

# Augu sabiedrību klasifikācija Klāsteranalīze

# Klasifikācija

**Mērķis:** vienkāršot oriģinālos datus, reducējot tos uz atsevišķām parauglaukumu grupām jeb klāsteriem, kurus ir vienkāršāk raksturot un vieglāk analizēt

# Klāsteranalīzes metodes

- Hierarhiskā klāsteranalīze
  - hierarhiska, daloša, kvantitatīva
- *k-means* klāsteranalīze
  - nehierarhiska, daļēji apvienojoša, kvantitatīva

## Atšķirības (*dissimilarities*)

- Statistiski rādītāji, kas atspoguļo divu objektu (parauglaukumu) atšķirības, balstoties uz to pazīmēm (sugu sastopamību)
- Ekoloģijā pielieto klasifikācijā, ordinācijā,  $\beta$  daudzveidības novērtēšanā

# Eiklīda (*Euclidean*) distance

Atšķirības novērtē, pielietojot Pitagora teorēmu

$$d_{jk} = \sqrt{\sum_{i=1}^N (x_{ij} - x_{ik})^2}$$

kur  $x_{ij}$ ,  $x_{ik}$  ir sugas  $x$  sastopamība parauglaukumos  $j$  un  $k$

Bez iepriekšējas datu transformācijas ekoloģiskos pētījumos parasti nepielieto

# Bray – Curtis dissimilarity

- Viens no ekoloģijā biežāk pielietotajiem atšķirības rādītājiem
- Vērtība robežās no 0 līdz 1

$$d_{jk} = \frac{\sqrt{\sum_{i=1}^N |x_{ij} - x_{ik}|}}{\sqrt{\sum_{i=1}^N (x_{ij} + x_{ik})}}$$

# Citi atšķirības rādītāji

- Jaccard
- Gower
- Cao
- u.c.

# Hierarhiskā klāsteranalīze

Trīs klāsteru veidošanas metodes.

Visas sākumā apvieno divus tuvākos parauglaukumus, taču atšķiras pēc tā kā apvieno klāsterus vai tiem pievieno jaunus parauglaukumus

R vidē pielieto funkciju *hclust*

# Klāsteru veidošanas metodes

- *Single linkage*
  - apvienošanas kritērijs – mazākā atšķirība starp divu klāsteru tuvākajiem parauglaukumiem
  - funkcijas *hclust* parametrs *method* = “*single*”
- *Complete linkage*
  - apvienošanas kritērijs – mazākā atšķirība starp divu klāsteru tālākajiem parauglaukumiem
  - *method* = “*complete*”

# Klāsteru veidošanas metodes

- *Average linkage*
  - apvienošanas kritērijs – mazākā atšķirība starp divu klāsteru centroīdiem
  - *method* = “*average*”
  - *method* = “*median*”
  - *method* = “*centroid*”

## *K – means* metode

- Nodala iepriekš izvēlētu klāsteru skaitu, balstoties uz klāsteru centroīdiem
- Pielieto biežāk nekā hierarhisko klāsteranalīzi
- funkcija *kmeans*

# Nodarbibā izmantojamie faili

\\Priede\kurvis\Botaanikas\_un\_ekologijas\_katetra\Fitocenologija\Nodarbibiba\_02

Uzdevums: **uzdevums\_3.pdf**