

Fiche TD avec le logiciel  : tdr62

Pratique de l'analyse des correspondances

A.B. Dufour & D. Chessel

La fiche illustre la pratique de l'analyse des correspondances en écologie (ordination indirecte), sur des tables de contingence pour lesquelles elle est une analyse canonique et sur les tableaux de fréquences alléliques qui demandent l'introduction de la version intra-classes.

Table des matières

1	L'ordination directe	2
2	AFC et ordination indirecte	7
3	Fer à cheval	9
4	Corrélation canonique	16
5	Une curiosité	20
6	AFC comme analyse canonique	22
7	Fréquences alléliques	28
	Références	33

1 L'ordination directe

Comme méthode sous-évaluée pour les uns (Hill 1974 [9]), surévaluée pour d'autres ("*Correspondence analysis rapidly gained popularity, first in France, where classical statistics did not exist*") (Heijden 1987 [27]), l'AFC est un lieu de débat.

Dans la documentation de `corresp` (bibliothèque MASS) :

`nf` : The number of factors to be computed. Note that although 1 is the most usual, one school of thought takes the first two singular vectors for a sort of biplot.

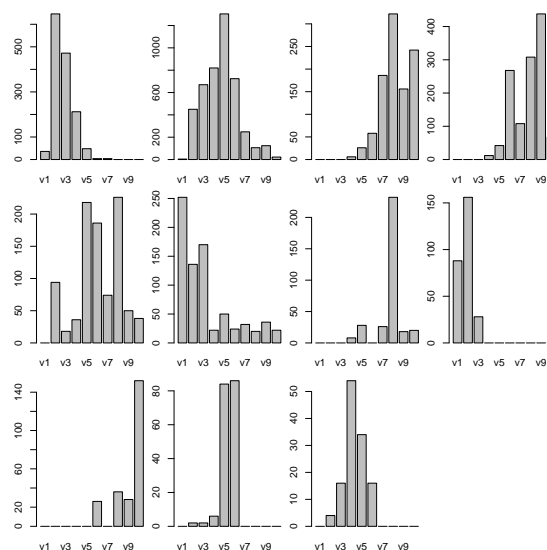
On peut parler ici de cette école qui garde plusieurs facteurs. Utiliser un jeu de données classiques (Gauch et al. 1974 [8]) :

```
library(ade4)
data(santacatalina)
santacatalina
```

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10
Pinus.ponderosa	36	646	472	212	48	4	4	0	0	0
Quercus.hypoleucoides	4	450	670	820	1304	724	248	106	124	22
Pinus.cembroides	0	0	0	6	26	58	186	322	156	242
Arctostaphylos.pringlei	0	0	0	12	42	268	108	308	438	68
Quercus.arizonaca	0	94	18	36	218	186	74	226	50	38
Quercus.rugosa	252	136	170	22	50	24	32	20	36	22
Garrya.wrightii	0	0	0	8	28	0	26	232	18	20
Pseudostuga.menziesii	88	156	28	0	0	0	0	0	0	0
Quercus.emoryi	0	0	0	0	0	26	0	36	28	152
Pinus.chihuahuana	0	2	2	6	84	86	0	0	0	0
Arbutus.arizonaca	0	4	16	54	34	16	0	0	0	0

Densité à l'hectare de 11 espèces d'arbres par classe de valeurs de l'humidité du sol (moyenne sur plusieurs stations par classe). Tracer les courbes de réponse des espèces le long du gradient :

```
g <- t(santacatalina)
par(mfrow = c(3, 4))
par(mar = c(2, 2, 1, 0.2))
apply(g, 2, barplot)
```



Les courbes de réponses sur un gradient sont en général unimodales et présentent un mode (*preferendum*) et un étalement (amplitude) caractéristique de l'espèce. L'ordination gaussienne ajuste à ces observations des courbes du type :

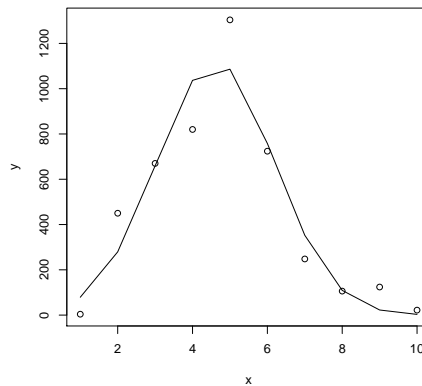
$$y = A \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right)$$

Ajuster aux moindres carrés les 11 courbes.
On essaye quelque chose de très élémentaire !

```
deb <- function(x, y) {
  A <- max(y)
  mu <- sum(x * y)/sum(y)
  sigma2 <- sum((x - mu)^2 * y)/sum(y)
  return(list(A = A, mu = mu, sigma2 = sigma2))
}
x <- 1:10
y <- g[, 2]
nls(y ~ A * exp(-(x - mu)^2/2/sigma2), start = deb(x, y))
Nonlinear regression model
model: y ~ A * exp(-(x - mu)^2/2/sigma2)
data: parent.frame()
      A      mu  sigma2
1119.443  4.615  2.462
residual sum-of-squares: 152035
Number of iterations to convergence: 6
Achieved convergence tolerance: 3.61e-06
```

Et cela marche !

```
nls0 <- nls(y ~ A * exp(-(x - mu)^2/2/sigma2), start = deb(x, y))
plot(x, y)
lines(x, predict(nls0))
```

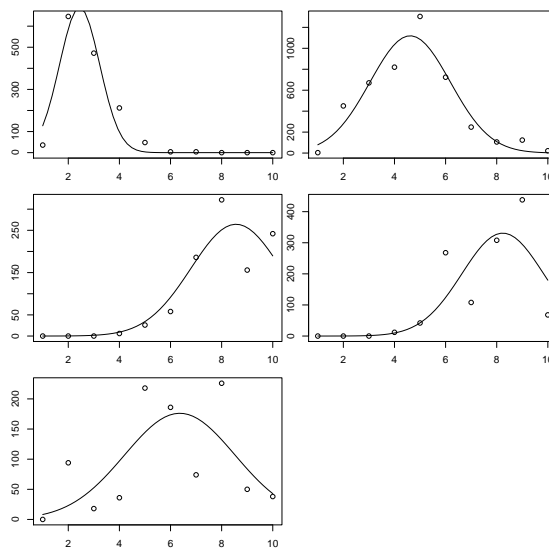
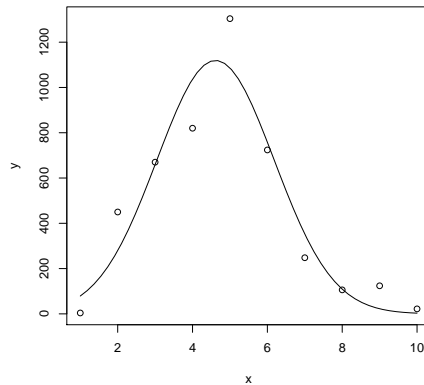


L'essentiel du travail est fait. On peut affiner :

```
fun1 <- function(x, y) {
  deb <- function(x, y) {
    A <- max(y)
    mu <- sum(x * y)/sum(y)
    sigma2 <- sum((x - mu)^2 * y)/sum(y)
    return(list(A = A, mu = mu, sigma2 = sigma2))
  }
  nls0 <- nls(y ~ A * exp(-(x - mu)^2/2/sigma2), start = deb(x,
y))
}
```

```

plot(x, y)
xnew <- seq(min(x), max(x), le = 50)
lines(xnew, predict(nls0, newdata = list(x = xnew)))
}
fun1(x, y)
    
```



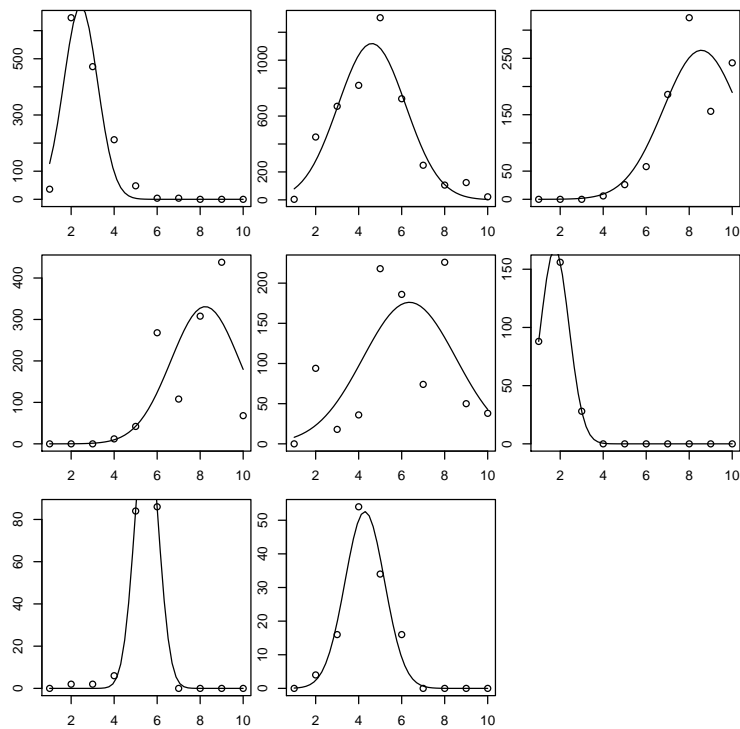
La procédure a échoué sur la sixième courbe.

Erreur dans $\text{nls}(y \sim A * \exp(-(x - \mu)^2/2/\sigma^2), \text{start} = \text{deb}(x, y))$:
le nombre d'itérations a dépassé le maximum de 50

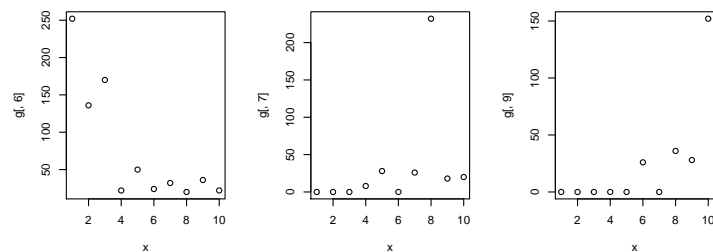
Elle échoue également sur 6, 7 et 9.

```

par(mfrow = c(3, 3))
par(mar = c(2, 2, 1, 0.2))
apply(g[, -c(6, 7, 9)], 2, fun1, x = x)
    
```

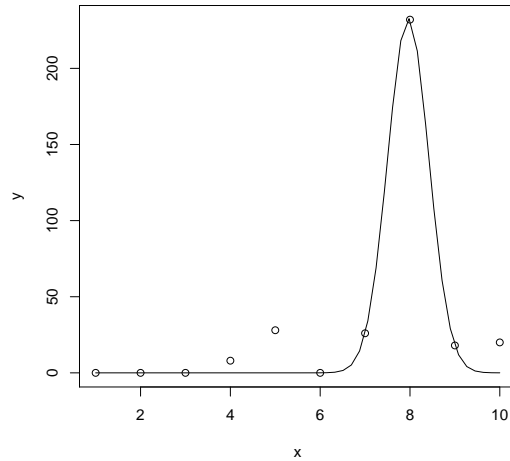


```
par(mfrow = c(1, 3))
plot(x, g[, 6])
plot(x, g[, 7])
plot(x, g[, 9])
```



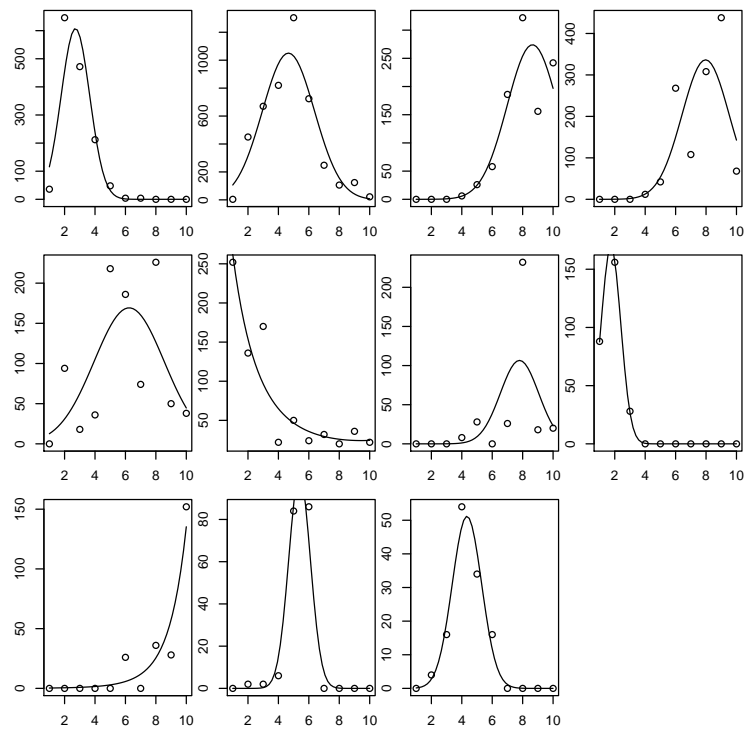
Le problème est résolu pour celle du milieu par (à mettre au bon endroit) :

```
mu <- x[which(y==max(y))]
```



Celles qui restent ont un optimum manifestement en dehors de l'intervalle d'étude.
Pour contourner cette difficulté :

```
fun2 <- function(x, y) {  
  glm0 <- glm(y ~ x + I(x^2), family = poisson)  
  plot(x, y)  
  xnew <- seq(min(x), max(x), le = 50)  
  lines(xnew, predict(glm0, type = "response", newdata = list(x = xnew)))  
}  
par(mfrow = c(3, 4))  
par(mar = c(2, 2, 1, 0.2))  
apply(g, 2, fun2, x = x)
```

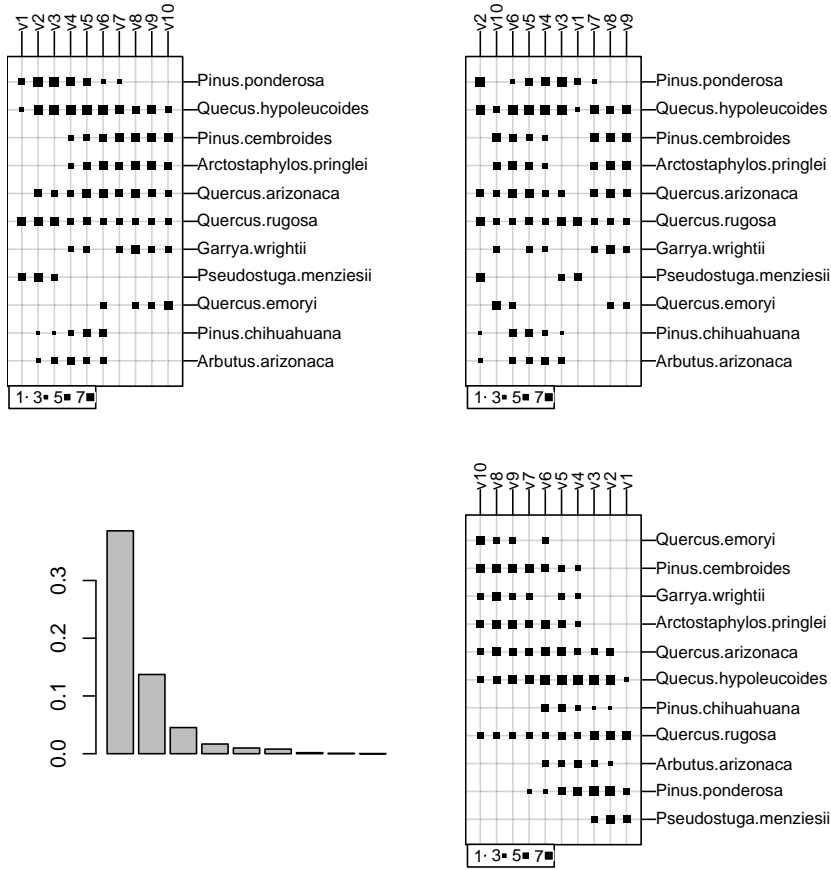


Ces techniques sont dites d'ordination directe. Beaucoup plus difficile est le cas où le gradient n'est pas connu mais à rechercher.

2 AFC et ordination indirecte

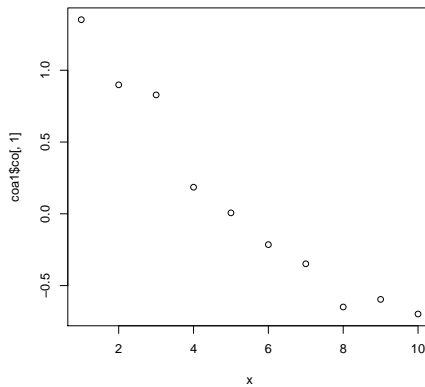
Ceci équivaut à traiter un tableau dont l'ordre des sites n'a aucun sens a priori :

```
par(mfrow = c(2, 2))
table.value(log(santacatalina + 1), grid = T)
table.value(log(santacatalina + 1)[, sample(10)], grid = T)
coa1 <- dudi.coa(log(santacatalina + 1), scann = F)
barplot(coa1$eig)
table.value(log(santacatalina + 1)[order(coa1$li[, 1]), order(coa1$co[,
1])], grid = T)
```



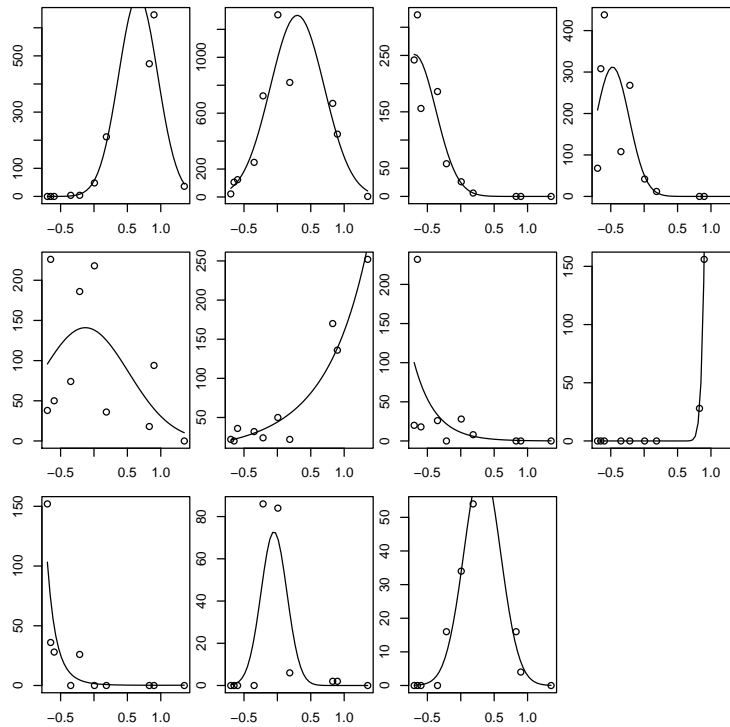
A gauche, le tableau des données : les sites (colonnes) sont les classes du gradient. En haut à droite, le tableau dont les colonnes sont dans un ordre arbitraire. En bas, le tableau traité : les sites et les espèces sont ordonnés

```
plot(x, coa1$co[, 1])
```



```

par(mfrow = c(3, 4))
par(mar = c(2, 2, 1, 0.2))
apply(g, 2, fun2, x = coa1$co[, 1])
    
```



L'analyse a trouvé le gradient dans les données (dans l'ordre inverse, ce qui est purement aléatoire). *L'analyse des correspondances d'un tableau floro-faunistique peut être considérée comme l'estimation des gradients qui sous-tendent leur organisation interne* (Ter Braak 1985 [23]).

3 Fer à cheval

```
santa.coa <- dudi.coa(santacatalina, scan = F)
```

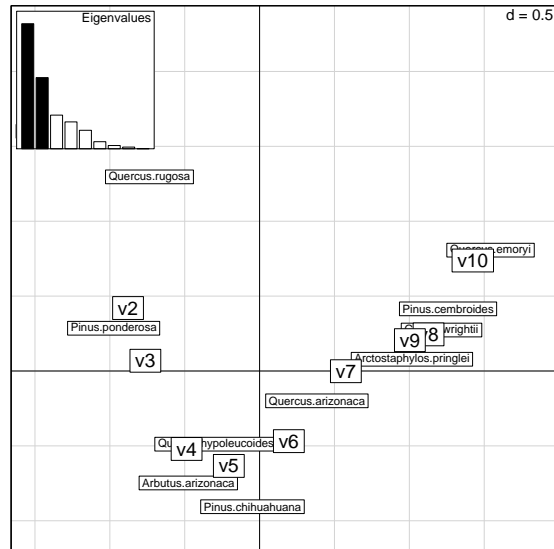
santa.coa est une liste :

```

santa.coa
Duality diagramm
class: coa dudi
$call: dudi.coa(df = santacatalina, scanf = F)
$nf: 2 axis-components saved
$rank: 9
eigen values: 0.6046 0.344 0.1627 0.13 0.08943 ...
  vector length mode  content
1 $cw   10      numeric column weights
2 $lw   11      numeric row weights
3 $eig   9      numeric eigen values

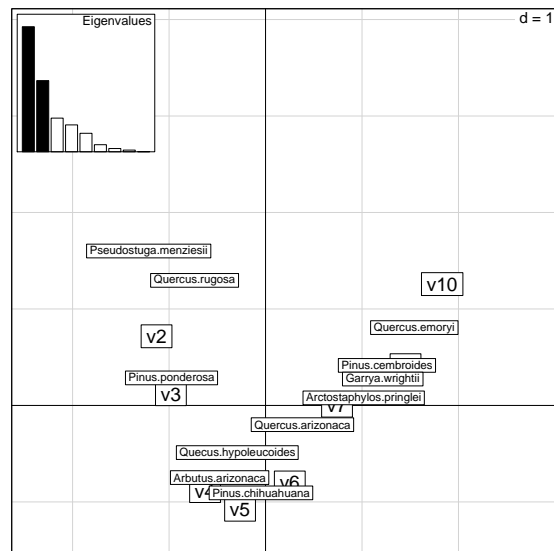
  data.frame nrow ncol content
1 $stab   11    10  modified array
2 $li     11     2   row coordinates
3 $l1     11     2   row normed scores
4 $co     10     2   column coordinates
5 $c1     10     2   column normed scores
other elements: N
    
```

`scatter(santa.coa)`

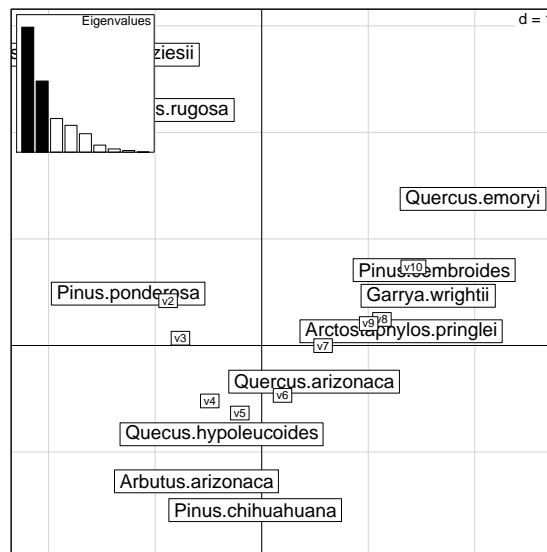


`args(scatter.coa)`

`scatter(santa.coa, met = 2)`



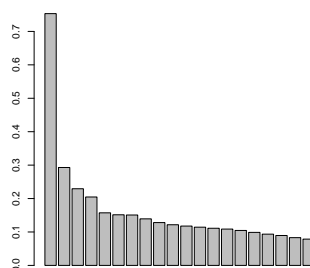
`scatter(santa.coa, met = 3)`



On a trois représentations fondamentales : la carte classique, les deux cartes doubles.

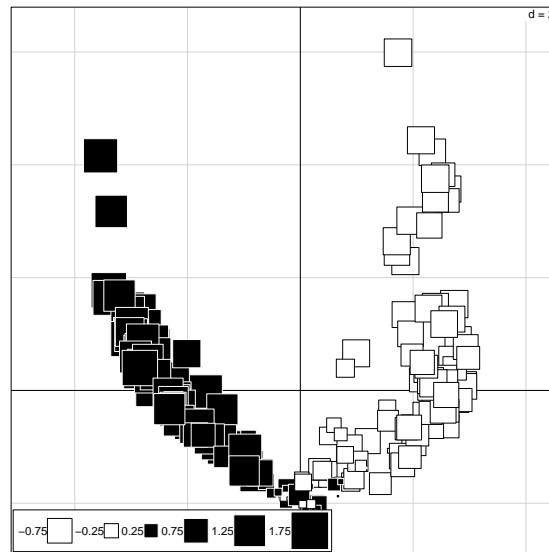
Dans la première, on a la moitié de la seconde et la moitié de la troisième. Dans la seconde, les sites sont positionnés avec des scores de variance unité et les espèces à la moyenne de leur distribution dans les relevés ont des coordonnées de variance λ . Dans la troisième, les espèces sont positionnées avec des scores de variance unité et les sites à la moyenne de leur distribution en espèces ont des coordonnées de variance λ . Les trois ont un sens (Oksanen, J. 1987 [18]). On reproche foncièrement à l'AFC de plier les gradients sur deux dimensions.

```
data(rpjd1)
coa2 <- dudi.coa(rpjd1$fau, scan = F)
barplot(coa2$eig[1:20])
```

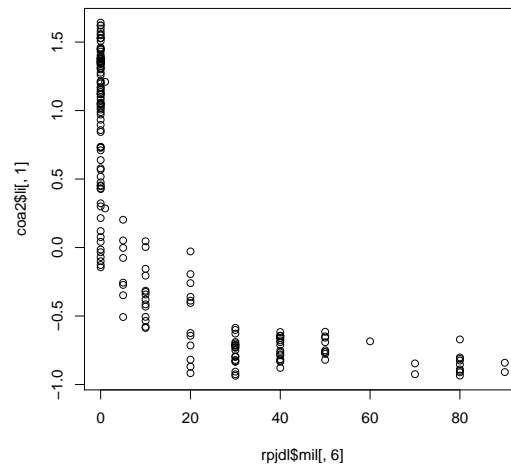


Cette configuration des valeurs propres est caractéristique d'un gradient sous-jacent. Reprendre l'analyse du tableau de milieu qui est dans le même jeu de données (pps048).

```
millog <- log(rpjd1$mil + 1)
pcamil <- dudi.pca(millog, scann = F)
s.value(pcamil$i, coa2$i[, 1])
```

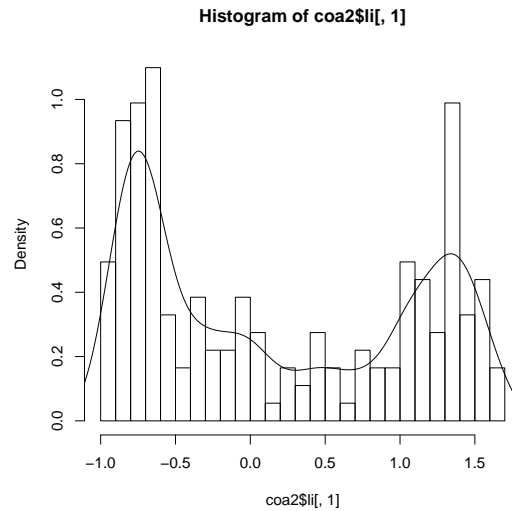


```
plot(rpjdl$mil[ 6], coa2$li[ 1])
```



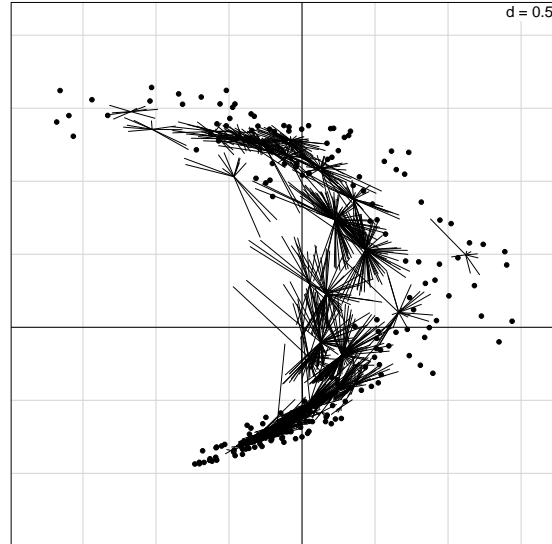
Quelle signification a ce graphe ?

```
hist(coa2$li[ 1], nclass = 20, proba = T)
lines(density(coa2$li[ 1], adj = 0.5))
```



Les coordonnées factorielles, comme scores optimaux, peuvent être utilisées comme des variables quantitatives. Ici l'AFC indique la présence de deux univers (avec ou sans arbres).

```
s.distri(coa2$li, rpjdl$fau, 2, 1, cstar = 0.3, cell = 0)
```

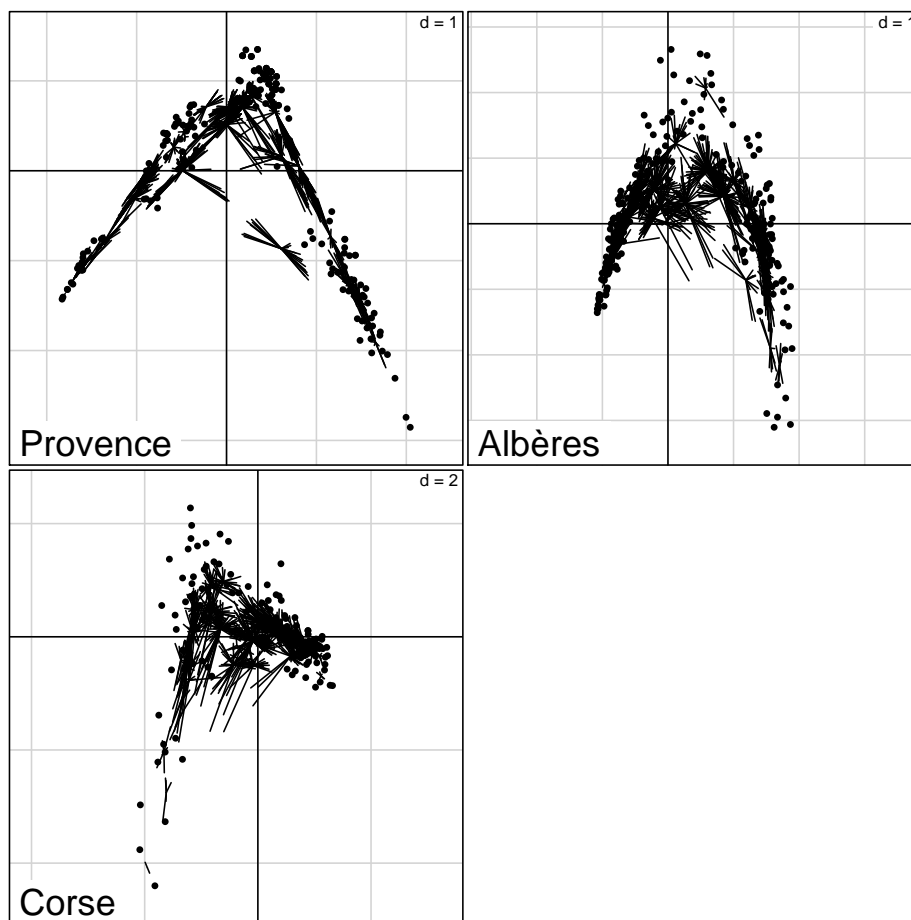


Il n'y a pratiquement pas d'espèces spécialistes du milieu du gradient (identifier l'exception). Deux pôles ont leur propre cortège et plusieurs espèces de grande amplitude font le raccord.

Les figures en arc sur les cartes factorielles sont fréquentes (effet Guttman, Horseshoe, arch effect). Elles passent pour des artefacts et certains s'en débarrassent (detrending). Il arrive qu'elles soient largement artefactuelles, il arrive qu'elle

mettent en évidence des propriétés particulières des données. S'en débarrasser "à l'aveugle" est une erreur.

```
data(avimedi)
provence <- avimedi$fau[avimedi$plan$reg == "Pr", ]
corse <- avimedi$fau[avimedi$plan$reg == "Co", ]
alberes <- rpjdl$fau
par(mfrow = c(2, 2))
s.distri(dudi.coa(provence, scan = F)$l1, provence, cstar = 0.3,
  cell = 0, sub = "Provence", csub = 2)
s.distri(dudi.coa(alberes, scan = F)$l1, alberes, cstar = 0.3, cell = 0,
  sub = "Albères", csub = 2)
s.distri(dudi.coa(corse, scan = F)$l1, corse, cstar = 0.3, cell = 0,
  sub = "Corse", csub = 2)
```



On fait la comparaison des structures internes de 3 tableaux faunistiques portant sur la relation avifaune-structure de la végétation en zone méditerranéenne et vue par la représentation des espèces à la moyenne des relevés. A gauche : plan 1-2 de l'AFC d'un tableau de 178 relevés comptant 45 espèces d'oiseaux. Les relevés sont dispersés sur un gradient d'ouverture du milieu (matorral très bas, bas et haut, taillis, futaie) dans plusieurs sites de Provence. En bas : plan 1-2 de l'AFC d'un tableau de 124 relevés étagés sur le même gradient en Corse et comportant 43 espèces ... (Blondel 1985 Ch.5 [3]) A droite : plan 1-2 de l'analyse des correspondances d'un tableau de 182 relevés étagés sur le même gradient

d'ouverture dans les Albères (pelouse, maquis bas et haut, forêt) comportant 51 espèces (Prodon and Lebreton 1981 [20]).

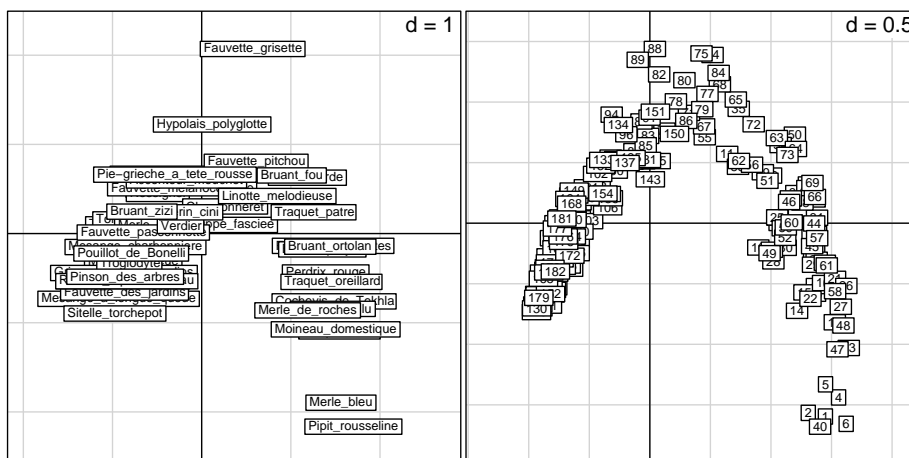
En dépit de variations considérables de la liste des espèces entre les trois observations, les structures globales sont conservées. Il y a deux gradients articulés (milieux ouverts à gauche, milieux fermés à droite), la même forme de nuage suggérant un effet Guttman et de fortes différences de structures internes. Les différences Provence-Corse ont été largement commentées dans le chapitre Biogéographie insulaire et le syndrome d'insularité (Blondel, op. cit. V-6 p. 146 [3]). La figure apporte un élément de poids dans la discussion qui hésite entre forme artificielle et information complémentaire entre les deux premiers facteurs.

Nevertheless, it is possible to neglect the other axes in the global gradient model, the unidimensional nature of which is an acceptable simplifying approximation since the non-correlation of the successive factors of C.A. prevent the contamination of F1 by other factors.

(Prodon et Lebreton, op. cit. p. 37 [20]), tandis que l'introduction de l'analyse canonique des correspondances porte à dire le contraire : *on aurait tort de considérer comme artificielle la relation entre F1 et F2, en la réduisant à un effet Guttman, et de vouloir la supprimer* (Lebreton et al. 1988 [12]).

C'est le mode d'organisation de l'amplitude des espèces le long des branches et entre les deux branches qui fait l'essentiel de l'information. Dans tous les cas la présence de deux gradients est indiscutable et c'est la seule quantification de la hauteur de la végétation qui laisse croire qu'il existe un gradient de hauteur de végétation. Il y a une faune des milieux ouverts et une faune des milieux boisés. Dans chacune, il y a étagement. En Provence la structure est forte sur les deux gradients et l'ambiguïté est faible à la jonction. Dans les Albères ces deux gradients sont articulés fortement par un paquet d'espèces à amplitude large ou distribution bimodale. En Corse la dilatation des niches augmente encore et la structure est affaiblie sur les deux branches. Pour préciser :

```
par(mfrow = c(1, 2))
s.label(coa2$co, clab = 0.6, lab = rpjdl$frlab)
s.label(coa2$li, clab = 0.6)
```



L'exemple montre la fonction essentielle de l'analyse : décrire la structure d'un tableau, c'est-à-dire faire le bilan de l'ensemble des ressemblances et différences entre les objets.

4 Corrélation canonique

On a l'âge (en année, de 1 à 8) et la fécondité (en nombre de chatons produits dans l'année, 0 pour aucun, 1 ou 2, 3 ou 4, ..., 13 ou 14) pour 350 chattes domestiques sous la forme d'une table de contingence (Legay and Pontier 1985 [13]) :

```
data(chats)
chats
      f0 f12 f34 f56 f78 f9a fbc fcd
age1  8  15  44  11  7  4  0  0
age2  6  12  36  21  11  6  1  1
age3  4  7  18  13  12  4  2  2
age4  2  8  7  3  7  5  1  0
age5  2  3  5  3  4  6  0  0
age6  1  0  5  3  2  2  1  1
age7  0  0  3  2  5  4  1  1
age8  2  2  5  1  7  4  1  0

age <- 1:8
fecon <- c(0, 1.5, 3.5, 5.5, 7.5, 9.5, 11.5, 13.5)
head(cbind(as.vector(row(as.matrix(chats))), as.vector(col(as.matrix(chats)))))

      [,1] [,2]
[1,]    1    1
[2,]    2    1
[3,]    3    1
[4,]    4    1
[5,]    5    1
[6,]    6    1

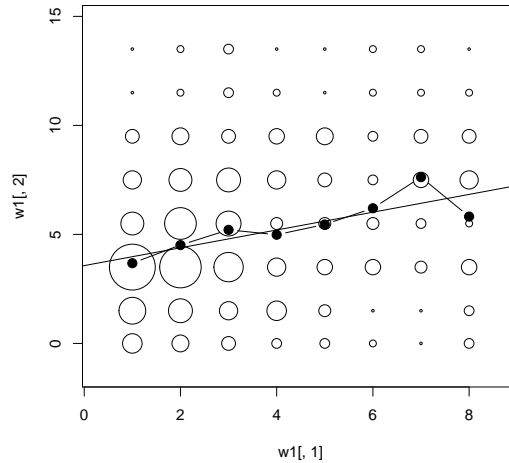
w1 <- cbind(as.vector(row(as.matrix(chats))), as.vector(col(as.matrix(chats))))
w1[, 1] <- age[w1[, 1]]
w1[, 2] <- fecon[w1[, 2]]
head(w1)

      [,1] [,2]
[1,]    1    0
[2,]    2    0
[3,]    3    0
[4,]    4    0
[5,]    5    0
[6,]    6    0

w1 <- data.frame(w1)
names(w1) <- c("age", "fecond")
w1$eff <- as.numeric(unlist(chats))
head(w1)

  age fecond eff
1   1      0   8
2   2      0   6
3   3      0   4
4   4      0   2
5   5      0   2
6   6      0   1

symbols(w1[, 1], w1[, 2], sqrt(w1[, 3]), inch = 0.25, col = grey(0.3))
abline(lm(rep(w1$fecond, w1$eff) ~ rep(w1$age, w1$eff)))
points(age, tapply(rep(w1$fecond, w1$eff), as.factor(rep(w1$age,
w1$eff)), mean), type = "b", pch = 20, cex = 2)
```



Valeur de la droite de régression :

```
fecrep <- rep(w1$fecond, w1$eff)
agerep <- rep(w1$age, w1$eff)
summary(lm(fecrep ~ agerep))
Call:
lm(formula = fecrep ~ agerep)
Residuals:
    Min       1Q   Median       3Q      Max
-6.835 -2.110 -0.477  1.523  9.115
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.56879    0.27504  12.975 < 2e-16 ***
agerep       0.40823    0.07425   5.498 7.38e-08 ***
---
Signif. codes:  0

Residual standard error: 2.889 on 352 degrees of freedom
Multiple R-squared: 0.07909, Adjusted R-squared: 0.07648
F-statistic: 30.23 on 1 and 352 DF, p-value: 7.38e-08
```

Valeur de la courbe de régression :

```
summary(lm(fecrep ~ as.factor(agerep)))
Call:
lm(formula = fecrep ~ as.factor(agerep))
Residuals:
    Min       1Q   Median       3Q      Max
-6.2000 -1.9565 -0.1798  1.8202  8.9894
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.6798    0.3053  12.055 < 2e-16 ***
as.factor(agerep)2  0.8309    0.4259   1.951  0.05190 .
as.factor(agerep)3  1.5299    0.4764   3.211  0.00144 **
as.factor(agerep)4  1.3051    0.5869   2.224  0.02683 *
as.factor(agerep)5  1.7767    0.6736   2.638  0.00873 **
as.factor(agerep)6  2.5202    0.8038   3.135  0.00186 **
as.factor(agerep)7  3.9452    0.7820   5.045 7.34e-07 ***
as.factor(agerep)8  2.1384    0.6857   3.119  0.00197 **
---
Signif. codes:  0

Residual standard error: 2.88 on 346 degrees of freedom
Multiple R-squared: 0.1005, Adjusted R-squared: 0.08233
F-statistic: 5.524 on 7 and 346 DF, p-value: 4.858e-06
```

Test de linéarité :

```

anova(lm(fecrep ~ agerep + as.factor(agerep)))
Analysis of Variance Table
Response: fecrep

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
agerep	1	252.32	252.32	30.4243	6.809e-08 ***
as.factor(agerep)	6	68.39	11.40	1.3743	0.2241
Residuals	346	2869.51	8.29		

```

---
Signif. codes:  0

```

La fécondité croit linéairement avec l'âge ?

```

coachats <- dudi.coa(chats, scann = F)
coachats
Duality diagram
class: coa dudi
$call: dudi.coa(df = chats, scannf = F)
$nf: 2 axis-components saved
$rank: 7
eigen values: 0.121 0.0526 0.01524 0.01291 0.005409 ...
vector length mode content
1 $scw 8 numeric column weights
2 $slw 8 numeric row weights
3 $eig 7 numeric eigen values
data.frame nrow ncol content
1 $stab 8 8 modified array
2 $li 8 2 row coordinates
3 $li 8 2 row normed scores
4 $co 8 2 column coordinates
5 $c1 8 2 column normed scores
other elements: N

```

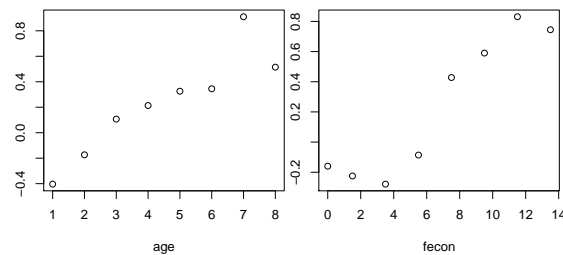
On a toujours $r_{xy}^2 \leq \eta_{xy}^2 \leq \lambda_1$ (à démontrer)

On gagne 0.02 à chaque fois. Le second n'est pas significativement différent du premier. Et le troisième ? L'AFC recode les variables :

```

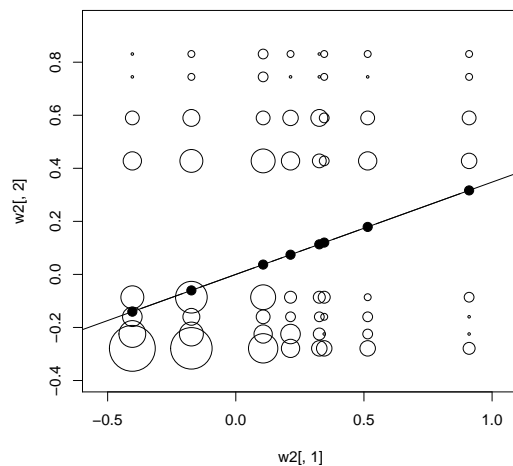
coachats$li
      Axis1      Axis2
age1 -0.4036788 0.08844887
age2 -0.1733273 -0.11129392
age3  0.1067175 -0.21186920
age4  0.2135553  0.37117124
age5  0.3258428  0.36005933
age6  0.3450007 -0.50947663
age7  0.9103606 -0.21661476
age8  0.5146002  0.28652452
coachats$co
      Comp1      Comp2
f0 -0.15975157 0.12528621
f12 -0.22461117 0.29042413
f34 -0.27871752 -0.04588939
f56 -0.08594866 -0.27536672
f78  0.42802304 0.06298687
f9a  0.59002314 0.26352503
fbc  0.83090261 -0.37584944
fcd  0.74468956 -1.09969996
par(mfrow = c(1, 2))
par(mar = c(5, 2, 1, 0.2))
plot(age, coachats$li[, 1])
plot(fecon, coachats$co[, 1])

```



```

w2 <- cbind(as.vector(row(as.matrix(chats))), as.vector(col(as.matrix(chats))))
w2[, 1] <- coachats$li[w2[, 1], 1]
w2[, 2] <- coachats$co[w2[, 2], 1]
w2 <- data.frame(w2)
names(w2) <- c("pseudo.age", "pseudo.fecond")
w2$eff <- as.numeric(unlist(chats))
symbols(w2[, 1], w2[, 2], sqrt(w2[, 3]), inch = 0.25, col = grey(0.3))
abline(lm(rep(w2$pseudo.fecond, w2$eff) ~ rep(w2$pseudo.age, w2$eff)))
x1 <- coachats$li[, 1]
y1 <- tapply(rep(w2$pseudo.fecond, w2$eff), as.factor(rep(w2$pseudo.age,
w2$eff)), mean)
points(x1[order(x1)], y1[order(y1)], type = "b", pch = 20, cex = 2)
    
```

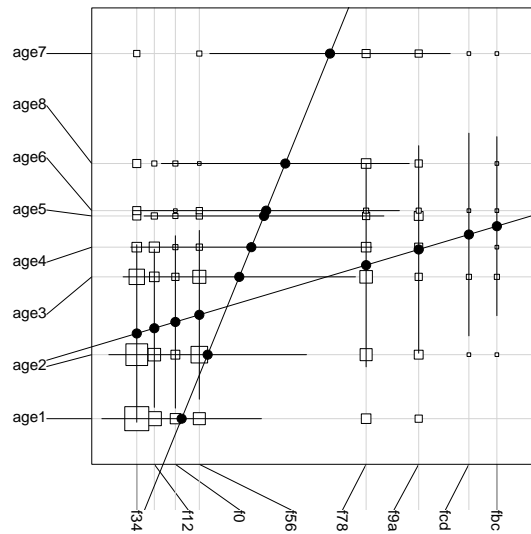


On note deux différences essentielles avec le graphe naturel. La première est mathématique. La régression est linéaire et doublement linéaire (Hirschfeld, H. O. 1935 [10]). La seconde concerne les données : on augmente la corrélation en respectant l'âge et en faisant deux classes de fécondité (0-6, 7-14). **Pourquoi ?**

On refait directement toutes ces opérations avec

```

table.cont(chats, coachats$co[, 1], coachats$li[, 1], abmean.x = T,
abmean.y = T, abline.x = T, abline.y = T)
    
```

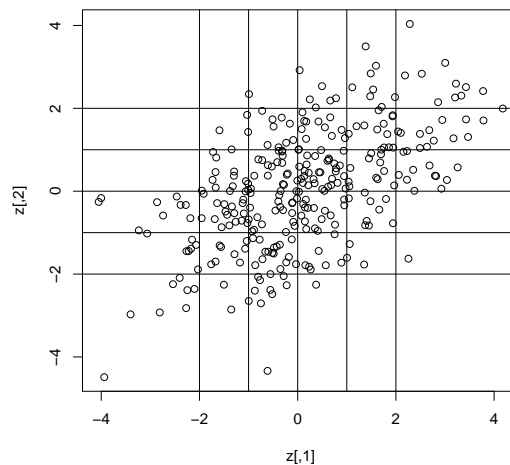


Hommage à H.O. Hirschfeld, un génial précurseur

Sur les vraies tables de contingence, l'AFC est irremplaçable.

5 Une curiosité

```
library(MASS)
z <- mvrnorm(300, mu = c(0, 0), Sigma = matrix(c(2, 1, 1, 2), 2,
2))
plot(z)
abline(h = seq(-2, 2, by = 1))
abline(v = seq(-2, 2, by = 1))
```



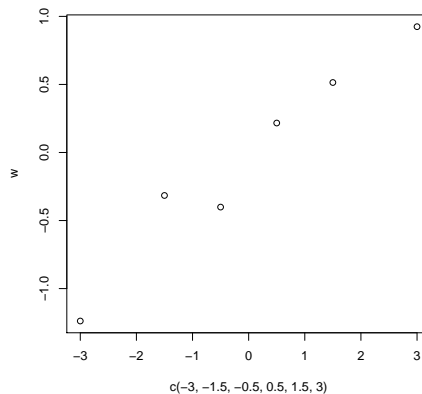
```
br0 <- c(-4, seq(-2, 2, by = 1), 4)
table(cut(z[, 1], br0), cut(z[, 2], br0))
```

```

      (-4,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,4]
(-4,-2]      7      7      7      0      0      0
(-2,-1]      2      8     18     13     4      0
(-1,0]     10     16     19     20     9      1
(0,1]       1      7     17     31    13     6
(1,2]       0      3      9     11    13     8
(2,4]       0      1      0     12    12     9
tab <- table(cut(z[, 1], br0), cut(z[, 2], br0))
as.matrix(tab)
      (-4,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,4]
(-4,-2]      7      7      7      0      0      0
(-2,-1]      2      8     18     13     4      0
(-1,0]     10     16     19     20     9      1
(0,1]       1      7     17     31    13     6
(1,2]       0      3      9     11    13     8
(2,4]       0      1      0     12    12     9
head(data.frame(tab))
  Var1   Var2 Freq
1 (-4,-2] (-4,-2] 7
2 (-2,-1] (-4,-2] 2
3 (-1,0] (-4,-2] 10
4 (0,1] (-4,-2] 1
5 (1,2] (-4,-2] 0
6 (2,4] (-4,-2] 0
tabnew <- tab[sample(6), sample(6)]
tabnew
      (-2,-1] (1,2] (0,1] (-1,0] (2,4] (-4,-2]
(0,1]      7     13     31     17      6      1
(1,2]      3     13     11      9      8      0
(-4,-2]     7      0      0      7      0      7
(-2,-1]     8      4     13     18      0      2
(2,4]      1     12     12      0      9      0
(-1,0]     16      9     20     19      1     10
dudi.coa(data.frame(unclass(tabnew)), scann = F)$li
      Axis1      Axis2
(0,1] -0.2165159 -0.22307523
(1,2] -0.5142370  0.12195775
(-4,-2] 1.2381187  0.57783120
(-2,-1] 0.3157390 -0.35973725
(2,4] -0.9243955  0.37632993
(-1,0] 0.4011442  0.03497339
dudi.coa(data.frame(unclass(tab)), scann = F)$li
      Axis1      Axis2
(-4,-2] -1.2381187  0.57783120
(-2,-1] -0.3157390 -0.35973725
(-1,0] -0.4011442  0.03497339
(0,1]  0.2165159 -0.22307523
(1,2]  0.5142370  0.12195775
(2,4]  0.9243955  0.37632993

w <- dudi.coa(data.frame(unclass(tab)), scann = F)$li[, 1]
plot(c(-3, -1.5, -0.5, 0.5, 1.5, 3), w)

```



L'AFC est une méthode d'estimation des centres des classes d'une statistique groupée binormale (Kendall and Stuart 1961 p. 536-591 [11])

```
sigma = matrix(c(2, 1, 1, 2), 2, 2)
cor(z)
      [,1] [,2]
[1,] 1.0000000 0.5458469
[2,] 0.5458469 1.0000000
sqrt(dudi.coa(data.frame(unclass(tab)), scann = F)$eig)
[1] 0.56228115 0.27434988 0.12644226 0.04510831 0.01616807
```

6 AFC comme analyse canonique

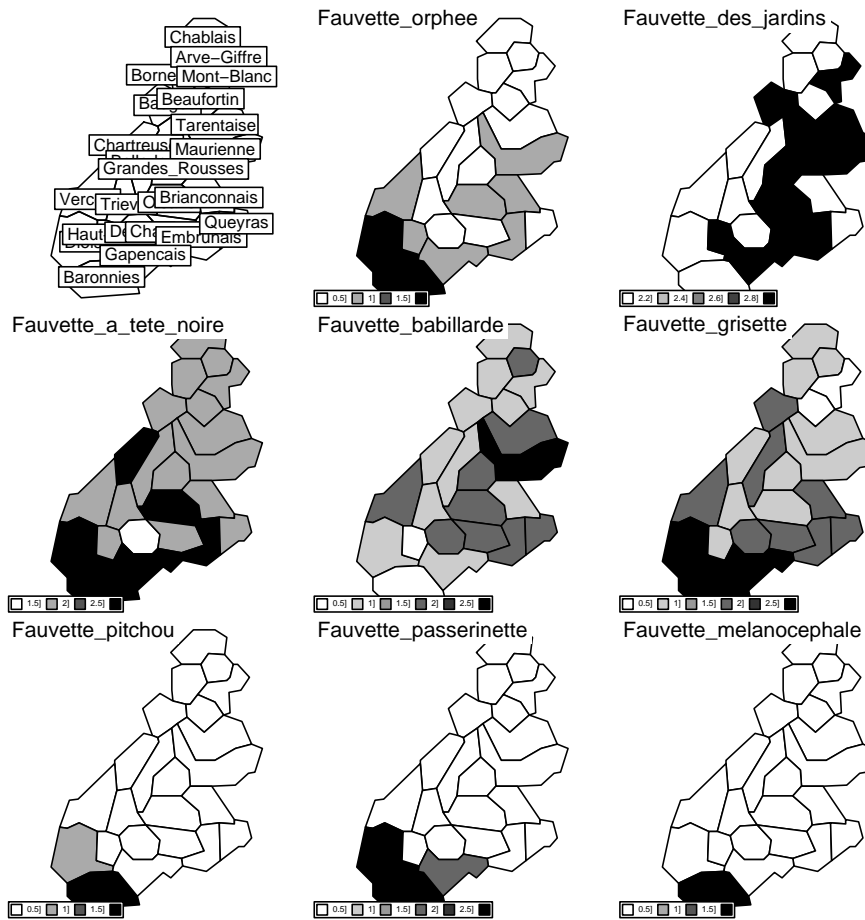
Pour disposer d'un petit tableau :

```
data(atlas)
w <- data.frame(t(atlas$birds[, 1:8]))
w
```

	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15
Fauvette_orphee	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1
Fauvette_des_jardins	2	2	2	3	3	2	2	2	3	3	3	2	2	3	2
Fauvette_a_tete_noire	2	2	2	2	2	2	3	2	2	2	2	2	2	3	2
Fauvette_babillarde	1	1	2	1	1	1	1	1	2	3	2	2	1	2	1
Fauvette_grisette	1	1	1	0	2	0	1	2	1	1	1	2	1	1	2
Fauvette_pitchou	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fauvette_passerinette	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fauvette_melanocephale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	S16	S17	S18	S19	S20	S21	S22	S23							
Fauvette_orphee	2	1	0	0	1	0	2	1							
Fauvette_des_jardins	2	3	2	3	3	2	2	3							
Fauvette_a_tete_noire	3	2	1	2	3	2	3	3							
Fauvette_babillarde	1	0	2	2	2	2	0	1							
Fauvette_grisette	3	1	2	2	2	2	3	3							
Fauvette_pitchou	1	0	0	0	0	0	2	0							
Fauvette_passerinette	3	0	0	0	0	0	3	2							
Fauvette_melanocephale	0	0	0	0	0	0	2	0							

Pour savoir ce qu'il y a dedans :

```
par(mfrow = c(3, 3))
area.plot(atlas$area, lab = atlas$names.district, clab = 1.5)
for (i in 1:8) area.plot(atlas$area, , as.numeric(w[i, ]), sub = row.names(w)[i],
  csub = 2)
sum(w > 0)
[1] 103
```

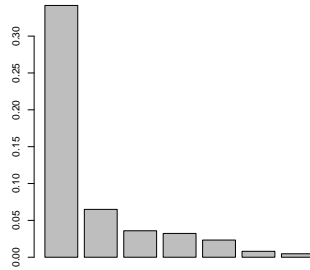


Le tableau contient 103 correspondances, ou cases non nulles, ou présences d'une espèce dans un district. Ces 103 correspondances se regroupent en 23 paquets-colonnes (les espèces présentes dans un site) et 8 paquets-lignes (les sites qui contiennent une espèce). Le poids d'une correspondance est la note d'abondance de "son" espèce dans "son" site.

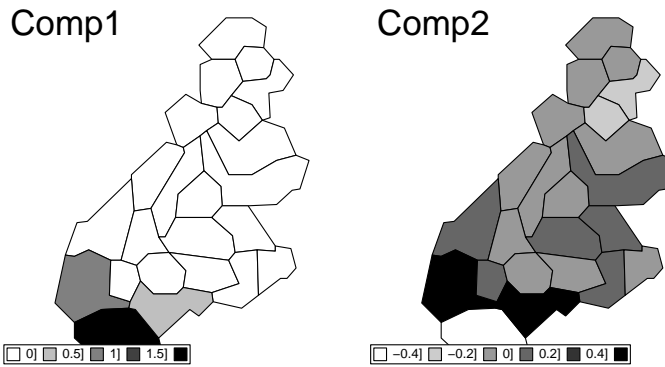
Le véritable individu dans une AFC est la correspondance qui lui donne son nom.

Fondamentalement, l'analyse des correspondances donne des scores numériques à ces correspondances.

```
w.coa <- dudi.coa(w, scann = F)
barplot(w.coa$eig)
```



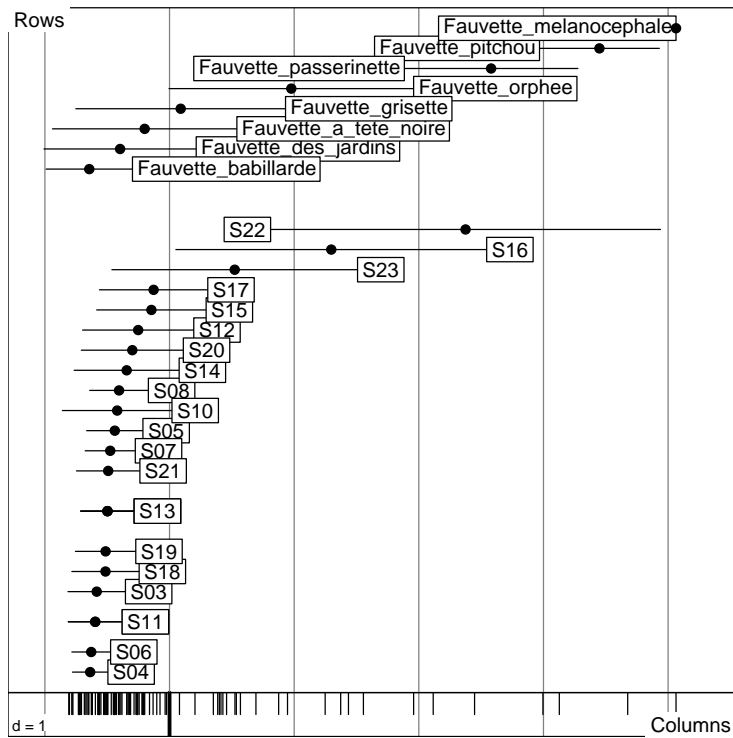
```
par(mfrow = c(1, 2))
for (i in 1:2) area.plot(atlas$area, , as.numeric(w.coa$co[, i]),
  sub = names(w.coa$co)[i], csub = 2)
```



L'arch effect est lié à la continuité entre un gradient Nord-Sud et la coupure de l'influence méditerranéenne.

Comme il y a un nombre impressionnant de correspondances, on ne s'en occupe généralement pas. Le graphe ci-dessous les positionne sur un axe avec une variance totale de 1. Chaque paquet-ligne et chaque paquet-colonne est un sous-ensemble pondéré de ces correspondances qui donne une moyenne (le point) et une variance (le trait vaut 1.5 écart-type).

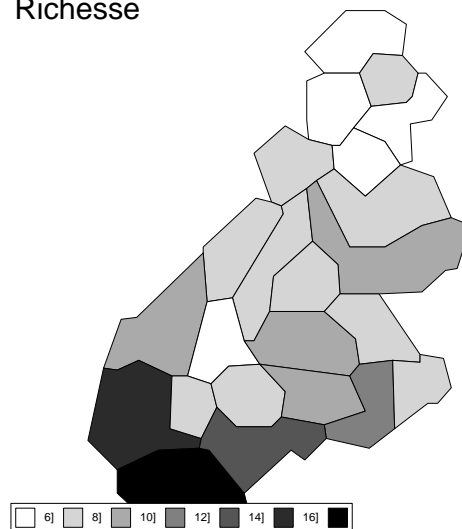
```
score(w.coa)
```



L'amplitude d'habitat de la Fauvette mélanocéphale est nulle (elle n'habite qu'un seul district). La variance est la mesure de l'amplitude d'habitat vue dans l'analyse. Mais en plus, à la même échelle, la variance d'un bloc-site est la diversité. Le 22 est le plus divers. Il contient presque toutes les espèces du cortège. On a une double approche diversité-amplitude coordonnée (Thioulouse and Chessel 1992 [24]).

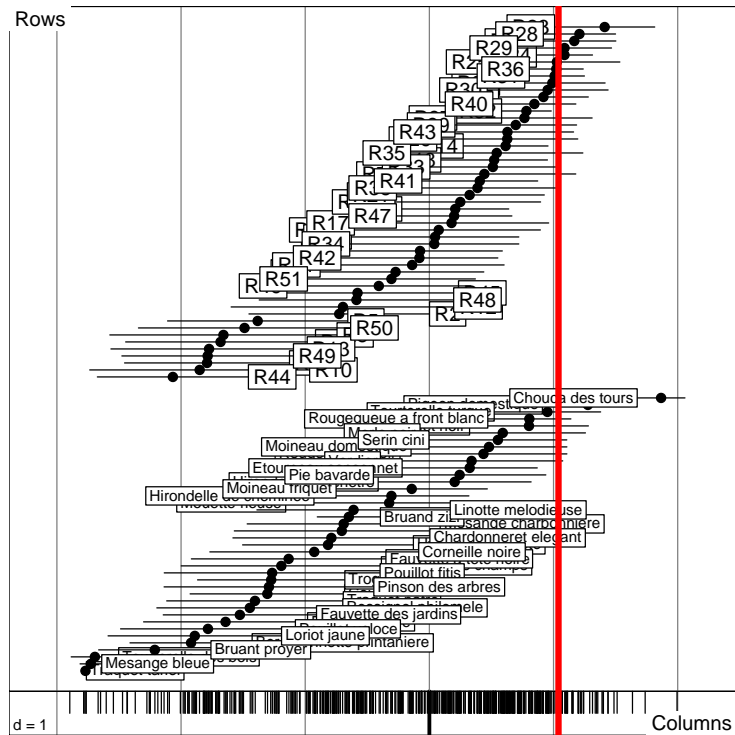
```
area.plot(atlas$area, , apply(w, 2, sum), sub = "Richesse", csub = 2)
```

Richesse



```

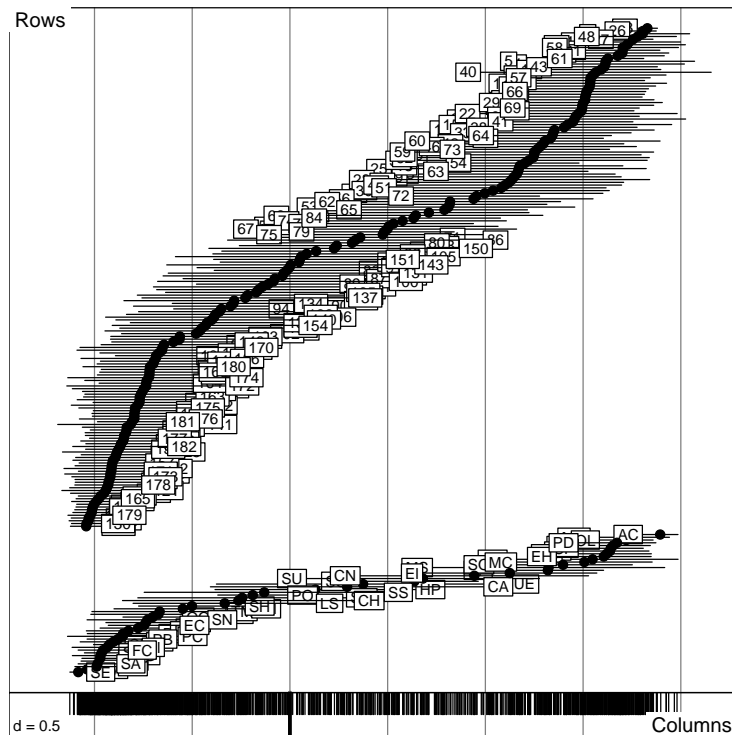
data(aviurba)
names(aviurba$fau) <- aviurba$species.names.fr
score(dudi.coa(aviurba$fau, scan = F), clab.c = 0.75)
abline(v = 1.04, lwd = 5, col = "red")
    
```



à droite les espèces n'entrant pas au contact de la ville, les relevés du milieu rural. à gauche le gradient de pénétration urbaine (centre - périphérie, au bout les nicheurs du centre-ville). Le trait définit une coupure entre milieux et une concentration d'espèces à niche large. L'analyse a repéré l'écotone (Données de Tatibouet et al. 1980 [22]).

```

score(dudi.coa(rpjdl$fau, scan = F), clab.c = 0.75, clab.r = 0.75)
    
```



On observe (ci-dessous) les deux concentrations de relevés avec leurs deux cortèges spécifiques et les relevés de transition de diversité stable associées aux espèces à large amplitude. Le score des correspondances est la variable canonique de rang 1 qui maximise simultanément et la variance inter-sites et la variance inter-espèces.

L'AFC est une méthode très particulière. Elle correspond exactement à la vision inférentielle d'une table de contingence, $\chi^2 = N(\lambda_1 + \dots + \lambda_r)$:

```
chisq.test(chats)
      Pearson's Chi-squared test
data:  chats
X-squared = 74.5351, df = 49, p-value = 0.01081
sum(dudi.coa(chats, scan = F)$eig) * sum(chats)
[1] 74.53507
```

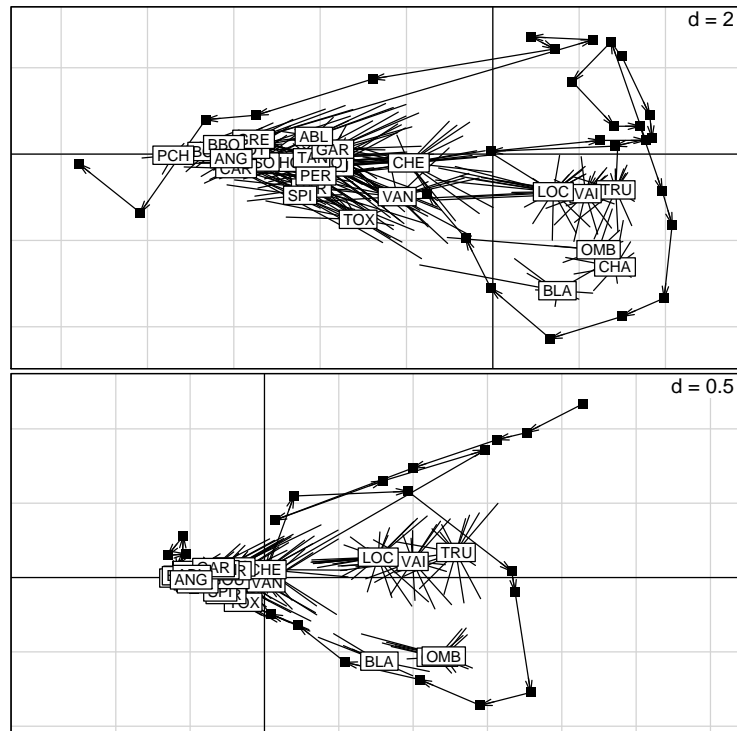
Ce calcul est toujours vrai : il n'a cependant de sens que pour une vraie table de contingence. L'argument, fort répandu, que l'AFC fonctionne bien, en écologie, du fait de l'introduction des métriques du Khi2, est largement une bêtise. Elle fonctionne bien comme analyse canonique (Estève 1978 [7]) et, par cohérence mathématique, peut aussi être vue comme une analyse d'inertie sur métriques du Khi2, ce qui pour les tableaux de présence-absence en particulier, n'a pas de sens précis.

L'AFC n'est pas *la* bonne méthode d'ordination. Elle est une méthode d'ordination et la seule qui relie diversité des sites et amplitude des espèces. Elle seule permet de voir une espèce comme distribution de son abondance entre sites *et* un site comme distribution de son contenu entre espèces. Elle s'impose quand on se pose deux questions en même temps :

- ★ quelle typologie entre sites est-elle induite par les espèces ?
- ★ quelle typologie entre espèces est-elle induite par les sites ?

Elle fait alors un compromis cohérent des réponses à ces deux questions. Si on ne se pose qu'une seule question, le compromis peut être une gêne. On l'utilise souvent pour éviter de savoir quelle question est effectivement posée. Sur un cas d'école (Verneaux 1973 [28]) :

```
data(doubs)
pcad <- dudi.pca(doubs$poi, scan = F)
coad <- dudi.coa(doubs$poi, scan = F)
par(mfrow = c(2, 1))
s.traject(pcad$li, clab = 0)
s.distri(pcad$li, doubs$poi, csta = 0.33, cell = 0, clab = 0.75,
  add.p = T)
s.traject(-coad$li, clab = 0)
s.distri(-coad$li, doubs$poi, csta = 0.33, cell = 0, clab = 0.75,
  add.p = T)
```



Identifier ce que l'AFC a perdu.

7 Fréquences alléliques

L'AFC est d'abord un programme dans lequel on envoie des tableaux de nombres positifs ou nuls.

Considérons des groupes d'individus identiques pour plusieurs variables qualitatives (Pontier 1984 [19]) :

```

data(chatcat)
w <- acm.disjonctif(chatcat$tab)
names(w) <- c(paste("A", 1:5, sep = ""), paste("B", 1:5, sep = ""),
             paste("C", 1:2, sep = ""))
w <- t(w * chatcat$num) %*% as.matrix(w)
w <- data.frame(w)
w

```

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2
A1	32	0	0	0	0	7	23	1	1	0	25	7
A2	0	47	0	0	0	3	24	14	4	2	17	30
A3	0	0	18	0	0	0	4	8	6	0	4	14
A4	0	0	0	19	0	0	8	5	4	2	3	16
A5	0	0	0	0	18	2	6	5	5	0	5	13
B1	7	3	0	0	2	12	0	0	0	0	12	0
B2	23	24	4	8	6	0	65	0	0	0	41	24
B3	1	14	8	5	5	0	0	33	0	0	1	32
B4	1	4	6	4	5	0	0	0	20	0	0	20
B5	0	2	0	2	0	0	0	0	0	4	0	4
C1	25	17	4	3	5	12	41	1	0	0	54	0
C2	7	30	14	16	13	0	24	32	20	4	0	80

Ce tableau est une juxtaposition de tables de contingences.

```

coachat <- dudi.coa(w, scan = F)
head(coachat$l1)

```

	RS1	RS2
A1	1.4747478	-0.3042755
A2	-0.0991037	0.8645123
A3	-1.0405188	-1.7359176
A4	-0.8299285	1.6882651
A5	-0.4464488	-1.7625433
B1	1.8387875	-1.3588561

```

tail(coachat$l1)

```

	RS1	RS2
B2	0.6902131	0.45775889
B3	-1.1187489	-0.07525643
B4	-1.2612140	-1.60398270
B5	-1.1965777	5.27876525
C1	1.2678660	-0.08888907
C2	-0.8558096	0.06000012

```

acmchat <- dudi.acm(chatcat$tab, chatcat$num, scan = F)
head(acmchat$c1)

```

	CS1	CS2
age.1	1.4747478	-0.3042755
age.2.3	-0.0991037	0.8645123
age.4.5	-1.0405188	-1.7359176
age.6.7	-0.8299285	1.6882651
age.>8	-0.4464488	-1.7625433
feco.1.2	1.8387875	-1.3588561

```

tail(acmchat$c1)

```

	CS1	CS2
feco.3.6	0.6902131	0.45775889
feco.7.8	-1.1187489	-0.07525643
feco.9.12	-1.2612140	-1.60398270
feco.13.14	-1.1965777	5.27876525
nport.1	1.2678660	-0.08888907
nport.2	-0.8558096	0.06000012

L'AFC d'un tableau de BURT est partiellement l'analyse des correspondances multiples d'un tableau de variables qualitatives.

```

w <- acm.disjonctif(chatcat$tab)
coa2 <- dudi.coa(w * chatcat$num, scan = F)
head(coa2$c1)

```

	CS1	CS2
age.1	1.4747478	-0.3042755
age.2-3	-0.0991037	0.8645123
age.4-5	-1.0405188	-1.7359176
age.6-7	-0.8299285	1.6882651
age.>=8	-0.4464488	-1.7625433
feco.1-2	1.8387875	-1.3588561

```

tail(coa2$c1)

```

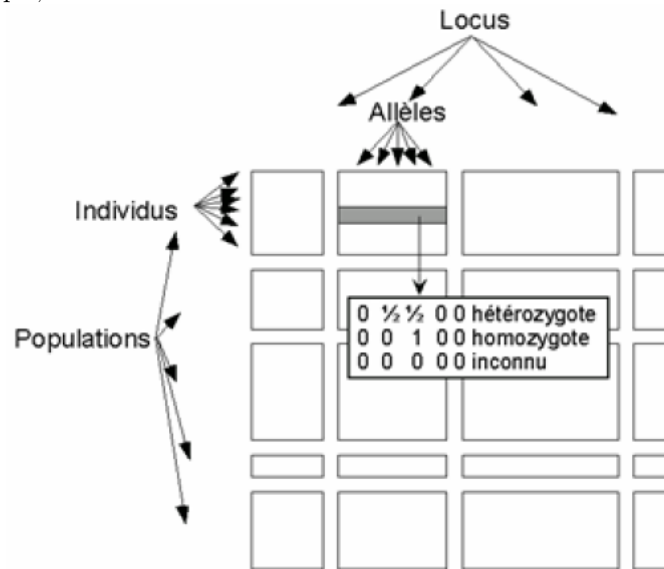
```

                CS1      CS2
feco.3-6      0.6902131  0.45775889
feco.7-8      -1.1187489 -0.07525643
feco.9-12     -1.2612140 -1.60398270
feco.13-14   -1.1965777  5.27876525
nport.1       1.2678660 -0.08888907
nport.2      -0.8558096  0.06000012

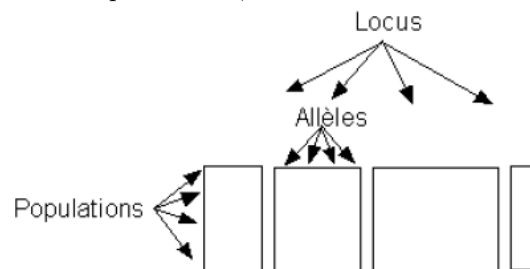
```

L'AFC d'un tableau disjonctif complet est encore partiellement l'analyse des correspondances multiples d'un tableau de variables qualitatives. Pour des raisons mathématiques, la même procédure est cohérente si on l'applique au tableau disjonctif et au tableau croisé.

En génétique, la structure de base des données se résume à :



Conformément aux concepts de base, les données traitées sont réduites à :



Chaque tableau est une table de contingence qui relève typiquement d'une analyse des correspondances. *Mais chaque tableau séparément relève d'une AFC.* On est donc très proche des tableaux-bandes sans y être vraiment.

```

w <- acm.burt(data.frame(chatcat$tab[, 1]), chatcat$tab[, 2:3],
  chatcat$num)
row.names(w) <- levels(chatcat$tab$age)
w

```

	feco.1-2	feco.3-6	feco.7-8	feco.9-12	feco.13-14	nport.1	nport.2
1	7	23	1	1	0	25	7
2-3	3	24	14	4	2	17	30
4-5	0	4	8	6	0	4	14
6-7	0	8	5	4	2	3	16
>=8	2	6	5	5	0	5	13

Ces deux tables de contingence juxtaposées ont même somme marginale sur les lignes. Ce sont les mêmes individus (par exemple, les chattes de 1 an) qui sont distribuées suivant le nombre de chatons et le nombre de portées. Dans un tableau de fréquences alléliques, ce n'est pas le cas dès qu'il y a des données manquantes, ce qui ne manque pas, ou même des échantillons différents ont pu fournir des tableaux accolés. On trouvera de telles données dans *microsatt*, préparée par D. Laloë (Moazami-Goudarzi et al. 1997 [17], Souvenir Zafindrajaona et al. 1999 [21], Moazami-Goudarzi et al. 2001 [16]) :

```
data(microsatt)
names(microsatt)
[1] "tab"          "loci.names"    "loci.eff"      "alleles.names"
dim(microsatt$tab)
[1] 18 112
microsatt$loci.eff
[1] 8 15 11 10 17 10 14 15 12
sum(microsatt$loci.eff)
[1] 112
row.names(microsatt$tab)
[1] "Baoule"      "Borgou"        "BPN"           "Charolais"     "Holstein"      "Jersey"
[7] "Lagunaire"   "Limousin"      "MaineAnjou"   "Mtbeliard"     "NDama"         "Normand"
[13] "Parthenais" "Somba"         "Vosgienne"    "ZChoa"         "ZMbororo"      "Zpeul"
```

18 races bovines, 9 loci totalisant 112 allèles.

```
loci <- rep(microsatt$loci.names, microsatt$loci.eff)
loci.fac <- as.factor(loci)
tapply(apply(microsatt$tab, 2, sum), loci.fac, sum)
ETH152 ETH225 INRA16 INRA32 INRA35 INRA5 INRA63 INRA72 INRAK
1494 1562 1548 1554 1548 1576 1538 1516 1456
```

Ici, le plan est très équilibré et les effectifs par sous-tableau sont voisins sans être égaux. On fait généralement l'impasse sur le problème.

Le tableau de départ contient K tableaux \mathbf{X}_k avec I lignes (populations) et J_k colonnes (allèles du locus k). x_{ij}^k est la valeur du tableau k à la ligne i ($1 \leq i \leq I$) et à la colonne j ($1 \leq j \leq J_k$). Les sommes marginales du tableau k sont :

$$\left. \begin{aligned} x_{\bullet j}^k &= \sum_{i=1}^I x_{ij}^k \\ x_{i\bullet}^k &= \sum_{j=1}^{J_k} x_{ij}^k \\ x_{\bullet\bullet}^k &= \sum_{j=1}^{J_k} x_{\bullet j}^k = \sum_{i=1}^I x_{i\bullet}^k \end{aligned} \right\} \Rightarrow \begin{cases} p_{\bullet j}^k = x_{\bullet j}^k / x_{\bullet\bullet}^k \\ p_{i\bullet}^k = x_{i\bullet}^k / x_{\bullet\bullet}^k \\ p_{ij}^k = x_{ij}^k / x_{\bullet\bullet}^k \end{cases}$$

Les analyses de chaque tableau sont définies par les triplets :

$$T_k = \left(\mathbf{P}_0^k = \left[\frac{p_{ij}^k}{p_{i\bullet}^k p_{\bullet j}^k} - 1 \right], \mathbf{D}_j^k = \text{diag}(p_{\bullet 1}^k, \dots, p_{\bullet J_k}^k), \mathbf{D}_I^k = \text{diag}(p_{1\bullet}^k, \dots, p_{I\bullet}^k) \right)$$

L'analyse intra-classe (Benzécri 1983 [1]) sur la partition des colonnes par tableaux introduit la coordination de ces triplets. La pondération commune à tous les tableaux est définie par :

$$p_i = \frac{\sum_{k=1}^K x_{i\bullet}^k}{\sum_{k=1}^K x_{\bullet\bullet}^k} \Rightarrow \mathbf{D}_I = \text{diag}(p_1, \dots, p_I)$$

tandis que le poids du tableau k est :

$$t_k = \frac{x_{\bullet\bullet}^k}{\sum_{k=1}^K x_{\bullet\bullet}^k}$$

L'analyse intra-classe est une méthode générale qui consiste à recentrer les nuages de points en amenant leur centre de gravité à l'origine. Elle a une version ACP (groupes d'individus) définie indépendamment en sciences sociales (Bourroche 1975 [4]) et en morphométrie (Thorpe 1983 [25], Thorpe and Leamy 1983 [26]). On fait cette opération sur le tableau général de l'AFC qui utilise :

$$x_{\bullet\bullet} = \sum_{k=1}^K x_{\bullet\bullet}^k \quad p_{ij}^{\bullet} = \frac{x_{ij}^{k(j)}}{x_{\bullet\bullet}^k} \quad p_i^{\bullet} = p_i = \frac{\sum_{k=1}^K x_{i\bullet}^k}{x_{\bullet\bullet}^k} \quad p_j^{\bullet} = \frac{x_{\bullet j}^{k(j)}}{x_{\bullet\bullet}^k}$$

où $k(j)$ est le numéro du tableau où se trouve la colonne j . L'analyse globale utilise le triplet :

$$T = \left(\mathbf{P}_0 = \begin{bmatrix} p_{ij}^{\bullet} \\ p_i^{\bullet} p_j^{\bullet} \end{bmatrix} - 1, \mathbf{D}_J^{\bullet} = \text{diag}(p_1^{\bullet}, \dots, p_J^{\bullet}), \mathbf{D}_I^{\bullet} = \text{diag}(p_1^{\bullet}, \dots, p_I^{\bullet}) \right)$$

Si on fait la moyenne des colonnes du tableau \mathbf{P}_0 pour le bloc de colonnes associé au tableau k on trouve à la position i la valeur :

$$\sum_{j=1}^{J_k} p_{\bullet j}^k \left(\frac{p_{ij}^{\bullet}}{p_i^{\bullet} p_j^{\bullet}} - 1 \right) = \frac{x_{i\bullet}^k x_{\bullet\bullet}^k}{x_{i\bullet}^k x_{\bullet\bullet}^k} - 1$$

Ceci met en évidence le tableau qui a autant de lignes que de populations et autant de colonnes que de tableaux et qui contient les effectifs marginaux par tableaux $x_{i\bullet}^k$. Si on note :

$$t_{k/i} = \frac{x_{i\bullet}^k}{x_{i\bullet}^k}$$

le triplet de l'analyse intra-classe est :

$$T_W = \left(\mathbf{Q}_0 = \begin{bmatrix} p_{ij}^{\bullet} & t_{k/i} \\ p_i^{\bullet} p_j^{\bullet} & t_k \end{bmatrix}, \mathbf{D}_J^{\bullet} = \text{diag}(p_1^{\bullet}, \dots, p_J^{\bullet}), \mathbf{D}_I^{\bullet} = \text{diag}(p_1^{\bullet}, \dots, p_I^{\bullet}) \right)$$

On voit immédiatement que l'AFC intra est exactement une AFC simple quand les profils marginaux sont constants ce qui est automatique dans l'AFC d'un tableau bande (Benzécri 1982 [2]) largement popularisée par Leclerc (1975 [14], 1976 [15]). L'AFC des tableaux doublement partitionnés dite Analyse des correspondances internes (Cazes et al. 1988 [5]) généralise cette approche.

Quand on extrait de ce schéma la part qui revient au tableau k , on garde :

$$T_W^k = \left(\mathbf{Q}_0^k = \begin{bmatrix} p_{ij}^{\bullet} & t_{k/i} \\ p_i^{\bullet} p_j^{\bullet} & t_k \end{bmatrix}, \mathbf{D}_{J_k}^{\bullet} = \text{diag}_{k(j)=k}(p_j^{\bullet}), \mathbf{D}_I^{\bullet} \right)$$

Ce n'est pas un triplet d'analyse des correspondances à cause du décentrage et du fait que la pondération des colonnes n'a pas une somme unité. $\mathbf{1}_{J_k}$ est encore axe principal pour la valeur propre nulle, $\mathbf{1}_I$ aussi. On a $\mathbf{D}_{J_k}^{\bullet} = t_k \mathbf{D}_{J_k}^k$ et

$\frac{p_{ij}^k}{p_{i\bullet}^k p_{\bullet j}^k} = \frac{p_{ij}^k t_k}{p_{i\bullet}^k t_k p_{\bullet j}^k} = \frac{p_{ij}^k}{p_{i\bullet}^k p_{\bullet j}^k}$. Au poids du tableau près, on a un triplet d'AFC sur modèles ... (Escofier 1984 [6]) qui s'écrit :

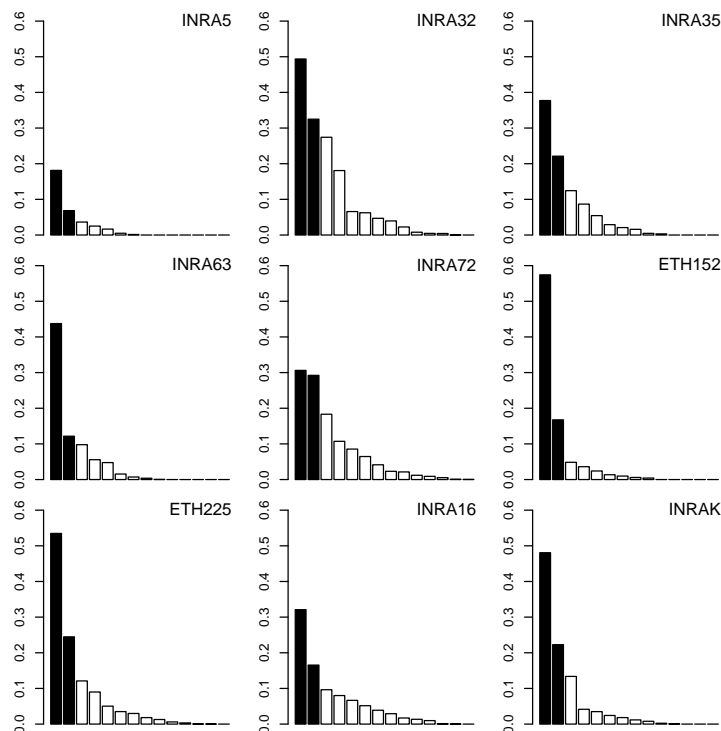
$$T_W^k = \left(\mathbf{Q}_0^k = \left[\frac{p_{ij}^k}{p_{i\bullet}^k p_{\bullet j}^k} - \frac{t_k/i}{t_k} \right], t_k \mathbf{D}_{J_k} \mathbf{D}_I^k \right)$$

Les valeurs propres ne peuvent dépasser t_k . Logiquement, on est conduit à retenir le schéma :

$$U_k = \left(\mathbf{Q}_0^k = \left[\frac{p_{ij}^k}{p_{i\bullet}^k p_{\bullet j}^k} - \frac{t_k/i}{t_k} \right], \mathbf{D}_{J_k} \mathbf{D}_I^k \right)$$

avec la sur-pondération t_k . L'idée est de considérer chaque marqueur comme étant capable de faire une typologie par l'analyse d'un schéma propre et de coordonner ces analyses dans l'une ou l'autre des méthodes spécifiques des K -tableaux. L'ensemble des U_k est un paquet d'analyses des correspondances avec une pondération commune des lignes, une pondération propre des colonnes, un décentrage coordonné. Chacune de ces analyses portera le poids t_k dans l'entreprise de comparaison. L'AFC intra-classe est alors une analyse factorielle multiple. Ce développement permet de passer dans la sphère du K -tableaux et d'abord de comparer directement les analyses par marqueurs :

```
w <- dudi.coa(data.frame(t(microsatt$stab)), scann = F)
wit <- within(w, loci.fac, scann = F)
microsatt.ktab <- ktab.within(wit)
plot(sepan(microsatt.ktab))
```



On accède directement à la forte hétérogénéité de la valeur typologique des marqueurs complètement cachée dans l'analyse globale. On y reviendra dans les méthodes K -tableaux.

Références

- [1] J.P. Benzecri. Analyse de l'inertie intraclasse par l'analyse d'un tableau de correspondance. *Les Cahiers d'Analyse des Données*, VIII(3) :351–358, 1983.
- [2] J.P. Benzecri. Sur la généralisation du tableau de burt et son analyse par bandes. *Les Cahiers de l'Analyse des Données*, VII :33–43, 1982.
- [3] J. Blondel. *Biogéographie évolutive*. Masson, Paris, 1985.
- [4] J.M. Bouroche. *Analyse des données ternaires. La double analyse en composantes principales*. PhD thesis, Paris VI, 1975.
- [5] P. Cazes, D. Chessel, and S. Doledec. L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie. *Revue de Statistique Appliquée*, 36(1) :39–54, 1988.
- [6] B. Escofier. Analyse factorielle en référence à un modèle. application à l'analyse des tableaux d'échanges. *Revue de Statistique Appliquée*, 32(4) :25–36, 1984.
- [7] J. Estève. Les méthodes d'ordination : éléments pour une discussion. In J.M. Legay and R. Tomassone, editors, *Biométrie et Ecologie*, pages 223–250. Société Française de Biométrie, Paris, 1978.
- [8] H.G. Jr Gauch, G.B. Chase, and R.H. Whittaker. Ordination of vegetation samples by gaussian species distributions. *Ecology*, 55 :1382–1390, 1974.
- [9] M.O. Hill. Correspondence analysis : a neglected multivariate method. *Applied Statistics - Journal of the Royal Statistical Society Series C*, 23(3) :340–354, 1974.
- [10] H.O. Hirschfeld. A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences*, 31 :520–524, 1935.
- [11] D.G. Kendall and A. Stuart. *The advanced theory of statistics. Vol 2 : Inference and relationships. Cha. 33 : Categorized data*. Griffin, London, 1961.
- [12] J.D. Lebreton, D. Chessel, R. Prodon, and N. G. Yoccoz. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. i. variables de milieu quantitatives. *Acta Oecologica, (Ecologia Generalis)*, 9 :53–67, 1988.
- [13] Ph. Lebreton, Ph. Lebrun, J.P. Martinot, A. Miquet, and H. Tournier. Approche écologique de l'avifaune de la vanoise. *Travaux scientifiques du Parc national de la Vanoise*, 21 :7–304, 1999.
- [14] A. Leclerc. L'analyse des correspondances sur juxtaposition de tableaux de contingence. *Revue de Statistique Appliquée*, 23(3) :5–16, 1975.
- [15] A. Leclerc. Une étude de la relation entre une variable qualitative et un groupe de variables qualitatives. *International Statistical Review*, 44 :241–248, 1976.

- [16] K. Moazami-Goudarzi, D.M.A. Belemsaga, G. Ceriotti, D. Laloë, F. Fagbohoun, N.T. Kouagou, I. Sidibé, V. Codjia, M.C. Crimella, F. Grosclaude, and S.M. Touré. Caractérisation de la race bovine somba à l'aide de marqueurs moléculaires. *Revue d'Élevage et de Médecine Vétérinaire des pays Tropicaux*, 54 :1–10, 2001.
- [17] K. Moazami-Goudarzi, D. Laloë, J.P. Furet, and F. Grosclaude. Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Animal Genetics*, 28 :338–345, 1997.
- [18] J. Oksanen. Problems of joint display of species and site scores in correspondence analysis. *Vegetatio*, 72 :51–57, 1987.
- [19] D. Pontier. *Contribution à la biologie et à la génétique des populations de chats domestiques (Felis catus)*. PhD thesis, 1984.
- [20] R. Prodon and J.D. Lebreton. Breeding avifauna of a mediterranean succession : the holm oak and cork oak series in the eastern pyrénées. 1 : Analysis and modelling of the structure gradient. *Oikos*, 37 :21–38, 1981.
- [21] P. Souvenir Zafindrajaona, V. Zeuh, K. Moazami-Goudarzi, D. Laloë, D. Bourzat, A. Idriss, and F. Grosclaude. Etude du statut phylogénétique du bovin kouri du lac tchad à l'aide de marqueurs moléculaires. *Revue d'Élevage et de Médecine Vétérinaire des pays Tropicaux*, 55 :155–162, 1999.
- [22] F. Tatibouet, D. Chessel, J. Broyer, and J.D. Lebreton. Etude des peuplements d'oiseaux nicheurs de la zone urbaine de lyon. Technical report, Ministère de l'Environnement, 1980.
- [23] C.J.F. Ter Braak. Correspondence analysis of incidence and abundance data : properties in terms of a unimodal response model. *Biometrics*, 41 :859–873, 1985.
- [24] J. Thioulouse and D. Chessel. A method for reciprocal scaling of species tolerance and sample diversity. *Ecology*, 73(2) :670–680, 1992.
- [25] R. S. Thorpe. A biometric study of the effects of growth on the analysis of geographic variation : Tooth number in green geckos (reptilia : Phelsuma). *Journal of Zoology*, 201 :13–26, 1983.
- [26] R. S. Thorpe and L. Leamy. Morphometric studies in inbred house mice (mus sp.) : Multivariate analysis of size and shape. *Journal of Zoology*, 199 :421–432, 1983.
- [27] P. G. M. van der Heijden. *Correspondence analysis of longitudinal categorical data*. DSWO Press, University of Leiden, 1987.
- [28] J. Verneaux. Cours d'eau de franche-comté (massif du jura). recherches écologiques sur le réseau hydrographique du doubs. essai de biotypologie. Thèse d'état, besançon, 1973.