

```
### Recomand mai iantai sa parcurgeti prezentarea 17. Asemanari, Similitudini, Distanta si
# prezentarea 13. Metode de Agregare si de Clasificare. De asemenea nu strica sa cititi si
# Capitolul 20. din Volumul de Ecologie Practica (referitor la construirea dendrogramelor, pp. 227-236)
```

```
# Astfel invatati si recapitulati totodata, ceea ce este benefic pentru voi!
```

```
# Aplicatiile se bazeaza pe sursele enumerate, dar si pe un capitol din volumul:
# A.M. Benedek, 2015, Ecologie - aplicatii si studii de caz. Cap. 11. Compararea, Ordonarea
# si clasificare sistemelor ecologice, pp. 92 - 102, care de asemenea a fost incarcat pe
# pagina materiei la tema 18 (Construirea dendrogramelor - aplicatii in R)
# Unele aplicatii si metode folosite aici, in R, difera de cele utilizate in manual, respectiv
# in capitolul mentionat. Raportati-va la acest script pentru rezultate si interpretari.
```

```
#### Pentru a ilustra modul de intocmire a dendrogramelor in R, vom utiliza setul
# de date BCI din pachetul vegan, pe care trebuie sa il incarcam cu functia library()
library(vegan)
# Intrucat setul de date nu se afla in pachetul de baza, el nu poate fi accesat direct,
# ci trebuie sa utilizam functia data().
data(BCI)
## si cadrul de date BCI a devenit vizibil in Global Environment.
```

```
## Dendrogramele sunt intocmite pentru a ilustra grafic relatiile dintre mai multe
# obiecte, relatii descrise prin indici de similitudine sau prin distante.
## Acum, vom intocmi dendrograma pe baza matricii de distante euclidiene, calculate
# pe numarul de indivizi din fiecare specie. Vom calcula aceste distante
# utilizand functia dist() din pachetul de baza. O alta posibilitate este utilizarea
# functiei vegdist() din pachetul vegan. In ambele cazuri trebuie sa includem in
# comanda atributul "euclidean"
euclid<-dist(BCI, method = "euclidean")
## vizualizam continutul obiectului creat
euclid # care este o matrice de distante, fara valorile de pe si deasupra diagonalei
summary(euclid) # functia summary() citeste aceasta matrice ca vector numeric
## Pentru intocmirea dendrogramei construim mai intai un obiect de tip hclust (de la
# hierarchical cluster = clasificare ierarhica), utilizand metode de agregare la distanta medie
dendro<-hclust(euclid, method="average")
# Vom ilustra dendrograma prin utilizarea functiei plot (o functie versatila,
# care aplicata asupra a diferite tipuri de obiecte creeaza diferite tipuri de grafice)
plot(dendro)
# Pentru a obtine un grafic mai estetic este necesar sa schimbam unele aspecte ale
# acestuia, prin utilizarea sau modificarea unor attribute. In primul rand, dorim ca toate elementele
# sa plece de la valoarea 0 iar etichetele lor sa fie la acelasi nivel.
# Pentru aceasta adaugam atributul h=-1
plot(dendro, h=-1)
# Indepartam titlul graficului
plot(dendro, h=-1, main="")
# Indepartam subtitlul graficului
plot(dendro, h=-1, main="", sub="")
# Inlocuim pe axa oY Height cu Distanta euclidiana (avem valorile distantelor)
plot(dendro, h=-1, main="", sub="", ylab="Distanta euclidiana")
# Inlocuim pe axa oX euclid cu Parcela
plot(dendro, h=-1, main="", sub="", ylab="Distanta euclidiana", xlab="Parcela")
## Aceasta va fi forma finala a graficului.
## Examinand aceasta dendrograma putem spune ca structura arboretului din aceste
# parcele este relativ omogena, intrucat diferenta dintre distanta la care se unesc
# cele mai asemanatoare parcele (28 cu 29 si apoi cu 22, 12 cu 17, 2 cu 8, 32 cu 33
```

```
# etc) si cea la care se unesc cele doua grupuri care contin majoritatea parcelor
# este relativ redusa. Singura parcela cu o structura semnificativ diferita a arboretului
# este parcela 35.
```

```
#### Problema rezolvata 1 (cap. 11, din Benedek, 2015) de la pag. 92 ####
```

```
### Atentie: urmatoarele seturi de date (tabele) le gasiti in Excelul denumit Dendrogame_BA1.xlsx ###
```

```
## REZOLVARE ##
```

```
## Am cosmetizat tabelul din carte in Excel, inainte de a-l importa in R
# Astfel, am inlocuit simbolurile (triunghiurile) cu 1 (pentru prezenta a speciilor)
# intr-un bazin hidrografic, si spatiile goale cu 0 (pentru absenta),
# Am prescurtat denumirile speciilor (utilizand doar initiala genului) si am
# eliminat diacriticele din denumirile raurilor.
# presupunem ca am facut deja asta - gasiti datele in tabelul "gastropode" din Dendrogame_BA1.xlsx #
```

```
## Importam apoi datele in R (copiati tabelul gastropode)
```

```
gastropode<-read.delim("clipboard")
```

```
## Problema este ca noul cadru de date gastropode, aparut in Global Environment,
# are 8 variabile, desi noi avem doar sapte bazine hidrografice. Aceasta este din
# cauza faptului ca denumirile speciilor sunt citite ca si variabila, nu ca denumiri
# ale observatiilor (randurilor). Avem doua posibilitati pentru a remedia aceasta
# problema. Fie ignoram complet prima variabila, fie ii cerem R-ului sa o citeasca
# ca denumiri ale speciilor (randurilor). Vom opta pentru cea de-a doua varianta, si
# pentru aceasta trebuie sa adaugam un atribut suplimentar la comanda de import,
# si anume row.names=1, adica prima coloana cuprinde denumirile randurilor.
## OBS. Este posibil ca intre timp comanda de copy in Excel sa se fi dezactivat,
# astfel incat datele sa nu mai fie pe clipboard. Inainte de a rula din nou comanda
# de import trebuie sa ne intoarcem in Excel sa verificam acest lucru si daca este
# cazul, sa copiem din nou datele.
```

```
gastropode<-read.delim("clipboard", row.names = 1)
```

```
## De data aceasta cadrul de date gastropode are doar sapte variabile, adica bazine
# hidrografice
```

```
# Vizualizam datele
```

```
gastropode
```

```
# si cerem sumarul
```

```
summary(gastropode)
```

```
## Totul pare in regula, astfel incat putem continua.
```

```
## Functiile din pachetul vegan (dar si spaa si altele) citesc
```

```
# coloanele ca fiind specii, si le includ ca atare in calculul diferitilor indici.
```

```
# In cazul nostru speciile sunt pe linii, prin urmare cadrul de date va trebui transpus
```

```
# Intrucat nu avem nevoie de setul de date original, cu speciile pe linii, vom
```

```
# suprascrie obiectul gastropode (ii vom inlocui continutul pastrandu-i numele)
```

```
# Pentru transpunere folosim functia t()
```

```
gastropode<-t(gastropode)
```

```
summary(gastropode)
```

```
## Pentru calculul indicelui Jaccard vom utiliza functia designdist() din pachetul vegan care permite
```

```
# calculul oricarui indice de similitudine binar, prin scrierea formulei care include
```

```
# a,b,c,d din tabelul de contingenta, unde a semnifica numarul de specii comune
```

```
# (adica prezente in ambele comunitati), b si c numarul de specii prezente numai
```

```
# intr-una dintre comunitati, iar d numarul de specii care lipsesc din ambele comunitati
```

```
# (in cazul indicelui Jaccard d lipseste).
```

```
designdist(gastropode, method="a/(a+b+c)", abcd=TRUE)
```

```

## Si astfel am obtinut matricea de similitudine pe baza indicelui Jaccard
## Cea mai mare asemanare din punct de vedere al faunei de gastropode este intre
# bazinele raurilor Olt si Mures (0.63), iar cea mai mica intre bazinele raurilor
# Ier si Crisul Alb (0.1)
## Pentru a vizualiza relatiile de similitudine dintre diferitele bazine sub aspectul
# faunei de gastropode, vom intocmi o dendrograma. R-ul porneste de la cea mai mica valoare,
# adica distanta la care se unesc cele mai asemanatoare elemente. Deci si R-ul porneste
# de la o matrice de distante, sau o matrice de disimilitudine, pe care trebuie sa
# o construim. Valoarea maxima a indicelui Jaccard este 1, prin urmare obtinem matricea
# de disimilitudine astfel:
d.jaccard<-1-designdist(gastropode, method="a/(a+b+c)", abcd=TRUE)
d.jaccard
## Efectuam analiza de clasificare ierarhica creand obiectul de tip hclust
dendro.jaccard<-hclust(d.jaccard, method="average")
# Ilustram dendrograma prin utilizarea functiei plot
plot(dendro.jaccard)
# Pentru a obtine un grafic mai estetic este necesar sa modificam unele aspecte ale
# acestuia, prin utilizarea unor atribute. In primul rand, dorim ca toate elementele
# sa plece de la valoarea 0. Pentru aceasta adaugam atributul h=-1
plot(dendro.jaccard, h=-1)
# Indepartam titlul graficului
plot(dendro.jaccard, h=-1, main="")
# Indepartam subtitlul graficului
plot(dendro.jaccard, h=-1, main="", sub="")
# Inlocuim pe axa oY Height cu 1-Jaccard (avem valorile complementare ale indicelui)
plot(dendro.jaccard, h=-1, main="", sub="", ylab="1-Jaccard")
# Inlocuim pe axa oX d.jaccard cu Bazinul hidrografic
plot(dendro.jaccard, h=-1, main="", sub="", ylab="1-Jaccard", xlab="Bazinul hidrografic")
# Daca dorim sa vizualizam intreaga axa oY, trebuie sa o indepartam pe cea originala
plot(dendro.jaccard, h=-1, main="", sub="", ylab="1-Jaccard",
      xlab="Bazinul hidrografic", axes=FALSE)
# pe care o vom inlocui cu o secventa de la 0 la 1, la o distanta de 0.1 (valori pe
# care le stocam in vectorul yticks)
yticks<-seq(0, 1, 0.1)
axis(2, at = yticks, labels = yticks) # in aceasta comanda 2 semnifica axa oY
## Aceasta va fi forma finala a graficului.
## Sub aspectul faunei de gastropode acvatice, cea mai mare asemanare se constata
# intre bazinele Olt si Mures, care, formeaza impreuna cu Ierul si Somesul un grup
# relativ compact. Crisul Negru si Crisul Repede, rauri situate la distanta mica unul
# de altul, si avand morfologie similara, formeaza un alt grup, in timp ce Crisul
# Alb, desi este situat in apropierea celor doua (la sud de Crisul Negru),
# prezinta cea mai distincta fauna de gastropode acvatice.

#### Problema 2 de la pag. 94 (Benedek, 2015) - capitol incarcat pe pagina materie####
# Gasiti setul de date in Dendrograme_BA1.xlsx sub numele bentos #
# Copiati tabelul bentos (dati Ctrl+A si Ctrl+C)

## Importam datele, considerand prima coloana ca denumirile randurilor (taxonilor)
bentos<-read.delim("clipboard", row.names = 1)
bentos
# Speciile (taxonii in acest caz) sunt din nou pe linie, prin transpunem datele
bentos<-t(bentos)
bentos
## Calculam distantele euclidiene, utilizand functia dist() din pachetul de baza
euclid<-dist(bentos, method = "euclidean")

```

```

euclid # matricea de distante euclidiene intre cele sase statii
## Intocmim dendrograma direct pe matricea de distante
dendro.euclid<-hclust(euclid, method="average")
plot(dendro.euclid)
## Cosmetizam graficul utilizand attributele prezentate anterior
plot(dendro.euclid)
plot(dendro.euclid, h=-1, main="", sub="", ylab="Distanta euclidiana", xlab="Statia")
## Sub aspectul structurii comunitatii bentonice, cele mai asemanatoare sunt statiile
# 4 si 6. La acestea se adauga statiile 3 si 2, impreuna cu care formeaza un grup
# compact. Un alt grup este format de statiile 5 si 7, intre care exista, de asemenea
# o asemanare ridicata. Cea mai distincta este statia 1, din zona de izvor.

#### Problema 3 de la pag. 96 ####

## Importam datele (copiem: Ctrl+A si apoi Ctrl+C, tabelul cu numele holandriana)
holandriana<-read.delim("clipboard", row.names = 1)
## In cazul acestui set de date nu avem specii, insa dorim sa calculam distantele
# dintre cele patru populatii de Holandriana holandrii, sub aspectul variabilitatii
# morfologice, prin urmare acestea vor trebui amplasate pe linii, iar caracterele
# evaluate, pe coloane (vor constitui echivalentul speciilor). Deci si acest set
# de date trebuie transpus
holandriana<-t(holandriana)
holandriana
## Cream matricea de distante euclidiene, dupa aceea obiectul care cuprinde rezultatul
# analizei de clasificare ierarhica
hol.euclid<-dist(holandriana, method = "euclidean")
hol.euclid
dendro.hol<-hclust(hol.euclid, method="average")
# si construim dendrograma
plot(dendro.hol, h=-1, main="", sub="", ylab="Distanta euclidiana", xlab="Populatia")
#### OBS. Desi valorile absolute ale distantelor euclidiene pe care le-am obtinut
# sunt diferite de cele din carte (acelea fiind calculate prin aplicarea variantei
# standardizate a formulei distatei eucliene), relatiile dintre populatii sunt
# aceleasi, forma dendrogramei fiind foarte asemanatoare.
## Interpretarea acestei dendrograme o gasiti la pag. 99.

## Alte aplicatii (dintre cele propuse în capitolul mentionat) - lucrati cu setul "habitat transpus"

#### a. Dendrograma pe baza trasaturilor calitative (variabile categoriale sau factoriale)

## Importam datele din tabelul habitat transpus (din Dendrograme.xlsx)
t.habitat<-read.delim("clipboard", row.names = 1)
summary(t.habitat)
## Marea majoritate a distantelor nu permit includerea de variabile calitative (factoriale sau
# de tip categorie) cum sunt, in cazul nostru, tipul de habitat, prezenta stratului
# arbustiv, prezenta apei in apropiere si nivelul impactului antropic.
## Una dintre distantele care permit acest lucru este distanta Gower. Ea poate fi
# calculata in R utilizand functia daisy() din pachetul cluster (sunt si alte metode)
## Incarcam pachetul cluster utilizand functia library().
# Daca nu il aveti instalat, va trebui mai intai sa il instalati, apoi sa il incarcati
library(cluster)
## Calculam distantele Gower pe baza variabilelor calitative (adica variabilele 2,3,5,6)
habitat.gower<-daisy(t.habitat[,c(2,3,5,6)], metric="gower")
# Vizualizam matricea de distante
habitat.gower

```

```

## Efectuam analiza de clasificare ierarhica
dendro.gower<-hclust(habitat.gower, method="average")
# Si construim dendrograma
plot(dendro.gower, h=-1, main="", sub="", ylab="Gower", xlab="Habitat")

### b. Dendrograma pe toate variabilele (atat cantitative cat si factoriale, impreuna)
# Distanta Gower se poate calcula nu numai pe variabile de tip categorie
# ci si pe o combinatie de variabile de tip categorie si numerice (sau numai numerice)
# Astfel, putem avea o imagine mai realista asupra relatiilor de similitudine intre
# conditiile de mediu din cele cinci habitate, daca luam in considerare si restul
# caracteristicilor (cele numerice, referitoare la altitudine, acoperirea stratului
# ierbos si adancimea solului) - deci vom utiliza intregul set de date
habitat.gower<-daisy(t.habitat, metric="gower")
# Vizualizam matricea de distante
habitat.gower
# Efectuam analiza de clasificare ierarhica
dendro.gower<-hclust(habitat.gower, method="average")
# Si construim dendrograma
plot(dendro.gower, h=-1, main="", sub="", ylab="Gower", xlab="Habitat")
## Cele doua dendrograme releva aceleasi relatii de similitudine intre habitate
# (cele mai asemanatoare sunt A si C, apoi D si E, B fiind mai apropiat de primul grup)
# Insa cea de-a doua dendrograma (pe toate variabilele) nuanteaza mult mai bine relatiile.
# Cu distanta Gower puteti compara obiecte descrise prin variabile heterogene, adica de tipuri diferite.

### c. Pe langa habitate, vom utiliza si tabelul "mamifere" pentru a obtine o matrice heterogena!
# Construim dendrograma pe baza distantelor euclidiene calculate pe urmasorii parametri:
# indicele Simpson, altitudine, acoperirea stratului ierbos si adancimea solului,
# nr de specii si nr de indivizi din cele cinci habitate
## Selectam din t.habitat cele trei variabile numerice si construim un nou obiect
habit<-t.habitat[,c(1,4,7)]
habit # este ceea ce dorim
# Aducam la acest cadru de date valorile indicelui Simpson, calculat cu functia
# diversity() tot din pachetul vegan, aplicata pe cadrul de date mamifere (care
# cuprinde datele de specii, avand aceeasi ordine a habitatelor)

## Importam (copiem si aducem in R) tabelul mamifere din Dendrograme_BA1.xlsx
mamifere<-read.delim("clipboard", row.names = 1)
# Transpunem cadrul de date pentru a avea speciile pe coloane
mamifere<-t(mamifere)
# Vizualizam datele
mamifere

# Calculam indicii specificati:
# mai intai indicele de diversitate Simpson
habit$Simpson<-1-diversity(mamifere, "simpson")
# Si numarul de specii determinat cu functia specnumber() din vegan
habit$S<-specnumber(mamifere)
# Si numarul total de indivizi, pe care il obtinem insumand valorile de pe linii
# cu functia rowSums()
habit$N<-rowSums(mamifere)
# Vizualizam noul continut al obiectului habit
habit
# Calculam distantele euclidiene
habit.euclid<-dist(habit, method="euclidean")
# Efectuam analiza de clasificare ierarhica

```

```
dendro.habit<-hclust(habit.euclid, method="average")
# Intocmim dendrograma
plot(dendro.habit, h=-1, main="", sub="", ylab="Distanța euclidiană", xlab="Habitat")

### Activitate independentă: construiți dendrograma stațiilor alese de-a lungul Cibinului, cu date din
# Tab. 11.8 (date după A. Curtean-Banaduc, 2000, ap. Benedek, 2015, pag. 101-102)
# descrie prin structura comunităților bentonice pe baza indicelui de abundență relativă,
# prin intermediul distanței euclidiene, metoda de grupare la distanță medie.
# Veti copia sau scrie datele într-un tabel excel, îl copiați, introduceți în R, și construiți
# dendrograma după cerințele redată mai sus. Încercați să interpretați în câteva fraze scurte.

# Ați reușit?! Felicitări, și nu uitați să trimiteți tema cu numele vostru: Nume_preunume_T11.R
```