
Advanced Biostatistics - ABSTAT17

IGC, 10th - 13th April, 2017

PROJECT - PCA and Bootstrap.

The dataset “meaudret” can be found in package **ade4**. It is a classical set of ecological information: three data frames containing environmental variables, Trichoptera abundance collected on 5 sites during 4 seasons along the Meaudret river (a river near Grenoble, France). The data set contains physico-chemical variables measured on 5 sites along the Meaudret River during four seasons (Pegaz-Maucet, 1980). Water samples were taken to study the following physico-chemical variables:

- **Temp** Water Temperature (in Celcius degree)
- **Flow** Flow (in l/s)
- **pH** pH
- **Cond** Conductivity (in μ S/cm)
- **Bdo5** Biological oxygen demand after days (in mg/l)
- **Oxyd** Oxygen (in mg/l of oxygen)
- **Ammo** Ammonium Hydroxide (in mg/l NH_4^+)
- **Nitr** Nitrate (in mg/l NO_3^-)
- **Phos** Phosphate (in mg/l PO_4^{--})

```
library(ade4)
data(meaudret)
names(meaudret)
[1] "env" "design" "spe" "spe.names"
names(meaudret$mil)
[1] "Temp" "Flow" "pH" "Cond" "Bdo5" "Oxyd" "Ammo" "Nitr" "Phos"
```

`meaudret$env` is a data frame containing all the environmental variables (9 variables in this case). `meaudret$design` is a data frame detailing the experimental design. `meaudret$spe` is a data frame providing information on species richness per site, per season.

```
table(meaudret$design$season)
[1] spring summer autumn winter
     5       5       5       5
table(meaudret$design$site)
[1] S1 S2 S3 S4 S5
     4  4  4  4  4
```

There is 4 measures (one per season) taken per site and 5 measures (one per site) for each season.

1. Start with a descriptive analysis of your data.
2. Use a Bootstrap approach for the following:
 - (a) Estimate the median for the pollution variables (`Phos`, `Nima`, `Ammo`) using a 99% CI.
 - (b) A factory near by the river states that the mean value of the oxygen of the river is higher than 3.5mg/l (values under this value will compromise the biodiversity). The data corroborate this statement?
 - (c) Analyze if there are significant differences between the pH mean values of the water in `winter` and `autumn`?
3. What findings can you get from the data using PCA analysis (e.g. are the pollution variables more related with a specific season, or site).

Additional information

- You will have 20' to make your presentation.
- After your presentation there will be a discussion of your results.

References

D. Pegaz-Maucet. Impact d'une perturbation d'origine organique sur la d'erive des macro-invertebres benthiques d'un cours d'eau. Comparaison avec le benthos. PhD thesis, 1980.