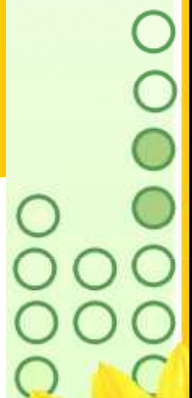
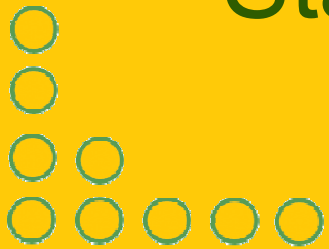




# Statistické hodnocení biodiverzity

Danka Haruštiaková, Jiří Jarkovský

Institut biostatistiky a analýz, Masarykova univerzita, 2015



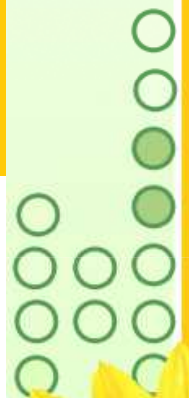
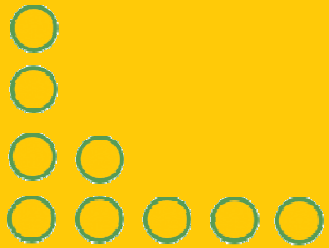
# Program kurzu

- 1.10. Biodiverzita jako pojem + Biodiverzita a biostatistika
- 8.10. Data pro analýzu biodiverzity a jejich vizualizace
- 15.10. Indexy diverzity a jejich statistická spolehlivost
- 22.10. Species-abundance křivky a stochastické modely
- 5.11. Niche-oriented species - abundance modely
- 12.11. Metody hodnocení  $\beta$  diverzity – podobnost společenstev
- 19.11. Metody hodnocení  $\beta$  diverzity – metody shlukové analýzy
- 26.11. Metody hodnocení  $\beta$  diverzity – ordinační metody
- 3.12. Metody hodnocení  $\beta$  diverzity – kanonické ordinační metody
- 10.12. Případová studie: Ptačí společenstva
- 17.12. Kolokvium



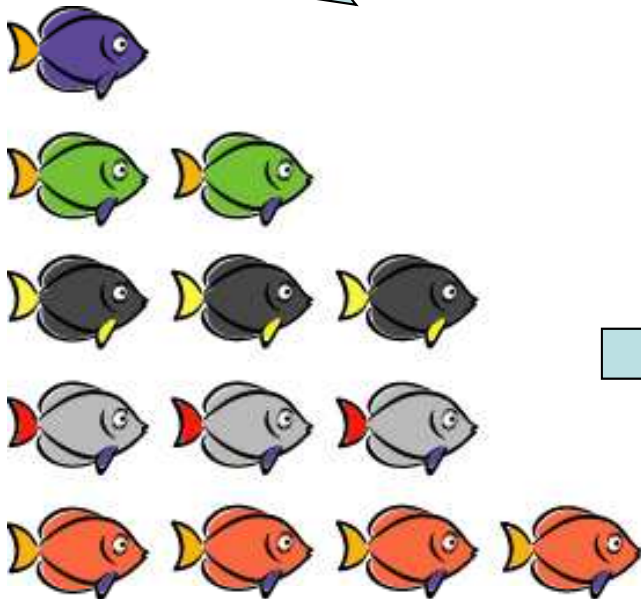
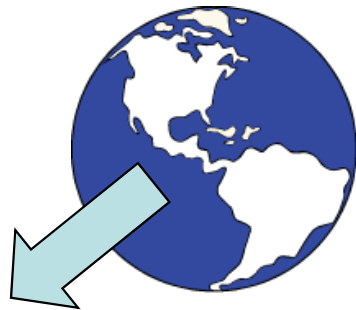


# I. Biodiverzita jako pojem



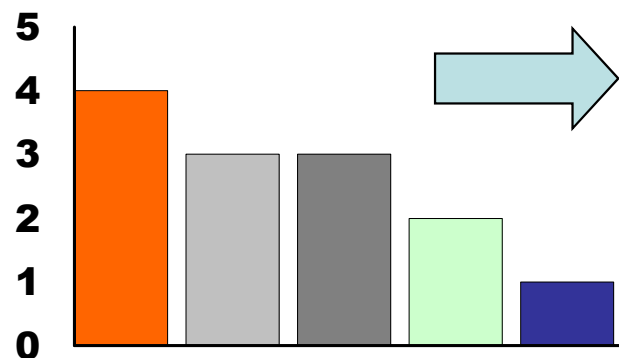
# Biodiverzita ?

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



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

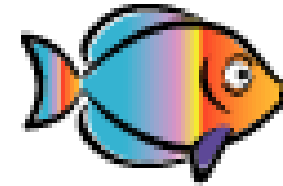
**5**



**Proč ?**



# Definice biodiversity I



- Mnoho definic biodiversity 😊
  - Nejjednodušší:

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

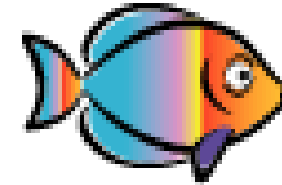
- Komplexní definice:

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)

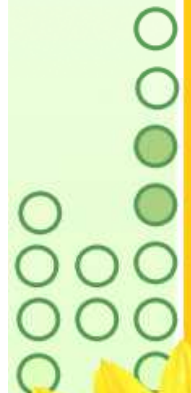




# Definice biodiverzity II

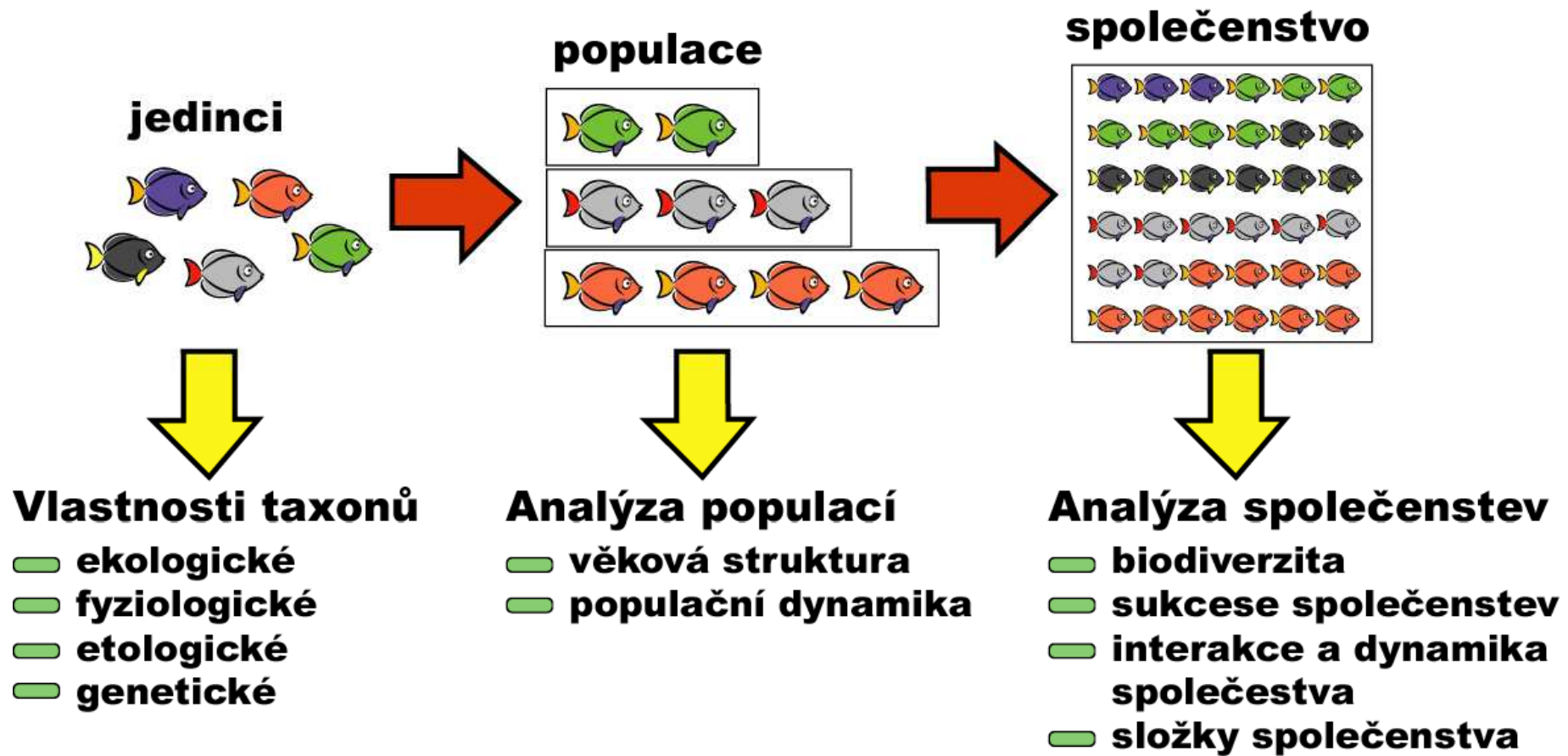


- Biodiverzita je odrazem pestrosti organismů
- V zcela nejjednodušším významu odpovídá biodiverzita počtu druhů
- V složitějším významu odráží také relativní abundance taxonů, tj. strukturu společenstva
- Kromě druhového složení je možné popisovat diverzitu i na nižší nebo vyšší úrovni než je organismus
  - 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



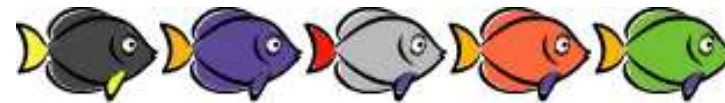
# Jedinci, populace a společenstva

- 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

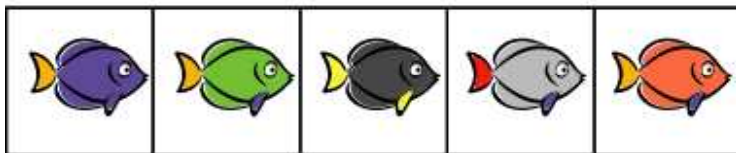




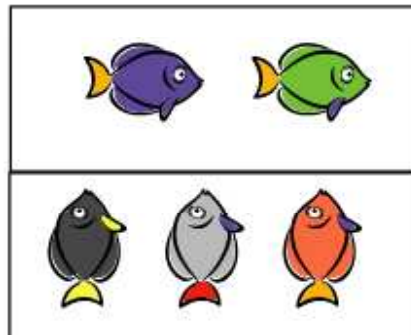
# Různorodost organismů (kvalita)

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

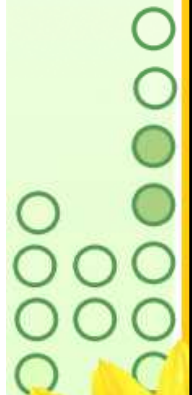
## Taxonomie



## Skupiny - ekologie, etologie, fyziologie



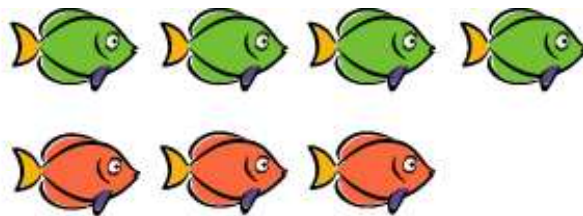
- ◆ Použití různých typů rozdělení organismů poskytuje různé pohledy na složení a typ jejich společenstva, proto je vhodné použít při hodnocení více možností kvalitativního přístupu
- ◆ V případě některých organismů lze použít pouze určité přístupy k jejich biodiverzitě (např. z důvodu nedostatku informací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



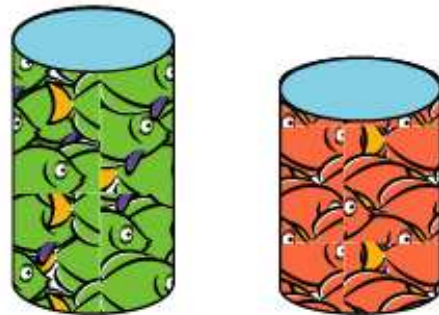
# Kvantita organismů

- Kvantita organismů může být měřena různými způsoby

## Počty



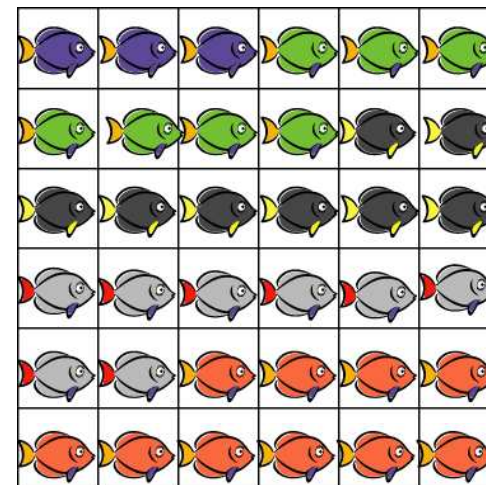
## Biomasa



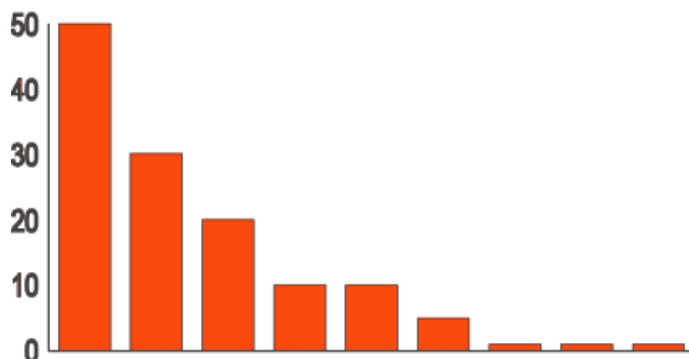
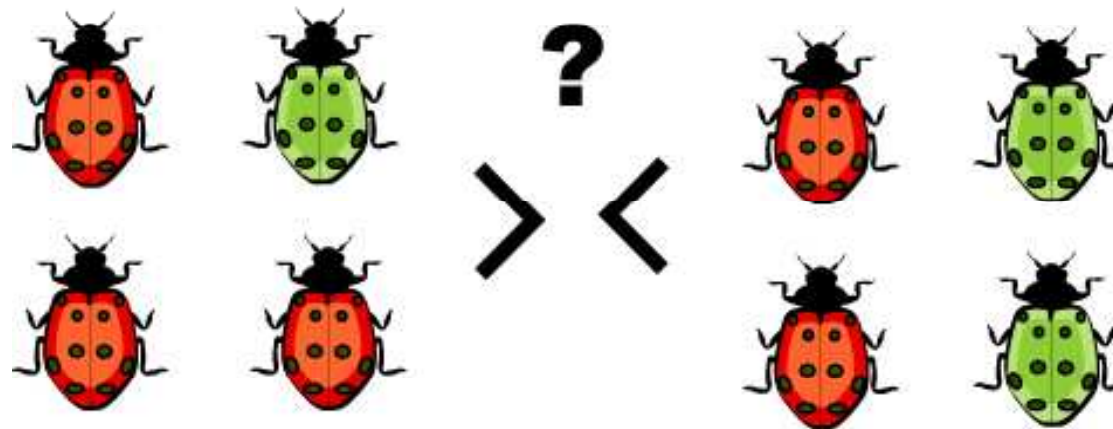
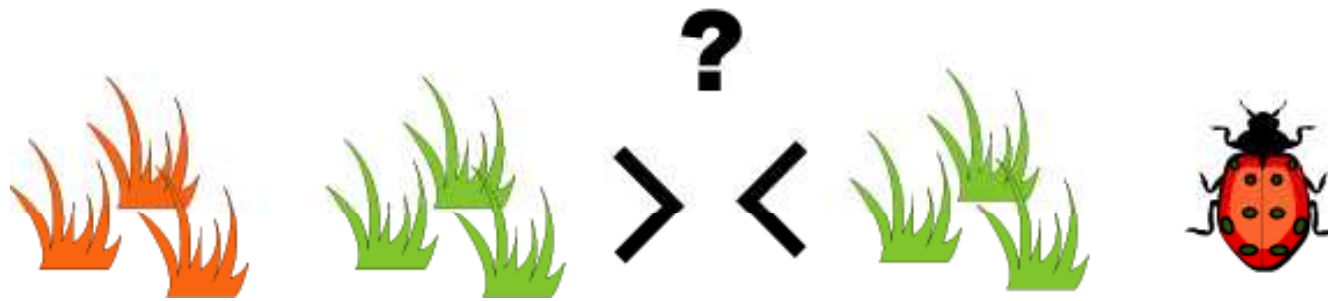
## Aktivita



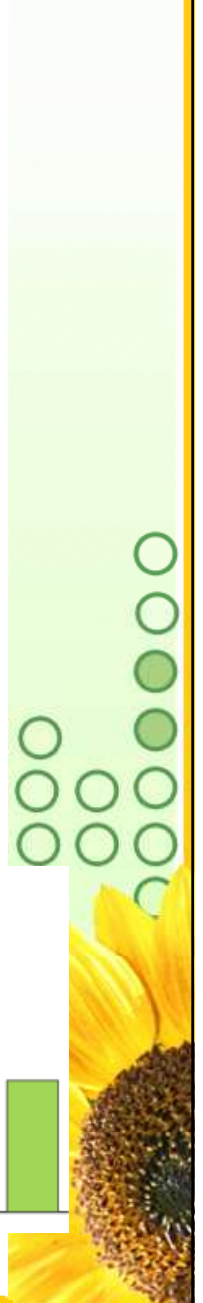
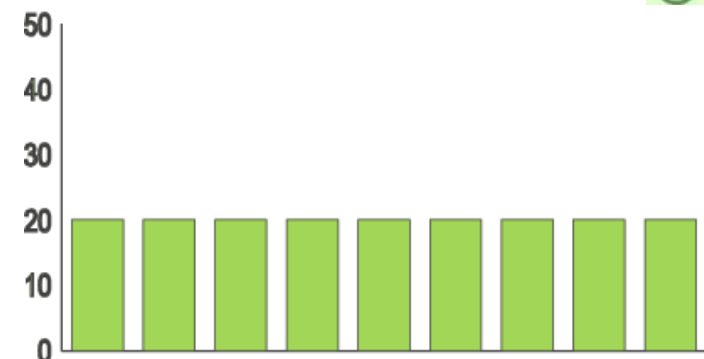
- ◆ Různé způsoby měření mají použití pro určité typy organismů nebo za určitých situací
- ◆ Kvantita organismů odráží obsazení prostředí organismy – lze analyzovat vzhledem k parametrům a historii daného prostředí



# Problémy s biodiverzitou – co znamená větší diverzitu ?



?



# Biodiverzita v současné biologii a environmentálních vědách

Biodiverzita je jedním z klíčových konceptů v biologii, který je definován na více úrovních ...

**Geny**

**Jedinci**

**Populace**

**Druhy**

**Společenstva**

**Ekosystémy**

... a velice vhodný pro bioindikaci

**Vlivy minulosti**

**Vlivy trvalé expozice**

**Následek expozice nízkými dávkami**

**Následek vztahů ve společenstvu**

**Biodiverzita nese velice užitečnou informaci i v případě, kdy „krátkodobé“ metody zlyhají (biotesty, chemické analýzy)**

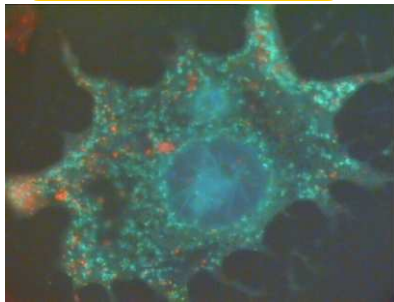




# Koncept biodiverzity využitelný na všech úrovních živé přírody

Biodiverzita jako jedno z nejkompexnějších a integrujících měření živé přírody

**Buňka**



**Organismus**



**Populace**



**Společenstvo**



Přírozený stresový faktor

Časová fluktuace

Prostorová heterogenita





# Biologická společenstva jako typický endpoint na úrovni ekosystému

## VÝHODY



## NEVÝHODY

Dlouhá „paměť“

Interpretace na úrovni ekosystému

Vztah k fungování a stabilitě ekosystému

End-point s evolučním pozadím

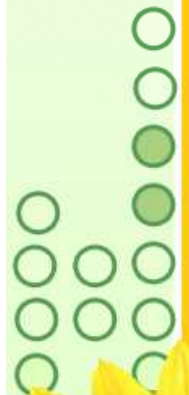
Časovo náročné měření

Problém s reprezentativním vzorkováním

Závislé na expertech (taxonomisté, ...)

Dlouhodobá reprezentativní data potřebná pro srovnávání

Diverzita = variabilita: mnoho maskujících efektů



# Biologická společenstva jako typický endpoint na úrovni ekosystému

## NEVÝHODY

Časovo náročné měření

Problém s reprezentativním vzorkováním

Závislé na expertech (taxonomisté, ...)

Dlouhodobá reprezentativní data potřebná pro srovnávání

Diverzita = variabilita: mnoho maskujících efektů

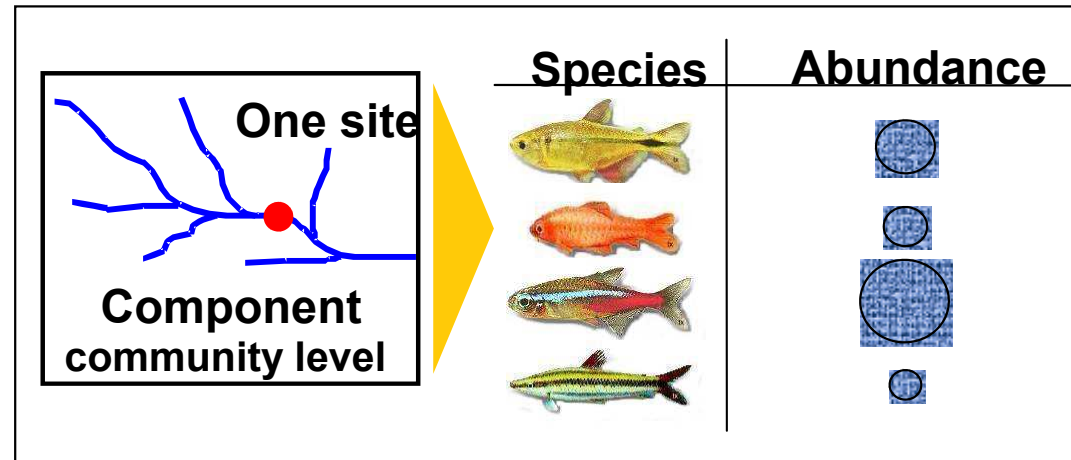
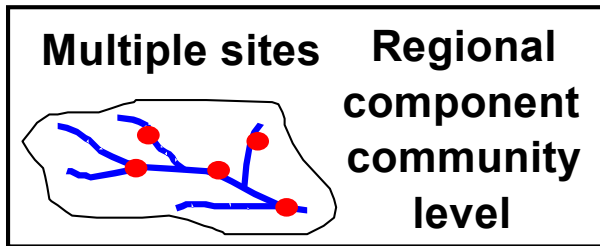
## NÁSLEDKY V BIOMONITORINGU

Trend zjednodušit dizajn studie nebo redukovat počet hodnocených druhů

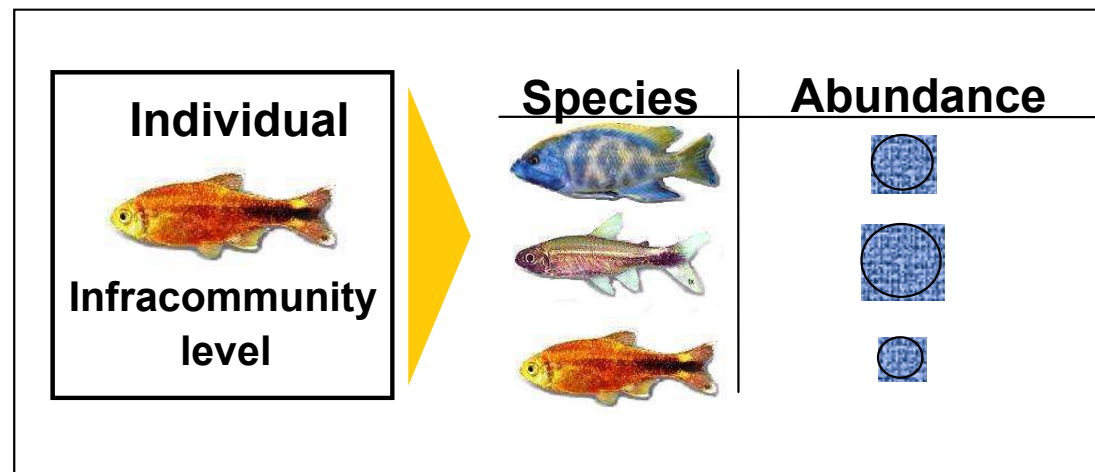
Hledání indikačních druhů složek společenstev, citlivých na stres



# Biodiverzita jako end-point v environmentálních studiích



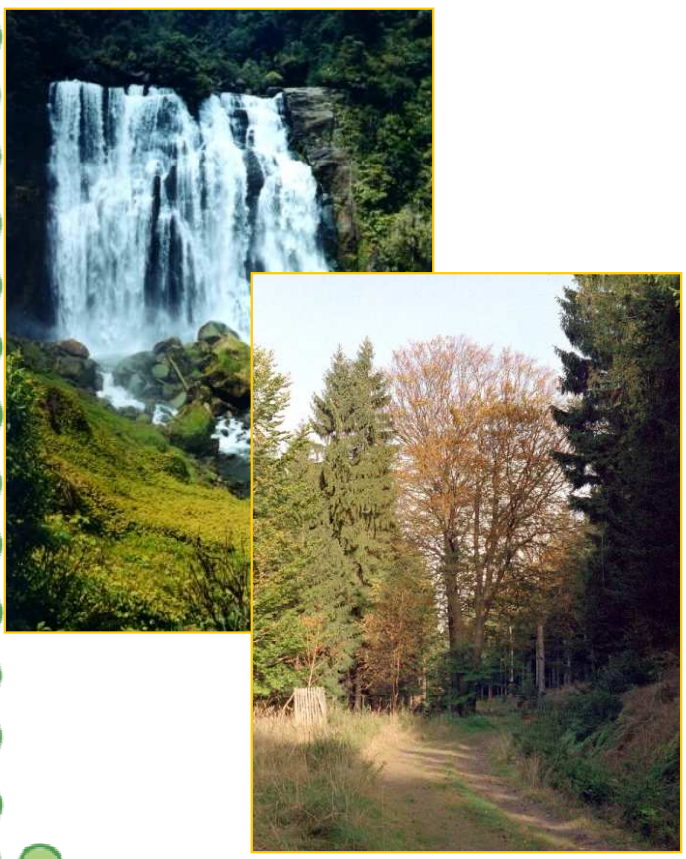
Objektivní stratifikace společenstev silně snižuje velikost vzorku !



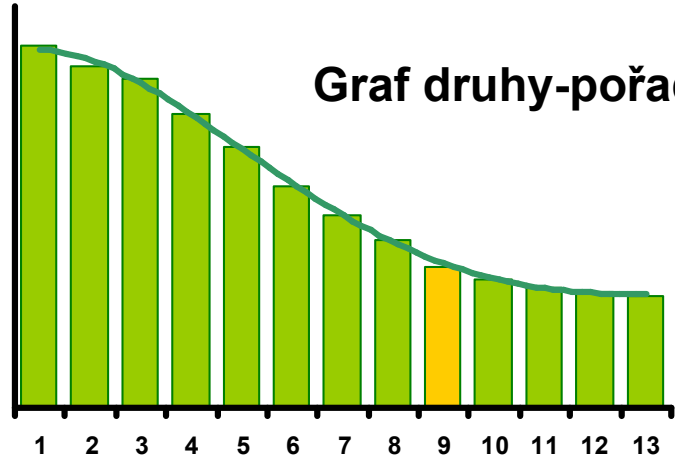
# Jak kvantifikovat a interpretovat komplexně hodnocený endpoint

Jako estetickou přírodu

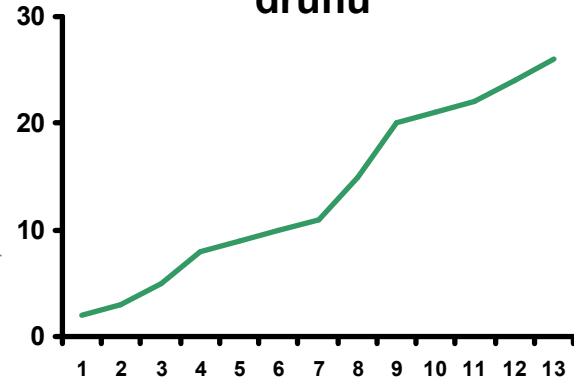
Jako parametricky standardizovaný endpoint



Počet jedinců



Kumulativní abundance druhů



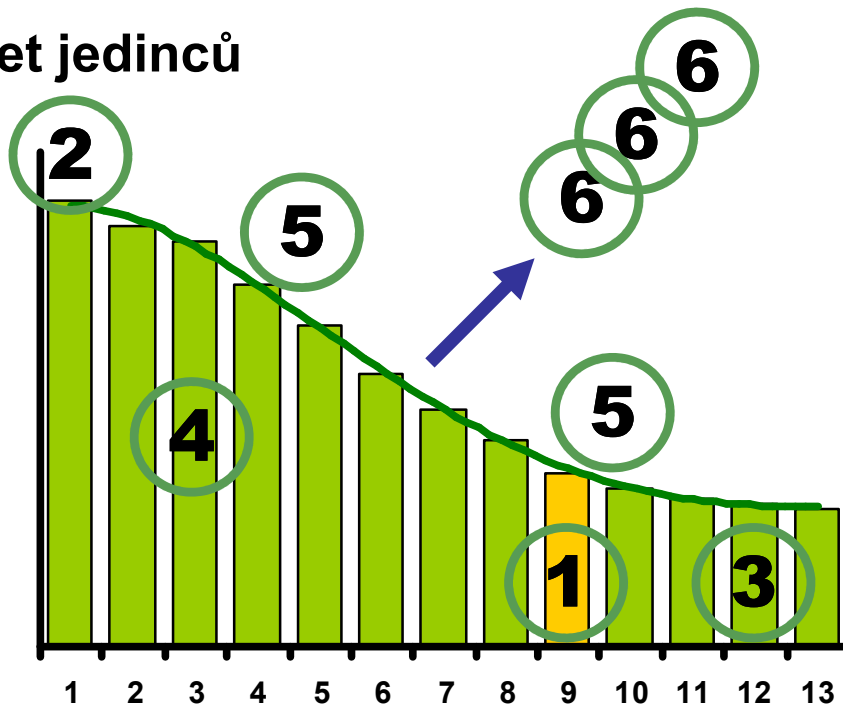
Pořadí druhů



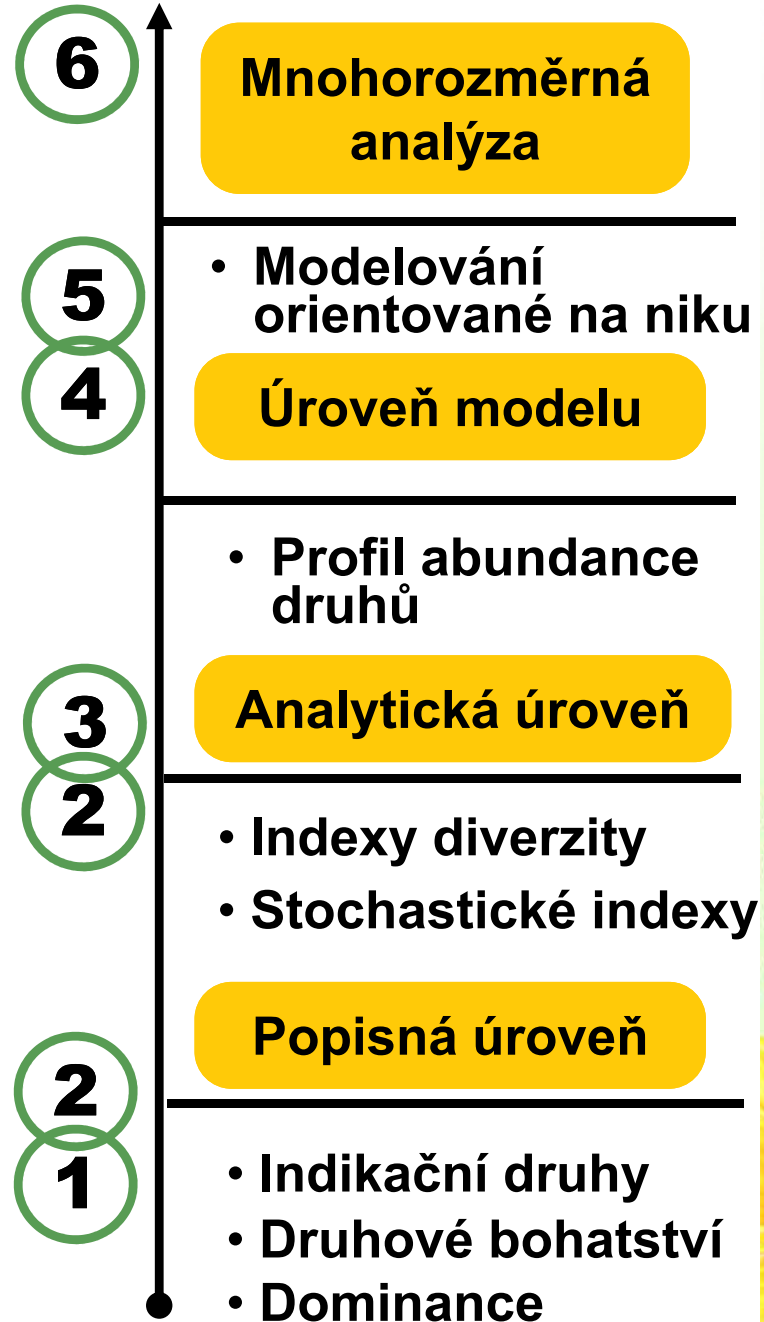
# Komplexní endpointy poskytují rozsáhlý seznam možných parametrů

**Profil abundance druhů jako standardní výstup**

Počet jedinců

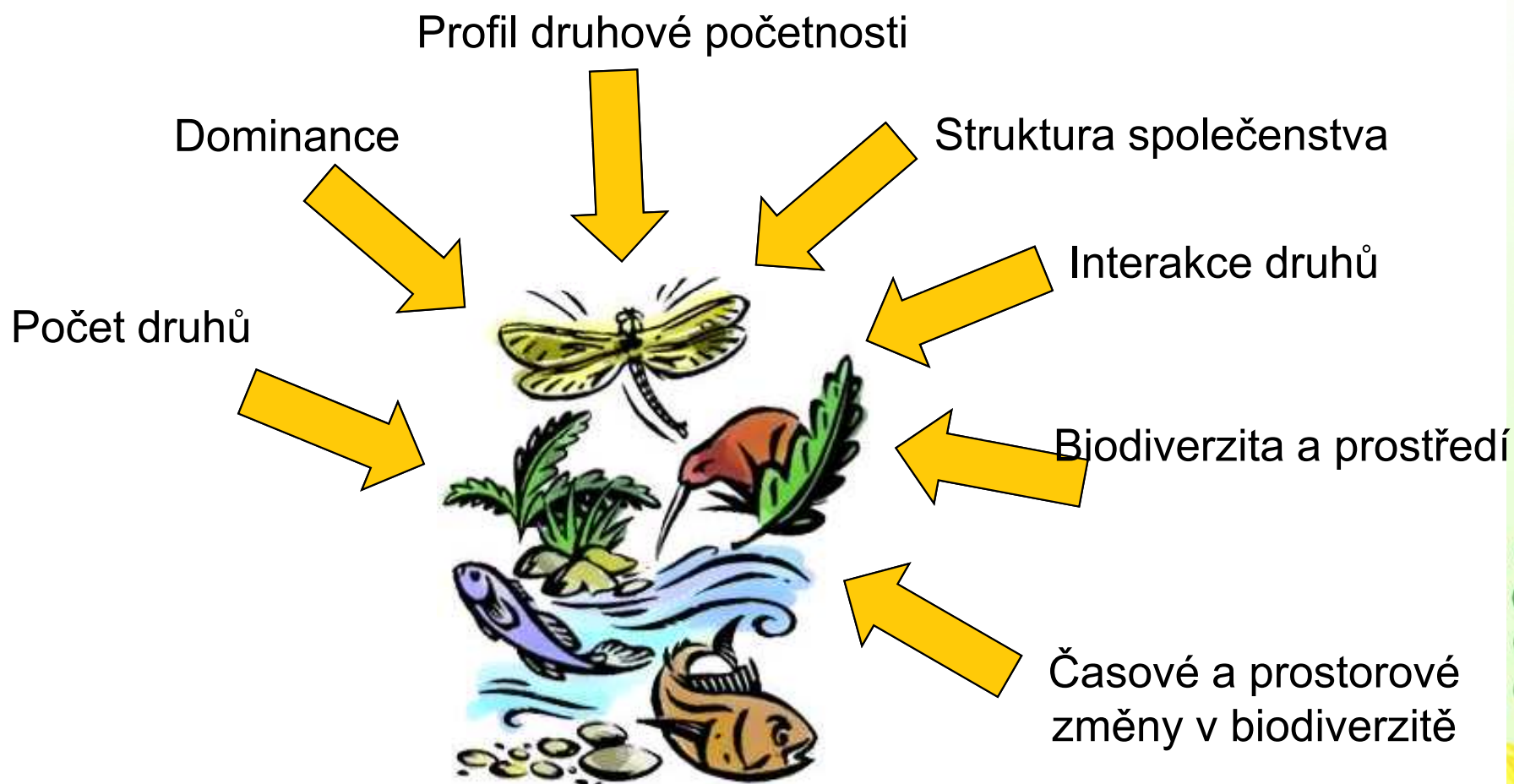


Pořadí druhu





# Otázky v analýze dat biodiverzity

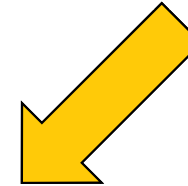
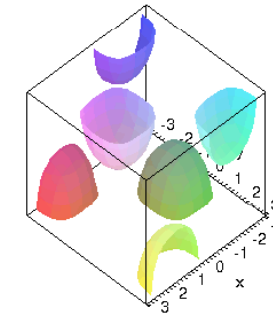


# Metody analýzy biodiverzity

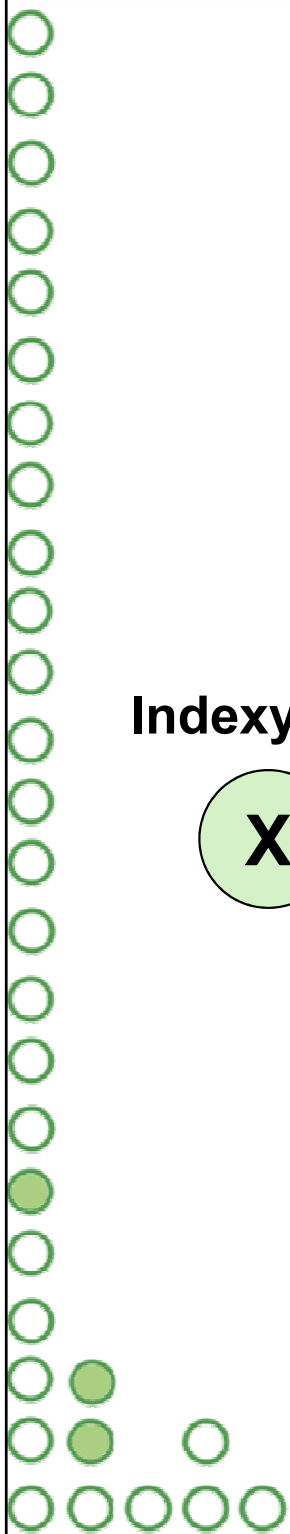
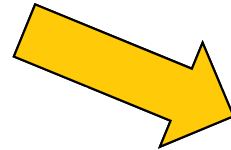
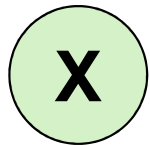
Modely druhové početnosti

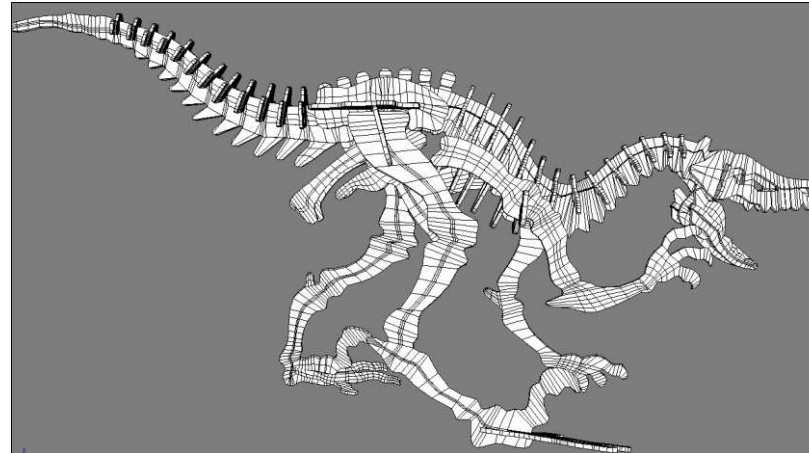


Vícerozměrná analýza

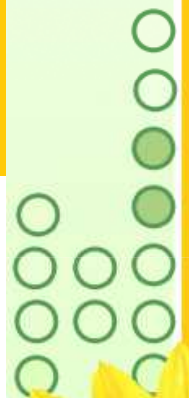
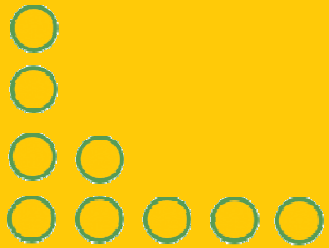


Indexy diverzity





## II. Biodiverzita a biostatistika



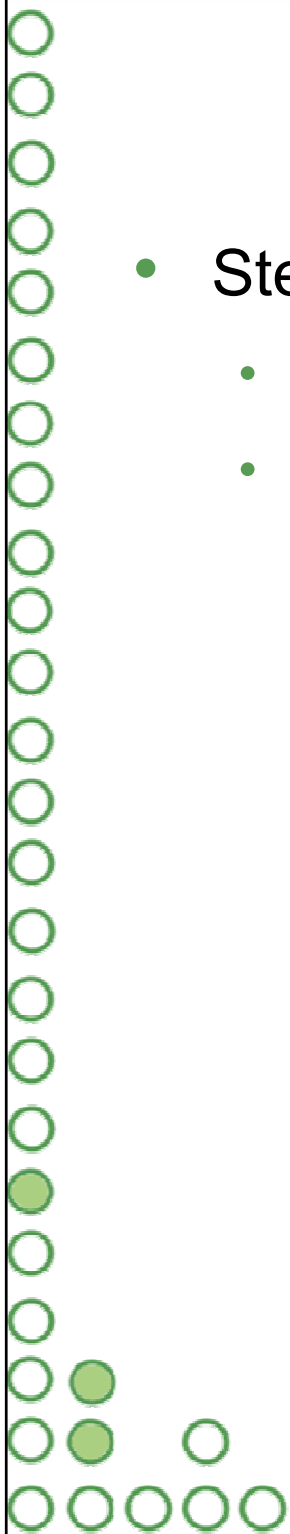
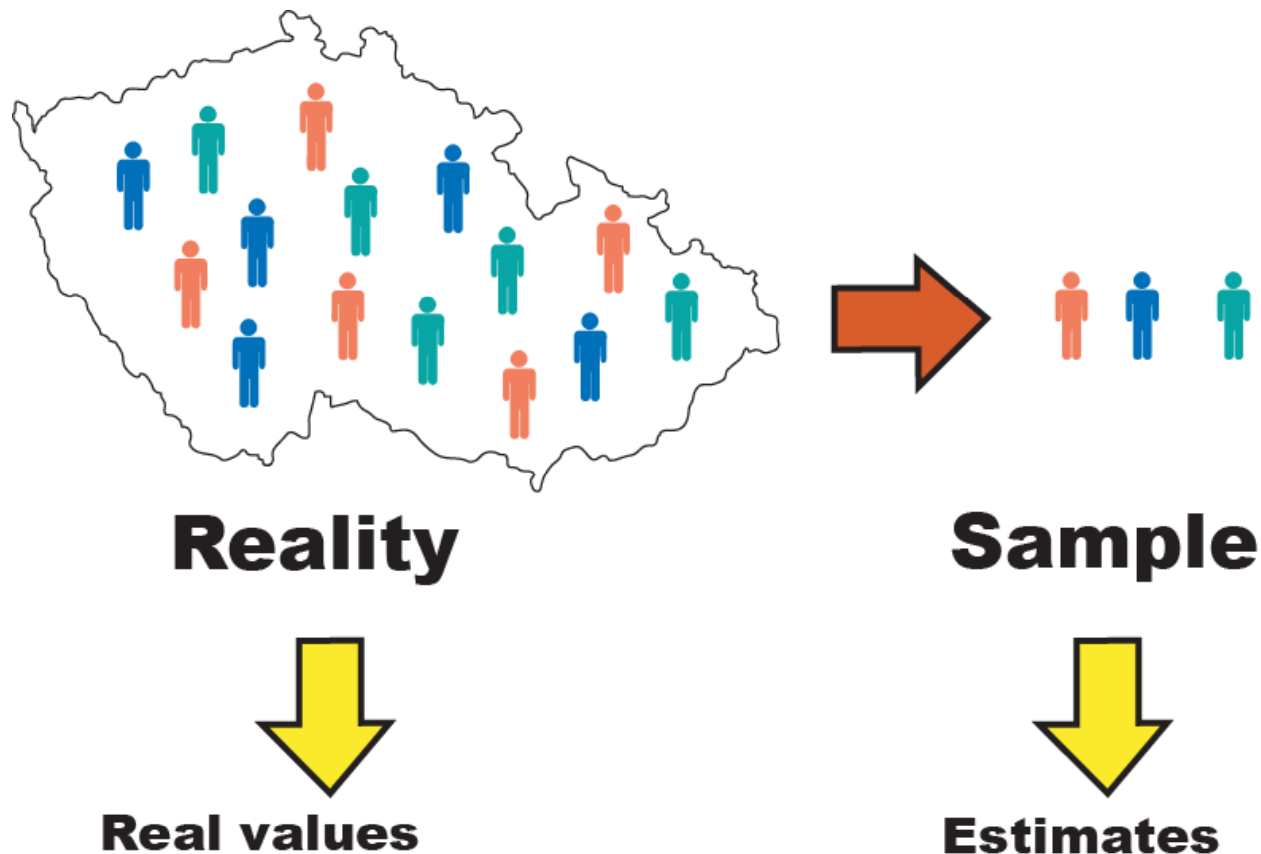
# Vazba mezi biodiverzitou a biostatistikou?

- 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



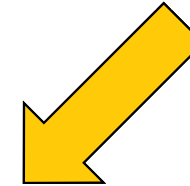
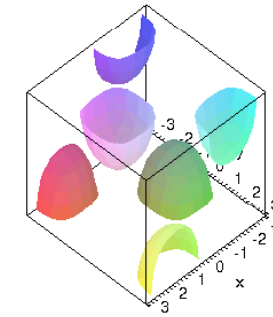


# Metody analýzy biodiverzity

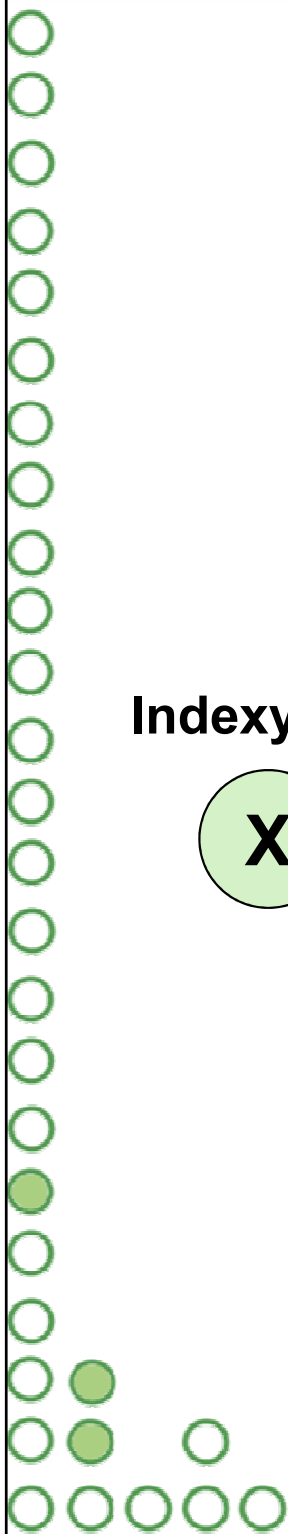
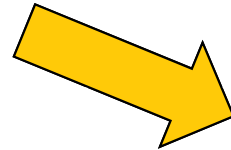
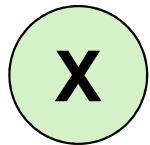
Species abundance models



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)



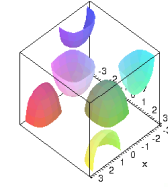
# Species abundance models: výhody a nevýhody



- 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

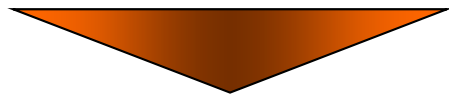


- Na data biodiverzity může být aplikována řada shlukovacích, ordinačních, regresních a klasifikačních vícerozměrných technik.
- Tyto metody hledají v rozsáhlých datech **vícerozměrné vzory společenstev** umožňující odpovědět na následující otázky:
  - Vztah druhů k prostředí
  - Prostorové vztahy
  - 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

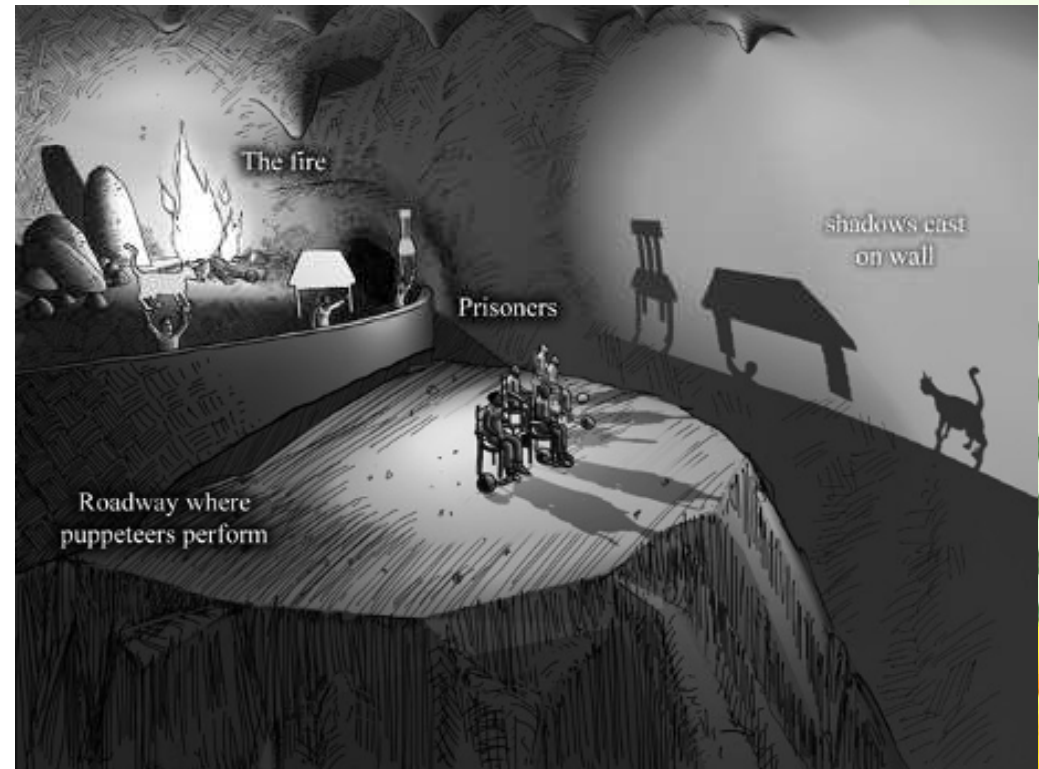


# Výzkum, realita, statistika

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



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





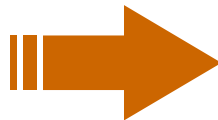
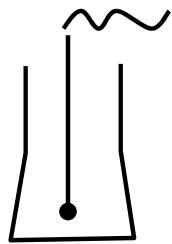
# Význam variability

- 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



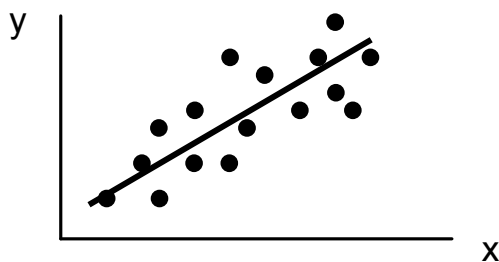
182 cm



163 cm

rozptyl znaku, přirozená variabilita

## Variabilita modelovaných dat



chyba = nepřesnost modelu

## Variabilita časových řad



fluktuace, časová proměnlivost

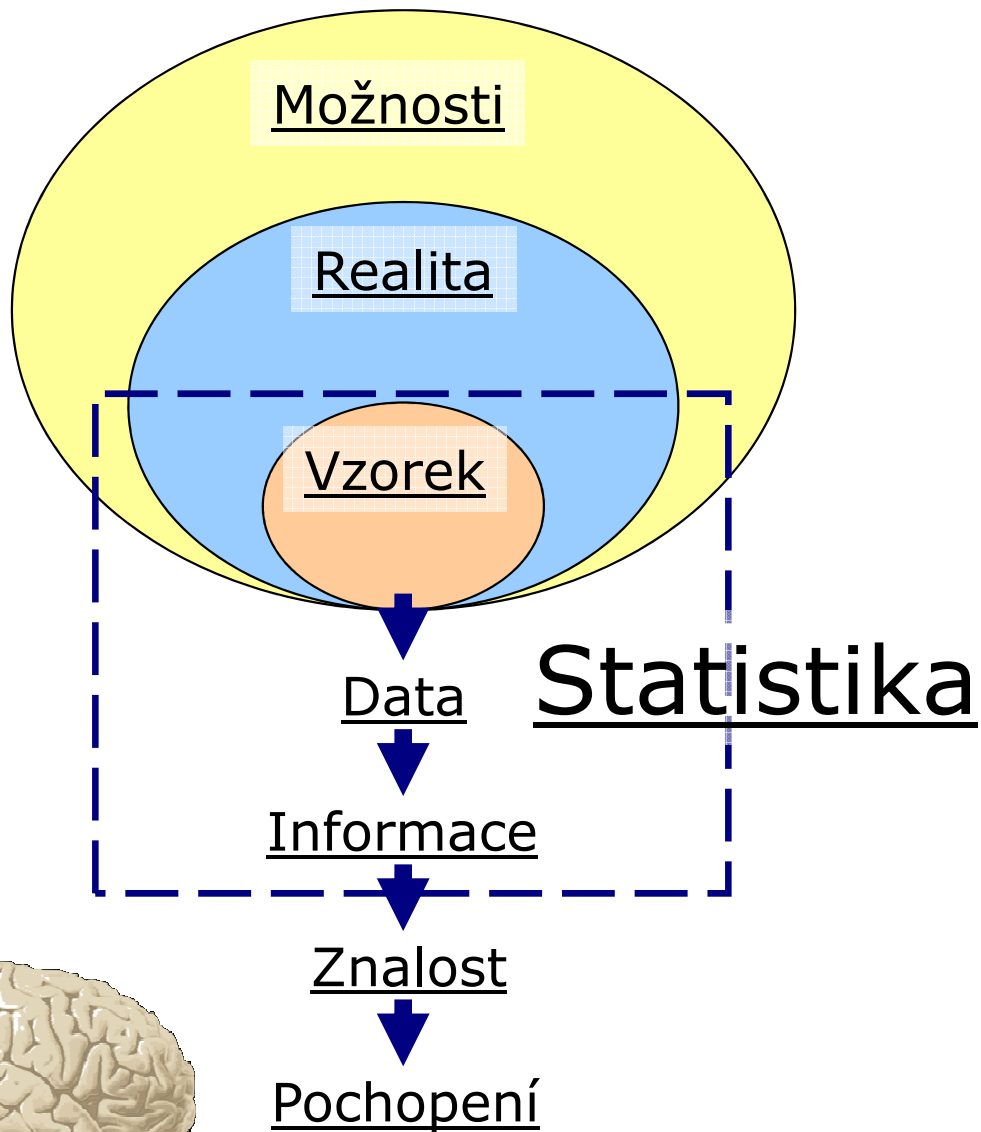
## Variabilita ve skladbě biologických společenstev

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



biodiverzita

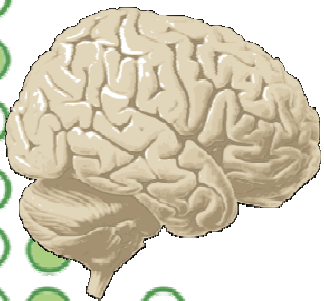
# Co může statistika říci o naší realitě?



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

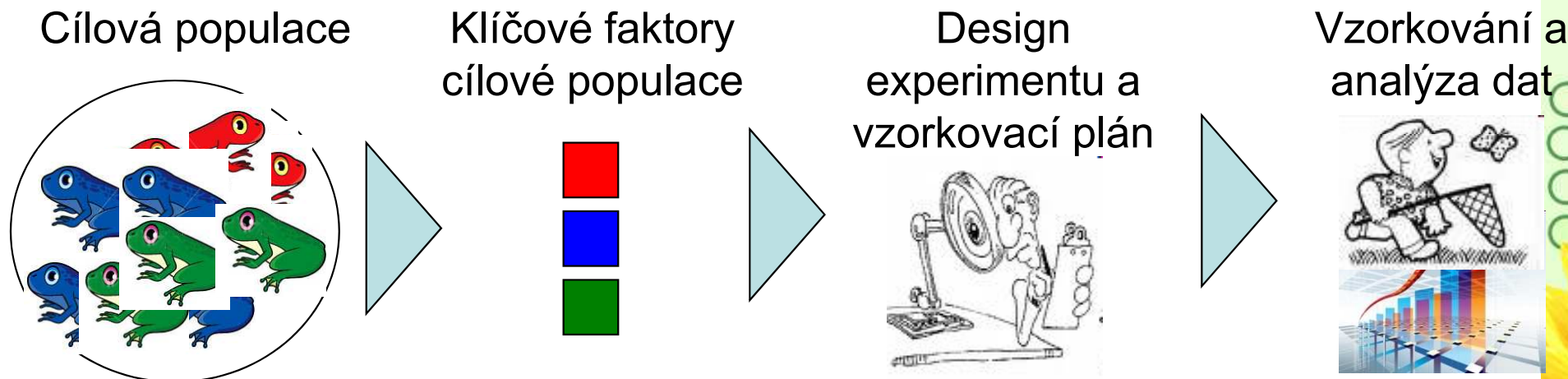
Statistika je nasazena v procesu získání informací z vzorkovaných dat a je podporou v získání naší znalosti a pochopení problému.

Statistika není náhradou naší inteligence !!!



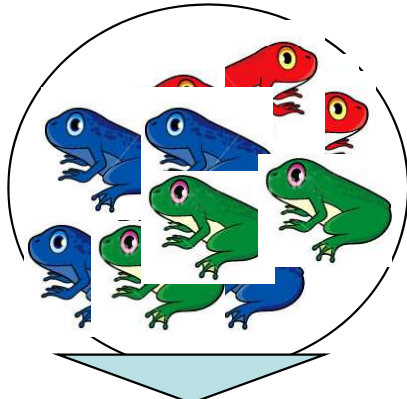
# Cílová populace

- **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





# Statistika a zobecnění výsledků



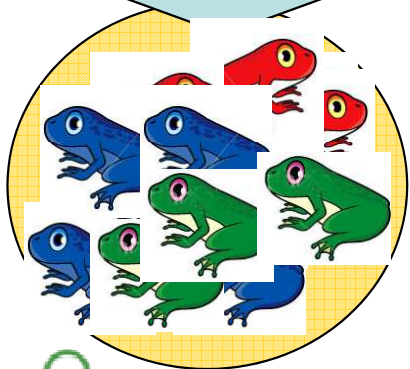
Neznámá  
cílová  
populace



Vzorek



Analýza



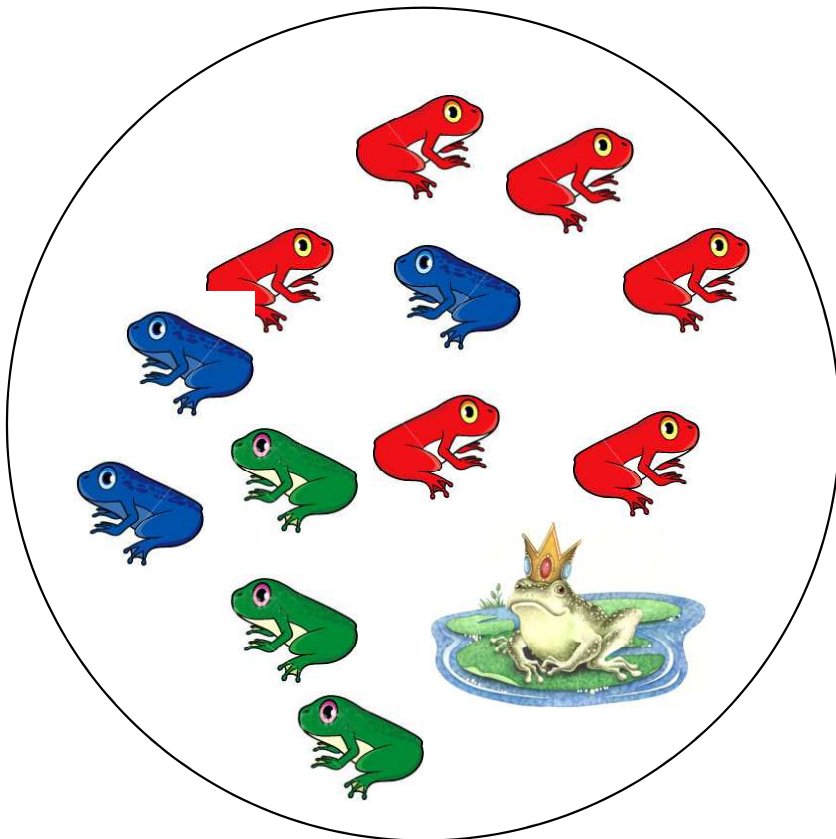
Díky zobecnění  
výsledků známe  
vlastnosti cílové  
populace

- 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

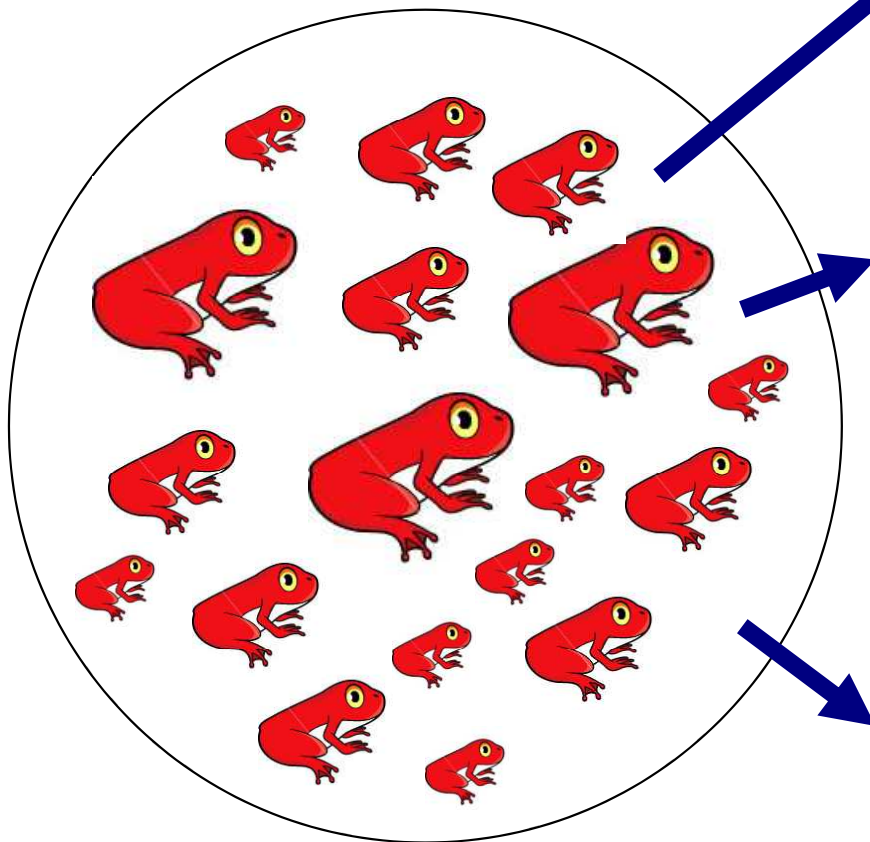


Nezávislost: několikanásobné vzorkování téhož objektu nepřináší ze statistického hlediska žádnou novou informaci



# Velikost vzorku a přesnost statistických výstupů

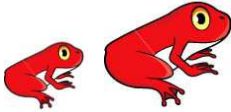
Existuje skutečné rozložení  
a skutečný průměr měřené  
proměnné



Z jednoho měření nezjistíme nic

Vzorek:  → ?????

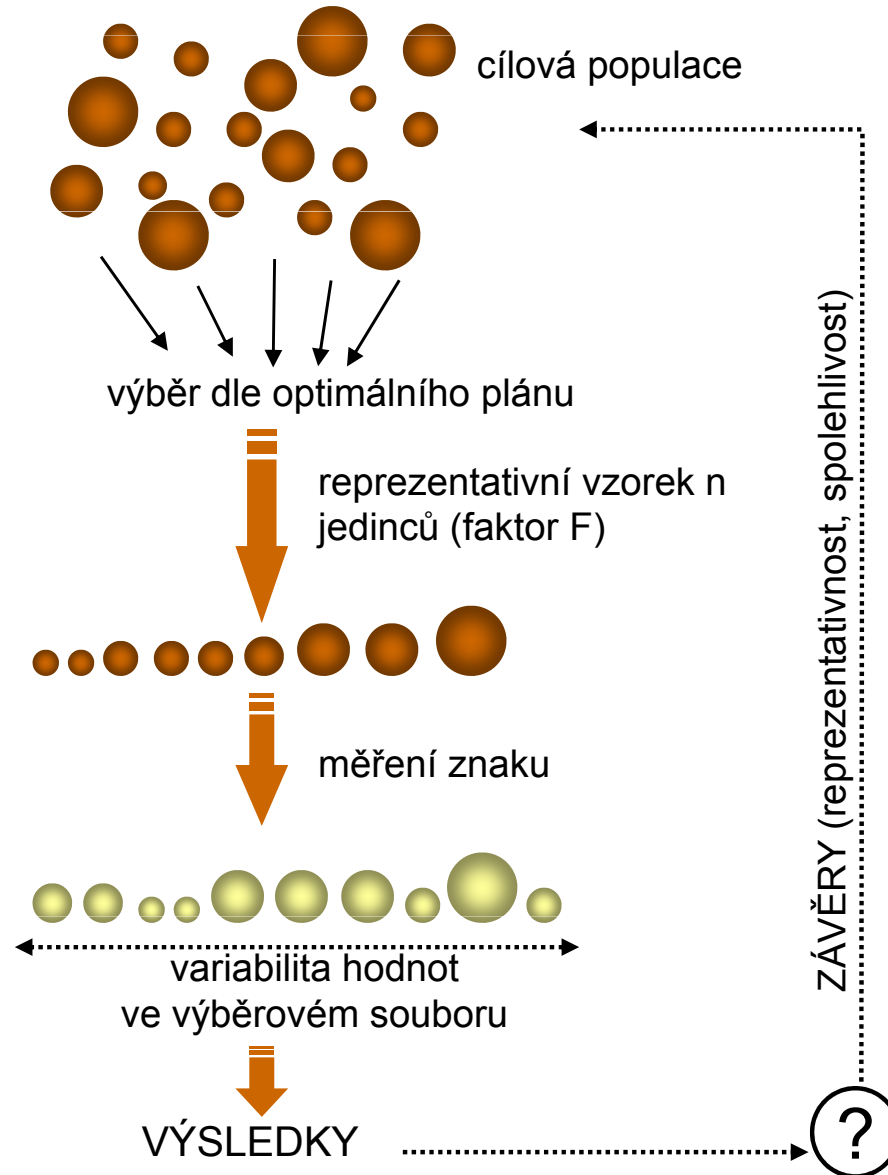
Vzorek určité velikosti poskytuje  
odhad reálné hodnoty s definovanou  
spolehlivostí

Vzorek:  → Odhad  
průměru  
atd.

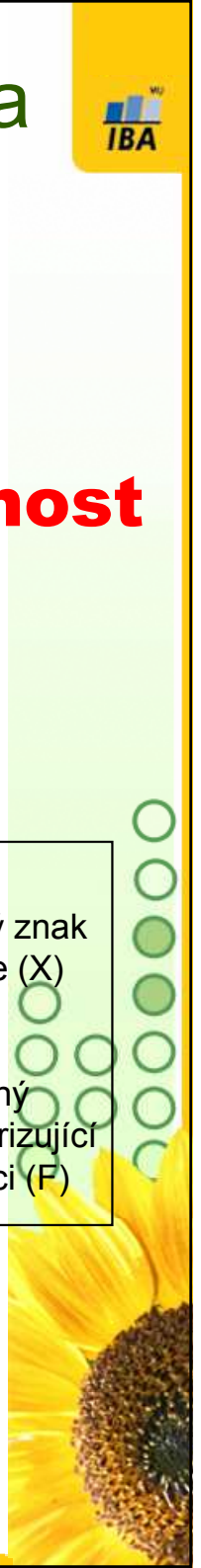
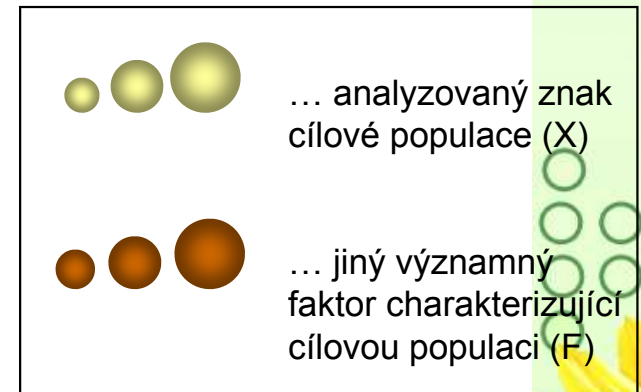
Vzorkování všech existujících  
objektů poskytne skutečnou  
hodnotu dané popisné statistiky,  
nicméně tento přístup je ve většině  
případech nereálný.

# Experimentální design: nezbytná výbava biologa

Účel analýzy:  
Popisný



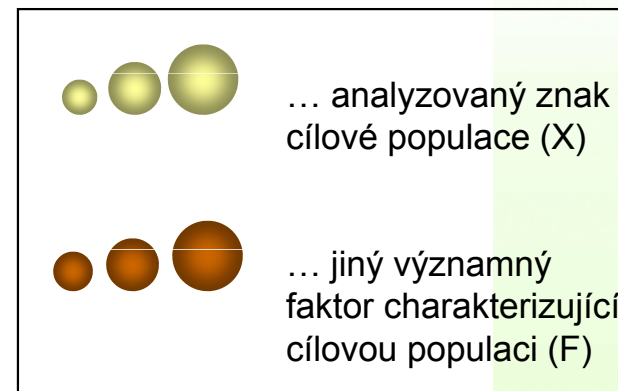
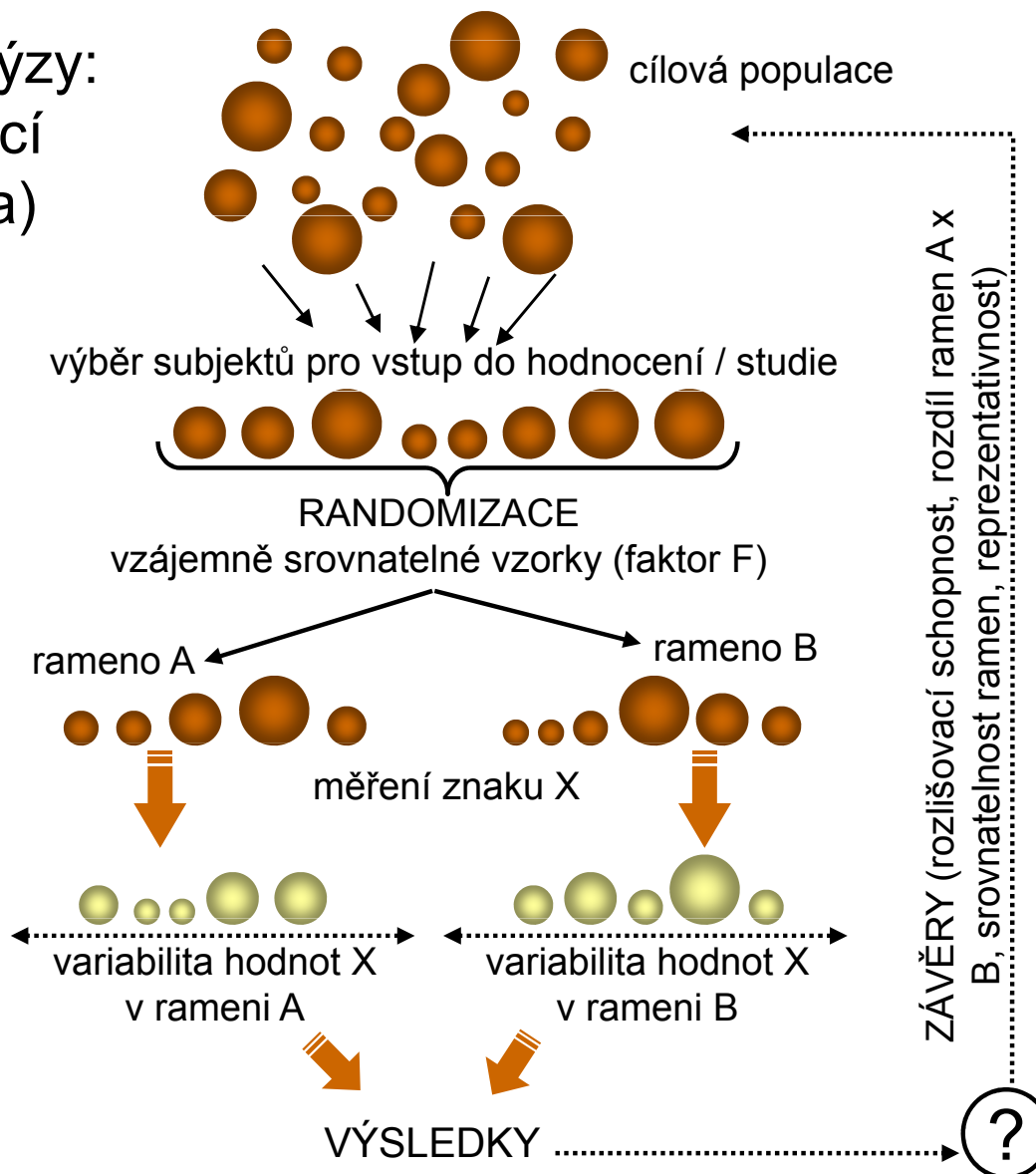
**?**  
**Reprezentativnost**  
**Spolehlivost**  
**Přesnost**





# 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

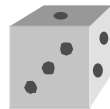
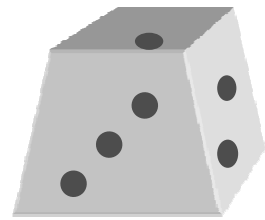
		Praktická významnost	
		ANO	NE
Statistická významnost	ANO	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr	Významný výsledek je statistický artefakt velkého vzorku, prakticky nevyužitelné
	NE	Výsledek může být pouhá náhoda, neprůkazný výsledek	OK, praktická i statistická významnost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



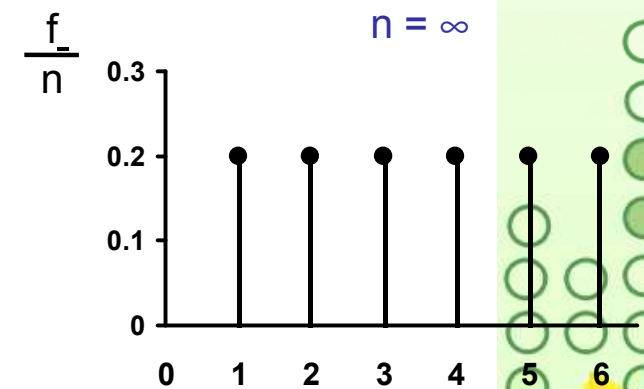
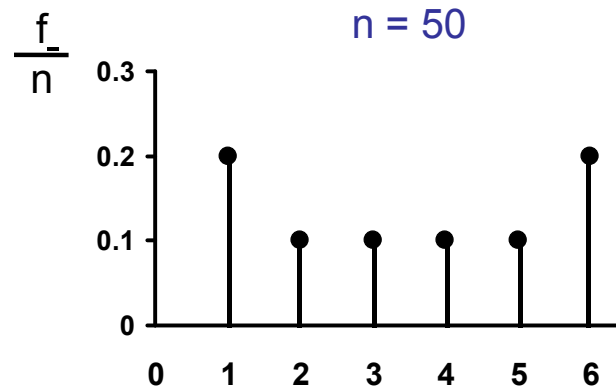
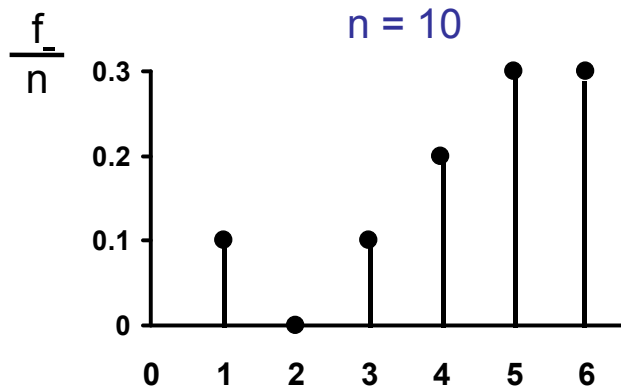
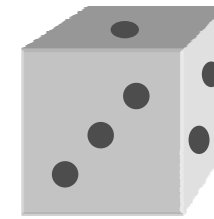
# JAK vznikají informace ?

„Empirical approach“

„Classical approach“



Empirický postup



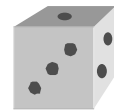
možné jevy: čísla 1 – 6

$n$  – počet hodů (opakování)

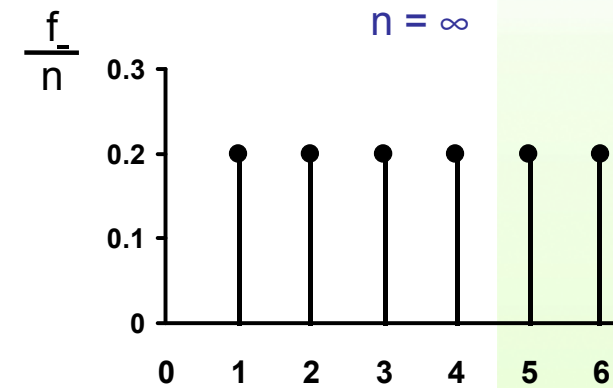
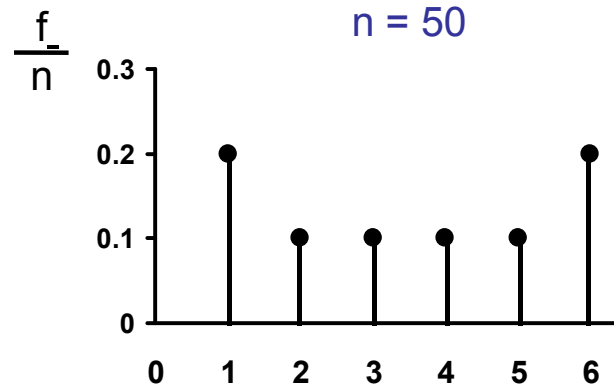
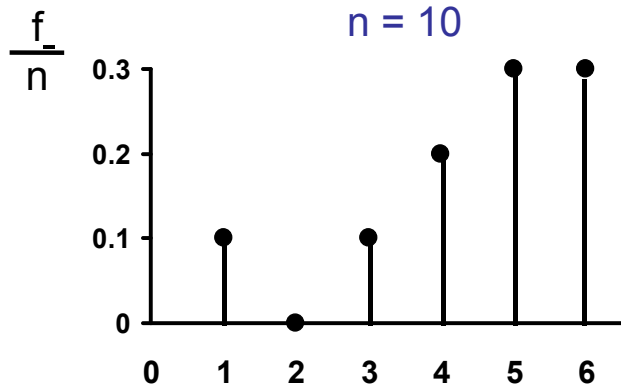
**U složitých stochastických systémů se pravda získá až po odvedení značného množství experimentální práce: musíme dát systému šanci se projevit**



# JAK vznikají informace ?



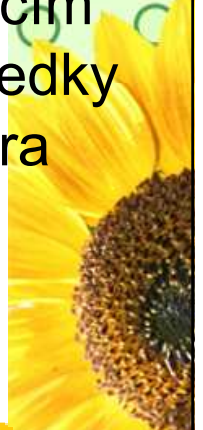
Empirický postup



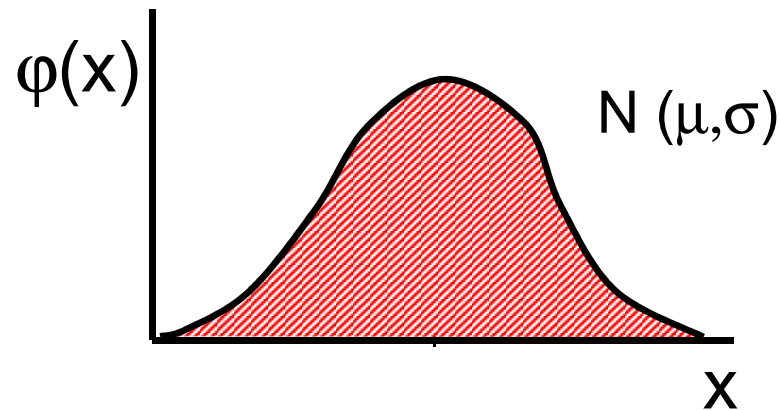
možné jevy: čísla 1 – 6

$n$  – 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



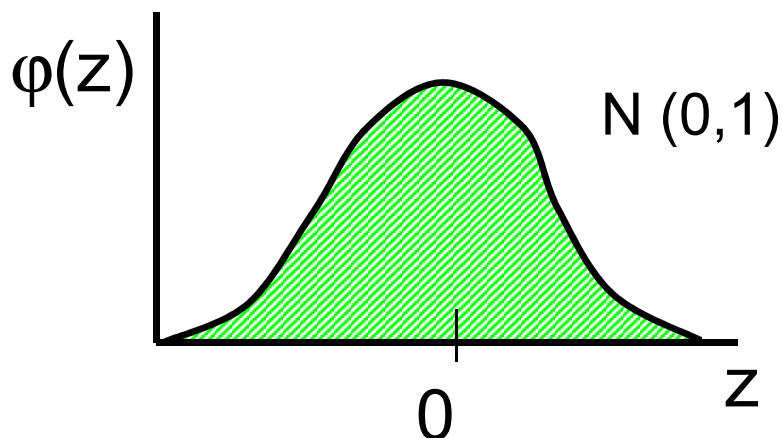
# Rozložení hodnot jako model: Normální rozložení



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

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

**Standardizovaná forma**



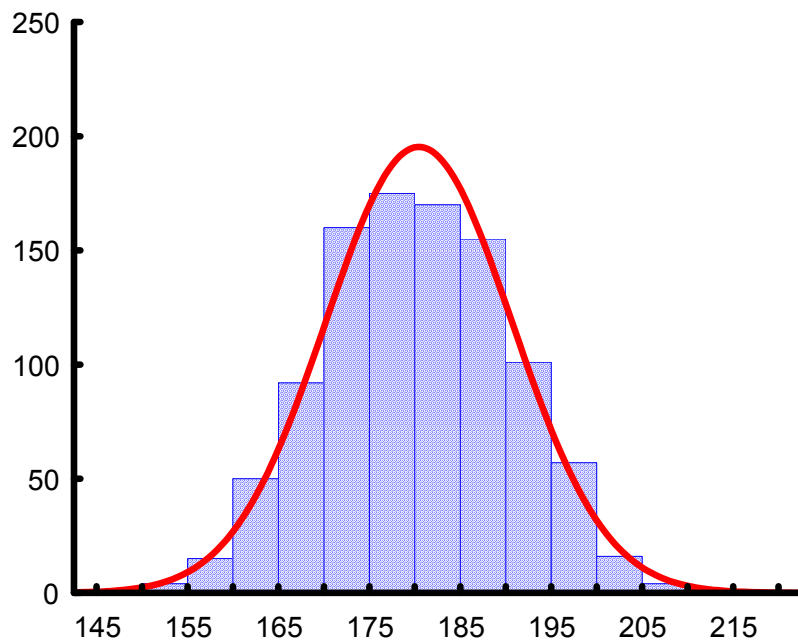
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

**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  $\chi^2$  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.

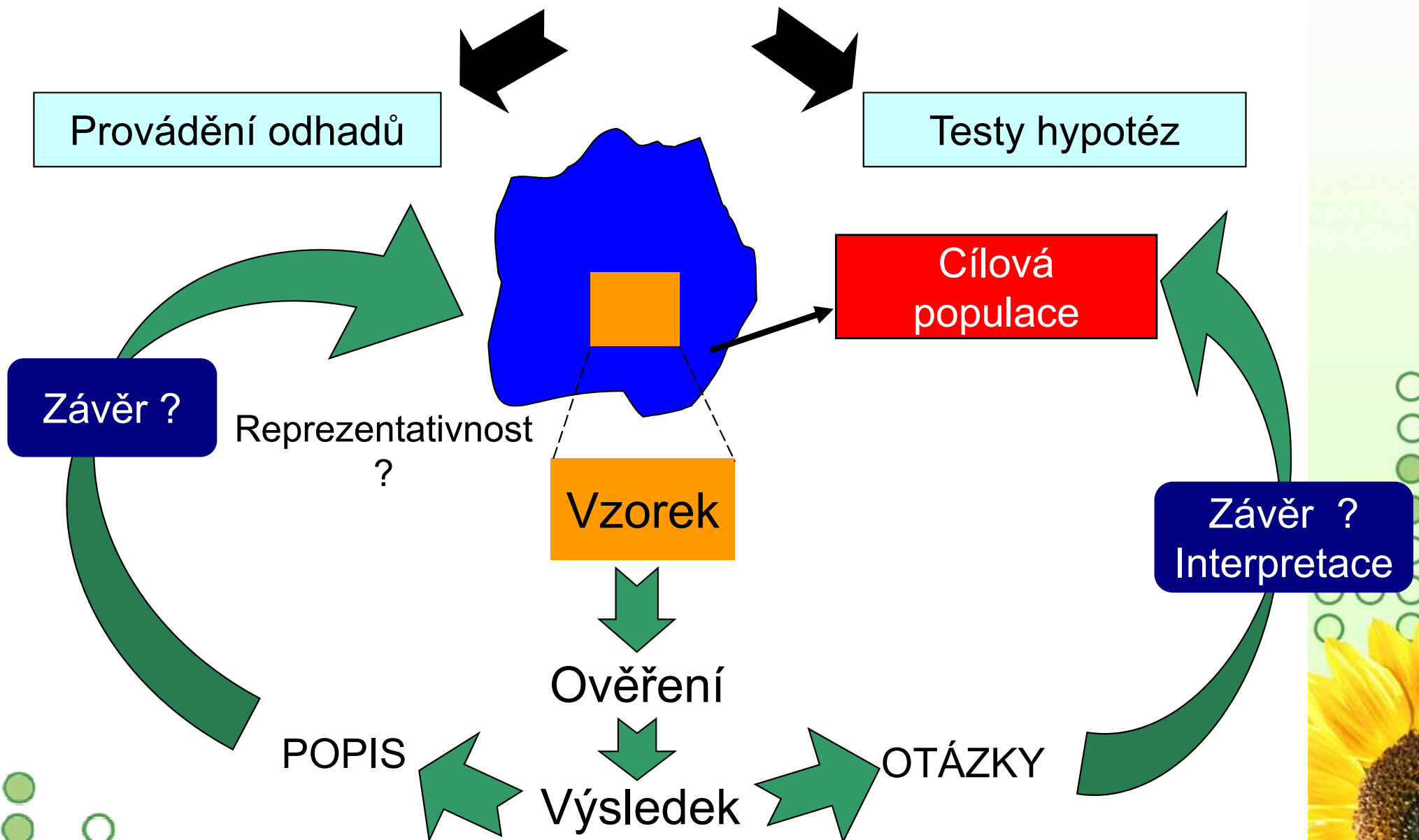




# Stručný přehled modelových rozložení I.

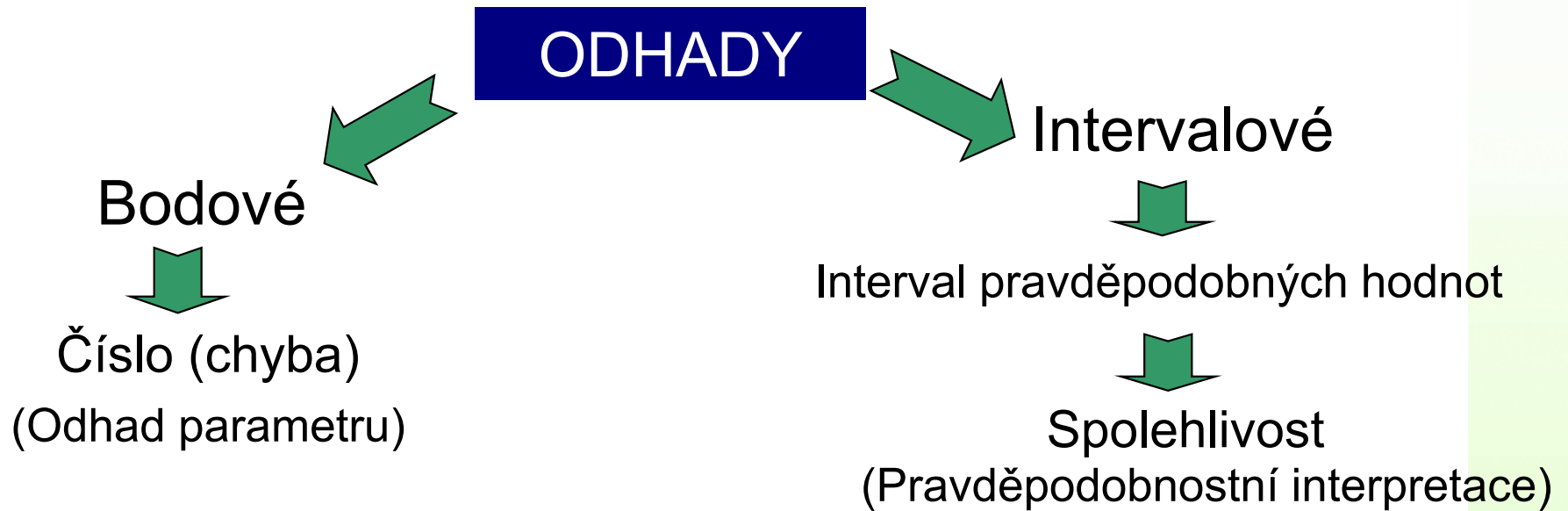
Rozložení	Parametry	Stručný popis
<b>Normální</b>	Průměr ( $\mu$ ) Rozptyl ( $\sigma^2$ )	Symetrická funkce popisující intervalovou hustotu četnosti; nejpravděpodobnější jsou průměrné hodnoty znaku v populaci.
<b>Log-normální</b>	Medián Geometrický průměr Rozptyl ( $\sigma^2$ )	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
<b>Weibullovo</b>	$\alpha$ - parametr tvaru $\beta$ - 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.
<b>Rovnoměrné</b>	Medián Geometrický průměr Rozptyl ( $\sigma^2$ )	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
<b>Triangulární</b>	$f(x) = [b - \text{ABS}(x - a)] / b^2$ $a - b < x < a + b$	Pravděpodobnostní funkce pro typ rozložení, kdy jsou střední hodnoty výrazně pravděpodobnější než hodnoty okrajové.
<b>Gamma</b>	Parametry distribuční funkce: $\alpha$ - parametr tvaru $\beta$ - parametr rozsahu hodnot	Umožňuje flexibilně modelování distribučních funkcí nejrůznějších tvarů. Např. $\chi^2$ 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

±

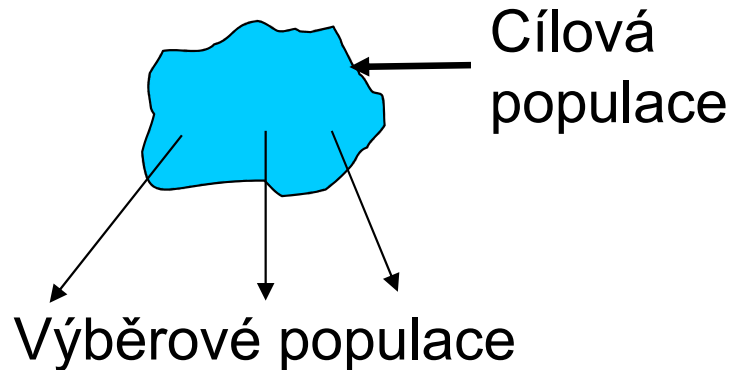
Kvantil  
modelového . SE (odhadu)  
rozložení

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



# Interval spolehlivosti odhadu průměru je pouze informací o přesnosti tohoto odhadu

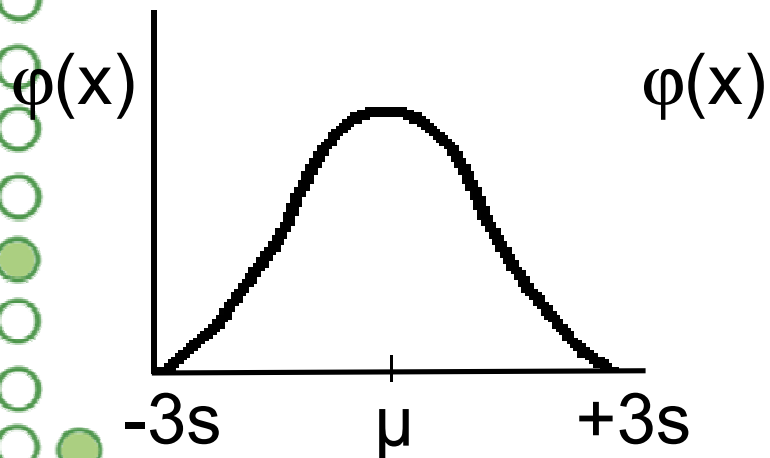
Interval spolehlivosti je hodnocen pro  $(1 - \alpha)$  procentní spolehlivost



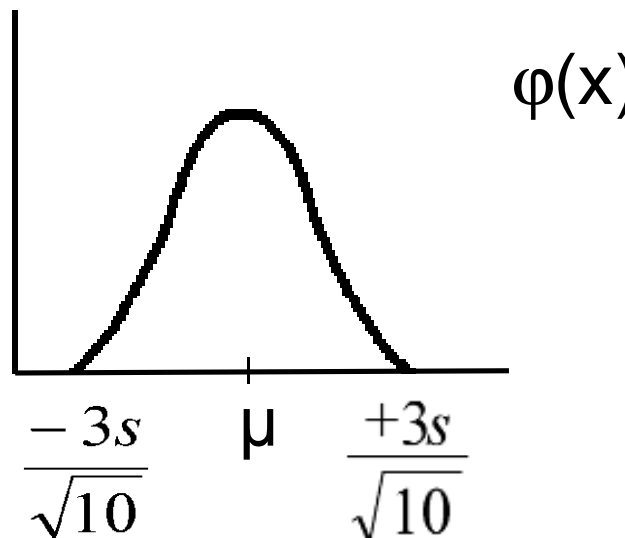
Šířku intervalu určuje:

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

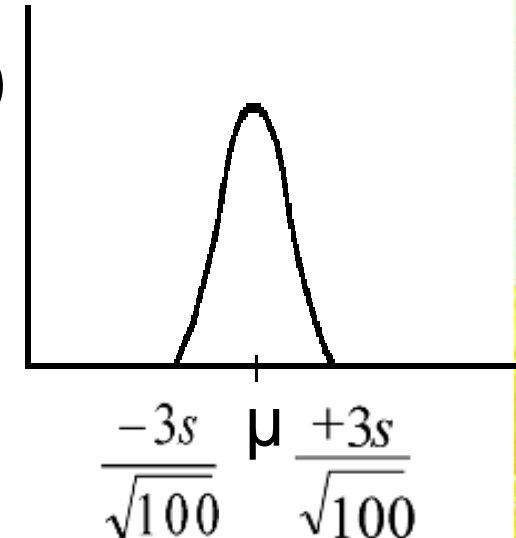
Původní proměnná  $x$



Výběr  $n=10$  pro odhad průměru



Výběr  $n=100$  pro odhad průměru

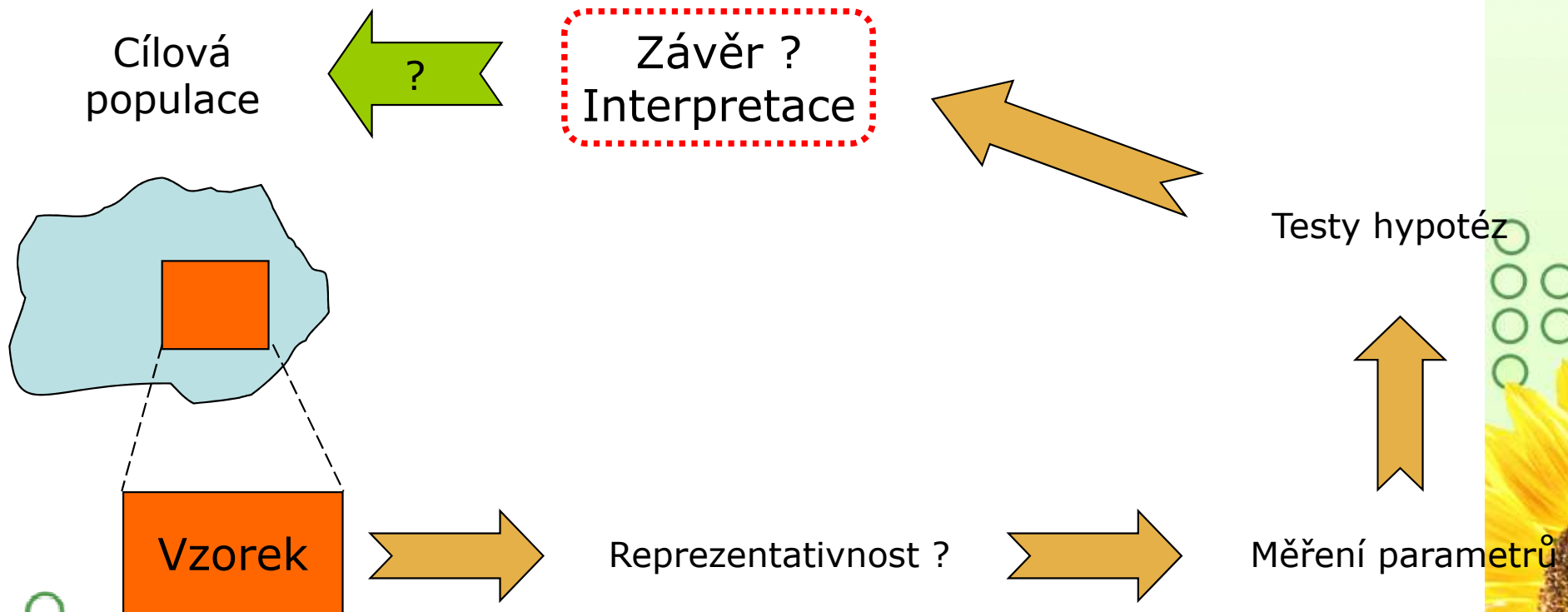






# Princip testování hypotéz

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



# Statistické testování – základní pojmy

➤ Nulová hypotéza  $H_0$

$H_0$ : sledovaný efekt je nulový

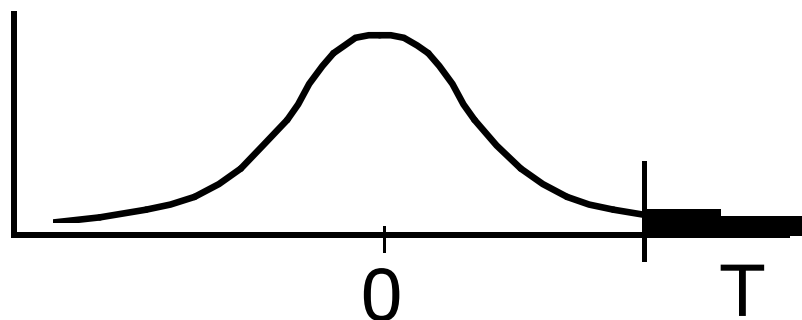
➤ Alternativní hypotéza  $H_A$

$H_A$ : sledovaný efekt je různý mezi skupinami

➤ Testová statistika

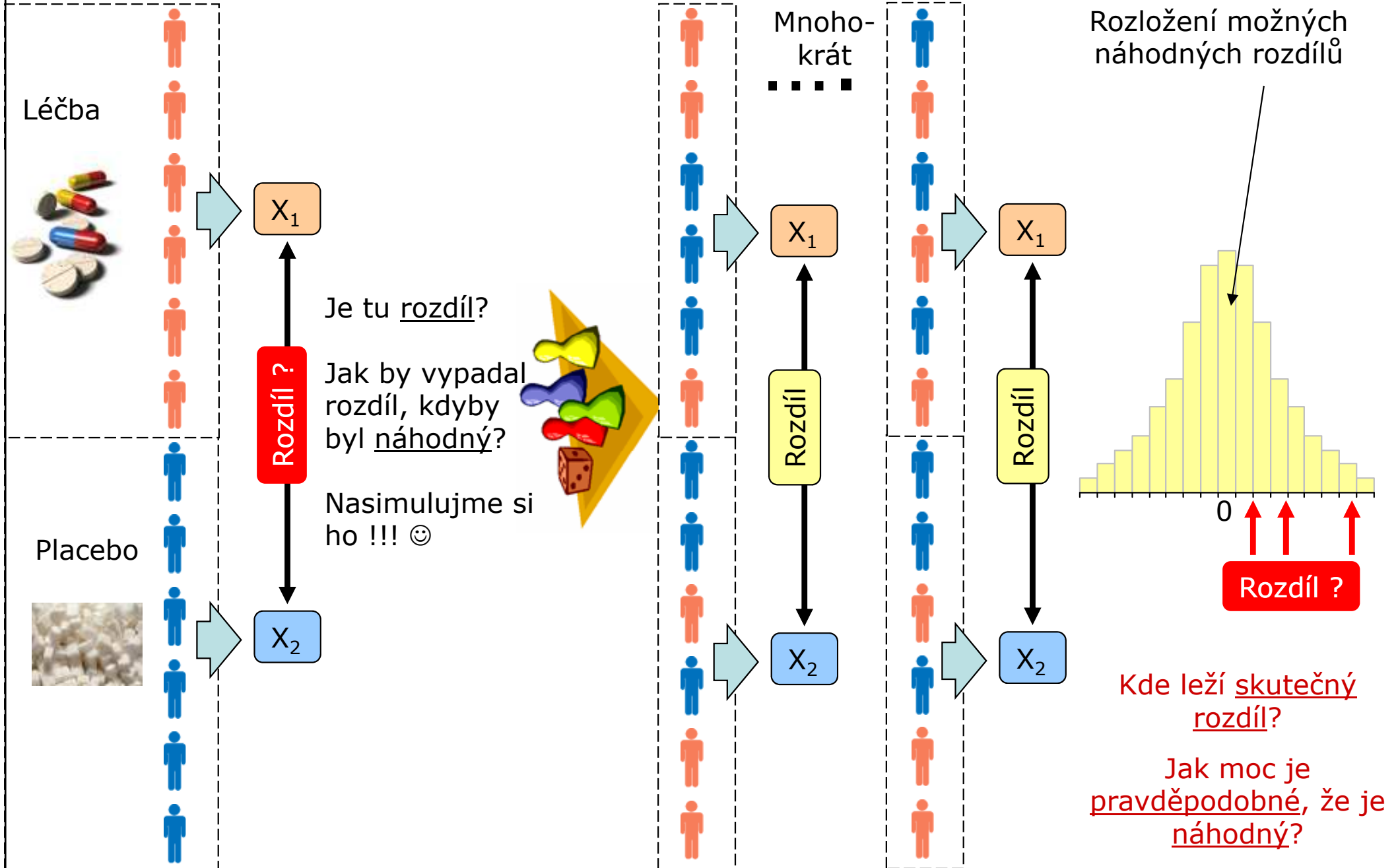
$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Variabilita dat}} * \sqrt{\text{Velikost vzorku}}$$

➤ Kritický obor testové statistiky



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

# Co znamená náhodný rozdíl?



# 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.

