

## Analyses discriminantes

### Plan

1.	LES IRIS DE FISHER				
	1.1.	La variable discriminante	4		
	1.2.	Comparaison de deux fonctions	5		
	1.3.	Comparaison des valeurs propres	8		
	1.4.	Tester les valeurs propres	9		
2.	DISC	RIMINATION PREDICTIVE	10		
3.	DISC	RIMINATION DESCRIPTIVE	14		
4.	L'ANA	ALYSE DES CORRESPONDANCES EST UNE ANALYSE DISCRIMINANTE	17		

### Sommaire

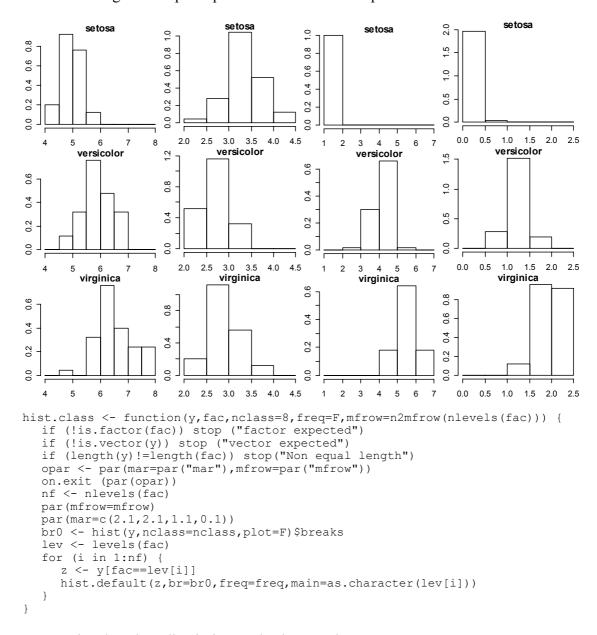
La fiche introduit à l'usage de la MANOVA et de l'analyse discriminante. On compare deux fonctions et leurs usages.

### 1. Les Iris de Fisher

C'est un des jeux de données les plus célèbres de la statistique.

```
> ?iris
> data(iris)
> names(iris)
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
> dim(iris)
[1] 150    5
> plot(iris)
```

Faire les histogrammes par espèce. Faire les ANOVA par variable.



Caractériser la valeur discriminante de chacune des mesures.

```
> apply(iris[,1:4],2,function(x) summary (lm(x~iris[,5])))
```

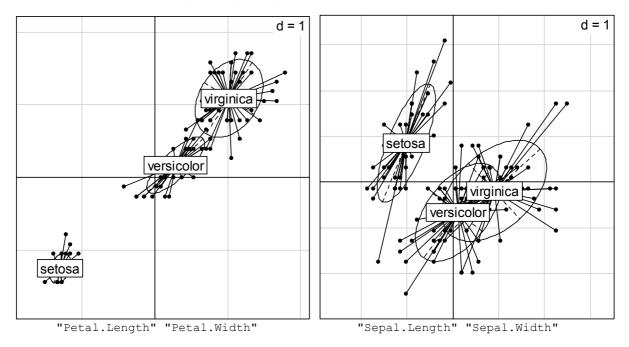
```
$Sepal.Length
Multiple R-Squared: 0.619, Adjusted R-squared: 0.614
$Sepal.Width
Multiple R-Squared: 0.401, Adjusted R-squared: 0.393

$Petal.Length
Multiple R-Squared: 0.941, Adjusted R-squared: 0.941

$Petal.Width
Multiple R-Squared: 0.929, Adjusted R-squared: 0.928
```

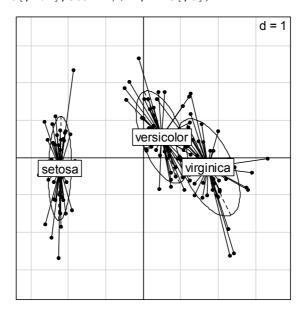
### Représenter les nuages bivariés par populations :

```
> s.class(scale.wt(iris[,3:4]),iris[,5])
> s.class(scale.wt(iris[,1:2]),iris[,5])
```



### La valeur discriminante d'un plan varie fortement dans $\mathbb{R}^4$ .

> s.class(dudi.pca(iris[,1:4],scan=F)\$li,iris[,5])



Proposer une bonne combinaison linéaire de variables permettant de discriminer les trois espèces.

### 1.1. La variable discriminante

Soit **X** le tableau de données (*n* individus et *p* variables) et **q** la variable qualitative qui définit les groupes.  $\mathbf{x}^k$  est la  $k^{\text{jéme}}$  colonne de **X**. **D** est la diagonale des poids des points,  $m_k = \left\langle \mathbf{x}^k \, \middle| \, \mathbf{1}_n \right\rangle_{\mathbf{D}}$  est la moyenne de  $\mathbf{x}^k$ ,  $\mathbf{x}_0^k = \mathbf{x}^k - m_k \, \mathbf{1}_n$  est la variable centrée,  $v_k = \left\| \mathbf{x}_0^k \, \right\|_{\mathbf{D}}^2$  est la variance de  $\mathbf{x}^k$ ,  $c_{jk} = \left\langle \mathbf{x}_0^j \, \middle| \, \mathbf{x}_0^k \, \middle|_{\mathbf{D}}^2$  est la covariance des variables j et k.  $\mathbf{X}_0$  est le tableau des variables centrées et  $\mathbf{C} = \left\lceil c_{jk} \right\rceil = \mathbf{X}_0^t \mathbf{D} \mathbf{X}_0$  est la matrice des covariances de  $\mathbf{X}$ .

La variable  $\mathbf{q}$  a g modalités et définit g groupes.  $\hat{\mathbf{x}}_0^k$  est la projection de  $\mathbf{x}_0^k$  sur le sous-espace engendré par les indicatrices des classes de  $\mathbf{q}$  (la variable centrée où chaque composante est remplacée par la moyenne des composantes de la même classe),  $b_k = \|\hat{\mathbf{x}}_0^k\|_{\mathbf{D}}^2$  est la variance inter-classe de  $\mathbf{x}^k$  (la variance des moyennes par classe pondérées par les poids des classes),  $w_k = \|\mathbf{x}_0^k - \hat{\mathbf{x}}_0^k\|_{\mathbf{D}}^2$  est la variance intra-classe de  $\mathbf{x}^k$  (la moyenne des variances dans chaque classe pondérées par les poids des classes) et

$$v_k = \|\mathbf{x}_0^k\|_{\mathbf{D}}^2 = \|\hat{\mathbf{x}}_0^k\|_{\mathbf{D}}^2 + \|\mathbf{x}_0^k - \hat{\mathbf{x}}_0^k\|_{\mathbf{D}}^2 = b_k + w_k$$

est la décomposition de la variance en variance inter-classe et variance intra-classe (théorème de Pythagore).

Montrer que ce résultat s'étend à deux variables par :

$$c_{jk} = \left\langle \mathbf{x}_0^j \, \middle| \, \mathbf{x}_0^k \, \middle\rangle_{\mathbf{D}} = \left\langle \hat{\mathbf{x}}_0^j \, \middle| \, \hat{\mathbf{x}}_0^k \, \middle\rangle_{\mathbf{D}} + \left\langle \mathbf{x}_0^j - \hat{\mathbf{x}}_0^j \, \middle| \, \mathbf{x}_0^k - \hat{\mathbf{x}}_0^k \, \middle\rangle_{\mathbf{D}} = b_{jk} + w_{jk}$$

 $b_{jk}$  est la covariance inter-classe des variables  $\mathbf{x}^j$  et  $\mathbf{x}^k$  (la covariance des moyennes par classe pondérées par les poids des classes) et  $w_{jk}$  est la covariance intra-classe des variables  $\mathbf{x}^j$  et  $\mathbf{x}^k$  (la moyenne des covariances dans chaque classe pondérées par les poids des classes)

On en déduit que :

$$\mathbf{C} = \begin{bmatrix} c_{jk} \end{bmatrix} = \mathbf{X}_0^t \mathbf{D} \mathbf{X}_0 = \begin{bmatrix} b_{jk} \end{bmatrix} + \begin{bmatrix} w_{jk} \end{bmatrix} = \mathbf{B} + \mathbf{W}$$

L'équation d'analyse de la variance s'étend à une matrice de covariances. On obtient directement ce résultat en utilisant le projecteur sur les indicatrices de classes (**H** est le tableau disjonctif complet associé à **q**):

$$\mathbf{P} = \mathbf{H} \left( \mathbf{H}^t \mathbf{D} \mathbf{H} \right)^{-1} \mathbf{H}^t \mathbf{D}$$

avec  $(\mathbf{Q} = \mathbf{I}_{n} - \mathbf{P})$ :

$$\mathbf{C} = \left[ \mathbf{c}_{jk} \right] = \mathbf{X}_0^t \mathbf{D} \mathbf{X}_0 = \mathbf{X}_0^t \left( \mathbf{P}^t + \mathbf{Q}^t \right) \mathbf{D} \left( \mathbf{P} + \mathbf{Q} \right) \mathbf{X}_0 = \mathbf{B} + \mathbf{W}$$

La variable  $\mathbf{x}^k$  a une valeur discriminante qui se mesure par :

$$\frac{b_k}{v_k} = \frac{b_k}{b_k + w_k} = \frac{1}{1 + \frac{w_k}{b_k}} = \frac{v_k - w_k}{v_k} = 1 - \frac{w_k}{v_k}$$

C'est un jeu d'écriture simplet qui a une grande importance pour comparer les programmes car on peut parler de grande valeur discriminante par *une grande valeur de inter/totale* ou une petite valeur de intra/inter ou *une grande valeur de inter/intra* ou une petite valeur de intra/totale.

Une combinaison linéaire des variables de départ s'écrit  $\mathbf{y} = \mathbf{X}_0 \mathbf{a}$ . Elle est centrée par combinaison linéaire de variables centrées et sa propre valeur discriminante dérive de :

$$v_{\mathbf{v}} = \mathbf{a}^{t} \mathbf{X}_{0}^{t} \mathbf{D} \mathbf{X}_{0} \mathbf{a} = \mathbf{a}^{t} \mathbf{C} \mathbf{a} = \mathbf{a}^{t} \left( \mathbf{B} + \mathbf{W} \right) \mathbf{a} = b_{\mathbf{v}} + w_{\mathbf{v}}$$

La variable discriminante est, si elle existe, celle qui maximisera  $\frac{b_y}{v_y}$  ou, ce qui est le même

problème, celle qui maximisera  $\frac{b_y}{w_y}$  ou, ce qui est le même problème, la variable qui maximisera  $b_y$  sous la contrainte  $v_y = 1$  ou, ce qui est le même problème celle qui maximisera  $b_y$  sous la contrainte  $w_y = 1$ . La solution, si elle existe, est celle d'un des deux problèmes :

$$\mathbf{a}^{t}\mathbf{C}\mathbf{a} = 1$$
  $\mathbf{a}^{t}\mathbf{B}\mathbf{a} = \mathbf{a}^{t}\mathbf{X}_{0}^{t}\mathbf{P}^{t}\mathbf{D}\mathbf{P}\mathbf{X}_{0}\mathbf{a} \ Max$   
 $\mathbf{a}^{t}\mathbf{W}\mathbf{a} = 1$   $\mathbf{a}^{t}\mathbf{B}\mathbf{a} = \mathbf{a}^{t}\mathbf{X}_{0}^{t}\mathbf{P}^{t}\mathbf{D}\mathbf{X}_{0}\mathbf{P}\mathbf{a} \ Max$ 

Cette solution existe comme premier facteur d'un des deux schémas :

Le premier diagonalise  $\mathbf{BC}^{-1}$  et le second diagonalise  $\mathbf{BW}^{-1}$ .

### 1.2. Comparaison de deux fonctions

Utiliser **lda** et consulter la documentation (dans MASS) :

```
> lda1
Call:
lda.data.frame(iris[, 1:4], iris[, 5])
Prior probabilities of groups:
    setosa versicolor virginica
    0.3333    0.3333
Group means:
```

```
Sepal.Length Sepal.Width Petal.Length Petal.Width
                  5.006
                           3.428 1.462 0.246
setosa
                              2.770
versicolor
                  5.936
                                            4.260
                                                        1.326
                  6.588
                              2.974
                                           5.552
virginica
                                                        2.026
Coefficients of linear discriminants:
                 LD1
                         LD2
Sepal.Length -0.8294 0.0241
Sepal.Width -1.5345 2.1645
Petal.Length 2.2012 -0.9319
Petal.Width 2.8105 2.8392
Proportion of trace:
   LD1
         LD2
0.9912 0.0088
```

### Utiliser **discrimin** et consulter la documentation :

```
> discrimin1 <- discrimin(dudi.pca(iris[,1:4],scan=F),iris[,5],scan=F)</pre>
> discrimin1
$eig
[1] 0.9699 0.2220
$nf
[1] 2
$fa
                 DS1
                          DS2
Sepal.Length -0.1200 0.01772
Sepal.Width -0.1169 0.83778
Petal.Length 0.6790 -1.46088
Petal.Width 0.3744 1.92177
discrimin(dudi = dudi.pca(iris[, 1:4], scan = F), fac = iris[,5], scannf = F)
attr(,"class")
[1] "discrimin"
```

Sans en avoir l'air, les deux fonctions sont cohérentes. La première donne une combinaison linéaire de variables de départ avec les coefficients

```
Sepal.Length -0.8294
Sepal.Width -1.5345
Petal.Length 2.2012
Petal.Width 2.8105
```

### Calculer cette combinaison:

```
> w1 <- as.vector(as.matrix(iris[,1:4])%*%lda1$scaling[,1])
> w1
  [1] -5.957 -5.024 -5.385 -4.708 -6.027 -5.597 -5.108 -5.500 -4.455 -5.238
...
[141] 8.758 7.211 7.613 8.901 8.952 7.750 7.285 7.073 7.991 6.788
```

La seconde donne une combinaison linéaire de variables normalisée (en 1/n) avec les coefficients :

```
Sepal.Length -0.1200
Sepal.Width -0.1169
Petal.Length 0.6790
Petal.Width 0.3744
```

### Calculer cette combinaison:

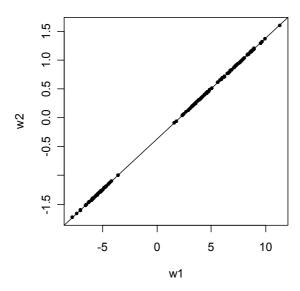
```
> w2 <- as.vector(scale.wt(iris[,1:4])%*%discrimin1$fa[,1])
> w2
```

```
[1] -1.41352 -1.24991 -1.31324 -1.19460 -1.42589 -1.35043 -1.26463 -1.33348 ...

[145] 1.20059 0.98977 0.90816 0.87102 1.03205 0.82112
```

### Comparer:

```
> plot(w1,w2,pch=20)
> abline(lm(w2~w1))
```



La seconde donne une combinaison linéaire de variance totale 1 (en 1/n):

```
> var(w2)*149/150
[1] 1
```

qui maximise la variance inter-classe (la première valeur propre) :

```
> discrimin1
$eig
[1] 0.9699 0.2220
...
> summary(lm(w2~iris[,5]))

Call:
lm(formula = w2 ~ iris[, 5])
...
Multiple R-Squared: 0.97, Adjusted R-squared: 0.969
```

### ou encore:

```
> var(predict(lm(w2~iris[,5])))*149/150 [1] 0.9699
```

La première donne une combinaison linéaire de variance intra-classe 1 :

```
> tapply(w1,iris[,5],var)
    setosa versicolor virginica
    0.7182    1.0736    1.2082
> mean(tapply(w1,iris[,5],var)) # les groupes ont les mêmes effectifs
[1] 1
```

qui maximise la variance inter-classe.

### 1.3. Comparaison des valeurs propres

Le lien entre les valeurs propres de discrimin et les contributions à la trace de 1da est plus caché. Notons que :

$$\mathbf{B}\mathbf{C}^{-1}\mathbf{u} = \lambda\mathbf{u} \Rightarrow \mathbf{C}\mathbf{B}^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{u} = \lambda\mathbf{C}\mathbf{B}^{-1}\mathbf{u} \Rightarrow \mathbf{u} = \lambda(\mathbf{B} + \mathbf{W})\mathbf{B}^{-1}\mathbf{u}$$
$$\Rightarrow \mathbf{u} = \lambda\mathbf{u} + \lambda\mathbf{W}\mathbf{B}^{-1}\mathbf{u} \Rightarrow \frac{\lambda}{1 - \lambda}\mathbf{W}\mathbf{B}^{-1}\mathbf{u} = \mathbf{u} \Rightarrow \mathbf{B}\mathbf{W}^{-1}\mathbf{u} = \frac{\lambda}{1 - \lambda}\mathbf{u}$$

Les deux matrices ont mêmes vecteurs propres. Si les valeurs propres de  $\mathbf{BC}^{-1}$  sont  $\lambda_k$  et les valeurs propres de  $\mathbf{BW}^{-1}$  sont  $\mu_k$ , alors :

$$\mu_k = \frac{\lambda_k}{1 - \lambda_k} \iff \lambda_k = \frac{\mu_k}{1 + \mu_k}$$

Les valeurs propres de discrimin sont :

Les valeurs propres de l'autre diagonalisation sont :

Ce qui donne des contributions à la trace de :

```
> w2/sum(w2)
[1] 0.991213 0.008787
```

exactement ce qu'on a dans :

La procédure discrimin utilise le schéma:

$$\begin{array}{ccc}
 & \mathbf{C}^{-1} \\
\hline
p & \rightarrow & p \\
\mathbf{X}_0^t \mathbf{P}^t \uparrow & & \downarrow \mathbf{P} \mathbf{X}_0 & [1] \\
\hline
n & \leftarrow & n \\
\mathbf{D} & & 
\end{array}$$

La procédure 1 da utilise le schéma :

$$\begin{array}{cccc}
\hline
p & \mathbf{W}^{-1} \\
 & \rightarrow & \boxed{p} \\
\mathbf{X}_0^t \mathbf{P}^t \uparrow & & \downarrow \mathbf{P} \mathbf{X}_0 & [2] \\
\hline
\underline{n} & \leftarrow & \boxed{n} \\
\mathbf{D} & & & \\
\end{array}$$

Les deux procédures donnent à une constante près la même fonction discriminante et les deux procédures ont des valeurs propres liées par une fonction simple. Sur cette base commune, les deux procédures ont des applications différentes.

### 1.4. Tester les valeurs propres

L'hypothèse nulle est que dans chaque groupe on a un échantillon aléatoire simple de la même loi normale multivariée. L'écart entre les moyennes par groupe n'est alors que la conséquence du hasard. C'est l'anova étendue au cas multivarié. Si l'hypothèse nulle est fausse, la valeur discriminante de la variable discriminante, la première valeur propre de l'analyse discriminante est anormalement grande. Les schémas ci-dessus indiquent l'existence d'une seconde variable discriminante, puis d'une troisième, ... qui maximise successivement le même critère sous contrainte d'orthogonalité.

La contrainte a une forte signification. Dans le schéma [1], soit  $\mathbf{a}_1$  le premier facteur : il est  $\mathbf{C}^{-1^{-1}} = \mathbf{C}$  normé ce qui signifie qu'il donne une combinaison de variance 1 :

$$\|\mathbf{a}_1\|_{\mathbf{C}}^2 = \mathbf{a}_1^t \mathbf{C} \mathbf{a}_1 = 1 = \mathbf{a}_1^t \mathbf{X}_0^t \mathbf{D} \mathbf{X}_0 \mathbf{a}_1 = \text{var}(\mathbf{X}_0 \mathbf{a}_1) = 1$$

Le second facteur  $\mathbf{a}_2$ , également  $\mathbf{C}$ -normé donne une autre combinaison de variance 1 et les deux variables discriminantes sont alors non corrélées car :

$$\langle \mathbf{a}_1 | \mathbf{a}_2 \rangle_{\mathbf{C}} = 0 \Rightarrow \mathbf{a}_1^t \mathbf{C} \mathbf{a}_2 = 1 = \mathbf{a}_1^t \mathbf{X}_0^t \mathbf{D} \mathbf{X}_0 \mathbf{a}_2 = \text{cov} (\mathbf{X}_0 \mathbf{a}_1, \mathbf{X}_0 \mathbf{a}_2) = 0$$

La deuxième combinaison est non corrélée avec la première et maximise à nouveau la variance inter-classe, le maximum atteint étant la deuxième valeur propre, etc jusqu'au nombre maximum possible, le plus petit du nombre de variables et du nombre de classes. Les valeurs propres de l'analyse permettent alors de tester la valeur discriminante de une ou toutes les combinaisons ainsi construites. C'est la MANOVA (Multivariate Analysis Of VAriance) qui généralise l'ANOVA. Evidemment, ici la MANOVA est extrêmement significative puisque les anova simples sont très significatives.

Le critère de Pillai est la somme des valeurs propres de l'analyse discriminante du schéma [1] .

```
Residuals 147
---
Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

Le critère de Wilks est le produit des pourcentages de variance intra-classes :

$$W = \prod_{k=1}^{r} (1 - \lambda_k)$$

qu'on teste avec  $-\log(W)$ :

Le critère de Hotelling-Lawley est la somme des valeurs propres de l'analyse discriminante du schéma 2 :

```
> w2 <- w1/(1-w1)
> w2
[1] 32.1919 0.2854
> sum(w2)
[1] 32.48
```

Le critère de Roy est la plus grande valeur propre du schéma 2.

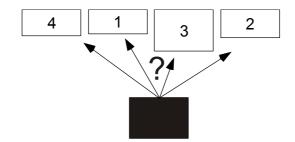
Ces tests s'étendent aux modèles linéaires quelconques comme l'anova. Pour en savoir plus :

Tomassone, R., M. Danzard, J. J. Daudin, and J. P. Masson. 1988. Discrimination et classement. Masson, Paris.

Krzanowski, W. J., and F. H. C. Marriot. 1994. Multivariate analysis. Part 1 Distributions, ordination and inference. Edward Arnold, London.

### 2. Discrimination prédictive

La fonction lda est centrée sur la question de l'affectation d'un individu à une classe. La question intervient dans le schéma de principe :



Le but de l'exercice est de diviser au hasard le tableau de données en deux parties, la première pour chercher une fonction discriminante, la seconde pour déterminer l'espèce à l'aide de cette fonction. On comparera ensuite le résultat obtenu et les vraies valeurs.

```
> echa <- sample(1:150,50)</pre>
> tabref <- iris[echa,1:4] # selection de 50 iris
> espref <- iris[echa,5] # noms d'espèces de la sélection</pre>
> tabsup <- iris[-echa,1:4] # tableau des 100 autres</pre>
> espsup <- iris[-echa,5] # nom de l'espèce des 100 autres
> lda2 <- lda(tabref,espref)</pre>
> 1da2
Call:
lda.data.frame(tabref, espref)
Prior probabilities of groups:
    setosa versicolor virginica
      0.32
                 0.38
                             0.30
Group means:
         Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa 5.000 3.425 1.475 0.2813 versicolor 5.953 2.863 4 274 1 200
                                              4.274 1.3421
5.547 2.0600
virginica
                   6.560
                                3.033
Coefficients of linear discriminants:
                  LD1
Sepal.Length -0.8711 0.2820
Sepal.Width -1.2152 1.6382
Petal.Length 2.7376 -1.3427
Petal.Width 1.8913 3.3693
Proportion of trace:
  LD1 LD2
0.992 0.008
> espestim <- predict(lda2,tabsup) $class # fonction générique utilise predict.lda
> table(espestim, espsup)
          espsup
           setosa versicolor virginica
espestim
             34 0 0
  setosa
                 0
  versicolor
                              31
                                          1
                                         34
  virginica
```

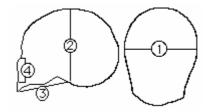
A l'aide de 50 plantes on récupère pratiquement sans erreur le nom des 100 inconnues. Le cas est très favorable. Étudier un cas qui l'est beaucoup moins avec les données 'skulls' proposées par Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215. L'exemple est traité pp. 6, 13, 51, 64, 72, 107, 112 et 117.

Les mesures concernent 5 groupes de 30 crânes égyptiens. Les classes sont :

- 1 période prédynastique ancienne (4000 avant JC)
- 2 période prédynastique récente (3300 avant JC)
- 3 12 et 13 ème dynastie (1850 avant JC)
- 4 période de Ptolémée (200 avant JC)

### 5 - période romaine (150 après JC)

Les variables sont définies par :



#### Les classes sont dans l'attribut date du data frame :

```
> attributes(skulls)
$names
[1] "V1" "V2" "V3" "V4"
$class
[1] "data.frame"
$row.names
 [1] "1"
            "2"
                  "3"
                        "4"
                              "5"
                                    "6"
                                           "7"
                                                 "8"
                                                       "9"
                                                             "10"
                                                                   "11"
                                                                         "12"
 [13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23"
                                                                         "24"
[145] "145" "146" "147" "148" "149" "150"
  [1] A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000}
  [9] A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000} A{-4000}
[145] E{150}
              E{150}
                       E{150}
                                E{150}
                                          E{150}
Levels: A{-4000} B{-3300} C{-1850} D{-200} E{150}
> date <- attr(skulls, "date")</pre>
> summary(date)
A{-4000} B{-3300} C{-1850} D{-200}
                                     E{150}
               3.0
                        30
                                 3.0
      30
```

Représenter les variables par groupes.

Faire les analyses de variance univariées. Les crânes sont-ils différents ?

Faire l'analyse en composantes principales normée du tableau et l'analyse de variance sur les coordonnées.

### Faire la MANOVA du tableau des 4 variables sur les groupes :

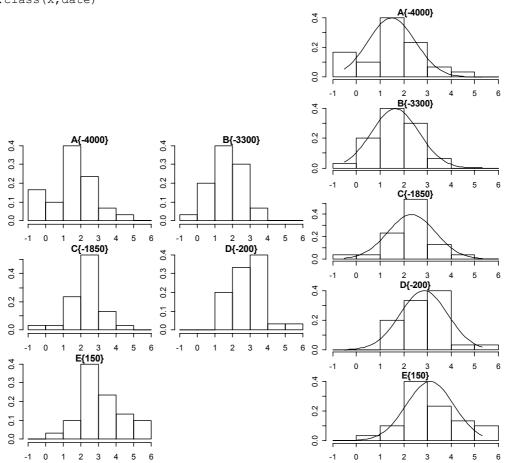
#### Faire l'analyse discriminante et réaffecter les individus à une classe :

```
> lda3 <- lda(skulls,date)
> table(date,predict(lda3,skulls)$class)
```

date	A{-4000}	B{-3300}	C{-1850}	D{-200}	E{150}
A{-4000}	12	8	4	4	2
B{-3300}	10	8	5	4	3
C{-1850}	4	4	15	2	5
D{-200}	3	3	7	5	12
E{150}	2	4	4	9	11

Faire une figure expliquant pourquoi le tableau des réaffectations n'est pas bon.

```
> x <- t(t(skulls)*lda3$scaling[,1])
> x <- apply(x,1,sum)
> hist.class(x,date)
```



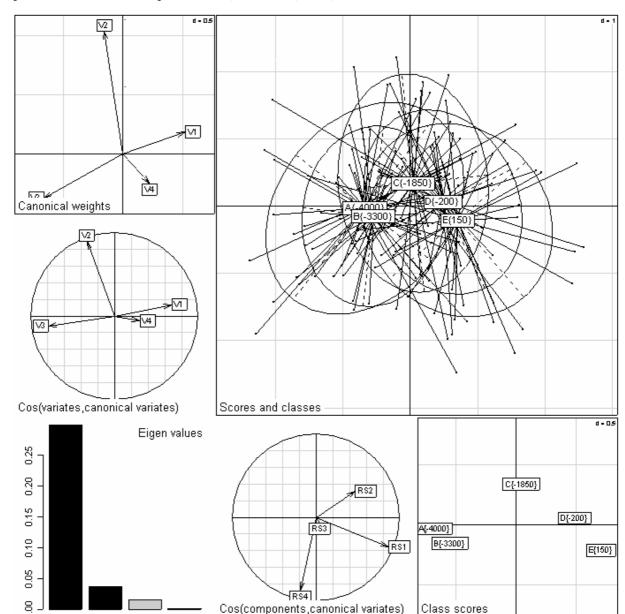
Quelle signification a le vecteur x?

### Commenter ce test:

Modifier la fonction pour obtenir une courbe de Gauss sur chacun des histogrammes. Doit-on prendre une variance commune ?

Une discriminante peut être très significative sans qu'elle serve à réaffecter des individus inconnus. La fonction lda contient bien d'autres possibilités pour cet objectif.

### 3. Discrimination descriptive



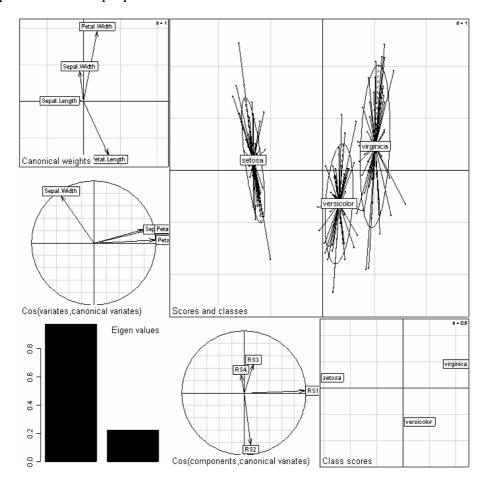
plot(discrimin(dudi.pca(skulls, scann=F), date, scann=F))

Le graphe associé à discrimin contient :

- 1 les poids canoniques (coefficients des combinaisons linéaires de variance unité et de variance inter maximales). Les variables utilisées sont les colonnes normalisées de l'analyse en composantes principales préalable.
- 2 les variables canoniques (combinaisons linéaires de variance unité et de variance inter maximales) et les groupes qui donnent le mode de discrimination opérée.
- 3 les corrélations entre variables canoniques et les variables de départ. Si les graphes 1 et 3 ne sont pas cohérents, c'est l'indice d'une instabilité numérique qui remet en cause l'analyse.
- 4 les corrélations entre les variables canoniques et le composantes principales de l'analyse de départ. On peut ainsi savoir si la discrimination se fait dans la partie interprétable de l'analyse

préliminaire (sinon il faut être méfiant, des variables discriminantes pouvant être non interprétables).

- 6 les moyennes des variables canoniques par classe.
- 5 le graphe des valeurs propres.



Refaire ce graphe et commenter.

Étudier le jeu de données Charolais-Zébus proposé par Tomassone, R., M. Danzard, J. J. Daudin, and J. P. Masson. 1988. Discrimination et classement. Masson, Paris. Tableau 8 p. 45. Les variables sont : poids vif, poids de la carcasse, poids de la viande de première qualité, poids de la viande totale, poids du gras et poids des os en kg pour 23 bovins, 12 Charolais et 11 Zébus.

```
> data(chazeb)
> chazeb
$tab
  vif carcasse viandel viandetot gras
                      79.1 6.0 14.9
73.4 9.7 16.4
            35.1
  395
         224
2
         232
              31.9
 410
22 400
         223
              26.8
                      70.3 13.5 16.2
                      70.4 12.1 17.5
23 400
         213
              25.8
$cla
 [20] zeb zeb zeb zeb
Levels: cha zeb
```

Il y a deux classes. Observer que le schéma 1 se transforme simplement :

Donner la définition des éléments g  $\mathbf{X}_0^g$  et  $\mathbf{D}_g$ . Montrer que  $\mathbf{1}_g$  est cofacteur. Quelle est la valeur propre associée ? En déduire que le nombre de variables discriminantes ne peut dépasser le nombre de groupe -1.

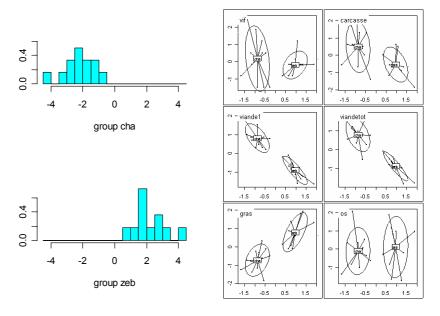
> summary(aov(as.matrix(chazeb\$tab)~chazeb\$cla))

### Qu'obtient-on avec cette commande?

```
> summary(manova(as.matrix(chazeb$tab)~chazeb$cla))
          Df Pillai approx F num Df den Df Pr(>F)
                                        16 1.1e-05 ***
chazeb$cla 1
               0.85
                       14.55
                                  6
Residuals
> summary(manova(as.matrix(chazeb$tab)~chazeb$cla),"Wilks")
         Df Wilks approx F num Df den Df Pr(>F)
                                       16 1.1e-05 ***
chazeb$cla 1 0.15
                      14.55
> summary(manova(as.matrix(chazeb$tab)~chazeb$cla),"Hotelling-Lawley")
           Df Hotelling-Lawley approx F num Df den Df Pr(>F)
chazeb$cla
                          5.46
                                 14.55
Residuals
```

### Pourquoi obtient-on le même résultat ?

```
> plot(lda(chazeb$tab,chazeb$cla))
> plot(discrimin(dudi.pca(chazeb$tab,scan=F),chazeb$cla,scan=F))
```



Pourquoi les fonctions ont-elles donner des résultats inhabituels ? Que représente ces graphes ? Résumer les données par le graphe le plus simple possible.

# 4. L'analyse des correspondances est une analyse discriminante

"Four category judgements of 200 judges in a blind tasting of five different types of claret (red wine from the Bordeaux area in the south western parts of France." in van Rijckevorsel, J. 1987. The application of fuzzy coding and horseshoes in multiple correspondence analysis. DSWO Press, Leiden. (p. 32)

### Vérifier l'identité des deux types d'information :

```
> table(fac.lig,fac.col)
                fac.col
 fac.lig
                 excellent good mediocre boring
                            68
30
 Bordeaux_d_origine
                                    74
                                          22
                        36
 Vin de marque
                                  111
> coa1 <- dudi.coa(bordeaux,scan=F)</pre>
> w1 <- model.matrix(~-1+fac.col)
> w1[sample(1000,5),]
   fac.colexcellent fac.colgood fac.colmediocre fac.colboring
789
                          0
                                        0
                0
                                                     1
22
                           0
469
                          1
                                                     0
511
                Λ
                                         1
                                                     0
```

### Caractériser l'objet w1.

```
> lda(w1,fac.lig)$cscaling
NULL
Warning message:
variables are collinear in: lda.default(w1, fac.lig)
```

#### Commenter.

Expliquer la relation entre les objets coa1\$c1 et discri1\$li.

```
> tapply(discri1$li[,1],fac.lig,mean)
    Cru_Bourgeois Grand_Cru_classé
                                     Vin_de_table Bordeaux_d_origine
         -0.7134
    Vin_de_marque
         0.5489
> tapply(discri1$li[,2],fac.lig,mean)
    Cru_Bourgeois Grand_Cru_classé
-0.06256 -0.26381
                                     Vin de table Bordeaux_d_origine
                                        -0.41671 -0.25424
    Vin de marque
         \overline{0}.48884
> coa1$li
                 Axis1
Axis2
Bordeaux_d_origine -0.1714 0.25424
```

L'AFC est bien une analyse discriminante. C'est surtout une double analyse discriminante. Faire l'autre. Pour approfondir la partie mathématique, considérer le triplet formé d'un tableau d'indicatrices de classe à c lignes et s colonnes, de la métrique canonique et d'une pondération quelconque des c lignes (c pour correspondances) qu'on notera  $(\mathbf{O}_c, \mathbf{I}_s, \Delta_o)$ . Monter que la matrice de covariances est de la forme :

$$\mathbf{V_O} = \mathbf{O}_c^t \Delta_o \mathbf{O}_c \mathbf{I}_s = \mathbf{O}_c^t \Delta_o \mathbf{O}_c = \frac{1}{c} \mathbf{O}_c^t \mathbf{O}_c = \Delta_s - \mathbf{f}_s \mathbf{f}_s^t$$

Cette matrice  $V_O$  n'est pas inversible (ce que signale la fonction lda) mais a un inverse généralisé  $V_O$  qui s'écrit ( $\mathbf{1}_{ss}$  est la matrice à s lignes et s colonnes dont toutes les valeurs égalent l'unité):

$$\mathbf{V_O}^- = \boldsymbol{\Delta}_s^{-1} - \mathbf{1}_{ss}$$

Il suffit de vérifier la définition :

$$\mathbf{V_O}\mathbf{V_O^-}\mathbf{V_O} = \mathbf{V_O} \quad \mathbf{V_O^-}\mathbf{V_O^-}\mathbf{V_O^-} = \mathbf{V_O^-}$$

Les inverses généralisés permettent d'étendre des procédures au cas des matrices de covariances singulières.