M2 Écologie,Évolution,Biométrie 
UE Description Statistique des Structures Biologiques

Gérer la redondance

D. Chessel

Notes de cours cssb4

Quelques conséquences de la corrélation entre variables. L'absence de redondance définit les bons prédicteurs. Sa présence massive est un parasite à éliminer. Les techniques de réduction de dimension permettent de la gérer. On peut mélanger les types de variables (numériques et facteurs).

Table des matières

1 Introduction

On utilisera les remarquables données de J.-M. Lascaux [4].

```
library(ade4)
data(lascaux)
names(lascaux)
```

```
[1] "riv"    "code"   "sex"    "meris"  "tap"    "gen"    "morpho" "colo"   "ornem"
```

On y trouve une information complexe décrite dans :

`http://pbil.univ-lyon1.fr/R/pps/pps022.pdf`

La morphométrie des truites est approchée par quatre types de variables et offre une diversité de propriétés d'un grand intérêt pédagogique. Préparer les variables quantitatives :

```
dim(lascaux$morpho)
```

```
[1] 306  37
```

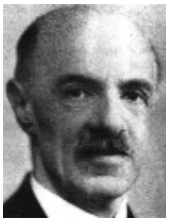
```
apply(lascaux$morpho, 2, function(x) sum(is.na(x)))
```

```
   LS     MD    MAD    MAN   MPEL   MPEC    DAD     DC    DAN   DPEL   DPEC    ADC
    0      0      1      0      0      0      0    127      0    127      0      0
 ADAN  ADPEL  ADPEC PECPEL  PECAN  PECC  PELAN  PELC    ANC   LPRO     DO   LPOO
    0    127      0      0    127    127      0    127      0      0      0      0
 LTET   HTET   LMAX    LAD     LD     HD     LC    LAN    HAN  LPELG  LPECG   HPED
    0    127      1      0      0      2      0      0      0      0      0    127
 ETET
    0
```

```
w <- apply(lascaux$morpho, 2, function(x) sum(is.na(x))) == 0
w["LMAX"] <- TRUE
morpho <- as.data.frame(lascaux$morpho[, w])
morpho[which(is.na(morpho$LMAX)), "LMAX"] <- mean(na.omit(morpho$LMAX))
```

2 Variable cachée et redondance parasite

La notion de variable cachée, ou variable latente, ou facteur (au sens de `factor analysis`) est une des sources de l'analyse des données. Quand un tableau $\mathbf{Y}$ est formé de p variables mesurées sur n individus, on le note : $\mathbf{Y} = \left[\mathbf{y}^1, \ldots, \mathbf{y}^j, \ldots, \mathbf{y}^p\right]$ Une variable latente est une variable inconnue $\mathbf{x}$ qui a la propriété de prédire les variables $\mathbf{y}^j$. C'est évidemment une hypothèse forte : chaque variable mesurée est prédictible par une variable commune.



Charles Spearman [1] a construit cette idée dans la domaine de la psychométrie. Les scores des enfants sur une large gamme de tests apparemment peu liés entre eux sont corrélés et on peut penser qu'ils dépendent d'une même aptitude que Spearman a baptisé facteur g ($g$eneral intelligence factor).

L'analyse factorielle est un monde. On trouve une version gaussienne dans la fonction `factanal` et plusieurs packages offre des fonctions d'analyse factorielle.

[1] Photo : `http://www.york.ac.uk/depts/maths/histstat/people/spearman.gif`

L'ACP donne une solution qui l'emporte par sa simplicité. Si on cherche une variable $\mathbf{x}$ qui prédit les variables $\mathbf{y}^j$, le critère de la qualité de cette prédiction peut être :

$$\sum_{j=1}^{j=p} \left(\text{cor}^2\left(\mathbf{x}, \mathbf{y}^j\right), \right)$$

L'ACP normée donne exactement la solution. Observer que trois fonctions donnent cette solution :

```
w1 <- dudi.pca(morpho, scal = T, scan = F)$l1[, 1]
w2 <- prcomp(morpho, scal = T)$x[, 1]
w3 <- princomp(morpho, cor = T)$scores[, 1]
cor(cbind(w1, w2, w3))
```

```
   w1 w2 w3
w1  1  1 -1
w2  1  1 -1
w3 -1 -1  1
```

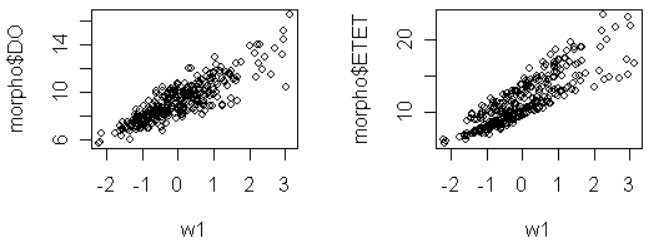
Observer combien chacune des variables est corrélée avec la variable cachée :

```
round(cor(morpho, w1)[, 1], dig = 2)
```

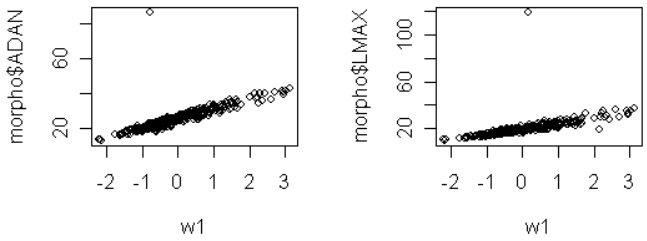
```
   LS     MD    MAN   MPEL   MPEC    DAD    DAN   DPEC    ADC   ADAN  ADPEC PECPEL
 0.99   0.99   0.99   0.98   0.95   0.96   0.98   0.96   0.96   0.80   0.97   0.93
PELAN    ANC   LPRO     DO   LPOO   LTET   LMAX    LAD     LD     LC    LAN    HAN
 0.92   0.97   0.95   0.86   0.97   0.98   0.62   0.89   0.96   0.92   0.95   0.93
LPELG  LPECG   ETET
 0.95   0.93   0.85
```

La plus corrélée est la longueur standard, la variable retenue pour caractériser la taille d'un poisson. `w1` est le facteur taille, la variable cachée est évidemment la taille globale (plus ou moins liée directement à l'âge). Examiner les mauvaises prédictions. Elles s'interprètent facilement.

```
par(mfrow = c(1, 2))
plot(w1, morpho$ADAN)
plot(w1, morpho$LMAX)
plot(w1, morpho$DO)
plot(w1, morpho$ETET)
```




```
par(mfrow = c(1, 2))
plot(w1, morpho$ADAN)
plot(w1, morpho$LMAX)
```



Justifier cette correction sans subtilité :

```
morpho$ADAN[71] <- mean(morpho$ADAN)
morpho$LMAX[155] <- mean(morpho$LMAX)
```

Dans ce tableau, la redondance est écrasante. C'est un parasite. Il faut le faire disparaître pour passer de la taille à la forme (voir [7]). On peut faire une ACP non centrée, qui n'est rien d'autre qu'une approximation de rang 1 de la matrice de départ au moindres carrés [2] :

```
w <- dudi.pca(morpho, cent = F, scal = F, scan = F)
morphomodel <- reconst(w, 1)
morphoresi <- morpho - morphomodel
```

`morpho` contient les données, `morphomodel` un modèle de rang 1 du type :

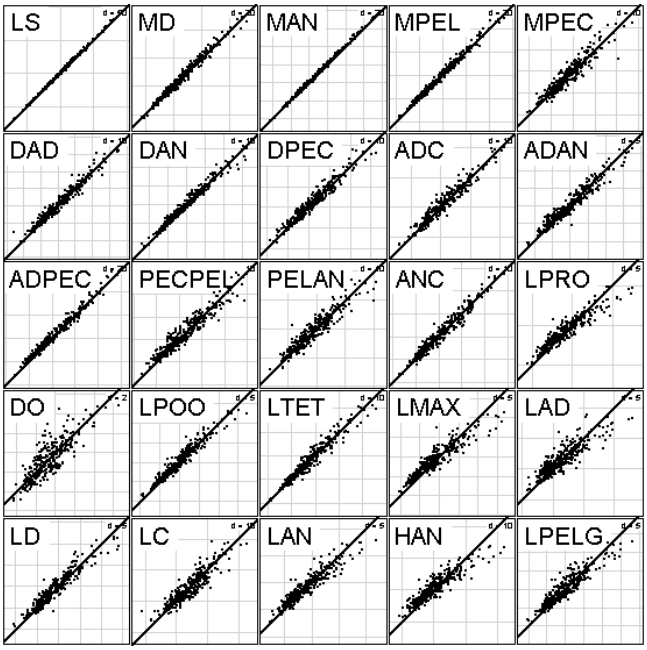
$$x_{ij} = \alpha_i \beta_j$$

et `morphoresi` les résidus autour de ce modèle. Pour une introduction pratique, voir :

`http://pbil.univ-lyon1.fr/R/fichestd/tdr51.pdf`

Pour voir la pertinence de la partie modèle :

```
par(mfrow = c(5, 5))
par(mar = rep(0, 4))
for (k in 1:25) {
    s.label(cbind.data.frame(morpho[, k], morphomodel[, k]), incl = F,
        clab = 0, pch = 20, csub = 3, sub = names(morpho)[k], possub = "topleft")
    abline(0, 1, lwd = 2)
}
```



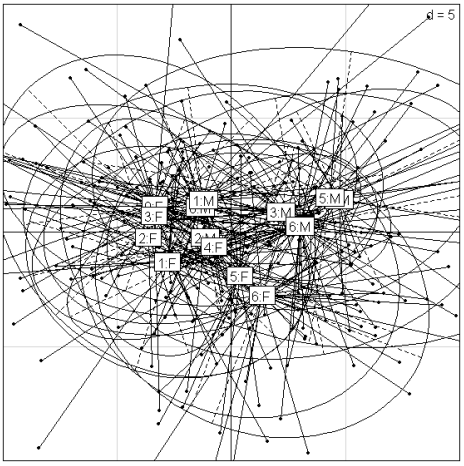
Pour voir que la forme a une composante génétique et une composante sexuelle :

```
resipca <- dudi.pca(morphoresi, scal = F, cent = F, scan = F)
s.class(resipca$li, lascaux$gen:lascaux$sex, xlim = c(-10, 10),
    ylim = c(-10, 10))
```

Faut-il vraiment une p-value ?

```
summary(manova(as.matrix(resipca$li) ~ lascaux$sex * lascaux$gen))
```

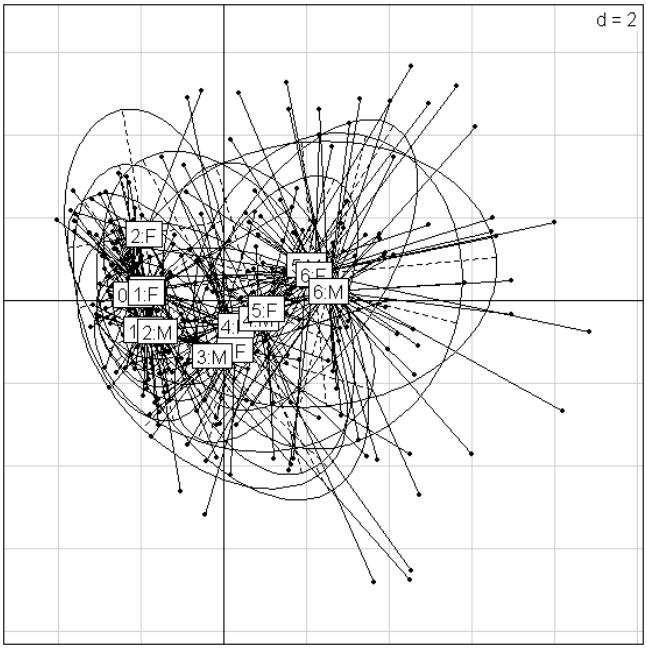
```
                        Df  Pillai approx F num Df den Df    Pr(>F)
lascaux$sex              1  0.1160  19.0860      2    291 1.627e-08 ***
lascaux$gen              6  0.2114   5.7525     12    584 1.940e-09 ***
lascaux$sex:lascaux$gen  6  0.0610   1.5311     12    584    0.1085
Residuals              292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Il est donc clair que l'analyse statistique vise ici à se débarrasser du rôle de la variable latente. C'est derrière cette trivialité que la description de structure commence.

3 Scores et structures de corrélation

En général, les variables latentes ne sont pas si simples et l'ACP les met en évidence. La variable latente est aussi appelée score quand on utilise ses valeurs plutôt que sa fonction (qui est de modéliser). Si une variable latente ne suffit pas, on en utilise plusieurs. Elles sont successivement non corrélées. Les scores sont aussi des coordonnées factorielles dans une vision géométrique.



Ce qui est essentiel en ACP est que le problème précédent, à savoir trouver une variable $\mathbf{x}$ qui maximise

$$\sum_{j=1}^{j=p} \left(\text{cor}^2\left(\mathbf{x}, \mathbf{y}^j\right)\right)$$

renvoie au problème trouver un ensemble de poids des variables (on dit aussi un axe) ω_j qui, sous la contrainte $\sum_{j=1}^{j=p} \omega_j^2 = 1$ (on dit que cet axe est unitaire), maximise la variance de la combinaison linéaire (on dit inertie projetée) :

$$\text{v}(\mathbf{z}) = \text{v}\left(\sum_{j=1}^{j=p} \omega_j \mathbf{y}^j\right)$$

$\mathbf{z}$ est exactement $\mathbf{x}$ à une constante près. Cette constante est la racine de la première valeur propre ou valeur singulière (voir `sdv`) :

$$\mathbf{z} = \sqrt{\lambda_1}\mathbf{x}$$

Dans un objet de la classe `dudi`, **z** est dans la composante `li` et **x** est dans la composante `l1`. Quand une variable latente (on dit aussi composante principale) ne suffit pas, on prend les suivantes qui sont non corrélées. Quand un axe ne suffit pas, on prend les suivants.

Ce qui est particulièrement caractéristique de l'ACP est que deux axes successifs, disons : $\mathbf{u}_1 = (\omega_{11}, \omega_{21}, \ldots, \omega_{p1})$ et $\mathbf{u}_2 = (\omega_{12}, \omega_{22}, \ldots, \omega_{p2})$ sont orthogonaux :

$$\langle \mathbf{u}_1 | \mathbf{u}_2 \rangle = \sum_{j=1}^{j=p} (\omega_{j1}\omega_{j2}) = 0$$

En même temps, les coordonnées sur ces deux axes sont orthogonales (non corrélées si il y a centrage) :

$$\mathbf{z}_1 = \mathbf{Y}\mathbf{u}_1 = \sum_{j=1}^{j=p} \omega_{1j}\mathbf{y}^j$$

$$\mathbf{z}_2 = \mathbf{Y}\mathbf{u}_2 = \sum_{j=1}^{j=p} \omega_{2j}\mathbf{y}^j$$

$$\langle \mathbf{Y}\mathbf{u}_1 | \mathbf{Y}\mathbf{u}_2 \rangle = \sum_{i=1}^{i=n} (\mathbf{z}_{1i}\mathbf{z}_{2i}) = 0$$

Une carte factorielle est la représentation des scores sur deux axes orthogonaux avec des coordonnées non covariantes (figure ci-dessus).

```
colopca <- dudi.pca(lascaux$colo, scan = F)
s.class(colopca$li, lascaux$gen:lascaux$sex)
summary(manova(as.matrix(colopca$li) ~ lascaux$sex * lascaux$gen))
```

```
                        Df  Pillai approx F num Df den Df Pr(>F)
lascaux$sex              1  0.0102   1.4955      2    291 0.2259
lascaux$gen              6  0.5656  19.1920     12    584 <2e-16 ***
lascaux$sex:lascaux$gen  6  0.0456   1.1367     12    584 0.3271
Residuals              292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour une introduction à la MANOVA, voir :

`http://pbil.univ-lyon1.fr/R/tdr6.html`

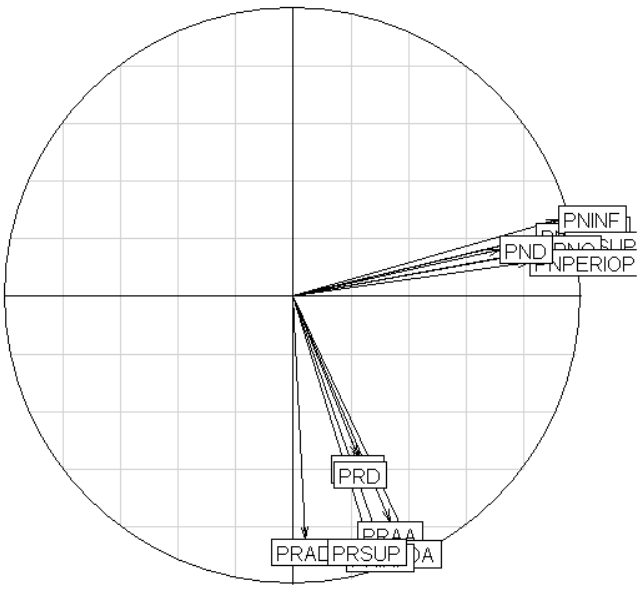
Les scores issus du tableau de coloration n'ont qu'une composante génétique. Mais que veulent dire ces scores ? Pour le savoir, on utilise les corrélations entre les scores et les variables qui valent exactement :

$$\text{cor}\left(\mathbf{z}_1, \mathbf{y}^j\right) = \sqrt{\lambda_1}\omega_{1j}$$

$$\text{cor}\left(\mathbf{z}_2, \mathbf{y}^j\right) = \sqrt{\lambda_2}\omega_{2j}$$

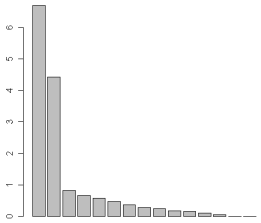
Les critères optimisés prennent pour valeurs optimales les valeurs propres qui jouent donc un rôle descriptif fondamental. Par symétrie avec ce qui se passe pour les individus (lignes), les coordonnées des variables (les corrélations des colonnes avec les scores) sont dans la composante `co` et les axes (les poids des variables ou loadings) sont dans `c1`.

```
s.corcircle(colopca$co)
```



De ceci on déduit qu'il y a un niveau global de coloration rouge, une coloration globale de coloration noire, qu'une seule de ces deux composantes a une source génétique. Le nombre de valeurs propres à utiliser se voit sur le graphe des valeurs propres :

```
barplot(colopca$eig)
```



mais évidemment tous les cas ne sont pas aussi jolis ! L'utilisation des scores à la place des données, appelée réduction de dimension, jouent un grand rôle en écologie comme en d'autres domaines. La plus célèbre colère des écologues contre ce point de vue est celle de Beals [1]. Un relevé écologique est-il un point de $\mathbb{R}^p$? Mais non ! C'est juste utile.

On pourra traîter le problème exposé dans :

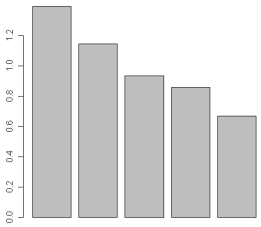
`http://pbil.univ-lyon1.fr/R/pps/pps034.pdf`

4 L'absence de redondance est une bonne propriété

Tout ensemble de variables ne contient pas forcément de réduction pertinente :

```
merispca <- dudi.pca(lascaux$meris, scan = F)
barplot(merispca$eig)
cor(lascaux$meris)
```

```
              rd          ra       rpelg      rpecg        caec
rd     1.00000000  0.31108004  0.02633742 0.07528987 -0.06038263
ra     0.31108004  1.00000000 -0.05677408 0.12549442 -0.11191258
rpelg  0.02633742 -0.05677408  1.00000000 0.04819827  0.10350996
rpecg  0.07528987  0.12549442  0.04819827 1.00000000  0.06108539
caec  -0.06038263 -0.11191258  0.10350996 0.06108539  1.00000000
```



Les variables peu liées sont de bons prédicteurs potentiels. Elles offrent de bonne garantie numérique dans une régression multiple :

```
summary(lm(colopca$l1[, 1] ~ rd + ra + rpelg + rpecg + caec, data = lascaux$meris))
```

```
Call:
lm(formula = colopca$l1[, 1] ~ rd + ra + rpelg + rpecg + caec,
    data = lascaux$meris)
Residuals:
    Min      1Q  Median      3Q     Max
-2.0317 -0.6301 -0.1338  0.5629  2.7634

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.349103   2.844886  -0.826   0.4096
rd           0.226837   0.087149   2.603   0.0097 **
ra           0.126315   0.096050   1.315   0.1895
rpelg        0.118570   0.281706   0.421   0.6741
rpecg       -0.033676   0.073090  -0.461   0.6453
caec        -0.041634   0.004855  -8.575 5.38e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8825 on 300 degrees of freedom
Multiple R-Squared: 0.2365,        Adjusted R-squared: 0.2237
F-statistic: 18.58 on 5 and 300 DF,  p-value: 4.411e-16
```

De même dans une analyse discriminante liée à la MANOVA :

```
summary(manova(as.matrix(lascaux$meris) ~ lascaux$gen * lascaux$sex))
```

```
                       Df Pillai approx F num Df den Df    Pr(>F)
lascaux$gen             6 0.3992   4.2232     30   1460 3.657e-13 ***
lascaux$sex             1 0.0054   0.3105      5    288    0.9065
lascaux$gen:lascaux$sex 6 0.0519   0.5104     30   1460    0.9874
Residuals             292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'abondance de cæca pyloriques est lié à la coloration noire et a comme elle une composante génétique.

On n'utilisera jamais dans un modèle linéaire, comme prédicteurs, un ensemble de variables corrélées. Soit on assure la réduction puis on travaille sur les scores (PCR : faire *GOOGLE* `principal components regression` puis ajouter `CRAN`). Prévoir du temps libre ! Ou passer dans le monde de la régression PLS (*Partial Least Squares Regression*). Pour une vue d'avion de la régression dans R :

`http://cran.r-project.org/doc/contrib/Ricci-refcard-regression.pdf`

Pour une introduction (et plus si affinités) au monde de la PLS voir l'ouvrage de M. Tenenhaus [5].

5 Redondance entre facteurs

C'est le domaine de l'Analyse des Correspondances Multiples (ACM). M. Tenenhaus en a fait le bilan théorique le plus complet dans [6]. Sa découverte est attribuée à R.A. Fisher [3]. Il y a plusieurs approches complémentaires. La plus simple étend l'idée des scores de l'ACP et on peut considérer la méthode comme une ACP normée sur variables qualitatives.

Il suffit de connaître la définition du rapport de corrélation entre une variable qualitative et une variable quantitative. Les détails sont dans :

`http://pbil.univ-lyon1.fr/R/cours/bs5.pdf`

Le pourcentage de la variance de la variable quantitative $\mathbf{x}$ expliquée par le facteur $\mathbf{F}$ est noté $\eta^2_{\mathbf{x}/\mathbf{F}}$. L'ACM est la méthode qui trouve, dans un tableau de v variables qualitatives $\mathbf{F}_j$ un ou plusieurs scores non corrélés qui maximisent successivement :

$$\sum_{j=1}^{j=v} \eta^2_{\mathbf{x}/\mathbf{F}_j}$$

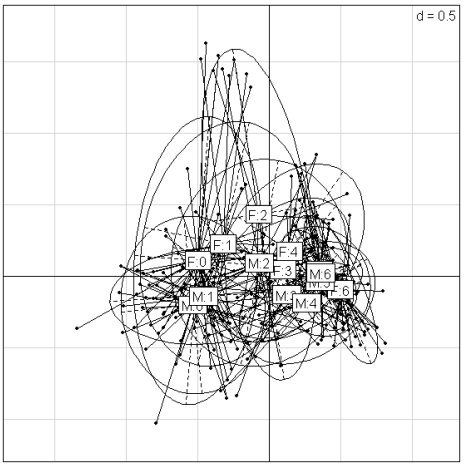
Quelques indications utiles sont dans :

`http://pbil.univ-lyon1.fr/R/fichestd/tdr54.pdf`

Les variables descriptives de la robe des truites sont dans la composante `ornem`.

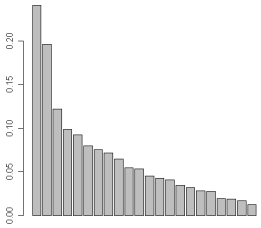
```
ornemacm <- dudi.acm(lascaux$ornem, scan = F)
s.class(ornemacm$li, lascaux$sex:lascaux$gen)
summary(manova(as.matrix(ornemacm$li) ~ lascaux$sex * lascaux$gen))
```

```
                       Df  Pillai approx F num Df den Df    Pr(>F)
lascaux$sex             1  0.0349   5.2603      2    291  0.005699 **
lascaux$gen             6  0.6551  23.7059     12    584 < 2.2e-16 ***
lascaux$sex:lascaux$gen 6  0.0716   1.8058     12    584  0.044063 *
Residuals             292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



La composante génétique est encore bien présente dans le premier score. La réduction à deux dimensions (le défaut) est valide :

```
barplot(ornemacm$eig)
```



Pour interpréter les scores, utiliser la propriété graphiquement après l'édition des rapports de corrélation :

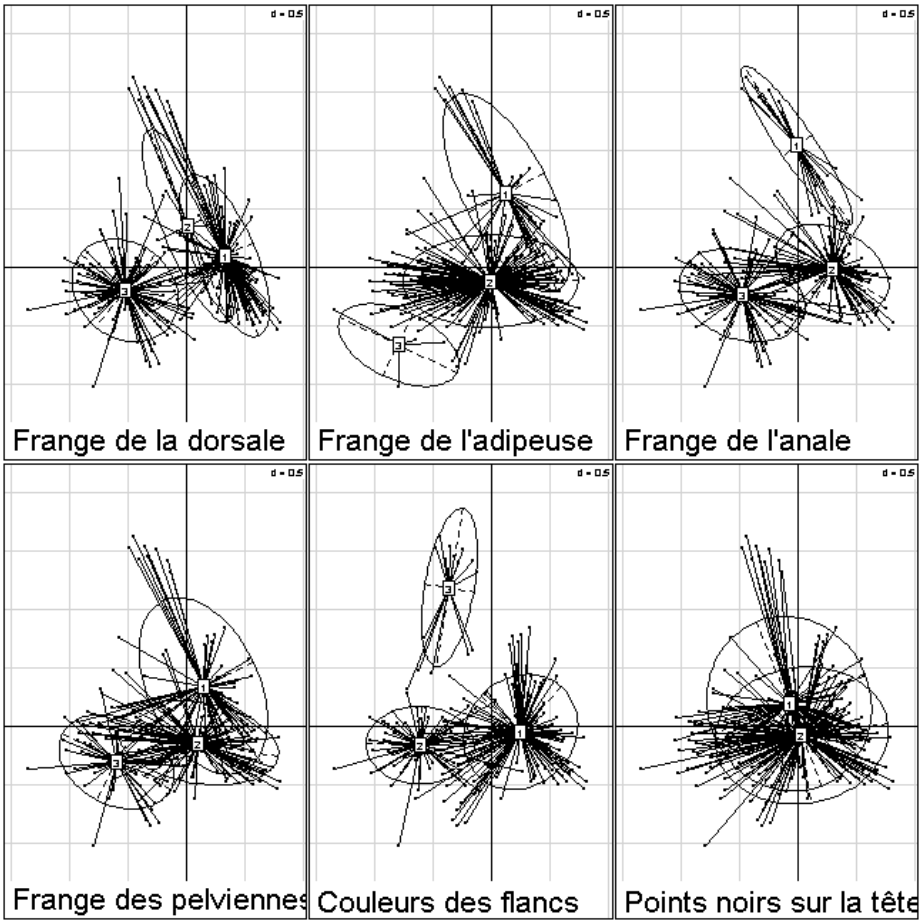
```
round(ornemacm$cr, dig = 2)
```

```
        RS1  RS2
ocpr   0.13 0.25
ocpn   0.16 0.18
maju   0.15 0.01
taop   0.32 0.11
pntet  0.01 0.08
frd    0.68 0.13
ptsad  0.01 0.05
frad   0.05 0.43
fran   0.53 0.59
frpel  0.31 0.34
frc    0.12 0.16
ptsdos 0.04 0.00
coufl  0.59 0.49
coupn  0.42 0.09
zeb    0.10 0.01
```

Le principe de ces graphiques est très accessible :

```
par(mfrow = c(2, 3))
s.class(ornemacm$li, lascaux$ornem$frd, sub = "Frange de la dorsale",
    csub = 3)
```

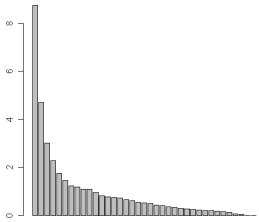
```
s.class(ornemacm$li, lascaux$ornem$frad, sub = "Frange de l'adipeuse",
    csub = 3)
s.class(ornemacm$li, lascaux$ornem$fran, sub = "Frange de l'anale",
    csub = 3)
s.class(ornemacm$li, lascaux$ornem$frpel, sub = "Frange des pelviennes",
    csub = 3)
s.class(ornemacm$li, lascaux$ornem$coufl, sub = "Couleurs des flancs ",
    csub = 3)
s.class(ornemacm$li, lascaux$ornem$pntet, sub = "Points noirs sur la tête",
    csub = 3)
```



6 Les mélanges sont-ils possibles ?

La réponse est oui.

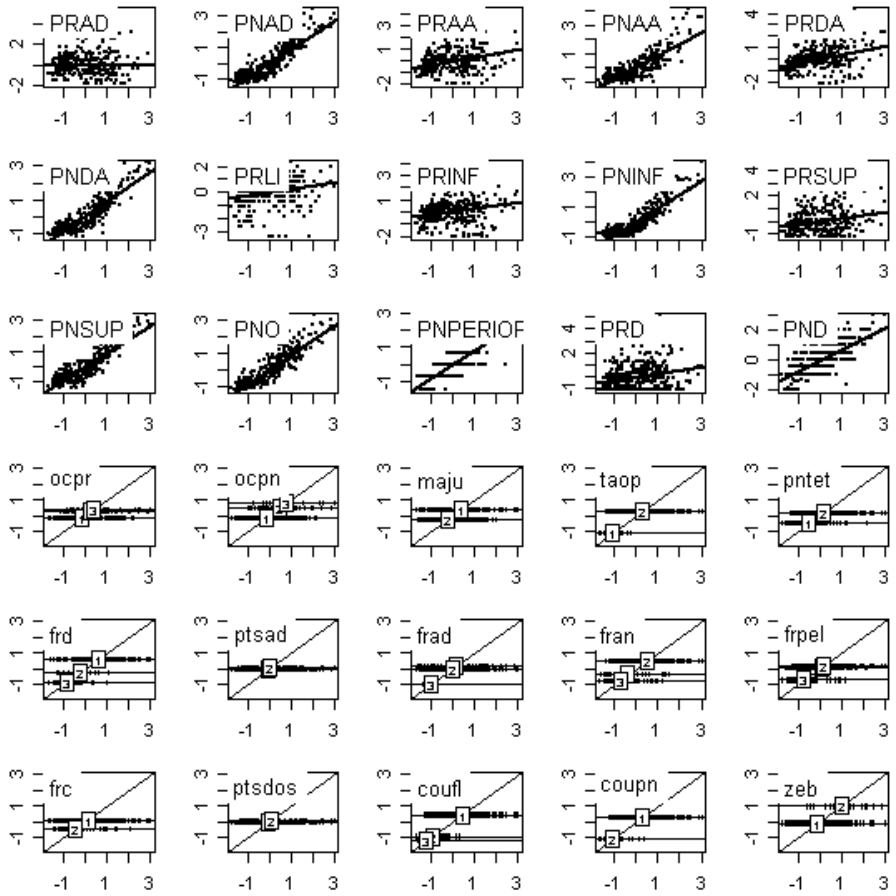
```
X <- cbind.data.frame(lascaux$col, lascaux$ornem)
Xmix <- dudi.mix(X, scan = F)
barplot(Xmix$eig)
```



La composante `cr` correspond au principe d'un score qui maximise :

$$\sum_{j=1}^{j=p} \text{cor}^2\left(\mathbf{x}, \mathbf{y}^j\right) + \sum_{k=1}^{k=v} \eta^2_{\mathbf{x}/\mathbf{F}_k}$$

Elle contient donc maintenant les carrés de corrélation et les rapports de corrélation suivant le type de variable.



On confirme que c'est toujours le premier score de synthèse qui exprime le rôle de la génétique.

```
anova(lm(Xmix$l1[, 1] ~ lascaux$gen * lascaux$sex))
```

```
Analysis of Variance Table
Response: Xmix$l1[, 1]
                       Df  Sum Sq Mean Sq F value Pr(>F)
lascaux$gen             6 193.561  32.260 85.0861 <2e-16 ***
lascaux$sex             1   0.058   0.058  0.1541 0.6950
lascaux$gen:lascaux$sex 6   1.669   0.278  0.7338 0.6228
Residuals             292 110.711   0.379
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

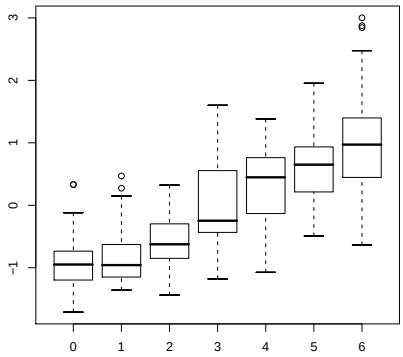
Pour interpréter les scores un par un, on dispose des fonctonction `score` (figure ci-dessus) :

```
score.mix(Xmix)
```

On y lit le mode de relation de chacune des variables en mélange avec le score qui est en abscisse dans chaque fenêtre. On comprend progressivement que l'accumulation des variables, en dépit des moyens de synthèse, butte sur l'objectif : rendre compte de la variabilité cohérente des variables, mais sans indiquer que le but est de retrouver la composante génétique (l'idée était de trouver des marqueurs directement accessibles sur le terrain).

Sans imposer cet objectif on a une bonne indication que l'ensemble des variables mesurées contient des marques de l'opposition méditerranéenne-atlantique, ancestrale-moderne, sauvage-domestique, 6 à 0.

```
plot(lascaux$gen, Xmix$l1[, 1])
```



Si on veut trouver dans cette jungle de variables le meilleur prédicteur on a encore à sa disposition les analyses inter-classes :

```
tot <- cbind.data.frame(morphoresi, lascaux$colo, lascaux$ornem,
    lascaux$meris)
totmix <- dudi.mix(tot, scan = F)
totbet <- between(totmix, lascaux$gen, scan = F)
```

En interprétant ce dernier essai (comme il est tentant de croire qu'avec la totalité on va remporter une victoire massive), on constatera que toute la structure, c'est-à-dire la redondance organisée, de ces données est liée à l'objectif.

La variabilité apparente des truites, tous marqueurs confondus, dans cette région est le résultat des introductions. Mais la variabilité individuelle, pour un type génétique donné, reste forte. Il est de toute évidence possible cependant de mettre une note à chaque individu et de caractériser ainsi avec précision un peuplement sur la base d'un échantillon.

A retenir : pour gérer la redondance `dudi.pca` sur variables numériques, `dudi.acm` sur facteurs, `dudi.mix` pour les mélanges, `dudi.fca` ou `dudi.fpca` pour les variables floues. Toutes ces fonctions ont un fond théorique et des formes de résultats communs.

Références

[1] E.W. Beals. Ordination : mathematical elegance and ecological naïveté. *Journal of Ecology*, 61 :23–35, 1973.

[2] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1 :211–218, 1936.

[3] J.C. Gower. Fisher's optimal scores and multiple correspondence analysis. *Biometrics*, 46 :947–961, 1990.

[4] J.M. Lascaux. *Analyse de la variabilité morphologique de la truite commune (Salmo trutta L.) dans les cours d'eau du bassin pyrénéen méditerranéen.* PhD thesis, 1996.

[5] M. Tenenhaus. *La Régression PLS. Théorie et pratique.* Technip, Paris, 1998.

[6] M. Tenenhaus and F.W. Young. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis ans other methods for quantifying categorical multivariate data. *Psychometrika*, 50 :91–119, 1985.

[7] N. G. Yoccoz. Morphométrie et analyses multidimensionelles. une revue des méthodes séparant taille et forme. In J.D. Lebreton and B. Asselain, editors, *Biométrie et Environnement*, pages 73–99. Masson, Paris, 1993.