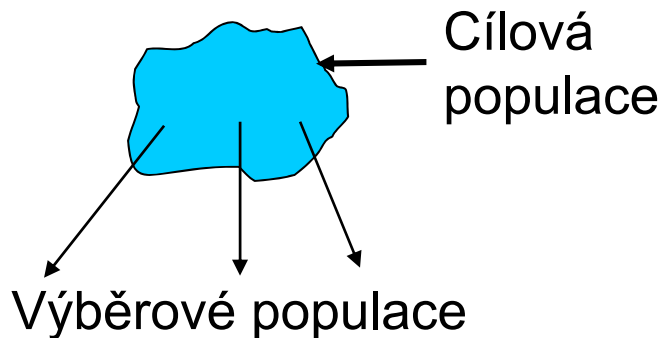
Odhadovaný parametr	\pm	Kvantil modelového rozložení	·	SE (odhadu)
---------------------	-------	------------------------------	---	-------------

K_V pro $(1 - \alpha/2)$



Interval spolehlivosti odhadu průměru je pouze informací o přesnosti tohoto odhadu

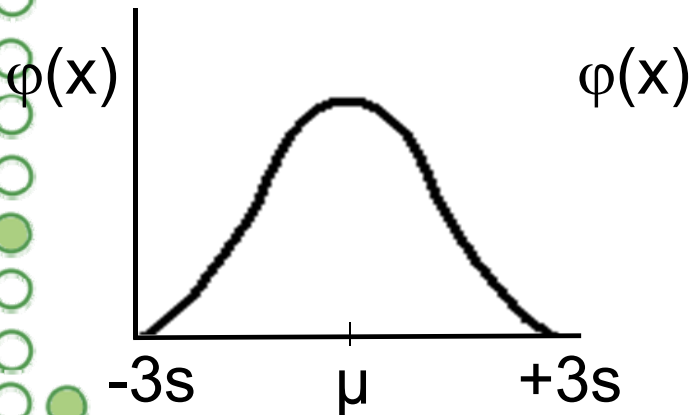
Interval spolehlivosti je hodnocen pro $(1 - \alpha)$ procentní spolehlivost



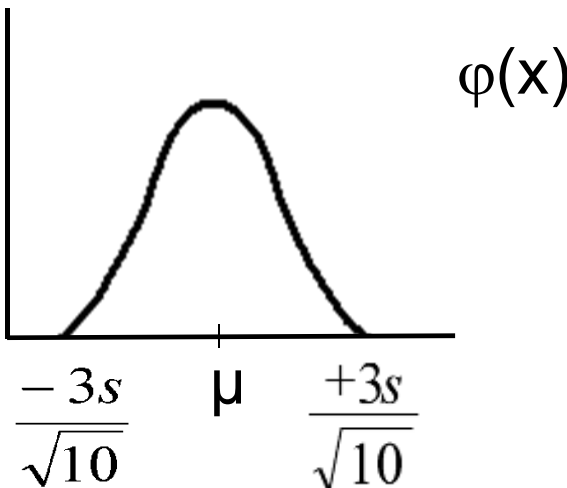
Šířku intervalu určuje:

- a) velikost vzorku
- b) rozptyl (variabilita) vzorku
- c) požadovaná spolehlivost

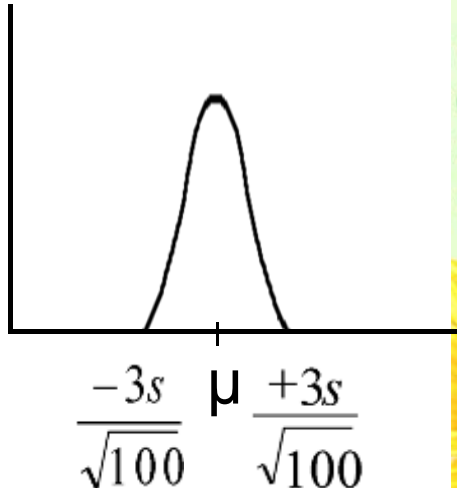
Původní proměnná x



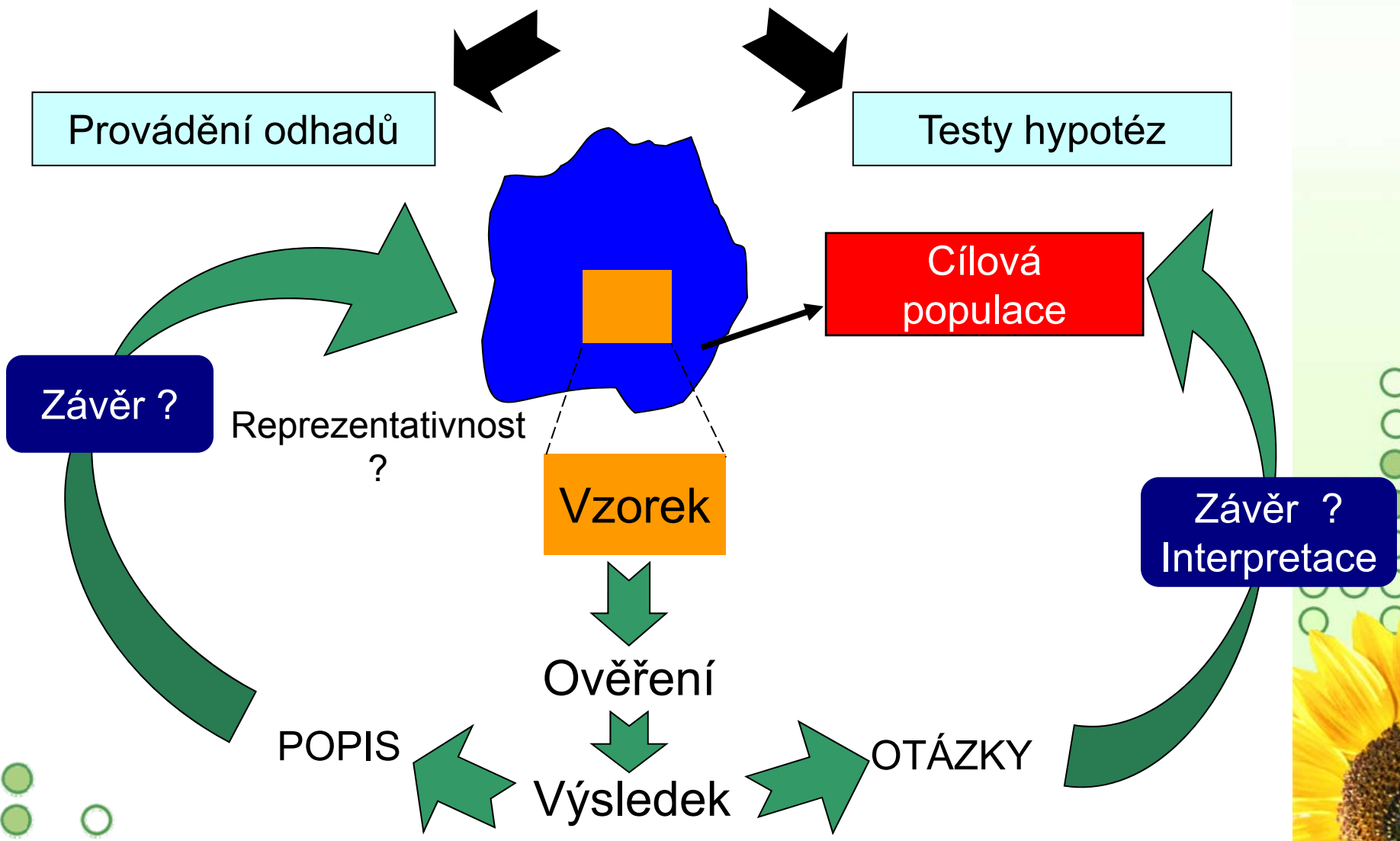
Výběr $n=10$ pro odhad průměru



Výběr $n=100$ pro odhad průměru

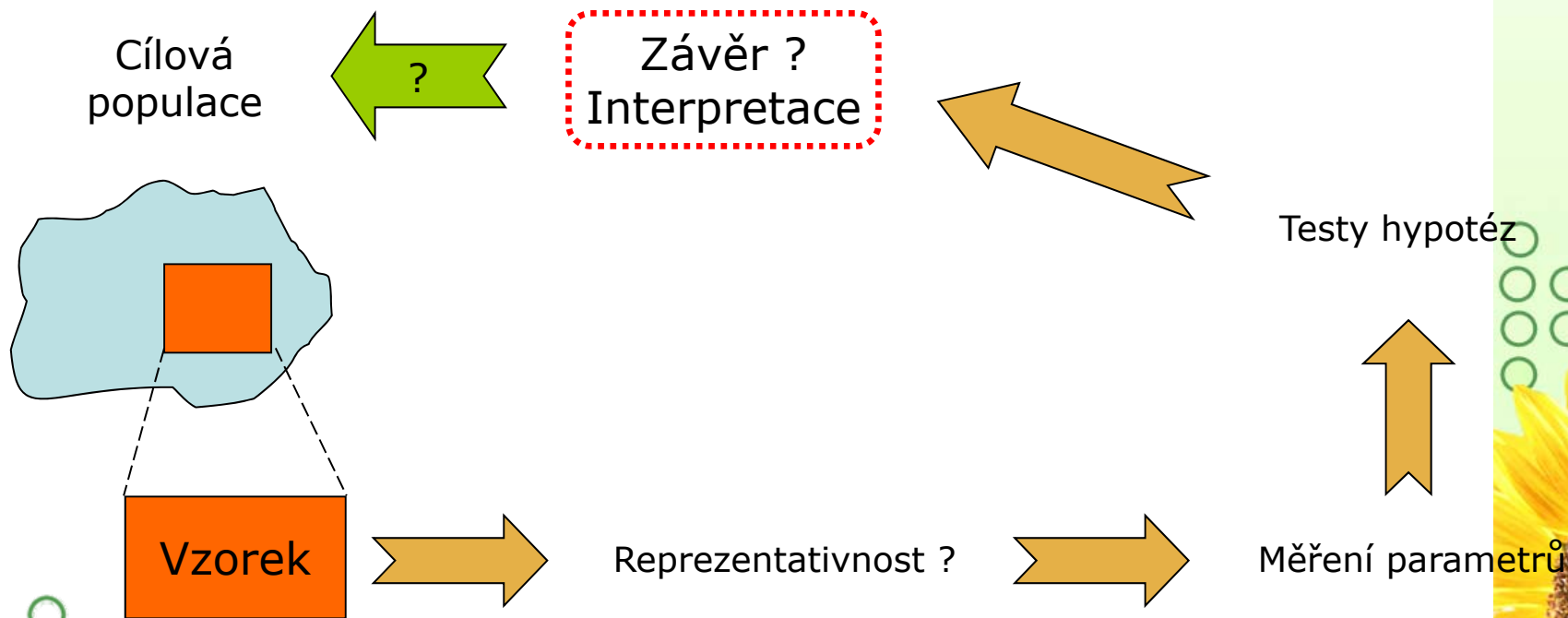


Statistika v průzkumném studiu



Princip testování hypotéz

- Formulace hypotézy
- Výběr cílové populace a z ní reprezentativního vzorku
- Měření sledovaných parametrů
- Použití odpovídajícího testu → závěr testu
- Interpretace výsledků



Statistické testování – základní pojmy

➤ Nulová hypotéza H_0

H_0 : sledovaný efekt je nulový

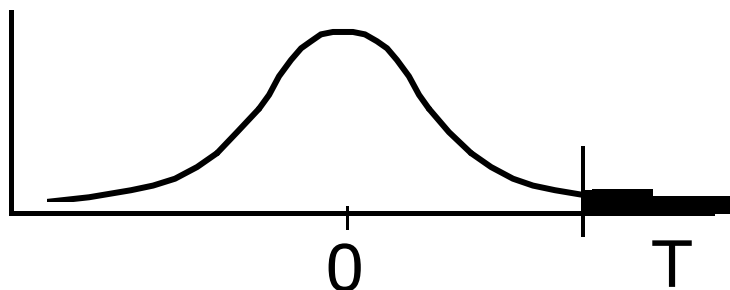
➤ Alternativní hypotéza H_A

H_A : sledovaný efekt je různý mezi skupinami

➤ Testová statistika

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Variabilita dat}} * \sqrt{\text{Velikost vzorku}}$$

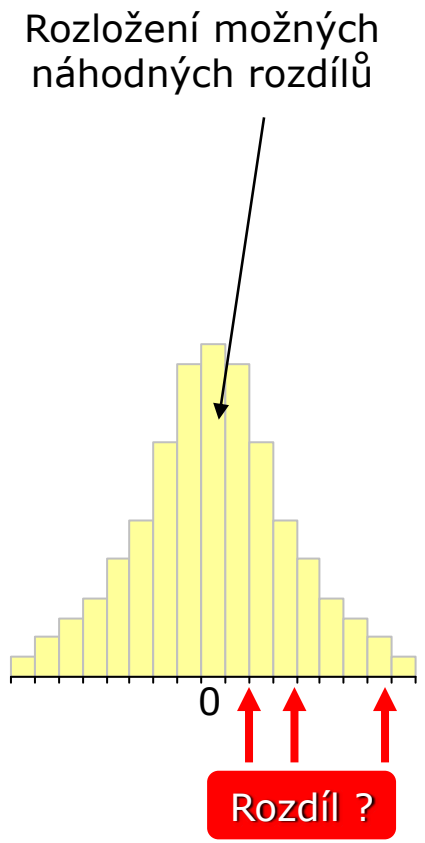
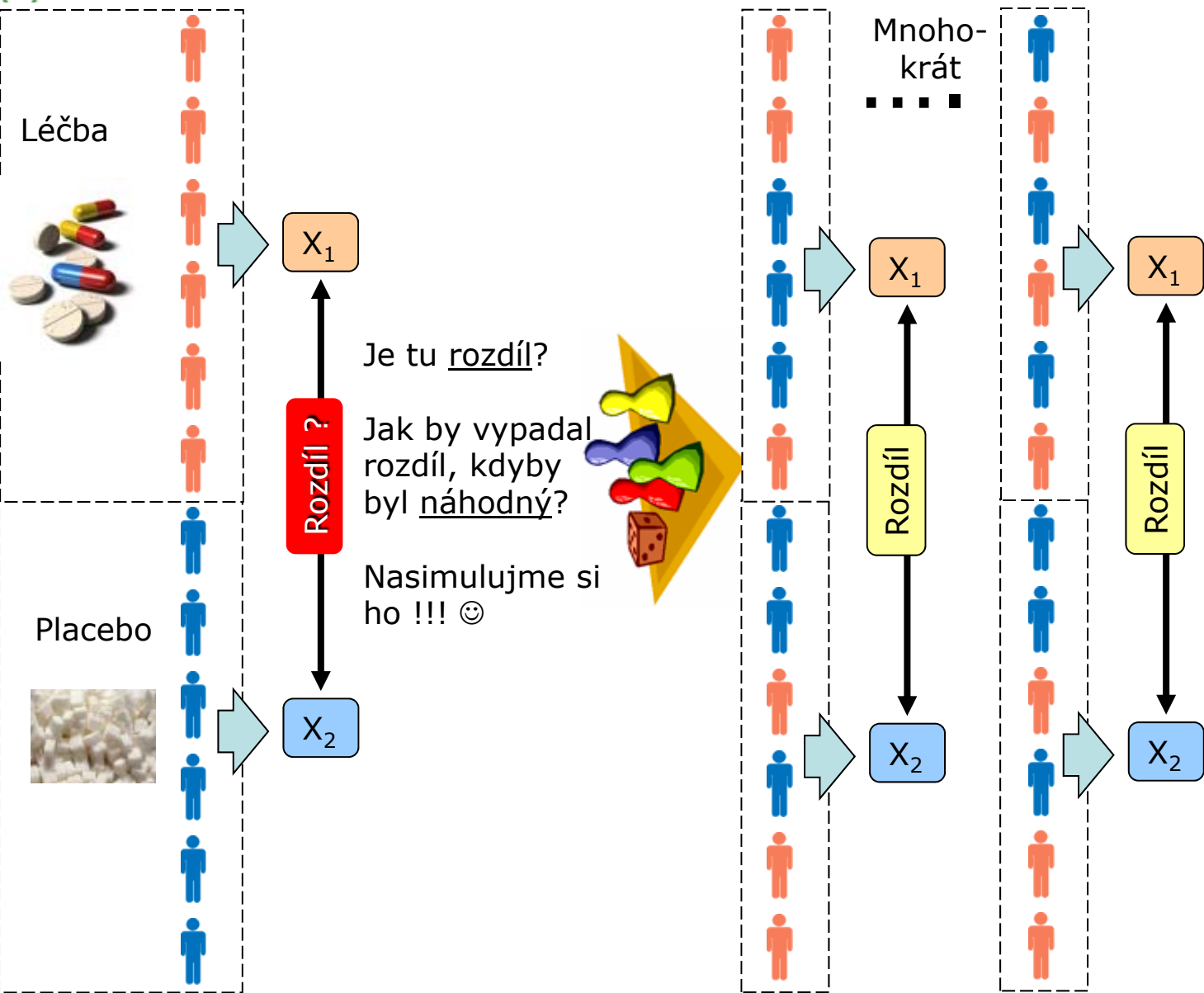
➤ Kritický obor testové statistiky



Statistické testování odpovídá na otázku zda je pozorovaný rozdíl náhodný či nikoliv. K odpovědi na otázku je využít statistický model – testová statistika.



Co znamená náhodný rozdíl?

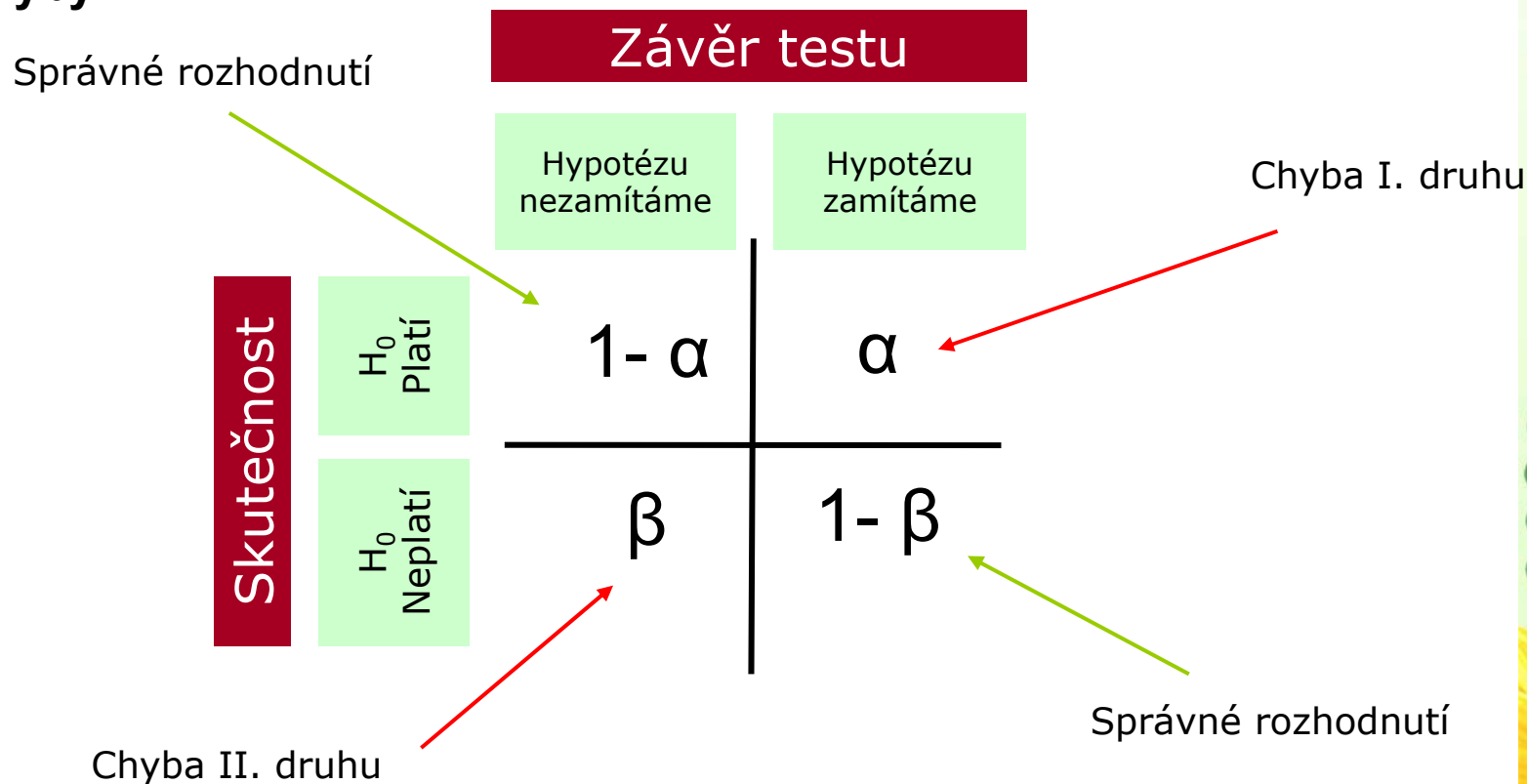


Kde leží skutečný rozdíl?

Jak moc je pravděpodobné, že je náhodný?

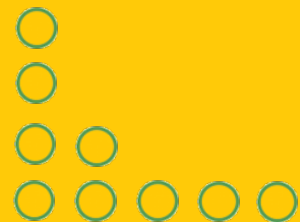
Možné chyby při testování hypotéz

- I přes dostatečnou velikost vzorku a kvalitní design experimentu se můžeme při rozhodnutí o zamítnutí/nezamítnutí nulové hypotézy dopustit chyby.





II-S1. Ukládání dat



DATA – ukázka uspořádání datového souboru

Parametry (znaky)



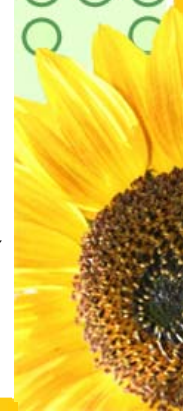
Opakování



Pacient	Clovek	aLeu cell.10 ⁶ /	aTy% %	aSe% %	aNeu% %	aLy% %	aTy cell.10 ⁶ /	aSe cell.10 ⁶ /	aNeu cell.10 ⁶ /	aLy cell.10 ⁶ /	aHtc %	aCLsk mV.s.10 ³	aCLNeus mV.s.10 ³	aCLOZ mV.s.10 ³	aCLNeuO mV.s.10 ³
3	1	4									33	72		32	
4	2	7,6	8	58	66	24	0,6	4,4	5,0	1,8	33	95	19	48	10
8	3	4	3	52	55	40	0,1	2,1	2,2	1,6	22	77	35	33	15
11	4	6,1	5	59	64	35	0,3	3,6	3,9	2,1	33	103	26	49	13
12	5	6,9	3	85	88	9	0,2	5,9	6,1	0,6	37	81	13	45	7
14	6	5,9	15	55	70	19	0,9	3,3	4,1	1,1	32	137	33	61	15
16	7	8	18	75	93	7	1,4	6,0	7,4	0,6	34	151	20	59	8
20	8	9,6	3	72	75	23	0,3	6,9	7,2	2,2	40	77	11	38	5
21	9	6	10	67	77	19	0,6	4,0	4,6	1,1	32	120	26	52	11
22	10	3,3	4	55	59	39	0,1	1,8	2,0	1,3	28	81	42	24	12
37	11	3,8	10	60	70	30	0,4	2,3	2,7	1,1	32	111	42	29	11
38	12	6,4	2	76	78	17	0,1	4,9	5,0	1,1	25	366	73	115	23
39	13	6,8	1	57	58	39	0,1	3,9	3,9	2,7	20	234	59	71	18
49	14	8,5	7	67	74	26	0,6	5,7	6,3	2,2	30	156	25	108	17
51	15	9,3	7	57	64	35	0,7	5,3	6,0	3,3	35	129	21	23	4
52	16	2,2	10	56	66	34	0,2	1,2	1,5	0,7	33	46	30	12	8
55	17	9,9	3	78	81	10	0,3	7,7	8,0	0,1	30	189	24	140	18
56	18	5	2	80	82	13	0,1	4,0	4,1	0,7	26	101	25	54	13
6	1	8,8	11	72	83	12	1,0	6,3	7,3	1,1	44	268	36,6	145	19,9
9	2	9,2	2	66	68	28	0,2	6,1	6,3	2,6	42	168	26,9	76	12,2
13	3	10,0	7	83	90	8	0,7	8,3	9,0	0,8	54	181	20,1	81	9
15	4	9,6	1	75	76	23	0,1	7,2	7,3	2,2	45	343	47	124	16,9
17	5	6,0									45	40		21	
19	6	7,2	2	78	80	18	0,1	5,6	5,8	1,3	44	103	17,8	63	10,9
24	7	8,2	1	72	73	25	0,1	5,9	6,0	2,1	41	209	34,9	57	9,6
26	8	10,3	1	85	86	3	0,1	8,8	8,9	0,3	41	364	41,1	112	12,6
29	9	5,0	1	74	75	21	0,1	3,7	3,8	1,1	39	83	22,1	32	8,5
30	10	11,9	1	51	52	47	0,1	6,1	6,2	5,6	33	83	13,4	52	8,4
31	11	7,2	3	53	56	29	0,2	3,8	4,0	2,1	28	109	27,1	63	15,5
32	12	10,8	36	50	76	8	3,9	5,4	9,3	0,9	27	146	15,7	106	11,4
33	13	11,8	22	54	76	16	2,6	6,4	9,0	1,9	45	246	27,4	63	7
34	14	17,0	1	82	83	16	0,2	13,9	14,1	2,7	34	440	31,2	119	8,4
40	15	10,0	8	72	80	4	0,8	7,2	8,0	0,4	37	176	22,0	52	6,5




Zásady pro ukládání dat

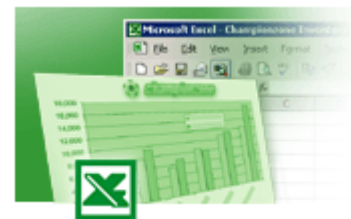
- Správné a přehledné uložení dat je základem jejich pozdější analýzy
- Je vhodné rozmyslet si předem jak budou data ukládána
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky
 - Každý sloupec obsahuje pouze jediný typ dat, identifikovaný hlavičkou sloupce
 - Každý řádek obsahuje minimální jednotku dat (např. pacient, jedna návštěva pacienta apod.)
 - Je nepřípustné kombinovat v jednom sloupci číselné a textové hodnoty
 - Komentáře jsou uloženy v samostatných sloupcích
 - U textových dat nezbytné kontrolovat překlepy v názvech kategorií
 - Specifickým typem dat jsou datумы u nichž je nezbytné kontrolovat, zda jsou datумы uloženy v korektním formátu
- Takto uspořádaná data je v tabulkových nebo databázových programech možné převést na libovolnou výstupní tabulku
- Pro základní uložení a čištění dat menšího rozsahu je možné využít aplikaci MS Office







Ukládání dat v MS Office

• MS Excel

-  Kontingenční tabulky – rychlá sumarizace rozsáhlých tabulek
-  Možnost výpočtů a grafových výstupů přímo v aplikaci
-  Visual Basic – složitější aplikace
- Omezení tabulky na 256x65536 buněk (do verze 2003)
- Omezená kontrola chyb při zadávání



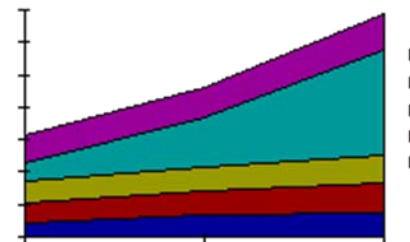
• MS Access

-  Plnohodnotná databáze vhodná pro velké množství dat, řádky omezeny v podstatě jen dostupnou pamětí
-  Kontrola typu dat
-  Relace tabulek – omezení velikosti souboru
-  Visual Basic a formuláře – složitější aplikace
- Omezení tabulky na 255 sloupců
- Výpočty a grafy jsou složitější než v Excelu



Možnosti MS Excel

- Správa a práce s tabulárními daty
- Řazení dat, výběry z dat, přehledy dat
- Formátování a přehledné zobrazení dat
- Zobrazení dat ve formě grafů
- Různé druhy výpočtů pomocí zabudovaných funkcí
- Tvorba tiskových sestav
- Makra – zautomatizování častých činností
- Tvorba aplikací (Visual Basic for Applications)



Počet z	Délka	Pohlaví
1	2	
2		
3		
4		
5		
6		
7	26	
8	106	
9	121	
10	160	
11	34	
12	45	
13	70	
14	72	
15	87	
16	Celkový součet	
17		

18		
19	10	2
20	12	3
21	5	4
22	8	5
23	4	8
24	7	9
25	9	11
26	suma součtinů řádků	310



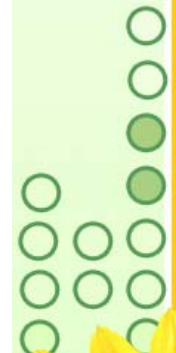
Import a export dat

- **Import dat**

- Manuální zadávání
- import – podpora importu ze starších verzí Excelu, textových souborů, databází apod.
- kopírování přes schránku Windows – vkládání z nejrůznějších aplikací – MS Office, Statistica atd.
- využití textových souborů jako kompatibilního formátu pro přenos dat mezi různými aplikacemi

- **Export dat**

- Ukládáním souborů ve formátech podporovaných jinými SW, časté jsou textové soubory, dbf soubory nebo starší verze Excelu
- Přímé kopírování přes schránku Windows



Tipy a triky

• **Výběr buněk**

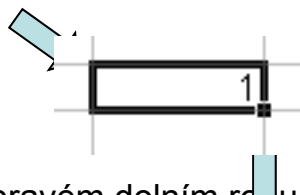
- CTRL+A – výběr celého listu
- CTRL + klepnutí myší do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myší na jinou buňku – výběr bloku buněk
- SHIFT + šipky – výběr sousedních buněk ve směru šipky
- SHIFT+CTRL+END (HOME) – výběr do konce (začátku) oblasti dat v listu
- SHIFT+CTRL+šipky – výběr souvislého řádku nebo sloupce buněk
- SHIFT + klepnutí na objekty – výběr více objektů

• **Kopírování a vkládání**

- CTRL+C – zkopírování označené oblasti buněk
- CTRL+V – vložení obsahu schránky – oblast buněk, objekt, data z jiné aplikace

• **Myš a okraje buňky**

- Chycení myší za okraj umožňuje přesun buňky nebo bloku buněk

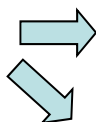


- Při chycení čtverečku v pravém dolním rohu výběru je tažením možno vyplnit více buněk hodnotami původní buňky (ve vzorcích se mění relativní odkazy, je také možné vyplnění hodnotami ze seznamu – např. po sobě jdoucí názvy měsíců).

Databázová struktura dat v Excelu

Sloupce tabulky = parametry záznamů, hlavička udává obsah sloupce – stejný údaj v celém sloupci

Jednotlivé záznamy
(taxon, lokalita atd.)

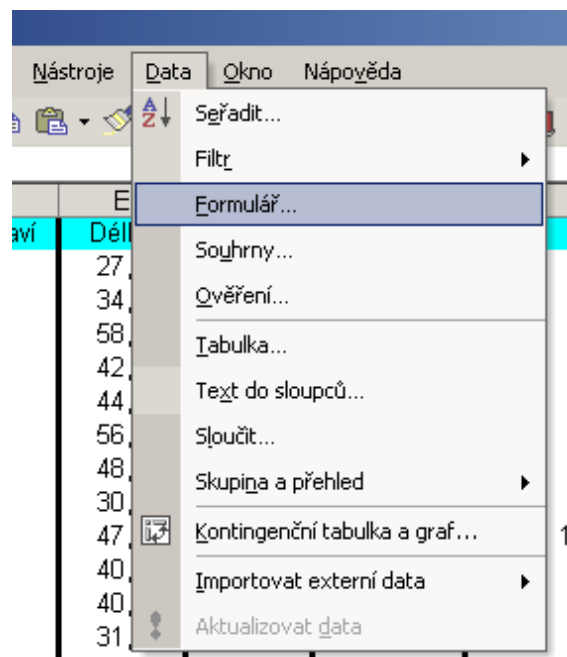


	A	B	C	D	E	F	G	H	I
1	Číslo	Značka	Společ	Pohlaví	Délka	Váha	P. anguillae	P. bini	
2	1	1	1	m	27,5	23,0	2	2	
3	2	2	2	f	34,0	62,5	0	2	
4	3	5	3	f	58,0	230,0	0	0	
5	4	6	4	f	42,0	155,0	0	0	
6	5	7	5	f	44,0	149,8	0	0	
7	6	8	6	f	56,0	323,0	0	1	
8	7	9	7	m	48,5	178,2	0	0	
9	8	10	8	f	30,5	47,7	4	6	
10	9	11	9	f	47,0	175,9	5	14	
11	10	12	10	f	40,0	85,1	5	9	
12	11	14	11	f	40,0	101,0	0	0	
13	12	15	12	f	31,0	84,0	15	9	
14	13	16	13	f?	22,0	9,0	0	0	
15	14	17	14	f	42,0	108,0	1	3	
16	15	18	15	f	44,0	130,0	0	0	
17	16	19	16	f	37,0	85,0	2	5	
18	17	20	17	f	50,0	212,0	1	8	



Automatický zadávací formulář

- Slouží k usnadnění zadávání dat do databázových tabulek
- Načítá automaticky hlavičky sloupců jako zadávané položky



The 'data' dialog box contains the following fields and controls:

- Číslo ryby: 1
- Značka ryby: 1
- Společ číslo: 1
- Pohlaví: m
- Délka: 27,5
- Váha: 23
- P. anguillae: 2
- P. bini: 2
- Buttons: Nový, Odstranit, Obnovit, Předchozí, Další, Kritéria, Zavřít
- Page indicator: 1 z 19

Nový záznam

Vyhledávání

Názvy sloupců

Obsah dané buňky - editovatelný



Automatické seznamy

- Vytváří se z hodnot buněk v daném sloupci a umožňují vložit hodnotu výběrem ze seznamu již zadaných hodnot – usnadnění zadávání

Sloupec z nějž je seznam vytvořen a pro který platí

Taxon	Abundance	Lokalita	etc.
[Empty Cell]			

Buňka, do níž se vloží vybraná hodnota

Context menu options:

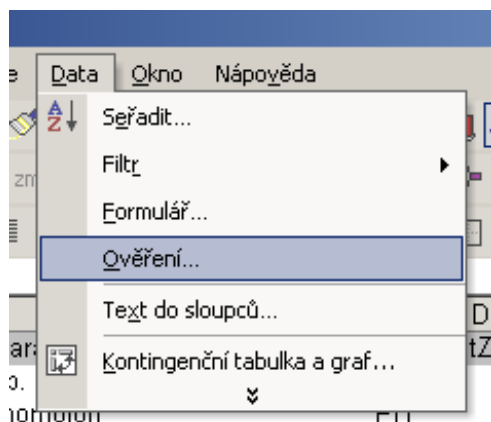
- Vymnout
- Kopírovat
- Vložit
- Vložit jinak...
- Vložit buňky...
- Odstranit...
- Vymazat obsah
- Vložit komentář
- Formát buněk...
- Vybrat ze seznamu...**
- Přidat kukátko
- Hypertextový odkaz...

Selected taxonomic names in the list:

- Piscicola geometra (Linnaeus, 1761)
- Acanthocephallus lucii (Müller, 1776)
- Apophallus mühlungi Jägerskiöld, 1899
- Argulus foliaceus (Linnaeus, 1758)**
- Caryophyllaeides fennica (Schneider, 1902)
- D. cabaleroi
- D. crucifer Wagener, 1857
- D. fallax Wagener, 1857
- D. nanus Dogiel et Bychowsky, 1934

Automatická kontrola dat

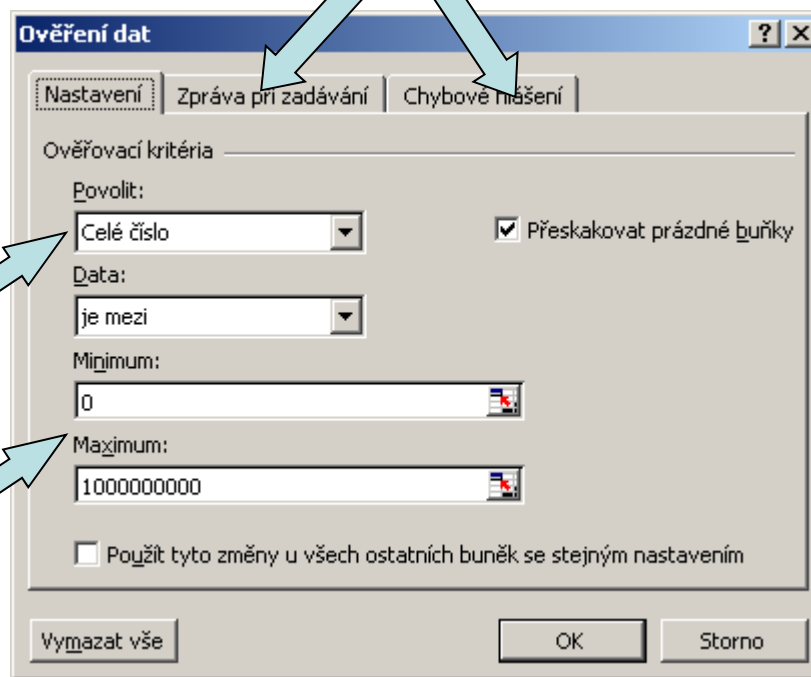
- Umožňuje ověřit typ, rozsah nebo povolit pouze určitý seznam hodnot zadávaných do sloupce databázové tabulky



Co je povoleno – definiční obory čísel, seznamy, vzorce atd.

Rozsahy hodnot, načtení seznamů apod.

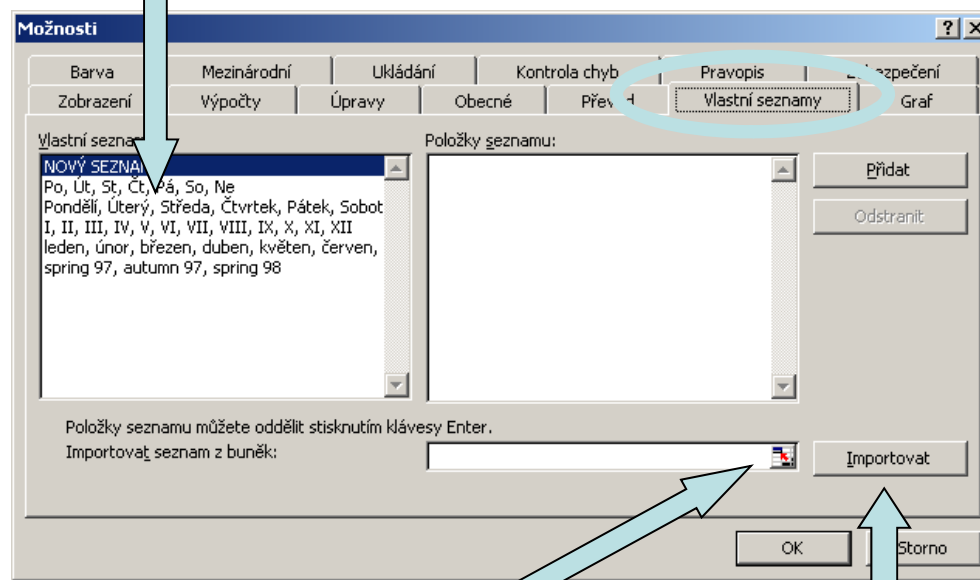
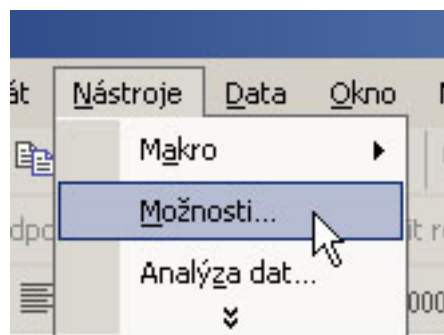
komunikace s uživatelem



Seznamy

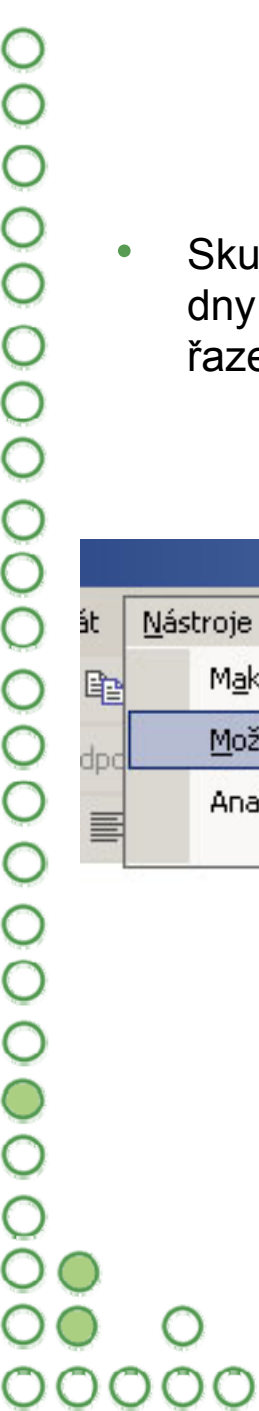
- Skupiny hodnot zachovávající logické pořadí, některé jsou zabudované (např. dny v týdnu, měsíce v roce), další je možné uživatelsky vytvořit, slouží pro účely řazení a automatického vyplňování dat

Existující seznamy



Výběr buněk pro nový seznam

Načtení nového seznamu



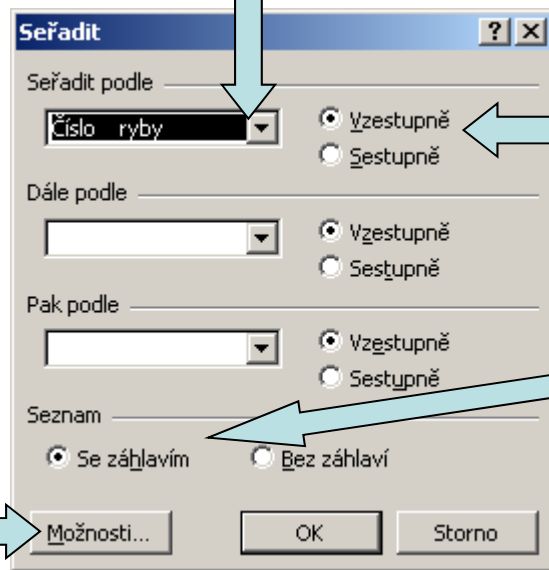
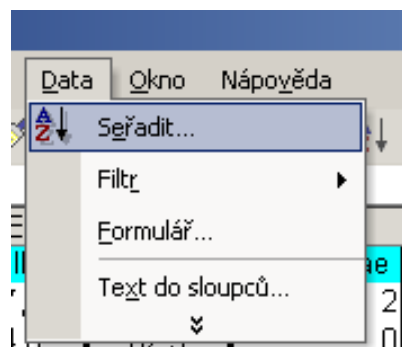
Řazení dat

- Řazení dat je nejjednodušším způsobem jejich zpřehlednění, užitečným hlavně u menších/výsledkových tabulek

Zkontrolujte, zda seřazení nezničí vazby mezi buňkami = kontrola oblasti, kterou řadíte.



Podle čeho řadit



Směr řazení – vzestupně, sestupně

Využít první řádek oblasti jako záhlaví

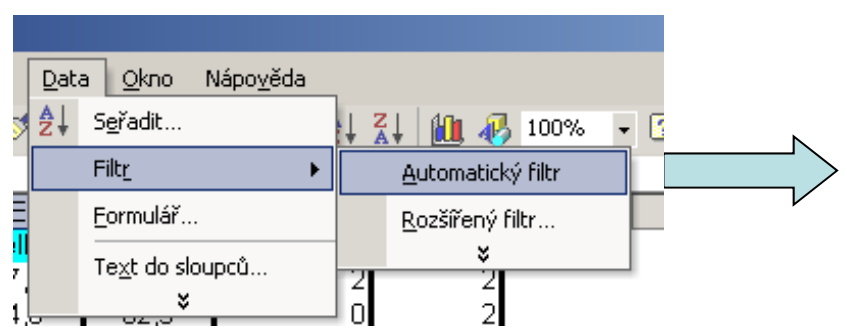
Další možnosti – řazení řádků, řazení podle seznamu



Automatický filtr

- Pomocí automatického filtru je snadné vybírat úseky dat pro další zpracování na základě hodnot ve sloupcích databázové tabulky, výběr je možný i podle více sloupců (např. určitá skupina pacientů)
- Funkce automaticky rozezná hlavičky sloupců v souvislé oblasti buněk
- U sloupců použitých pro filtraci jsou rozbalovací seznamy zbarveny modře
- **Výhodné pro čištění dat (vyhledávání překlepů, kombinace textu a čísel)**

Výběr hodnot pro filtraci

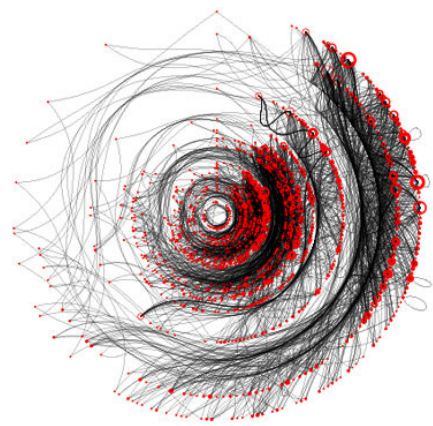


Rozbalení seznamu hodnot nalezených ve sloupci

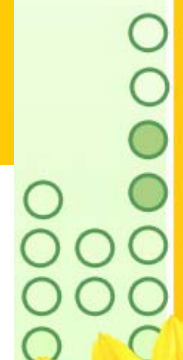
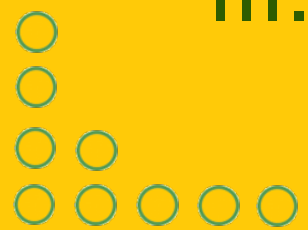
A screenshot of an Excel spreadsheet titled 'Microsoft Excel - ryby.xls'. The spreadsheet has columns labeled 'Číslo', 'Značka', 'Společ', 'Pohlav', and 'Délka'. The 'Společ' column has a dropdown menu open, showing a list of values: '(Vše)', '(Prvních 10...)', '(Vlastní...)', 'f', 'f?', and 'm'. The value 'f?' is currently selected. A blue arrow points from the 'Automatický filtr' menu in the previous image to this dropdown menu.

	A	B	C	D	E
1	Číslo	Značka	Společ	Pohlav	Délka
2	1	1	1	(Vše)	27,5
3	2	2	2	(Prvních 10...)	34,0
4	3	5	3	f	58,0
5	4	6	4	f?	42,0
6	5	7	5	m	44,0
7	6	8	6	f	56,0
8	7	9	7	m	48,5

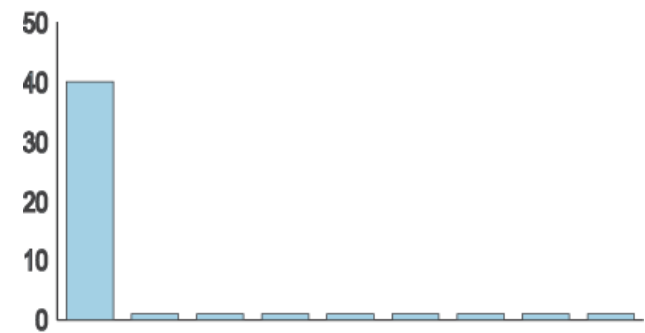
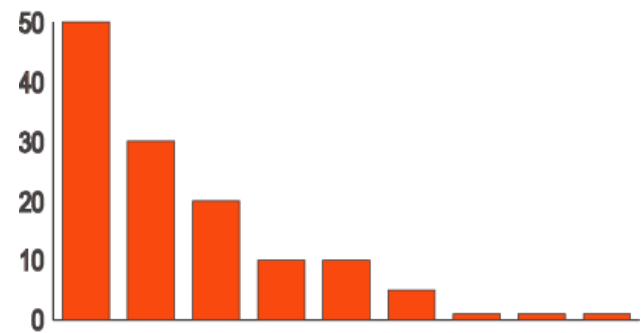




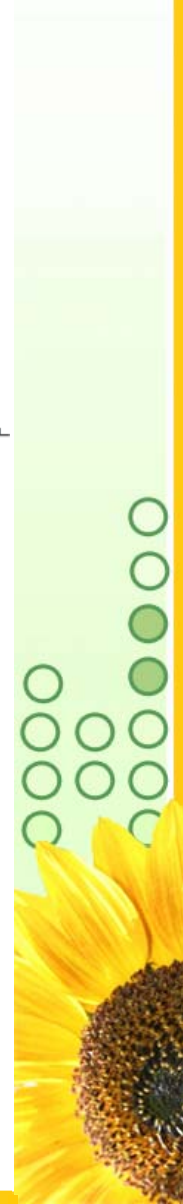
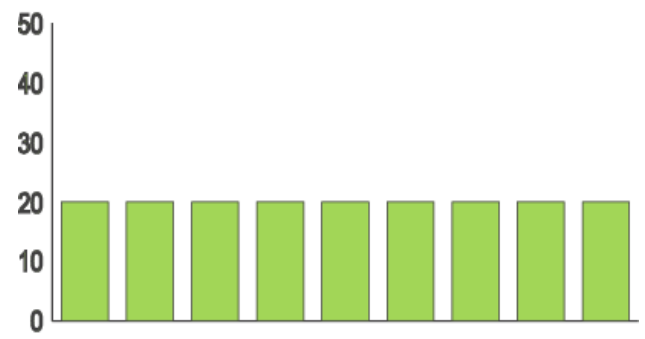
III. Vizualizace biodiverzitních dat



Species abundance curves

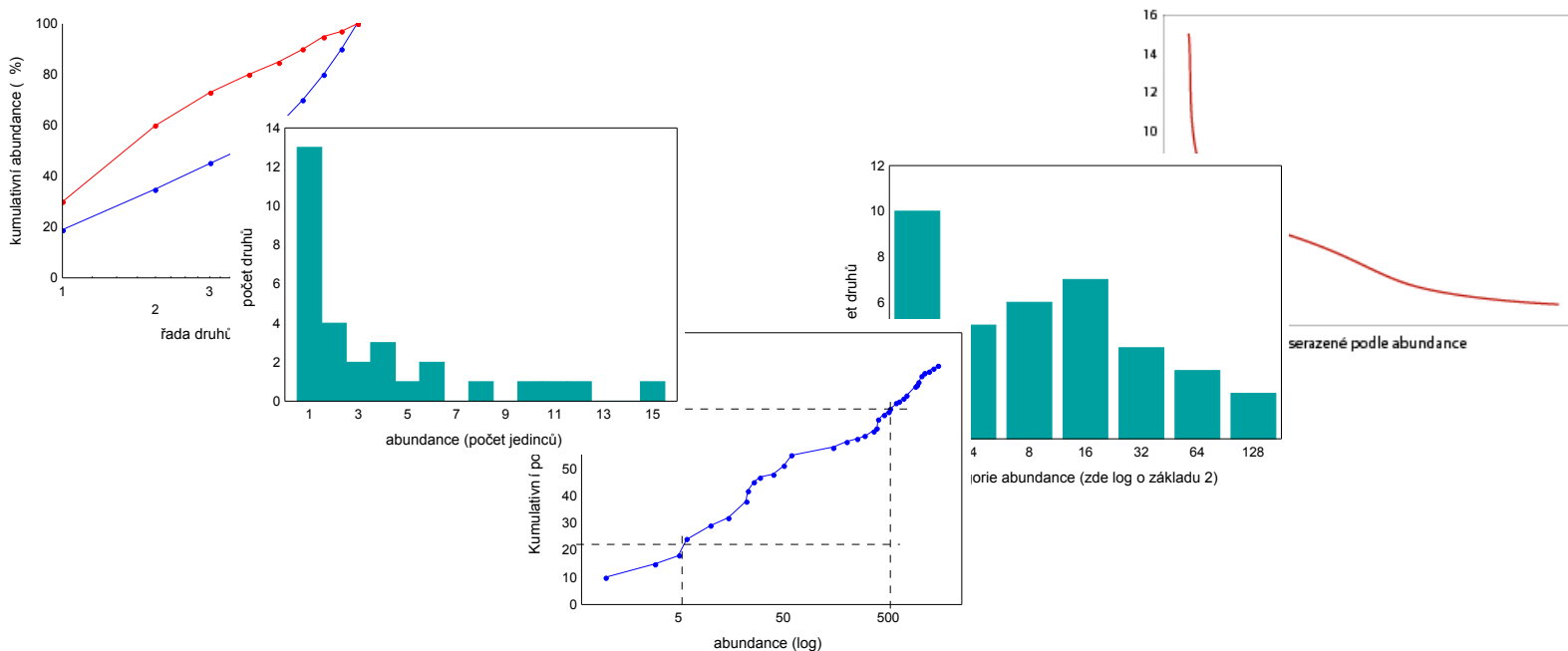


?



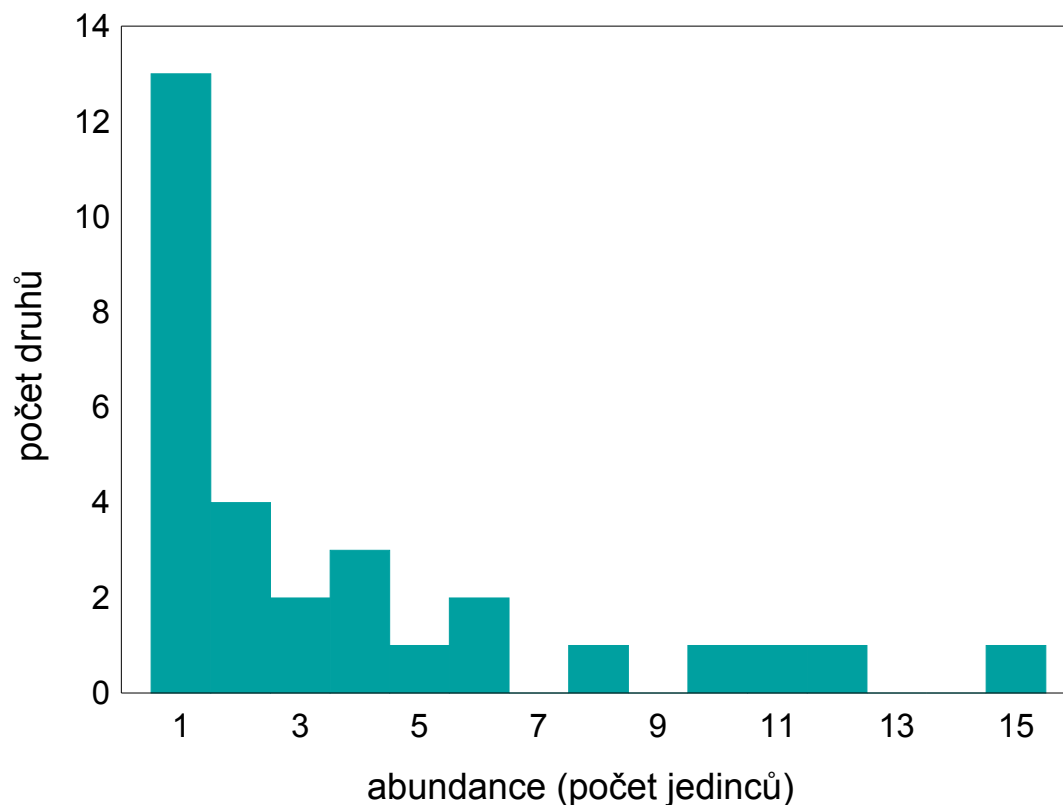
Zobrazení křivky společenstva

- Různé metody zobrazení – různé pohledy na společenstvo
- Jednoduché optické srovnání různých společenstev nebo modelových průběhů společenstev
- Jednoduchá tvorba v Excelu



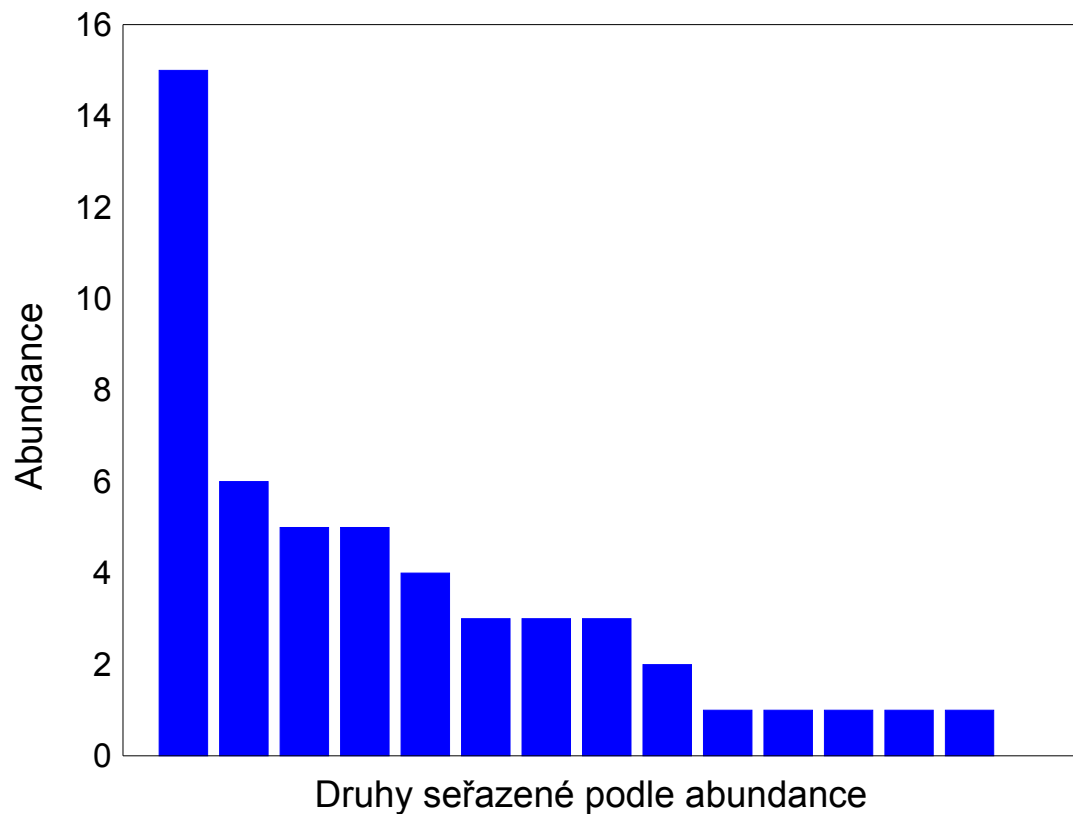
Četnost druhů

- Četnost druhů s danou abundancí je vynášena proti hodnotám nalezených abundancí. Graf poskytuje uživateli přehled o rozložení vzácných, středně početných a hojně se vyskytujících druhů.



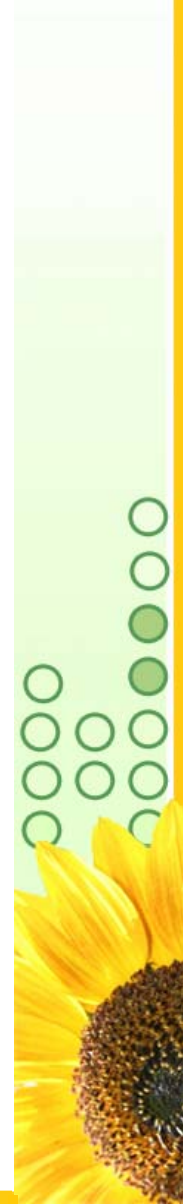
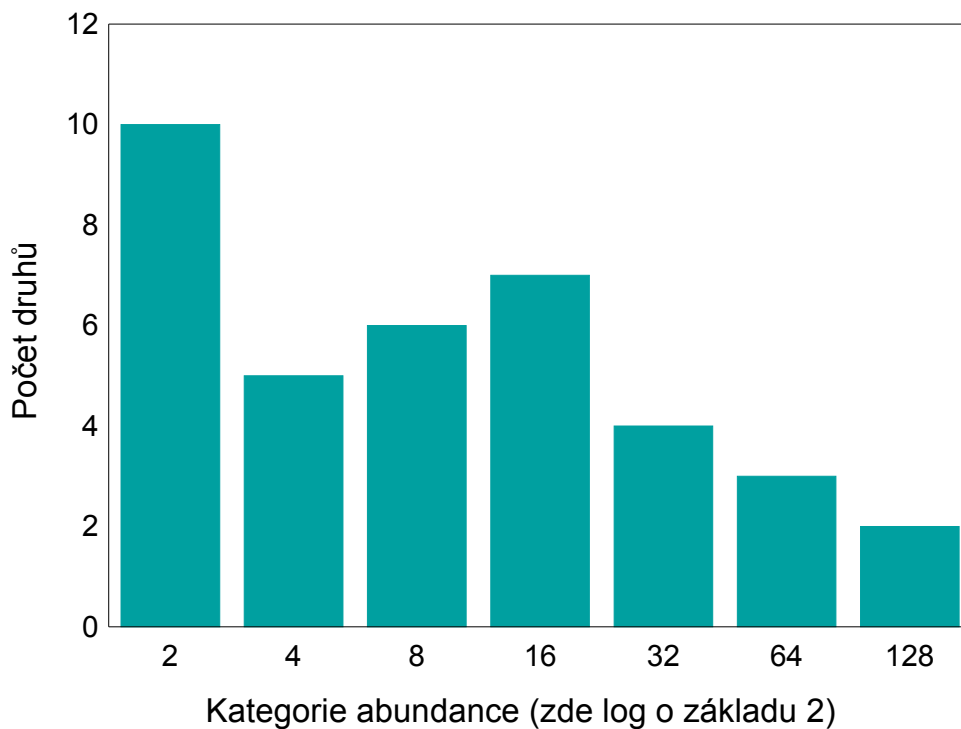
Rank abundance plot

- Graf zobrazuje abundanci druhů seřazených podle této abundance. Poskytuje uživateli přehled o tvaru společenstva – vyrovnanost abundancí, přítomnost „ocásku“ vzácných druhů apod.



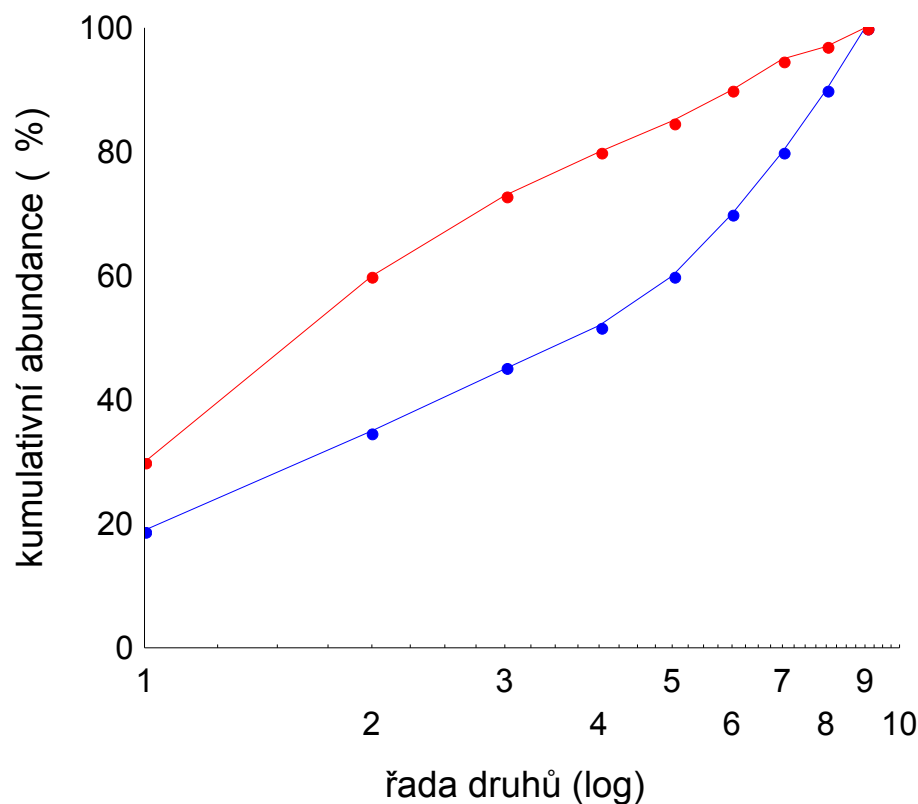
Graf kategorií abundance

- Je obdobou grafu četnosti druhů s určitou abundancí, namísto konkrétních hodnot zde jsou ale třídy abundance a četnost druhů k nim náležejících. Opět umožňuje sledovat relativní podíly vzácných a hojných druhů. Ve formě početnosti druhů v log třídách abundance jsou generovány výsledky některých species abundance models.



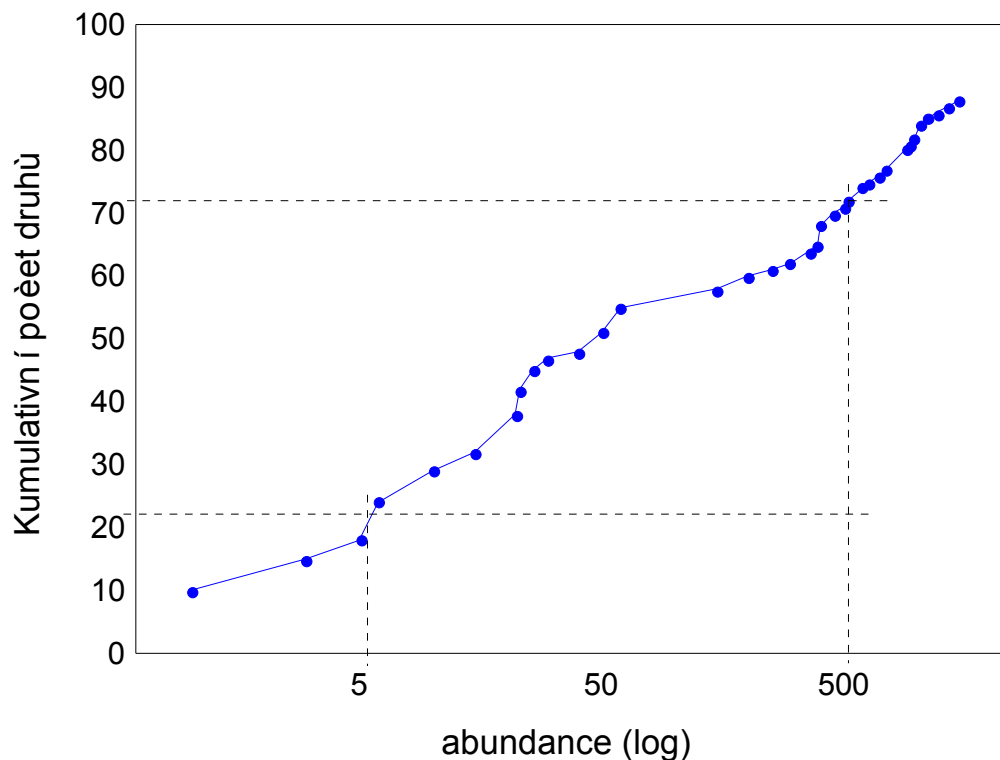
K- dominance plot

- Tento graf vynáší kumulativní abundanci druhů proti logaritmu druhové řady. Může být využit pro optické srovnání diverzity různých vzorků



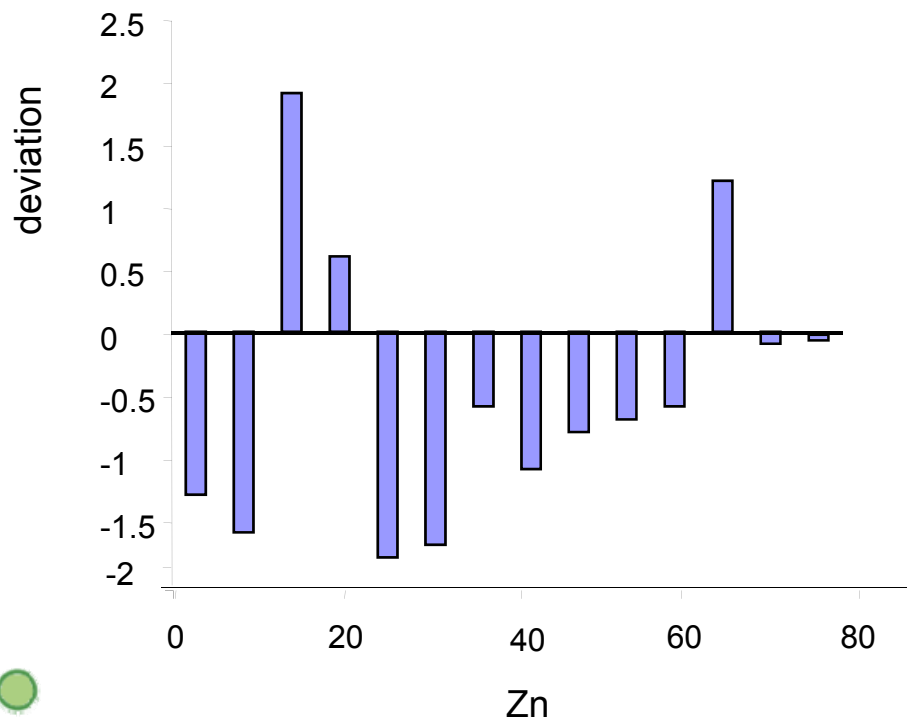
Kumulativní počet druhů

- Graf vynáší kumulativní počet druhů proti ose jejich logaritmované abundance. Graf slouží jako doplněk k výpočtu Q statistiky, který je založen na obdobně uspořádaných datech. Zobrazuje strmost narůstání počtu druhů se stoupající abundancí.

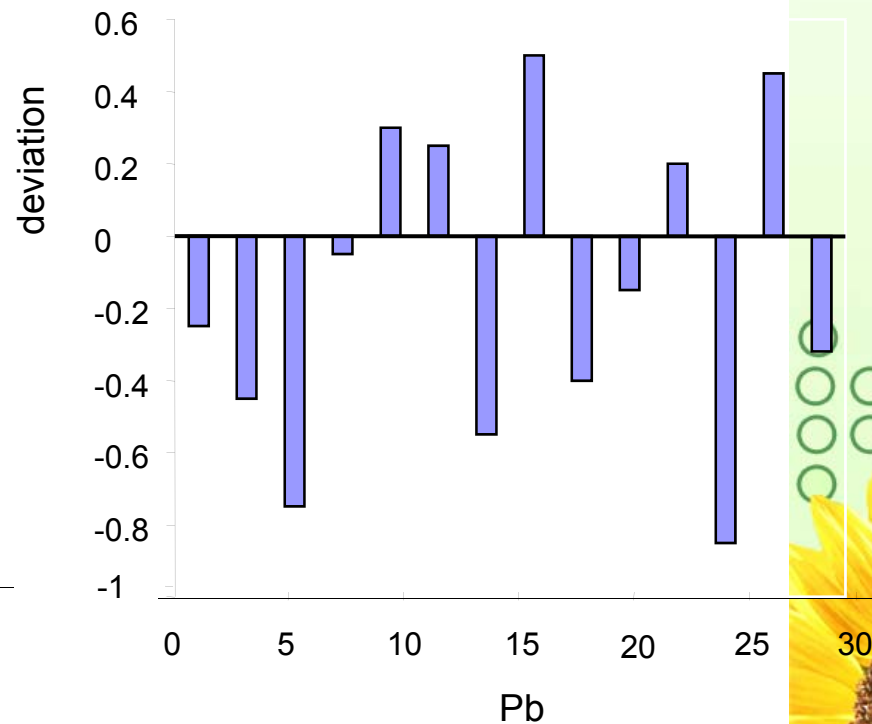


Grafická diagnostika rozložení

Rootgram

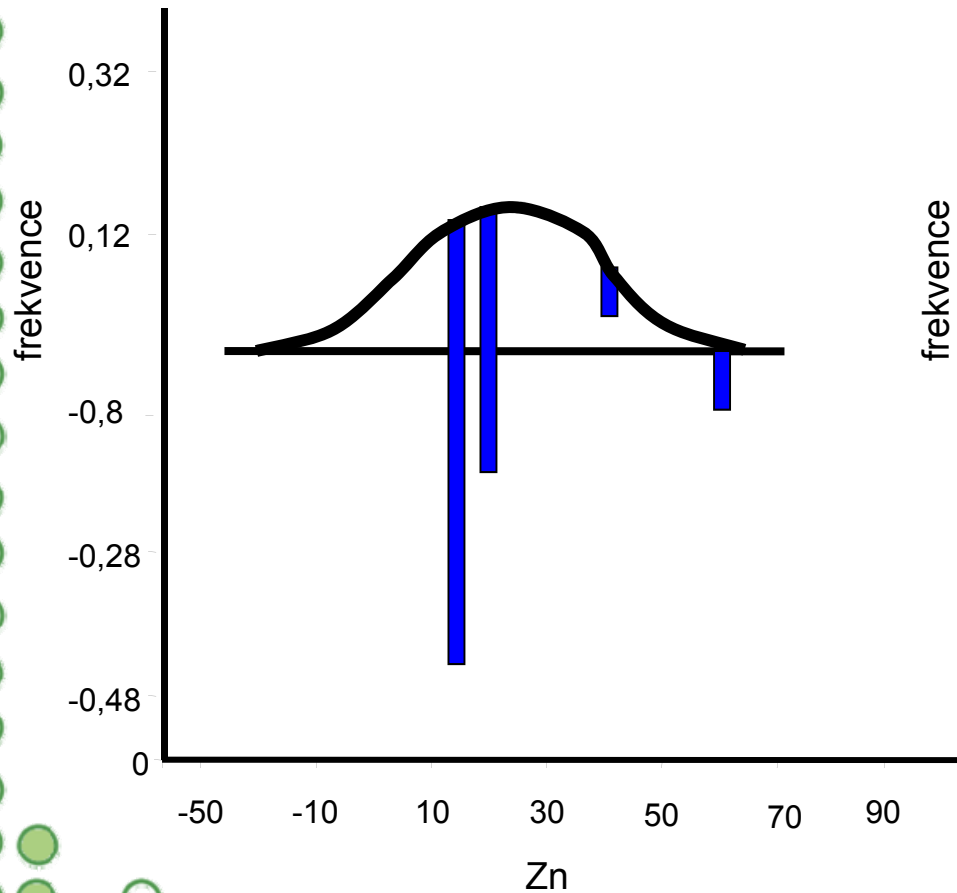


Rootgram

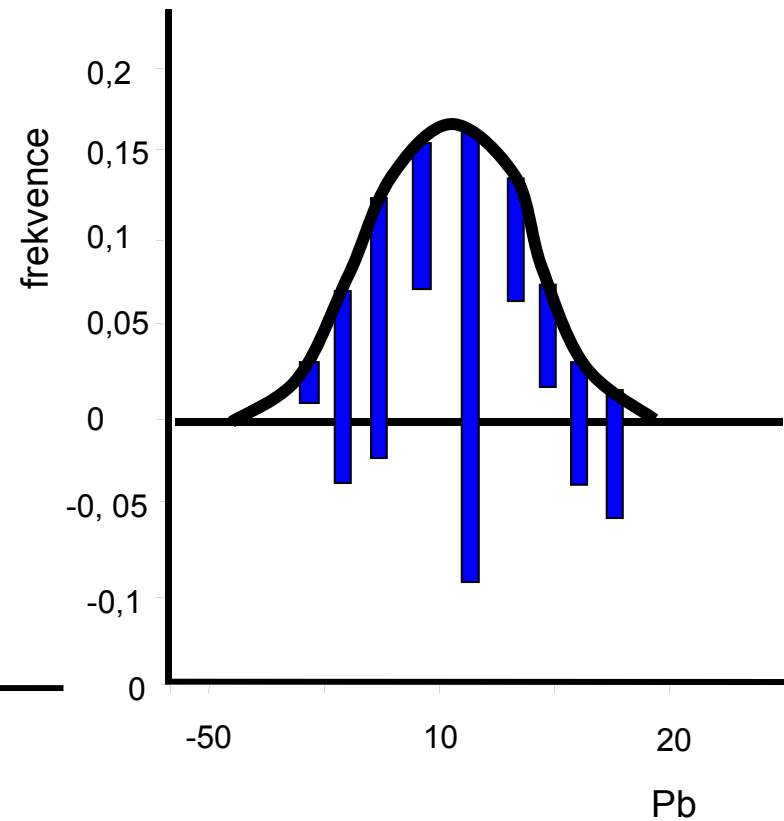


Grafická diagnostika rozložení

Hanging Histobars.

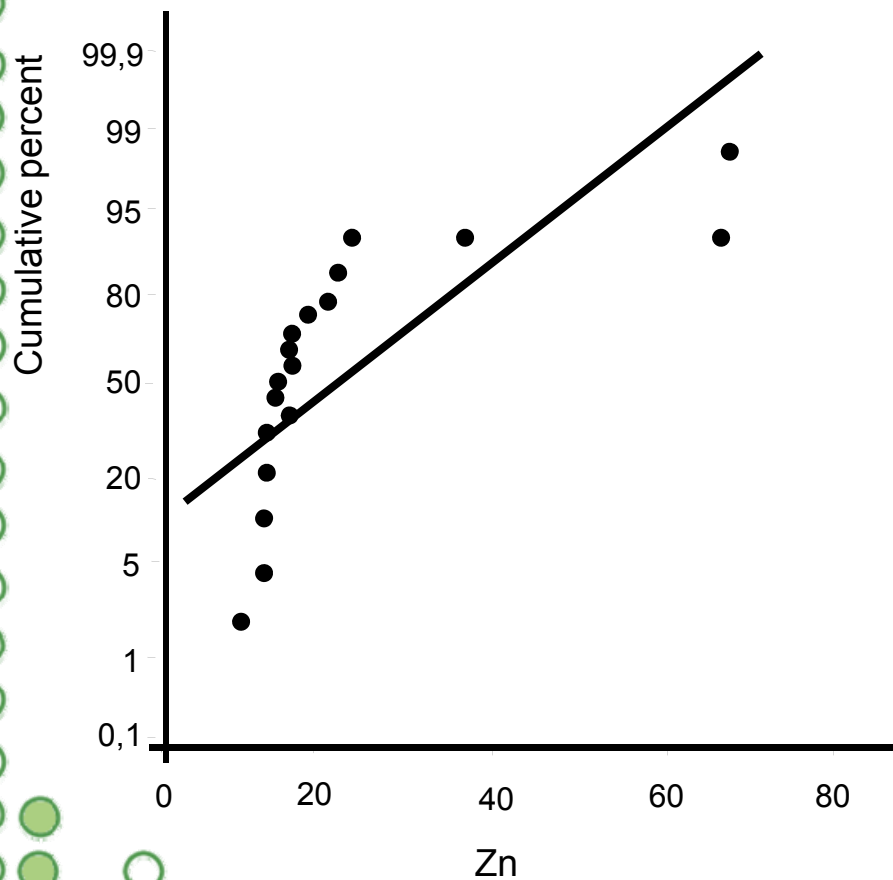


Hanging Histobars.

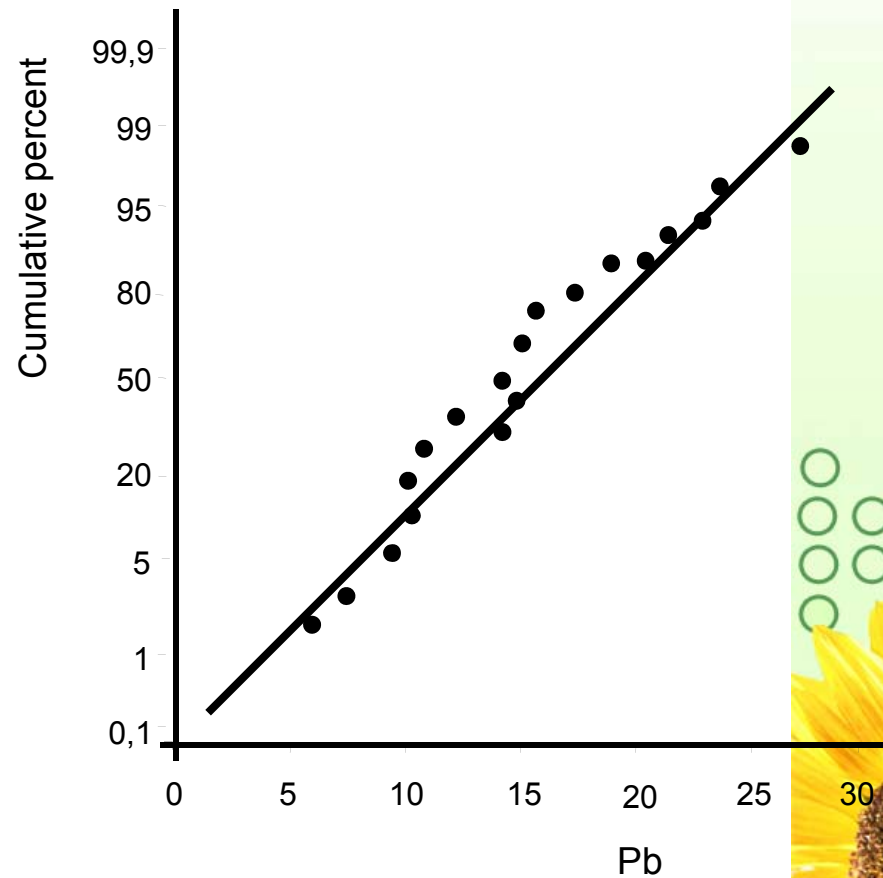


Grafická diagnostika rozložení

Normal Probability Plot

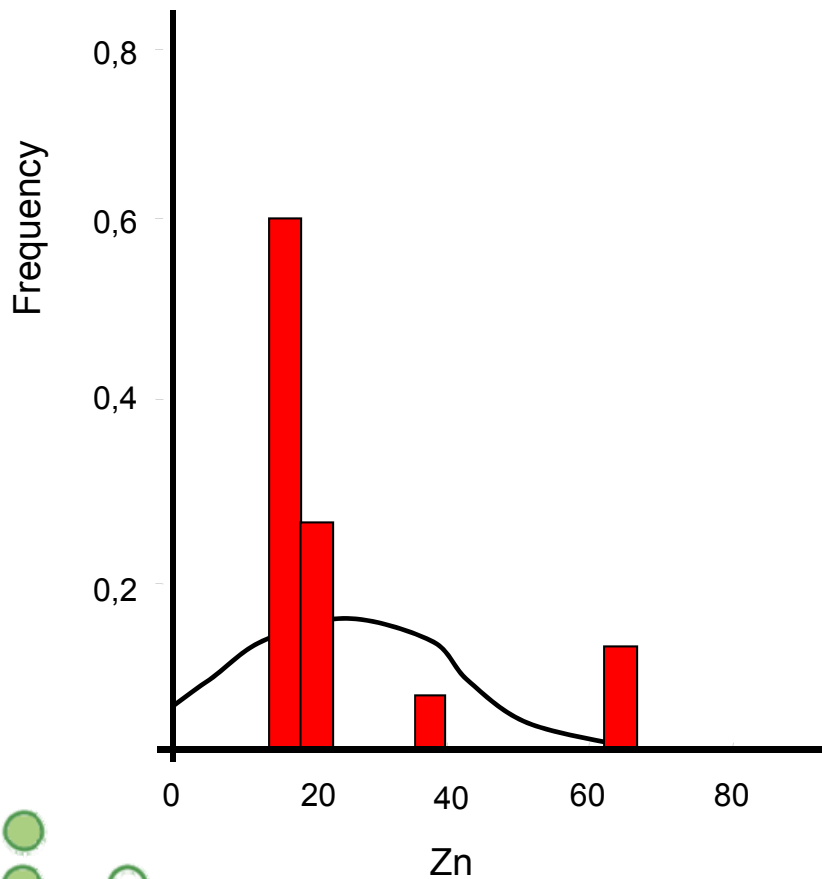


Normal Probability Plot

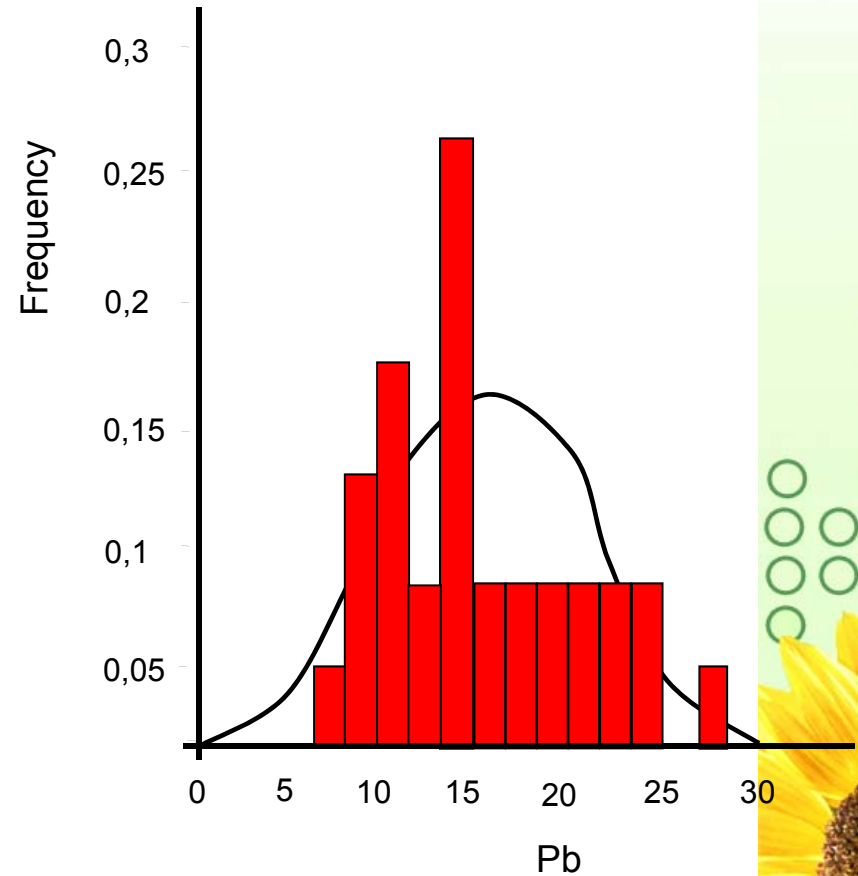


Grafická diagnostika rozložení

Frequency Histogram

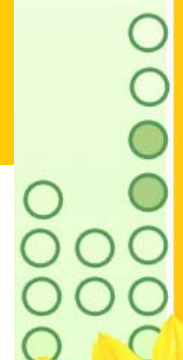
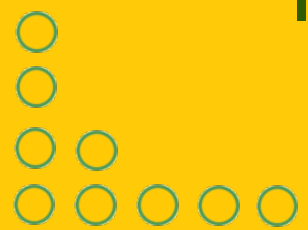


Frequency Histogram





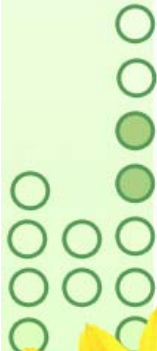
IV. Indexy biodiverzity a jejich spolehlivost



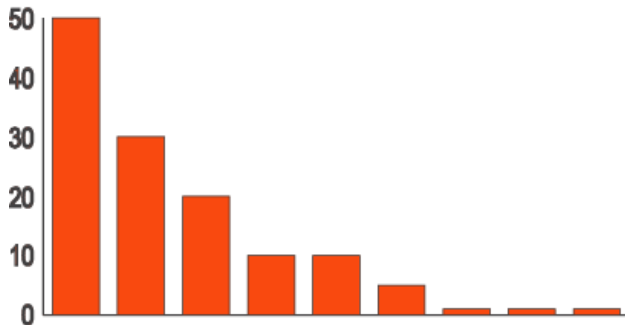
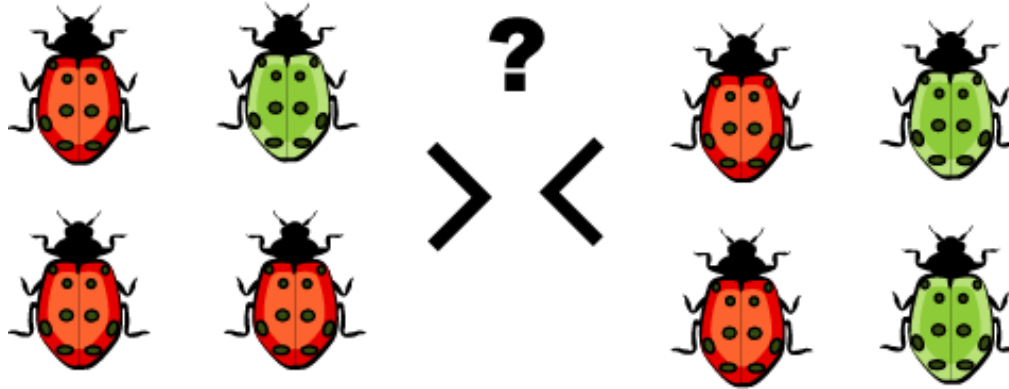
Indexy diverzity: výhody a nevýhody

X

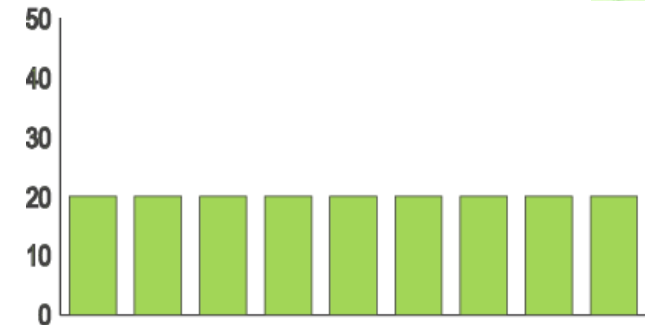
- Indexy diverzity je možné brát jako analogii k **popisné statistice**
- Celé společenstvo je agregováno **jediným číslem**, které reprezentuje **počet druhů a/nebo jejich dominanci ve společenstvu**
- Pro popisnou statistiku diverzity je možné získat intervaly spolehlivosti a dostupné jsou i statistické testy
- **Výhody:**
 - Měření diverzity v jediném čísle
- **Nevýhody:**
 - Redukce individuality taxonů
 - V některých případech nejasná interpretace (stejná hodnota indexu může být spočítána z velmi odlišných společenstev)



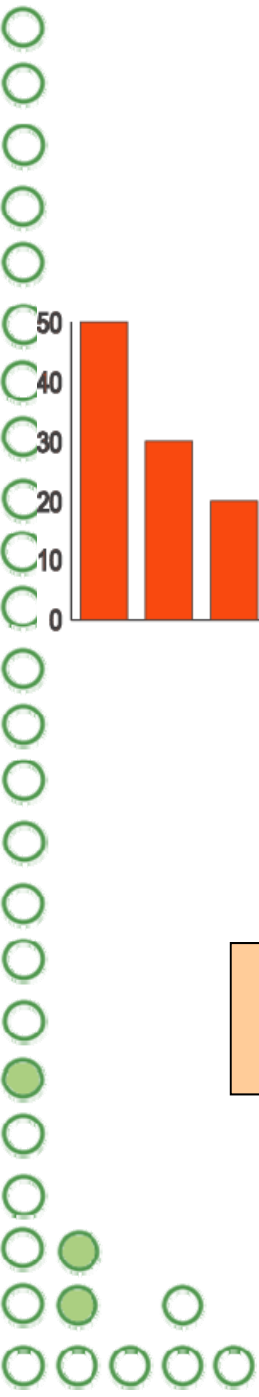
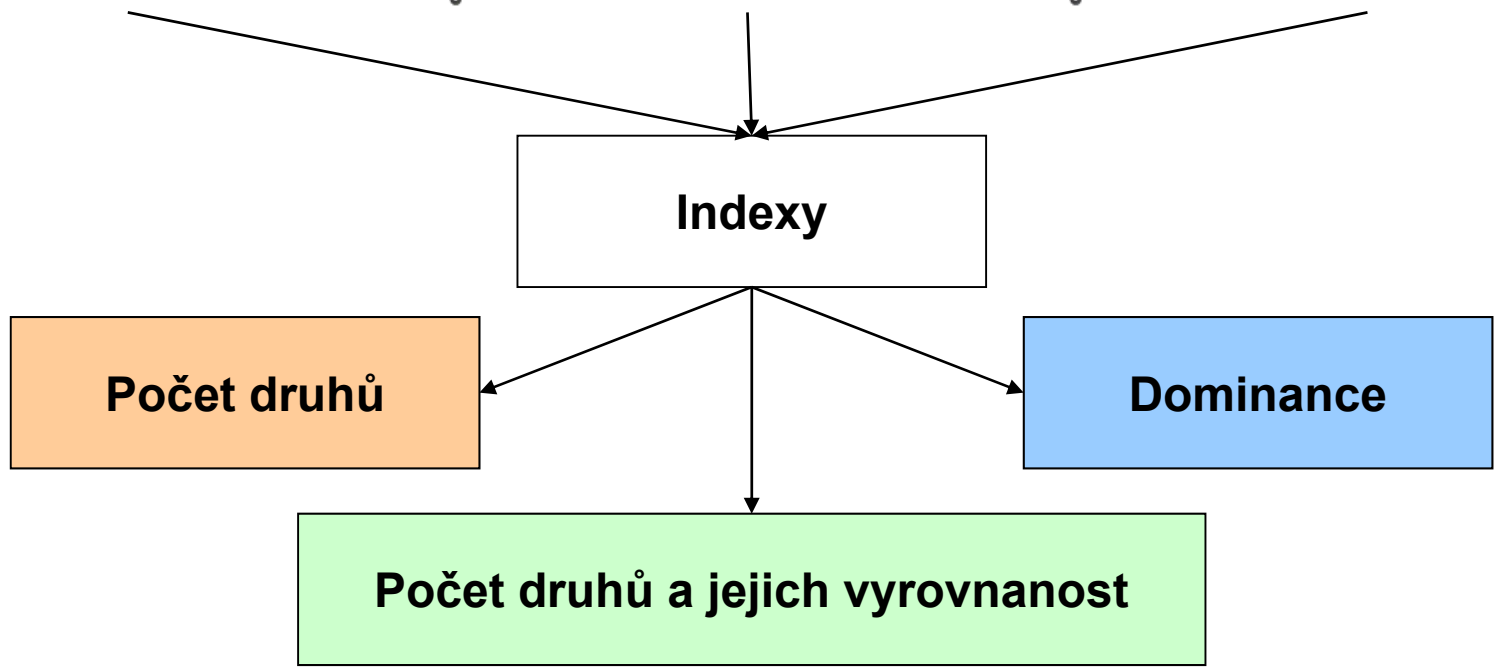
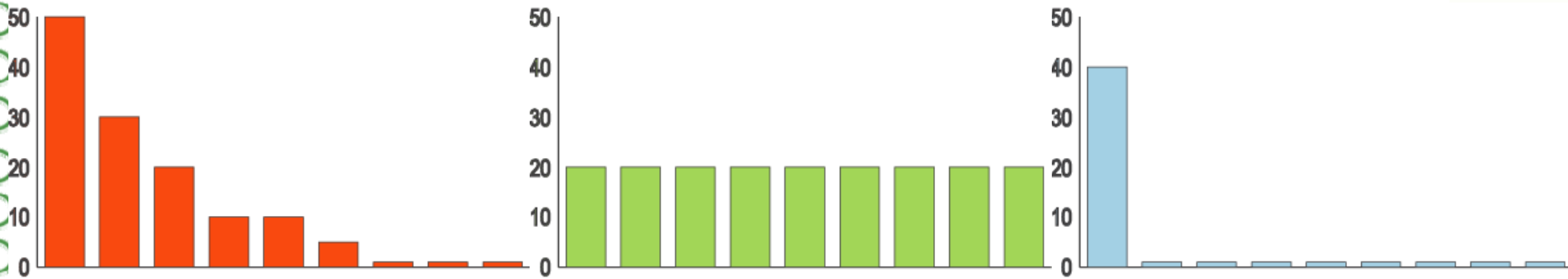
Problémy s biodiverzitou – co znamená větší diverzitu ?



?



Indexové hodnocení diverzity



Indexy diverzity I

S, 'species richness'

$$R1 = (S-1)/\ln N$$

$$R2 = S/\bar{O}N$$

$$H' = -\sum_i (p_i \cdot \ln p_i)$$

$$H'_{adj}$$

$$N1 = e^{H'}$$

$$N2 = 1/\sum_i p_i^2$$

$$E1 = H'/\ln S = H'/H'_{max}$$

$$E2 = N1/S$$

$$E3 = (N1-1)/(S-1)$$

$$E4 = N2/N1$$

$$E5 = (N2-1)/(N1-1)$$

$$D = (N - \sum_i n_i^2)/(N \cdot \bar{O}N)$$

$$DE = (N - \sum_i n_i^2)/[N - (N/\bar{O}S)]$$

McIntosh, 1967

Margalef, 1958

Menhinick, 1964

Shannon & Weaver, 1949

Hutcheson, 1970

Simpson, 1949

Hill, 1973

Pielou, 1969

Sheldon, 1969

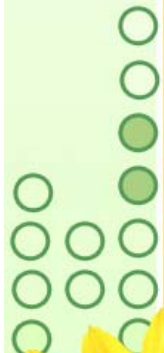
Heip, 1974

Hill, 1973

Alatalo, 1981

McIntosh, 1967

Pielou, 1969



Indexy diverzity II

$$PIE = 1 - \sum_i p_i^2$$

Hurlbert, 1971

$$NMS = \sum_i (R_i p_i - 1)$$

Fager, 1972

$$Q = (0.5n_{Q25} + \sum_r n_r + 0.5n_{Q75}) / \ln(Q75/Q25)$$

Kempton & Taylor, 1978

$$HB = (\ln N! - \sum_i \ln n_i!) / N$$

Brillouin, 1956

$$HBe = HB / HB_{\max}$$

Pielou, 1969

$$G = E5 \cdot (\arcsin E5/90)$$

Molinari, 1989

$$G = (E5)^3$$

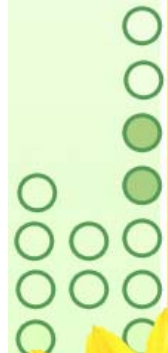
Molinari, 1989

$$1/d = N / n_{\max}$$

Berger & Parker, 1970

$$R100 = \sum_i \left\{ 1 - \left[\frac{\binom{N-n_i}{n}}{\binom{N}{n}} \right] \right\}$$

Hurlbert, 1971



Indexy druhové bohatosti

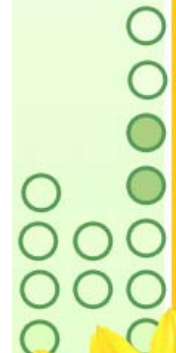
- Počet druhů (**druhová bohatost**) S
- Počet druhů vztažený na počet jedinců nebo biomasu
v případě, že není možné zajistit stejnou velikost vzorku (počet druhů narůstá s velikostí vzorku a vzorkovacím úsilím) => RAREFACTION – metoda výpočtu počtu druhů očekávaných ve vzorku standardní velikosti (např. 1000 jedinců)
- Hustota druhů – počet druhů v dané oblasti

Margalefův index $R1 = (S-1)/\ln N$

Menhinickův index $R2 = S/\sqrt{N}$

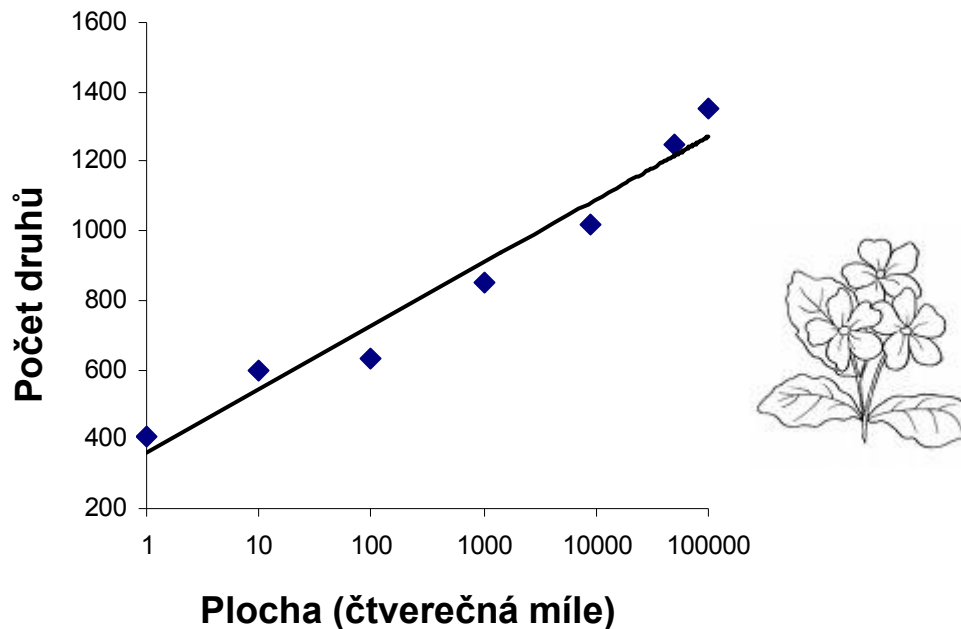
Počet druhů je vážen počtem jedinců.

- **Nevýhody indexů druhové bohatosti:** křivka společenstva není uvažována ve výpočtu



Druhová bohatost

- Druhová bohatost narůstá s velikostí vzorku.

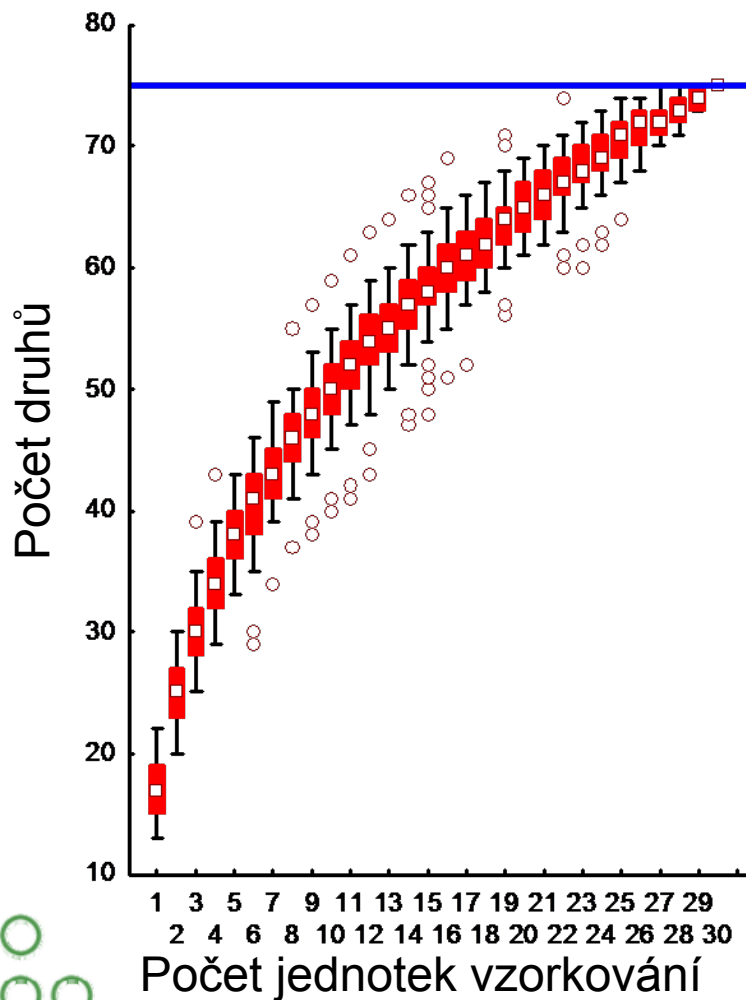


Relationship between number of species and area for flowering plants in England.
Redrawn from Krebs (1985) after Williams (1964).

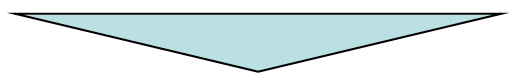


Velikost vzorku a počet druhů

- Počet druhů nelineárně závisí na počtu jedinců ve vzorku



- Jak srovnat druhovou bohatost ve vzorku různé velikosti?

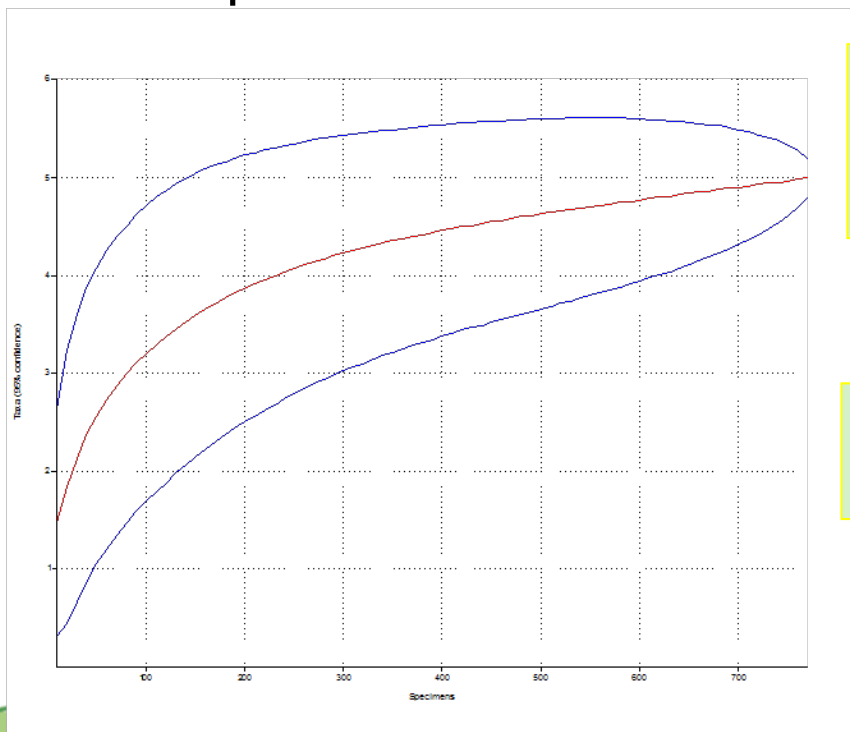


RAREFACTION



Rarefaction

- Metoda řešící problém srovnání druhové bohatosti ve vzorcích o různé velikosti
- Standardizuje oba vzorky na velikost menšího vzorku sample



Metoda výpočtu odhadu počtu druhů v náhodném výběru o n jedincích ze vzorku

Je možno spočítat jak bodové, tak intervalové odhady



Rarefaction: výpočet

$$E(\hat{S}_n) = \sum_{i=1}^s \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right]$$

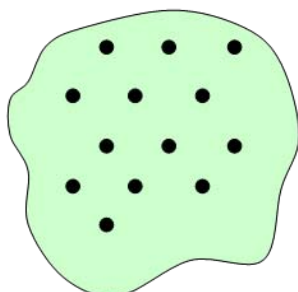
- Kde **E(S_n)** je odhadnutý počet druhů ve vzorku o **n** jedincích, **N** je celkový počet jedinců, **N_i** je počet jedinců druhu **i** a **n** je počet jedinců pro které je odhad počítán.
- Variabilita odhadu je vypočtena:

$$\text{var}(\hat{S}_n) = \binom{N}{n}^{-1} \left[\sum_{i=1}^s \binom{N - N_i}{n} \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right] + 2 \sum_{i=1}^{s-1} \sum_{j=i+1}^s \left[\binom{N - N_i - N_j}{n} - \frac{\binom{N - N_i}{n} \binom{N - N_j}{n}}{\binom{N}{n}} \right] \right]$$

Rarefaction: předpoklady

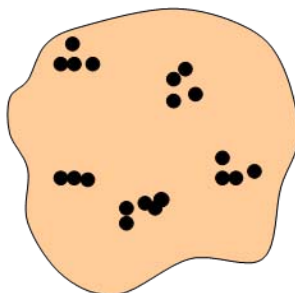
- Společenstva porovnávaná pomocí rarefaction mají mít podobné taxonomické složení
- Srovnatelné metody vzorkování
- Rarefaction křivka nemůže být extrapolována za sumu jedinců největšího vzorku
- Jedinci jsou ve společenstvu náhodně rozmístěni
 - Ve skutečnosti jsou častější shluky jedinců a za těchto podmínek dává rarefaction nadsazené odhady

$$\sigma^2 < \mu$$



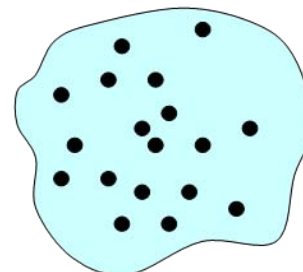
Uniform

$$\sigma^2 > \mu$$



Clustered

$$\sigma^2 = \mu$$

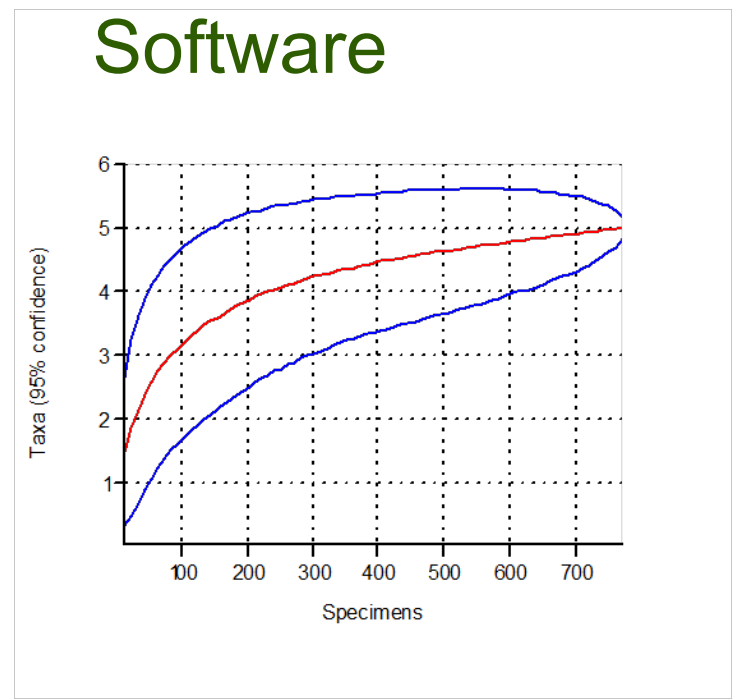


Random

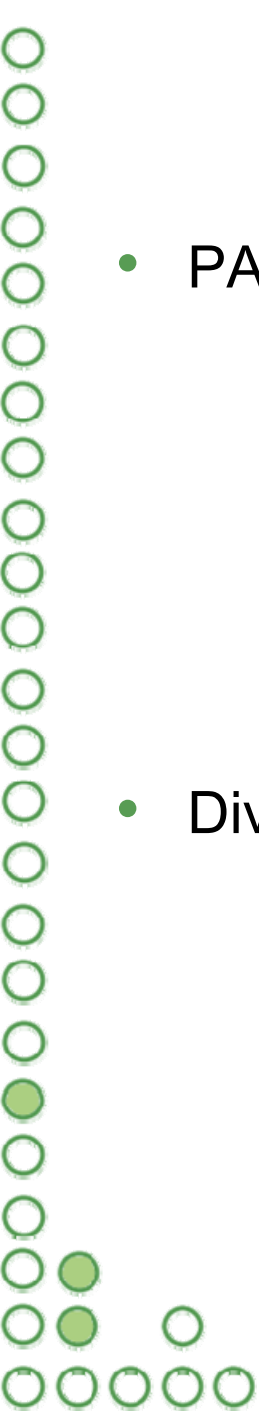


Software

- PAST software



- Diversity software od IBA



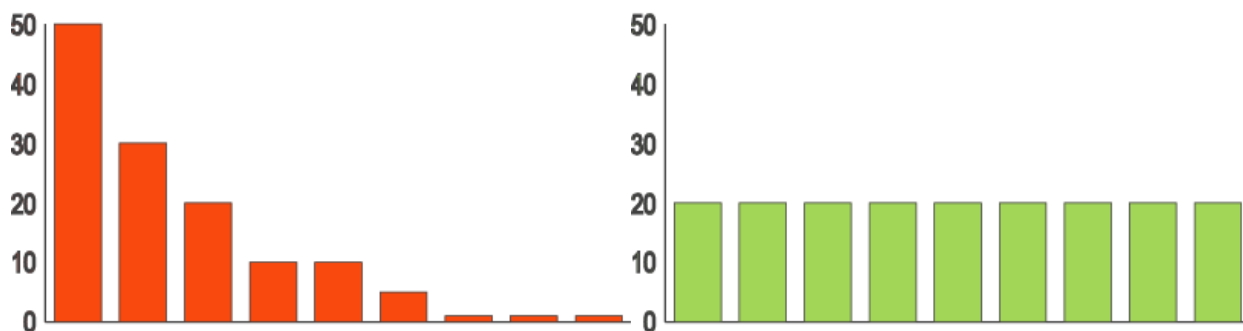
Příklady

1. Předpokládejte stejné vzorkovací úsilí a stejné přírodní podmínky u dat v souboru `example_set.xls`
2. Zjistěte největší velikost vzorku použitelnou pro výpočet rarefaction
3. Vypočtěte rarefaction odhady pro výběr 5 lokalit a interpretujte je



Vyrovnanost společenstva a počet druhů

- Doplnkem těchto indexů je tzv. evenness, která je počítána jako podíl indexu skutečného společenstva a teoretické maximální hodnoty pro daný počet druhů (tj. pokud by měly zcela vyrovnané abundance) – evenness vlastně vyjadřuje jak moc je reálné společenstvo vzdáleno od maximální vyrovnanosti



- Do této skupiny patří Shannonův a Brillouinův index, liší se od sebe použitím, Brillouinův index by měl být používán pouze pro skutečně vzorkovanou část společenstva (příkladem mohou být infrakomunity parazitů, kdy sesbíráme všechny parazity na hostiteli), Shannonův index při výpočtu uvažuje, že část společenstva vzorkována nebyla



Shannon index

- n_i je abundance i-tého druhu, N celkový počet jedinců a S počet druhů

Shannonův index $H' = -\sum p_i \ln p_i$ $p_i = \frac{n_i}{N}$

Shannonův index s korekcemi

$$H' = -\sum p_i \ln p_i - \frac{S-1}{N} + \frac{1 - \sum p_i^{-1}}{12N^2} + \frac{\sum (p_i^{-1} - p_i^{-2})}{12N^3}$$

Shannon evenness $E = \frac{H'}{H_{\max}} = \frac{H'}{\ln S}$



Shannon index

Shannon index

$$H' = -\sum p_i \ln p_i \quad p_i = \frac{n_i}{N}$$

n_i ... abundance i -tého druhu

N ... celkový počet jedinců



Species	Abundance	Shannon index of diversity		
		p_i	$\ln(p_i)$	$p_i * \ln(p_i)$
Baetis alpinus	736	0.9472	-0.0542	-0.0513
Rhithrogena semicolorata	28	0.0360	-3.3232	-0.1198
Epeorus sylvicola	8	0.0103	-4.5760	-0.0471
Baetis rhodani	4	0.0051	-5.2691	-0.0271
Ephemerella mucronata	1	0.0013	-6.6554	-0.0086
Total number of individuals	777			
Shannon index of diversity (using natural logarithm)				0.2539

Shannon index je mezi ekology nejpopulárnější index.





Shannon index

Shannon index

$$H' = -\sum p_i \ln p_i \quad p_i = \frac{n_i}{N}$$

log₂ ? ln ? log ?

Species	Abundance	Shannon index of diversity		
		$p_i \cdot \log_2(p_i)$	$p_i \cdot \ln(p_i)$	$p_i \cdot \log(p_i)$
 Baetis alpinus	736	-0.0741	-0.0513	-0.0223
Rhithrogena semicolorata	28	-0.1728	-0.1198	-0.0520
Epeorus sylvicola	8	-0.0680	-0.0471	-0.0205
 Baetis rhodani	4	-0.0391	-0.0271	-0.0118
Ephemerella mucronata	1	-0.0124	-0.0086	-0.0037
Total number of individuals	777			
Shannon index of diversity		0.3663	0.2539	0.1103

Báze logaritmu ↑ Shannon index ↓



Shannon index

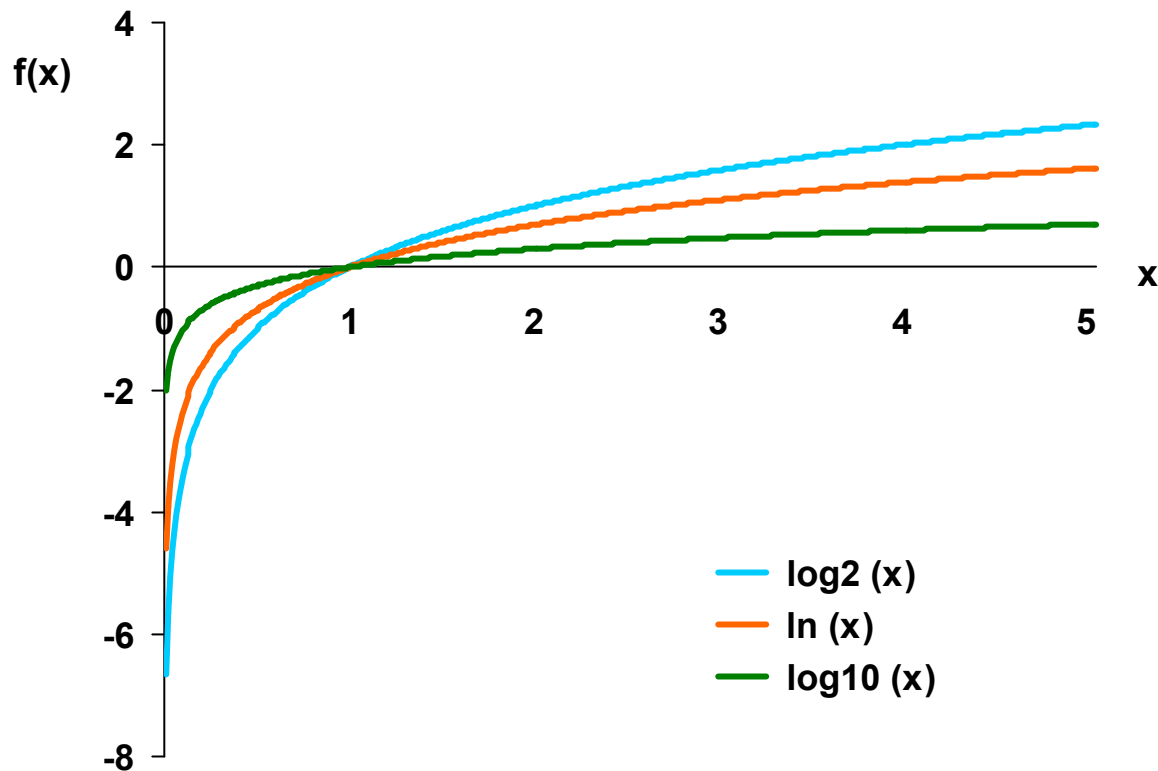
Shannon index

$$H' = - \sum p_i \ln p_i \quad p_i = \frac{n_i}{N}$$

Báze logaritmu



Shannon index



Brillouinův index

- Možným problémem při výpočtu je neschopnost Excelu spočítat faktoriál více než 160, lze obejít použitím logaritmu

Brillouinův index

$$HB = \frac{\ln N! - \sum \ln n_i!}{N}$$

Brillouin evenness

$$E = \frac{HB'}{HB_{\max}}$$

$$HB_{\max} = \frac{1}{N} \ln \frac{N!}{\left(\left[\frac{N}{S} \right]! \right)^{s-r} \left(\left(\left[\frac{N}{S} \right] + 1 \right)! \right)^r}$$

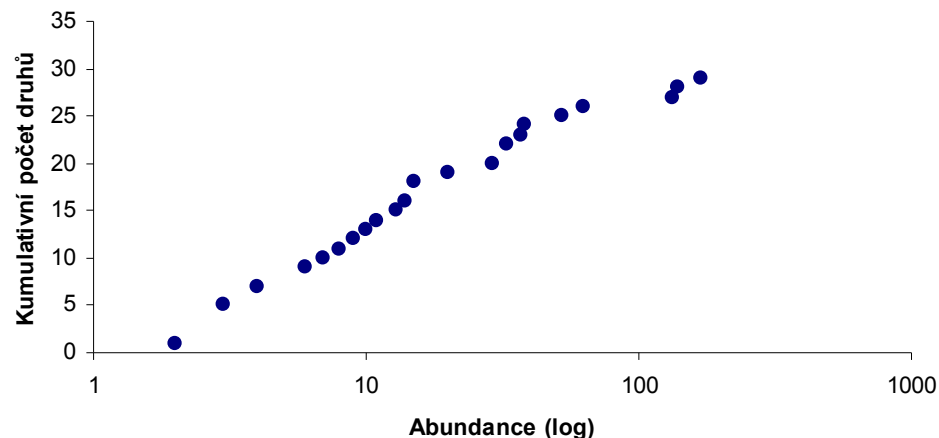
kde $\left[\frac{N}{S} \right]$ je celá část $\frac{N}{S}$ a $r = N - S \left[\frac{N}{S} \right]$



Q statistic I

- Měřítka sklonu křivky abundancí kumulativního počtu druhů

$$Q = \frac{\frac{1}{2}n_{R1} + \sum_{R1+1}^{R2-1} n_r + \frac{1}{2}n_{R2}}{\log\left(\frac{R2}{R1}\right)}$$



$\sum n_r$ – celkový počet druhů mezi kvartilů

S – celkový počet druhů ve vzorku

R1 a R2 – 25% a 75% kvartil

n_{R1} – počet druhů ve třídě, do níž spadá dolní kvartil počtu druhů

n_{R2} – počet druhů ve třídě, do níž spadá horní kvartil počtu druhů

R1 – počet jedinců ve třídě, do níž spadá dolní kvartil počtu druhů

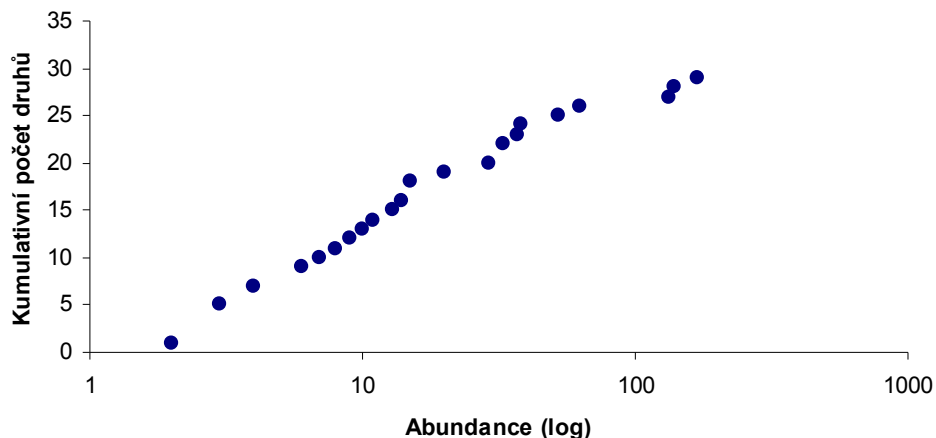
R2 – počet jedinců ve třídě, do níž spadá horní kvartil počtu druhů



Q statistic II

- Dalším způsobem výpočtu je odhad hodnoty Q

$$X_i = \frac{S_j - S_{j'}}{\log\left(\frac{N_j}{N_{j'}}\right)}$$



je počítán pro všechny páry S_j a $S_{j'}$ a N_j a $N_{j'}$ ($j > j'$, $j=1,2, \dots, r$)

S – kumulativní počet druhů

N – počet jedinců ve třídě

r – počet tříd a ($i=1,2, \dots, r(r-1)/2$)

Z vzniklé řady čísel je Q zjištěno jako medián nebo geometrický průměr



Dominance

- Tyto indexy zjišťují, zda jsou ve společenstvu přítomny silně dominantní druhy nebo je společenstvo spíše vyrovnané
- Hodnoty indexů jdou opačným směrem než v případě indexů počítajících s vyrovnaností a počtem druhů (Shannon, Brillouin) a proto se často používá jejich odpočet od jedné nebo převrácená hodnota

Simpson index

$$D = \sum \left(\frac{n_i(n_i - 1)}{N(N - 1)} \right)$$

Berger Parker index

$$d = \frac{N_{\max}}{N}$$



Berger – Parker index

Berger-Parker index

$$d = \frac{N_{\max}}{N}$$



Species	Abundance
Baetis alpinus	736
Rhithrogena semicolorata	28
Epeorus sylvicola	8
Baetis rhodani	4
Ephemerella mucronata	1
Total number of individuals	777
Berger-Parker index	0.9472

N_{\max} ... počet jedinců nejpočetnějšího druhu

N ... celkový počet jedinců

Shannon diversity ↑

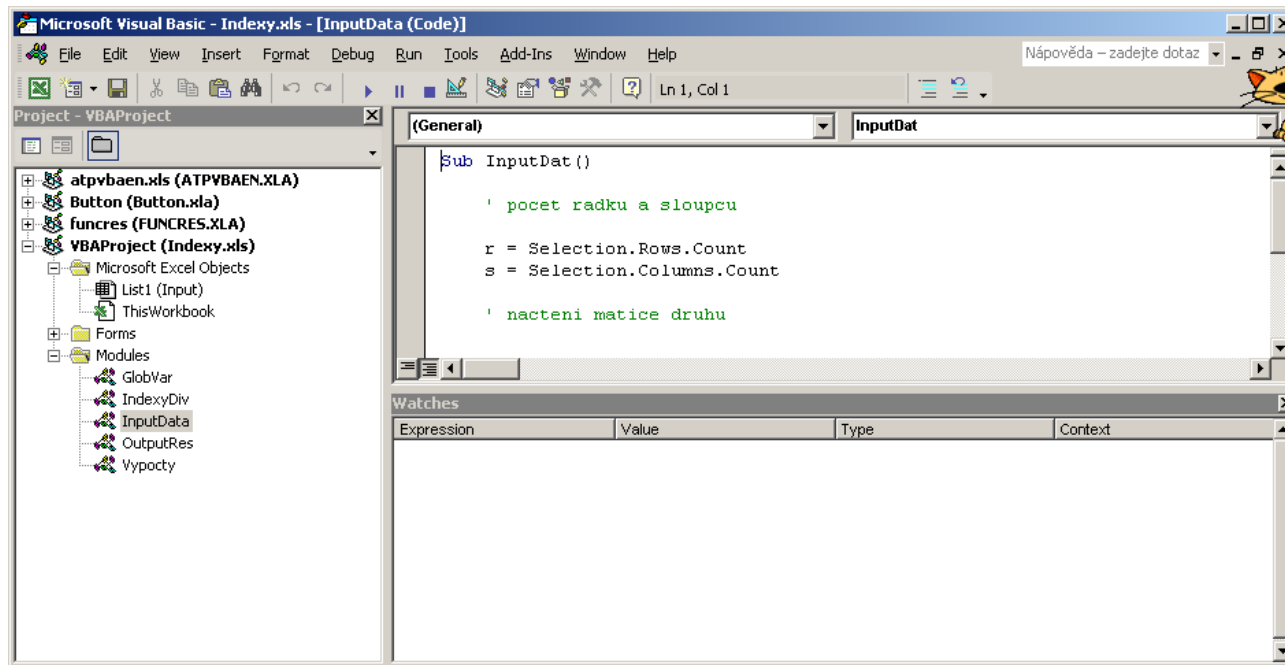
Berger-Parker index ↓

$1 - d$ je často použito namísto d



Algoritmy v MS Office

- Programy MS Office obsahují plnohodnotný programovací jazyk MS Visual Basic – v tomto jazyce jsou také psána makra
- Pomocí maker (Visual Basicu) lze ovládat všechny části programů Office – tabulky, databáze, grafy atd. a ve spojení s formuláři vytvářet i složitější aplikace
- Pro zjištění složitějších příkazů je vhodné analyzovat klasická makra



Proměnné

Proměnné slouží pro ukládání výsledků, dat atd., je vhodné je definovat, definice se skládá z jména proměnné a jejího typu

Základní typy proměnných:

Double – reálná čísla

Long – celá čísla

String – text

Objekty – jako proměnná mohou být definovány například grafy, listy Excelu atd., s každým objektem je spjata sada jeho parametrů, které je možné nastavit, např. data a formát grafu

Dim **jménoProměnné** As Double
definice proměnné

Dim **jménoPole()** As Double
definice dynamického pole (pole=maticе, tabulka), následuje jeho dimenzování
ReDim **jménoPole**(1 to 10, 1 to 2) – dvourozměrné pole 10x2 buněk

Základní příkazy

Ohraničení programu/makra

```
Sub jméno_Makra()  
...  
End Sub
```

Podmínka - větvení

```
If (podmínka) Then  
...  
Else  
...  
End if
```

Cyklus s počtem opakování

```
For i = 1 To r  
...  
Next i
```

Cyklus s podmínkou

```
Do  
...  
Loop Until (podmínka)
```

```
Do While (podmínka)  
...  
Loop
```

Načtení dat

```
Sub InputDat()
```

```
' pocet radku a sloupcu
```

```
r = Selection.Rows.Count
```

```
s = Selection.Columns.Count
```

```
' nacteni matice druhu
```

```
ReDim Matice(r + 1, s + 1)
```

```
For i = 1 To r
```

```
    For j = 1 To s
```

```
        Matice(i, j) = Selection.Cells(i, j)
```

```
    Next j
```

```
Next i
```

```
End Sub
```

- ◆ Data označená v listu jsou označena jako **Selection**, jednotlivé buňky oblasti mohou být adresovány pomocí **Selection.Cells(i, j)**, kde i a j jsou souřadnice ve vybraných datech (obdoba A1 adresování v listu Excelu) a načteny do normálních proměnných
- ◆ Velikost selekce jako základ informace pro její načtení je zajištěna příkazy

Selection.Rows.Count

Selection.Columns.Count



Výpis dat

```
Worksheets.Add after:=Worksheets("Input")
```

```
ActiveSheet.Name = "Indices,,
```

```
Cells(1, 1) = "Locality"
```

```
Cells(1, 2) = "Number of species,,
```

```
Cells(1, 3) = "Number of parasites,,
```

První dva příkazy vytváří v sešitu nový list a pojmenovávají jej, jednotlivé buňky listu jsou adresovány pomocí příkazu `Cells(i, j)`, kde `i, j` představují souřadnice buněk v listu (obdoba A1 adresování)

Příklady

1. Připravte data společenstev ze souboru 01 Excel Data.xls pro výpočet indexů diverzity a pro vizualizaci profilu společenstva
2. Vizualizujte křivku abundance druhů
3. Vypočtete diverzitu společenstev pomocí
 1. Maker v excelu
 2. Software PAST
4. Vizualizujte indexy diverzity
5. Porovnejte jednotlivé lokality v rámci typů lokalit
6. Srovnejte různé typy lokalit pomocí standardních statistických testů

