

Analyse des données expérimentales en Biologie

Cours introductif

Anne B Dufour

Plan du cours

- ✳ Qu'est-ce que la Biométrie ?
- ✳ Statistique vs Biologie
 - le programme classique du DEUG
- ✳ Biologie vs Statistique
 - exemple de la morphométrie
- ✳ Présentation d'un problème écologique

Biométrie

« La biométrie est la science de la variabilité,
des phénomènes qui s'y attachent et des problèmes
qui en découlent »

Schreider (1952)

Etat actuel de la biométrie

Comment les objets sont reliés entre eux ?

Quelle est leur diversité, leur complexité, leur stabilité
au niveau où on les approche ?

Legay (1986)

Biométrie

BIOLOGIE



MATHEMATIQUE
géométrie, analyse
algèbre,...
statistique

information structurée



– Logiciel de Statistique

- ✳ General Public License
- ✳ Version Actuelle 1.9.1
- ✳ Disponible dans les archives du réseau CRAN :
Comprehensive R Archive Network

Ross Ihaka and Robert Gentleman (1996)
R: A language for data analysis and graphics.

– Logiciel de Statistique

De nombreux packages
couvrent un large champ de la statistique
comme la modélisation linéaire et non linéaire,
la classification, les séries chronologiques...
et

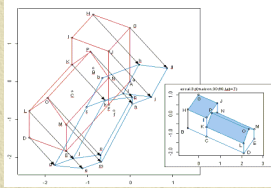
ade4

Analyses des Données Écologiques :
méthodes Exploratoires et Euclidiennes
en sciences de l'Environnement

Chessel, Dufour, Thioulouse (2004)
The ade4 package – I: one table methods.
R news

Enseignements de Statistique en Biologie

A.B. Dufour, D. Chessel, J. Lobry & J.Thioulouse
Notes de cours, illustrations, exercices, problèmes
Fiches de Travaux Dirigés, jeux de données



<http://pbil.univ-lyon1.fr/R/enseignement.html>

Description des données

Population étudiée : *Gerris costae* males

Variables mesurées : longueurs

tho = thorax

ant = segment de la première antenne

mf = fémur de la patte postérieure

hf = fémur de la patte du milieu



Provenance géographique : Alpes, Grèce et Pyrénées

```
> gerris = read.table("gerris.txt",h=TRUE)
```

Klingenberg et al (1996) Advances in Morphometrics

Statistique descriptive classique

Etude de la variable quantitative
longueur du thorax

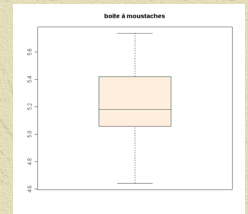
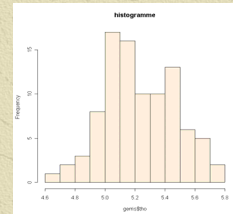
```
> summary(gerris$tho)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
4.640  5.060  5.180  5.228  5.420  5.740
```

```
> var(gerris$tho)           minimum, maximum
[1] 0.0557738              moyenne
> sd(gerris$tho)           quartiles (médiane)
[1] 0.2361648              variance, écart-type
```

Statistique descriptive classique

longueur du thorax

```
> hist(gerris$tho,main="histogramme",col="antiquewhite")
> boxplot(gerris$tho,main="boite à moustaches",col="antiquewhite")
```



Statistique descriptive classique

Etude de la variable qualitative
origine géographique

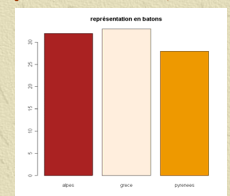
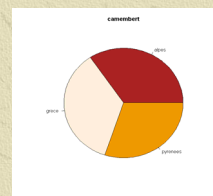
```
> summary(gerris$group)
  alpes  grece  pyrenees  Fréquence absolue
    32     33     28
```

```
> options(digits=4)
> summary(gerris$group)/sum(summary(gerris$group))
  alpes  grece  pyrenees  Fréquence relative
0.3441 0.3548 0.3011
```

Statistique descriptive classique

origine géographique

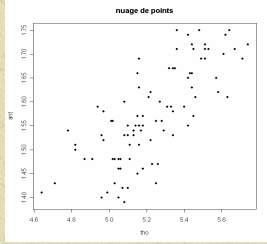
```
> pie(groupe,col=c("brown","antiquewhite","orange2"),main="camembert")
> barplot(groupe,col=c("brown","antiquewhite","orange2"),
          main="représentation en batons")
```



Statistique descriptive classique

Croisement de deux variables quantitatives
longueurs du thorax et du segment de la première antenne

```
> cor(tho, ant)
[1] 0.7580122
> plot(tho, ant, pch=20, cex=1.5)
```



Régression linéaire simple

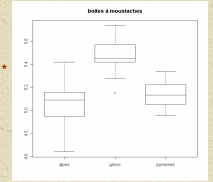
?

Statistique descriptive classique

Croisement d'une var. quantitative et d'une var. qualitative
longueur du thorax et origine géographique

```
> anova(lm(tho~gerris$group))
Analysis of Variance Table

Response: tho
          Df Sum Sq Mean Sq F value Pr(>F)
gerris$group  2   3.27   1.64   79.2 <2e-16 ***
Residuals    90   1.86   0.02
> boxplot(tho~gerris$group)
```



Analyse de la variance à un facteur

?

Analyse des données

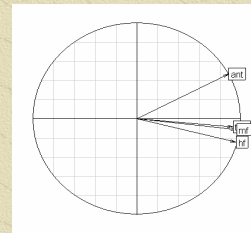
Etude simultanée des 4 variables morphométriques

```
> cov(gerris2)
      tho      ant      mf      hf
tho  0.05577380  0.01796751  0.05124743  0.07328205
ant  0.01796751  0.01007380  0.02116344  0.02680817
mf   0.05124743  0.02116344  0.07062821  0.09050722
hf   0.07328205  0.02680817  0.09050722  0.13230252

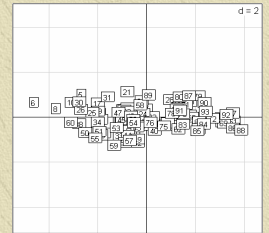
> cor(gerris2)
      tho      ant      mf      hf
tho  1.00000000  0.7580122  0.8165220  0.8530970
ant  0.7580122  1.00000000  0.7934153  0.7343220
mf   0.8165220  0.7934153  1.00000000  0.9362893
hf   0.8530970  0.7343220  0.9362893  1.00000000
```

Analyse en composantes principales

Analyse en composantes principales : Effet taille

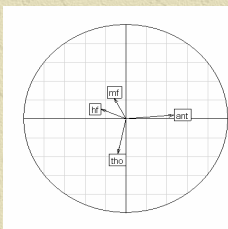


Cercle des corrélations

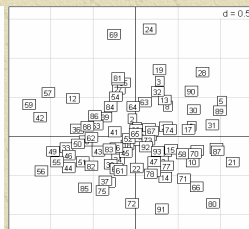


Représentation des gerris costae

A.C.P. : vecteur Forme

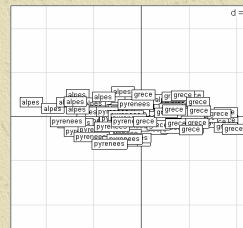


Cercle des corrélations

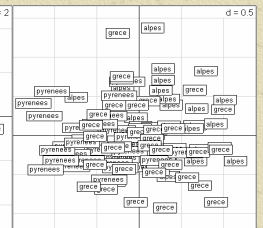


Représentation des gerris costae

Morphologie & origine géographique



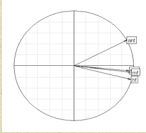
Axes 1 et 2



Axes 2 et 3

Problématique biologique

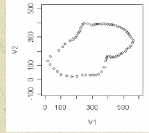
Effet taille
Vecteur forme



morphométrie

Développement
d'analyses multivariées
spécifiques

Landmarks



Jolicœur et Mosiman (1960)
Size and shape variation in the painted turtle. A principal component analysis

Bookstein (1991)
Morphometric tools for Landmark data: geometry and biology

Présentation d'un problème écologique : l'avifaune de la Tarentaise

Lebreton, Lebrun, Martinot, Miquet et Tourmier (1999) Approche écologique de l'avifaune de la Vanoise. Travaux scientifiques du Parc de la Vanoise, 21, 7-304

L'objet à étudier est une liste de 5 composantes :

