

Analyse en composantes principales - Doubs

Sarah Ouadah

Module DYST, M2 BEE, Septembre 2023

On considère différents sites le long du Doubs. Sur chacun des 30 sites d'observation, on a mesuré 11 variables désignant les caractères morphologiques d'une rivière (distance à la source, altitude, pente, débit) et la qualité de l'eau (pH, calcium, phosphates, nitrates, amoniaque, oxygène et demande biologique en oxygène).

Les données sont issues du package `ade4` et peuvent être récupérées via la commande `data(doubs)`. Elles contiennent diverses informations telles que `doubs$xy` pour les positions géographiques des sites considérés et `doubs$env` pour les mesures effectuées.

```
head(doubs$xy)
```

```
##      x  y
## 1  88  7
## 2  94 14
## 3 102 18
## 4 100 28
## 5 106 39
## 6 112 51
```

```
head(doubs$env)
```

```
##  dfs alt  slo flo pH har pho nit amm oxy bdo
## 1   3 934 6.176 84 79 45  1 20  0 122 27
## 2  22 932 3.434 100 80 40  2 20 10 103 19
## 3 102 914 3.638 180 83 52  5 22  5 105 35
## 4 185 854 3.497 253 80 72 10 21  0 110 13
## 5 215 849 3.178 264 81 84 38 52 20  80 62
## 6 324 846 3.497 286 79 60 20 15  0 102 53
```

Après avoir chargé les données, on pourra découvrir leurs caractéristiques (`str`, `summary`), faire des représentations graphiques (`plot`, `boxplot`, `hist`), commenter et décider d'une éventuelle normalisation des données.

L'objectif est d'établir une typologie des sites et de mettre en évidence quels groupements se font au vu des caractères morphologiques et de la qualité de l'eau des sites considérés.

On propose d'effectuer une ACP pour rendre compte des liens entre les différentes variables étudiées et visualiser le nuage d'individus dans un espace approprié. Avec le logiciel R, ceci est réalisable avec le package `FactomineR` et la commande `PCA`.

1. Combien de composantes principales peut-on garder?
2. Représenter les variables sur le cercle des corrélations. Discuter de la qualité de représentation des variables et interpréter les axes;
3. Représenter les individus, discuter de leur qualité de représentation et de leur contribution aux axes;
4. Est-ce que des groupes semblent se détacher de ce graphique ? Caractériser les au vu de la signification des axes.