



Statistické hodnocení biodiverzity

Jiří Jarkovský, Danka Haruštiaková

Institut biostatistiky a analýz, Masarykova univerzita, 2014



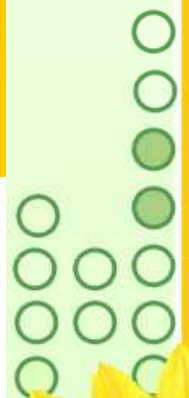
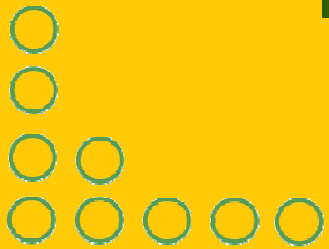
Program kurzu

- 25.9 Biodiverzita jako pojem + Biodiverzita a biostatistika
- 2.10. Vizualizace biodiverzity
- 9.10. Indexy diverzity a jejich statistická spolehlivost
- 16.10. Species-abundance křivky a stochastické modely
- 23.10. Niche-oriented species - abundance modely
- 30.10. Aplikovatelnost parametrických a neparametrických statistických technik při hodnocení biodiversity
- 6.11. Aplikovatelnost parametrických a neparametrických statistických technik při hodnocení biodiversity
- 13.11. Možnosti frakcionace biologických společenstev a následná analýza biodiversity získaných podjednotek
- 20.11. Parametrické hodnocení biodiversity ve vícerozměrných analýzách
- 27.11. Parametrické hodnocení biodiversity ve vícerozměrných analýzách
- 4.12. Případová studie: Parazitární společenstva
- 11.12. Případové studie: Lišejníky a znečištění ovzduší
- 18.12. předtermín





IV. Indexy biodiverzity a jejich spolehlivost



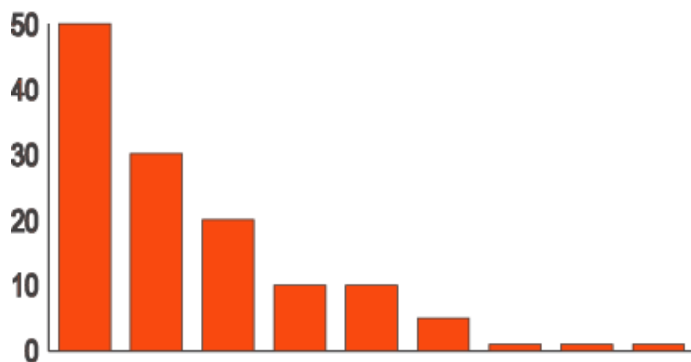
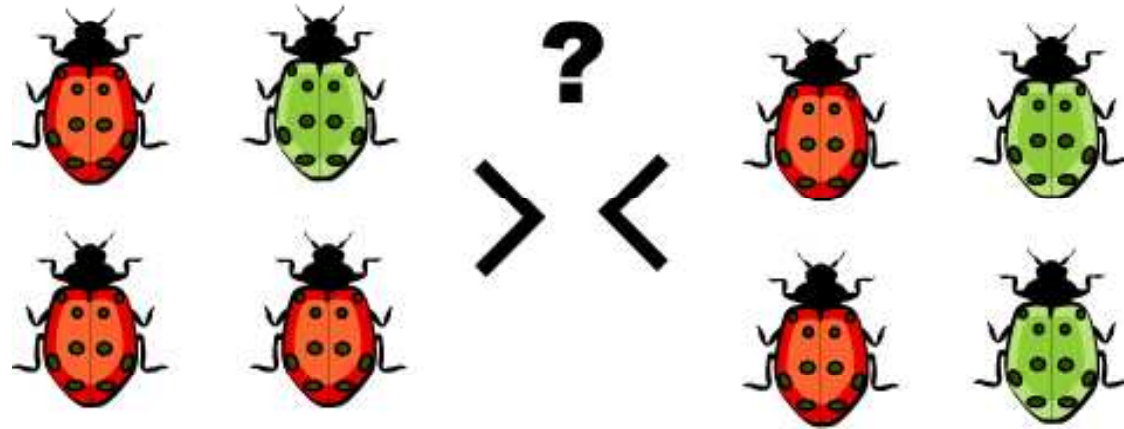
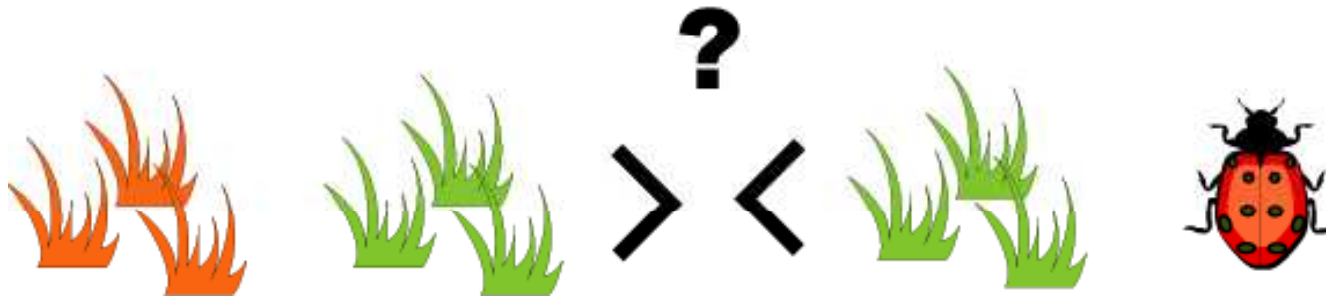
Indexy diverzity: výhody a nevýhody

X

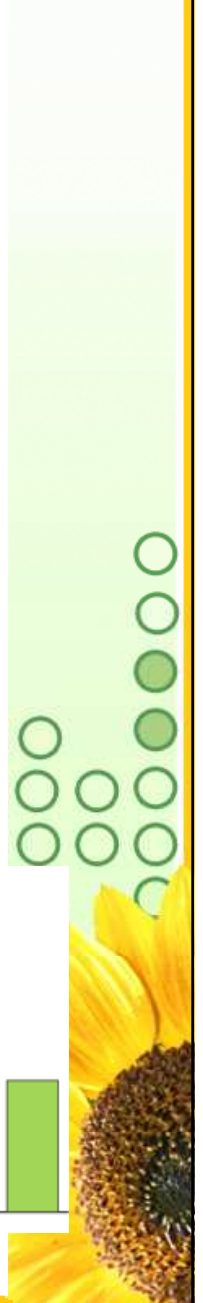
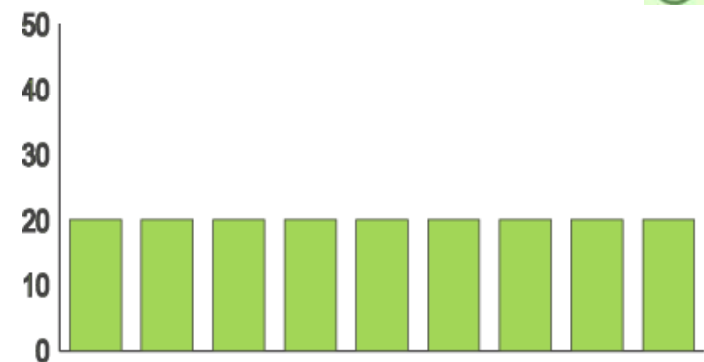
- Indexy diverzity je možné brát jako analogii k **popisné statistice**
- Celé společenstvo je agregováno **jediným číslem**, které reprezentuje **počet druhů a/nebo jejich dominanci ve společenstvu**
- Pro popisnou statistiku diverzity je možné získat intervaly spolehlivosti a dostupné jsou i statistické 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)



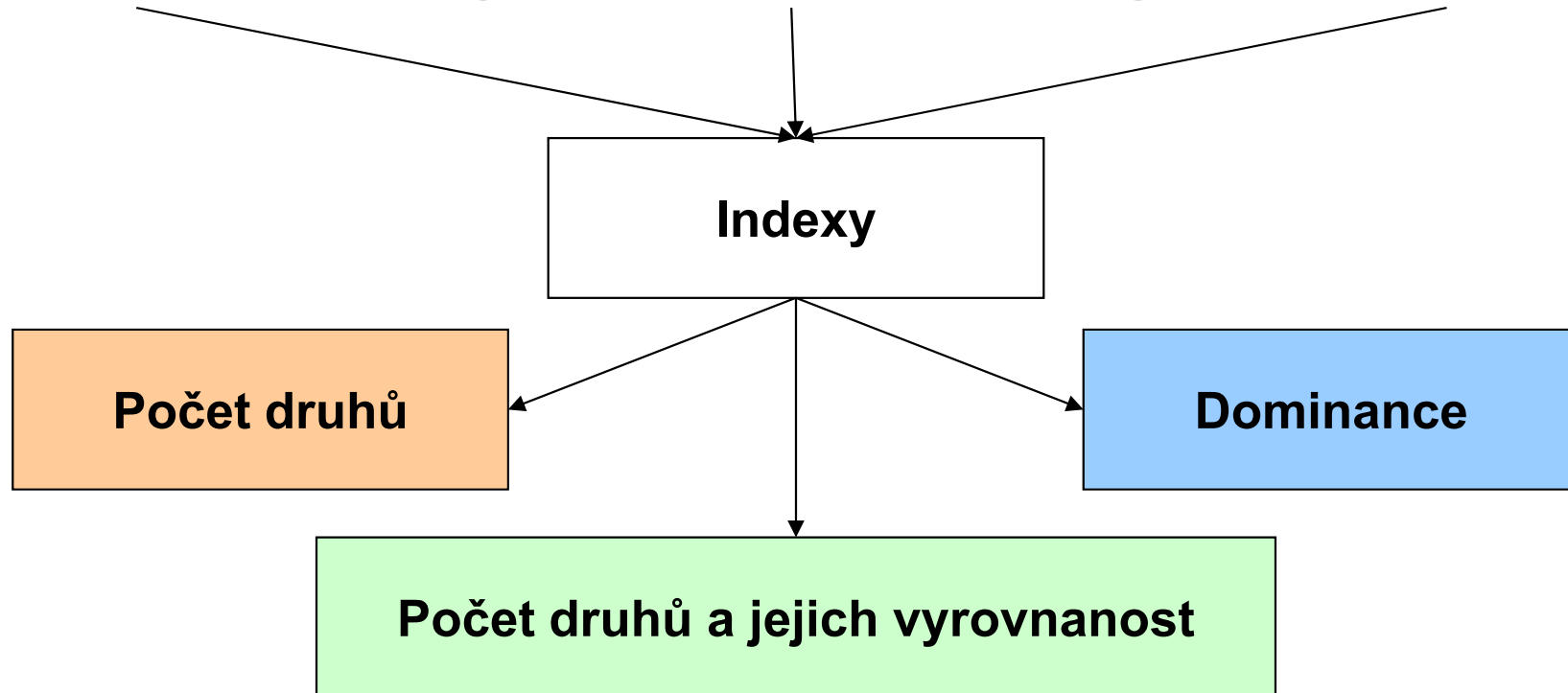
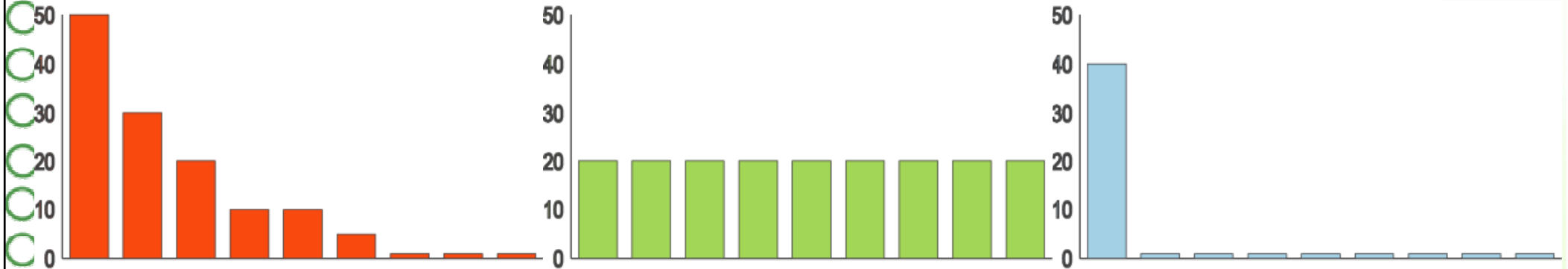
Problémy s biodiverzitou – co znamená větší diverzitu ?



?



Indexové hodnocení diverzity



Indexy diverzity I

S, 'species richness'

McIntosh, 1967

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

Margalef, 1958

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

Menhinick, 1964

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

Shannon & Weaver, 1949

$$H'_{adj}$$

Hutcheson, 1970

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

Simpson, 1949

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

Hill, 1973

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

Pielou, 1969

$$E2 = N1/S$$

Sheldon, 1969

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

Heip, 1974

$$E4 = N2/N1$$

Hill, 1973

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

Alatalo, 1981

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

McIntosh, 1967

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

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

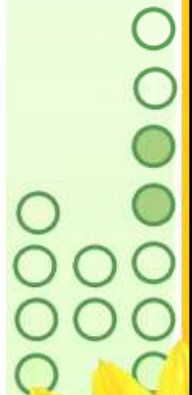
- Počet druhů (**druhová bohatost**) S
- Počet druhů vztažený na počet jedinců nebo biomasu
v případě, že není možné zajistit stejnou velikost vzorku (počet druhů narůstá s velikostí vzorku a vzorkovacím úsilím) => RAREFACTION – metoda výpočtu počtu druhů očekávaných ve vzorku standardní velikosti (např.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

Margalefův index $R1 = (S-1)/\ln N$

Menhinickův index $R2 = S/\sqrt{N}$

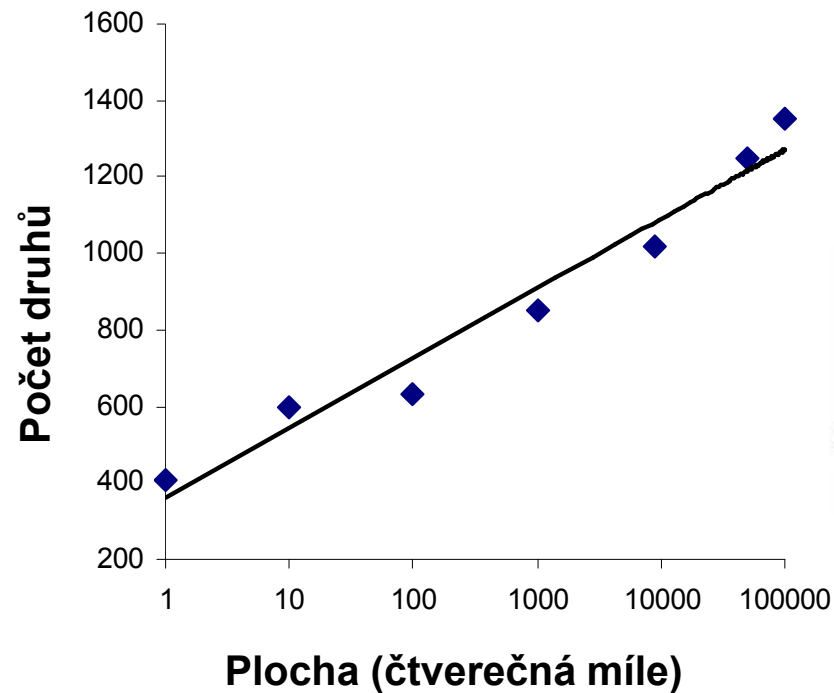
Počet druhů je vážen počtem jedinců.

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



Druhová bohatost

- Druhová bohatost narůstá s velikostí vzorku.

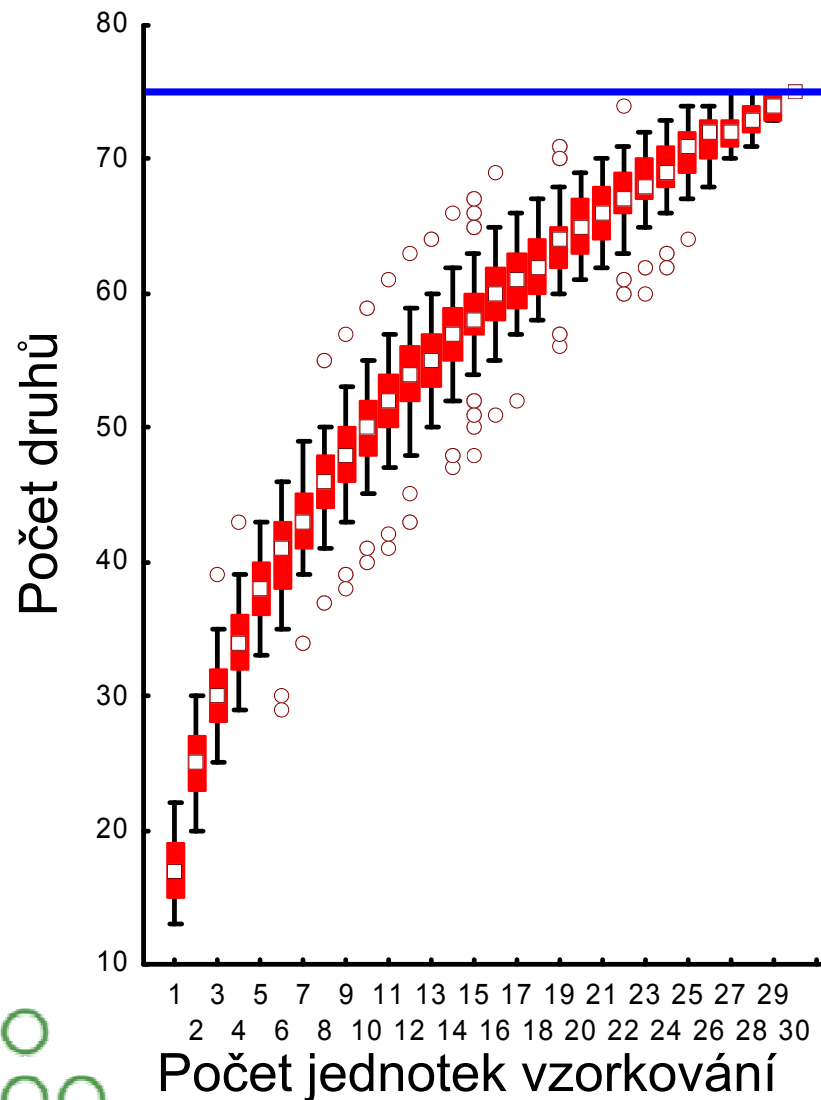


Relationship between number of species and area for flowering plants in England.
Redrawn from Krebs (1985) after Williams (1964).



Velikost vzorku a počet druhů

- Počet druhů nelineárně závisí na počtu jedinců ve vzorku



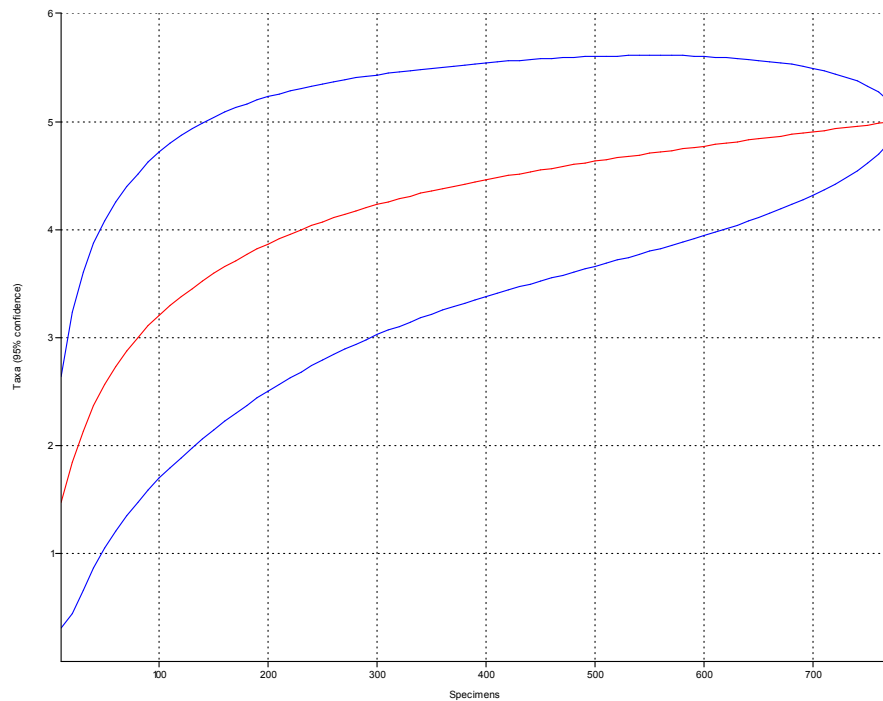
- Jak srovnat druhovou bohatost ve vzorku různé velikosti?

RAREFACTION



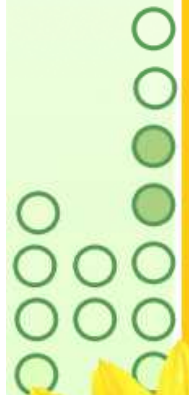
Rarefaction

- 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

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

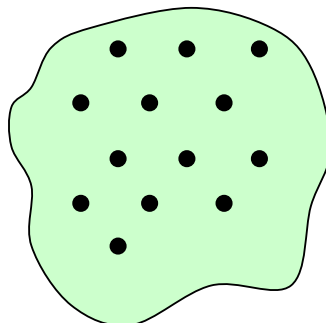
- Kde **E(S_n)** je odhadnutý počet druhů ve vzorku o **n** jedincích, **N** je celkový počet jedinců, **N_i** je počet jedinců druhu **i** a **n** je počet jedinců pro které je odhad počítán.
- Variabilita odhadu je vypočtena:

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

Rarefaction: předpoklady

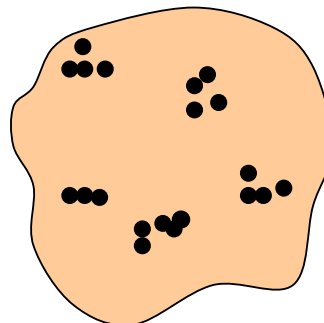
- Společenstva porovnávaná pomocí rarefaction mají mít podobné taxonomické složení
- Srovnatelné metody vzorkování
- Rarefaction křivka nemůže být extrapolována za sumu jedinců největšího vzorku
- Jedinci jsou ve společenstvu náhodně rozmístěni
 - Ve skutečnosti jsou častější shluky jedinců a za těchto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

$$\sigma^2 < \mu$$



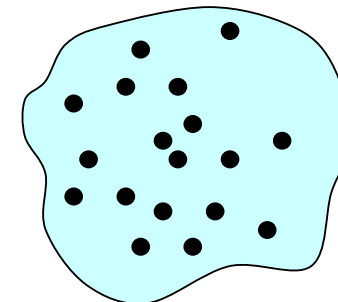
Uniform

$$\sigma^2 > \mu$$

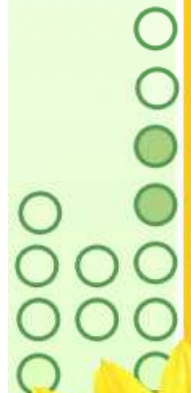


Clustered

$$\sigma^2 = \mu$$

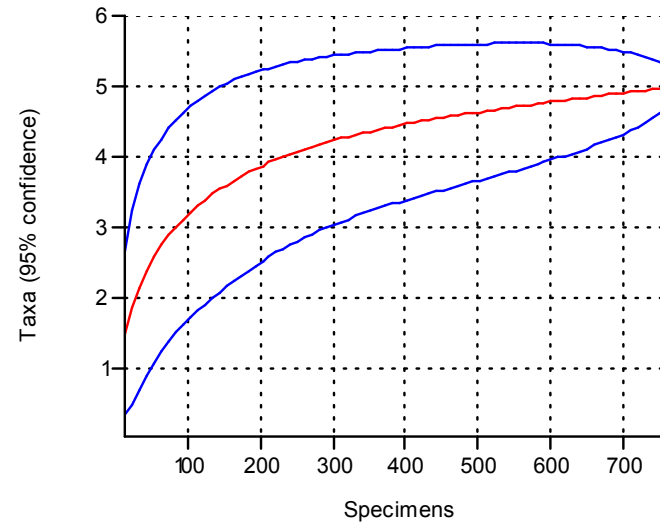


Random

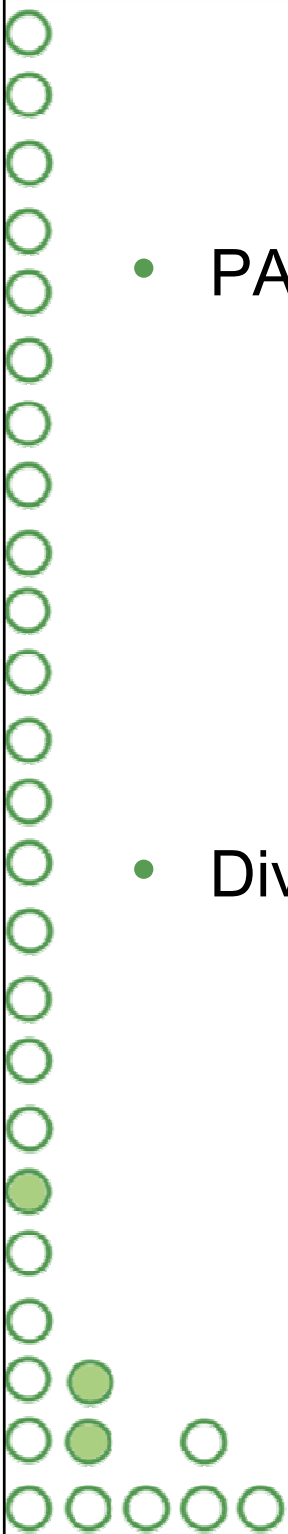


Software

- PAST software



- Diversity software od IBA



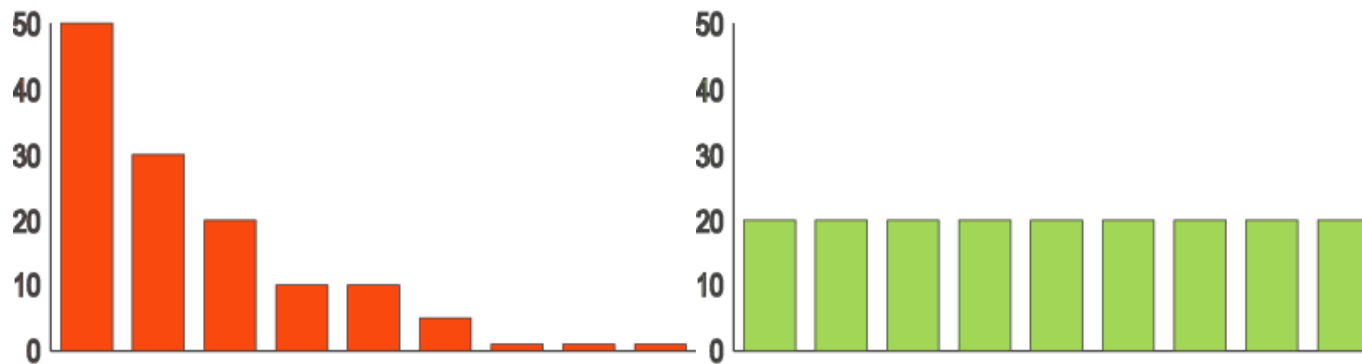
Příklady

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

- Doplňkem 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.





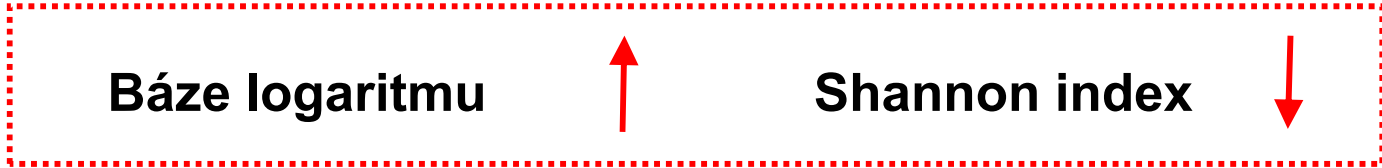
Shannon index

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		
		$p_i * \log_2(p_i)$	$p_i * \ln(p_i)$	$p_i * \log(p_i)$
 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



Shannon index

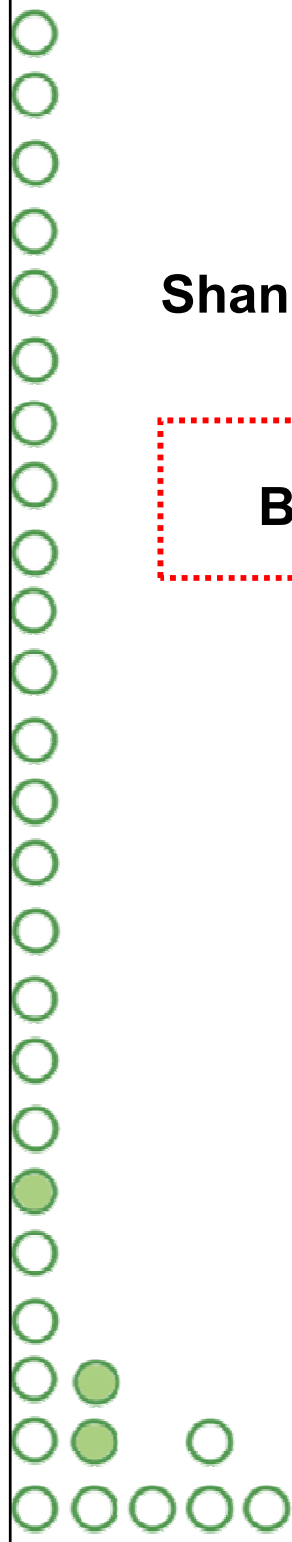
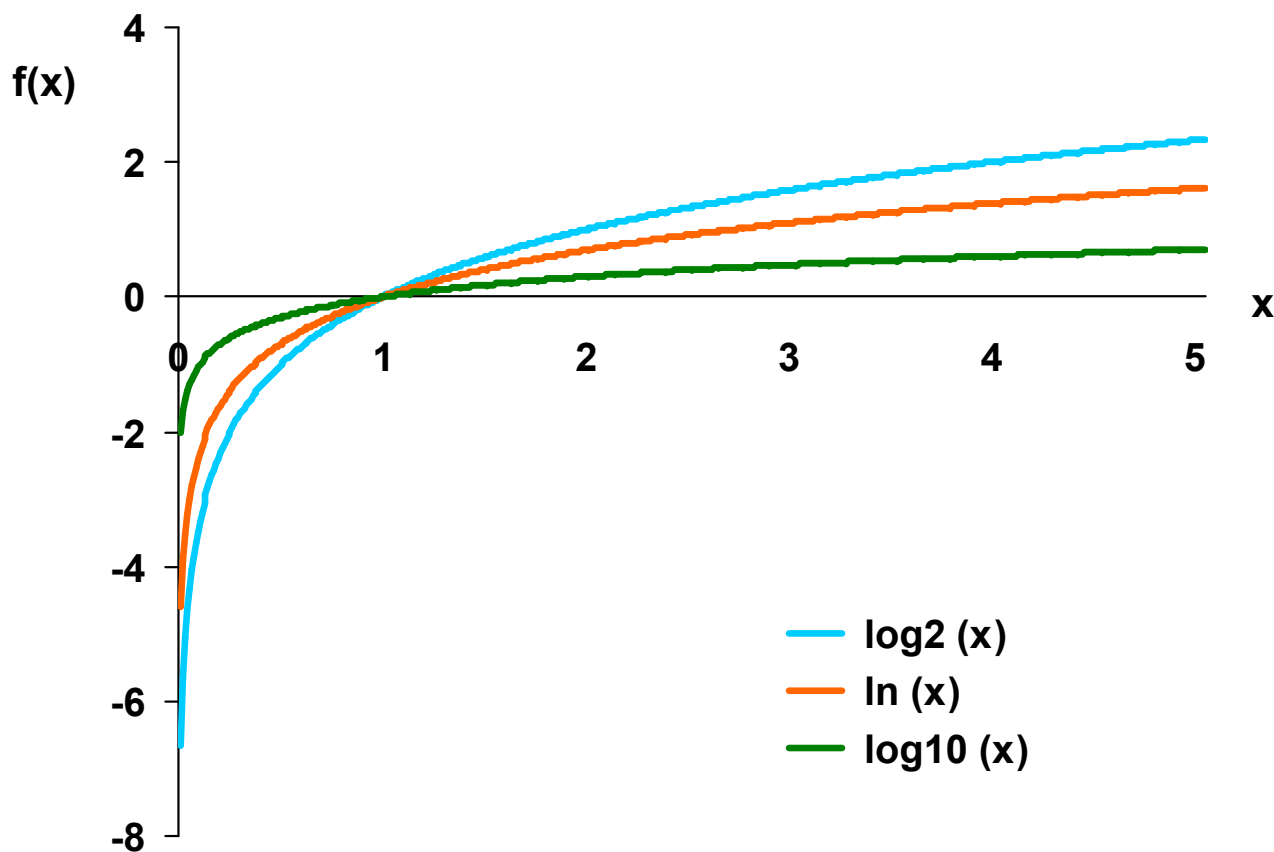
Shannon index

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

Báze logaritmu



Shannon index



Brillouinův index

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

Brillouinův index

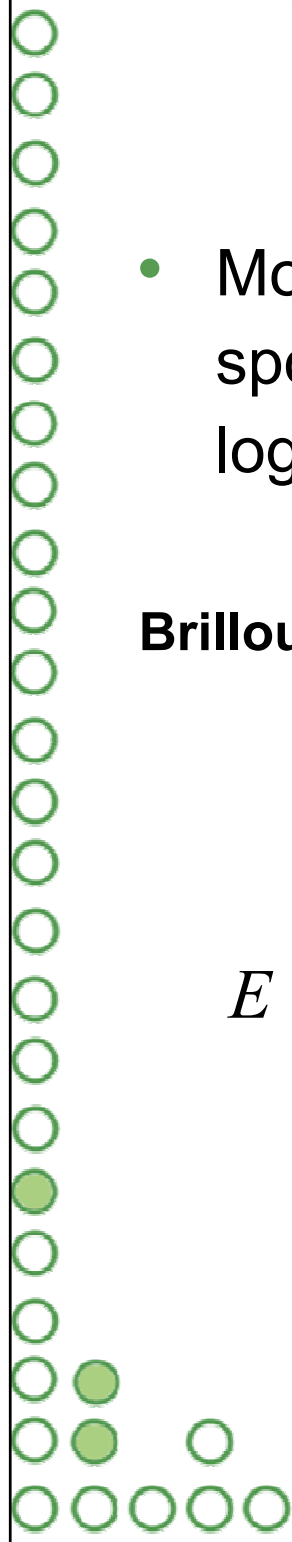
$$HB = \frac{\ln N! - \sum \ln n_i!}{N}$$

Brillouin evenness

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

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

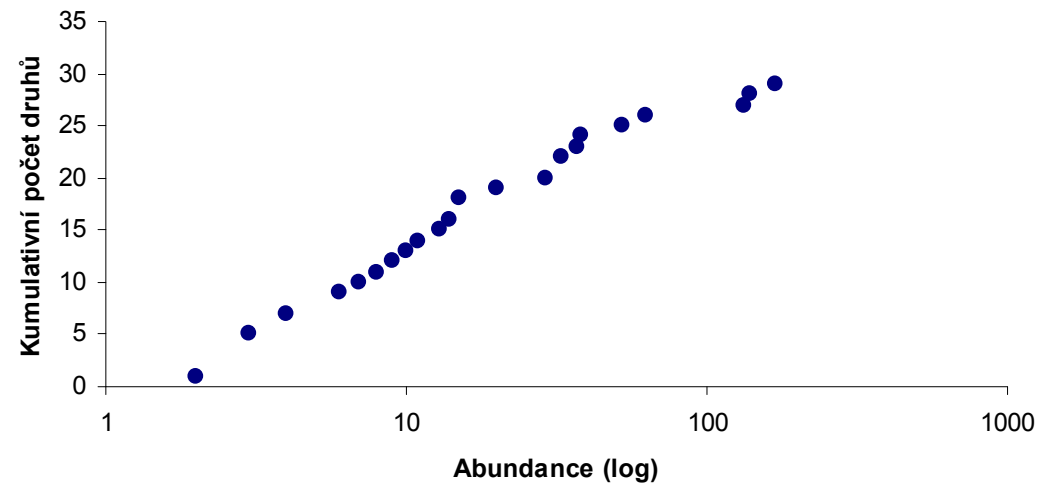
kde $\left[\frac{N}{S} \right]$ je celá část $\frac{N}{S}$ a $r = N - S \left[\frac{N}{S} \right]$



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

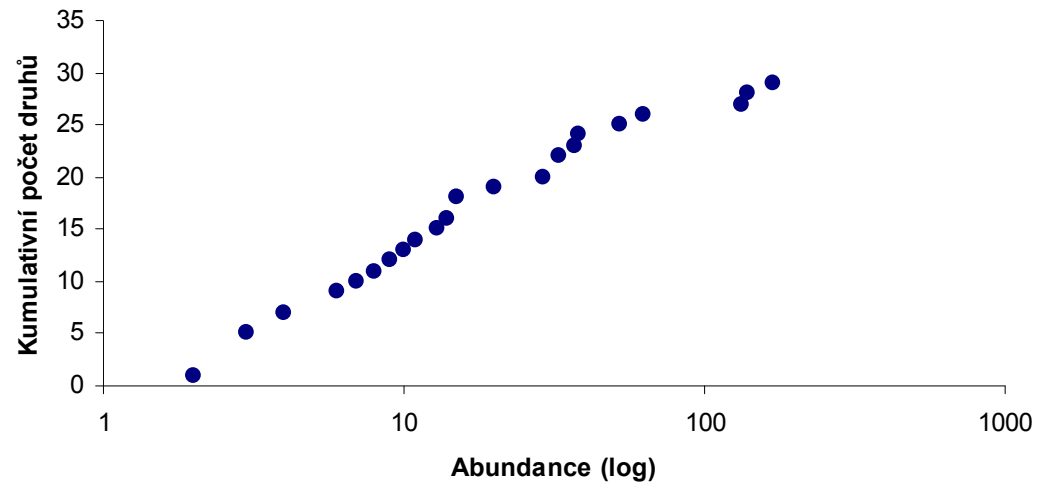
R2 – počet jedinců ve třídě, do níž spadá horní kvartil počtu druhů



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

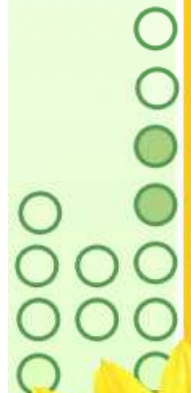
N_{\max} ... počet jedinců nejpočetnějšího druhu

N ... celkový počet jedinců

Shannon diversity ↑

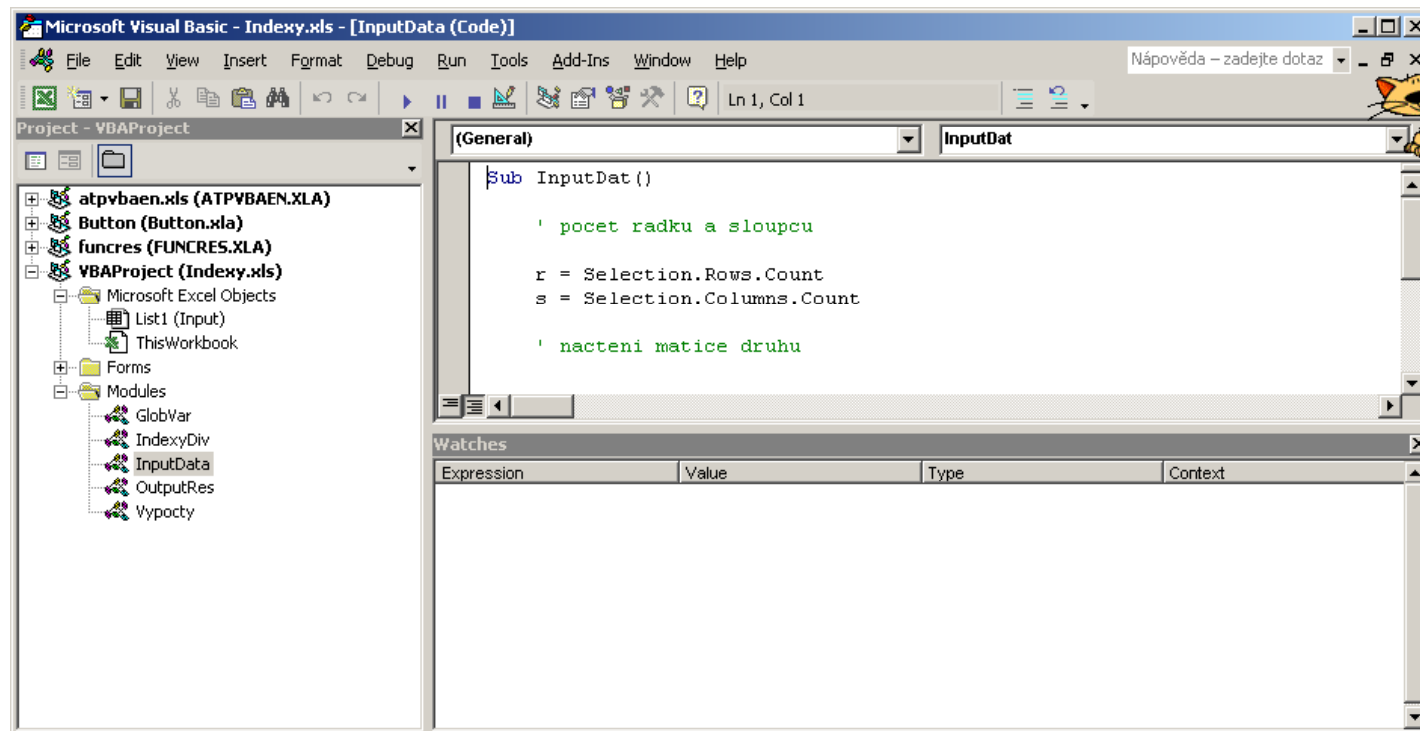
Berger-Parker index ↓

$1 - d$ 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



Proměně

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=maticice, 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

```



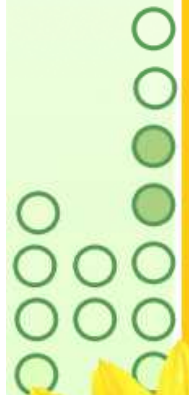
Načtení dat

```
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)

