

# Discrimin

Discrimin : Among_group_distances.....	2
Discrimin : Anova1-FF.....	4
Discrimin : Assignment_New.....	5
Discrimin : Assignment_Old.....	7
Discrimin : Between analysis/Run.....	8
Discrimin : Between analysis/Test.....	12
Discrimin : Discriminant analysis/Run.....	13
Discrimin : Discriminant analysis/Test.....	18
Discrimin : Initialize/LinkPrep.....	21
Discrimin : Supplementary rows.....	23
Discrimin : Within Analysis/Run.....	25
Discrimin : Within Parameters.....	31

## Discrimin : Among\_group\_distances



Calcul d'une matrice de distance.



L'option calcule la distance entre deux populations au sens de la métrique de Mahalanobis <sup>1</sup> (inverse de la matrice de covariances intra-classes). Voir la définition dans Manly 1994 <sup>2</sup>.



L'option utilise une seule fenêtre de dialogue :



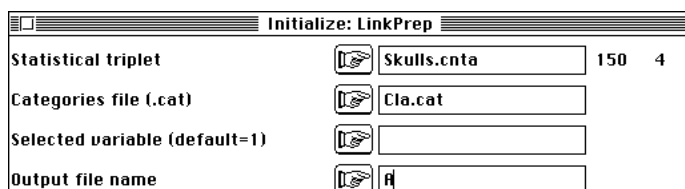
Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



Utiliser la carte Skulls (tableau 1.2 dans <sup>2</sup>) de la pile ADE-4•Data. Créer un fichier binaire Skulls (150-4) et un fichier Cla (150-1) lu par CategVar : Read Categ File. Effectuer l'ACP normée du tableau Skulls (PCA : Correlation matrix PCA) :



Initialiser (Discrimin : Initialize/LinkPrep) :



Utiliser la présente option :



```
W- distance matrix
Groups are defined by column 1 of file Cla
Input statistical triplet: table Skulls.cnta
Number of rows: 150, columns: 4
Output file: A_Maha
Number of rows: 5, columns: 5
Mahalanobis distances between class (W-)
```

Éditer le résultat :

A_Maha					
	1	2	3	4	5
1	0.0000	0.0942	0.9342	1.9460	2.7898
2	0.0942	0.0000	0.7541	1.6490	2.2507
3	0.9342	0.7541	0.0000	0.4584	0.9423
4	1.9460	1.6490	0.4584	0.0000	0.2268
5	2.7898	2.2507	0.9423	0.2268	0.0000

Les résultats sont légèrement différents de ceux de Manly (Tab p. 67). La différence vient comme d'habitude de l'estimation des covariances avec  $n - 1$ . Dans l'ouvrage de référence on trouve pour 2.7898 la valeur 2.697. Comme on a inversé la matrice de covariance, la liaison est simplement  $2.79 = 2.697 * 30 / 29$ .



Comme dans toutes les options, le calcul est étendu à tous les types d'analyses discriminantes. Un inverse généralisé est utilisé.



<sup>1</sup> Mahalanobis, P.C. (1936) On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* : 12, 49-55.

<sup>2</sup> Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215. p. 62-67.

## Discrimin : Anova1-FF



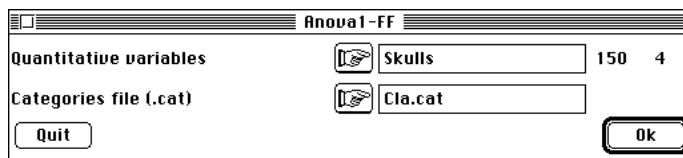
Tests élémentaires (analyse de variance à un facteur contrôlé)





L'option exécute toutes les ANOVA1 de chacune des variables d'un fichier quantitatif sur chacune des variables d'un fichier qualitatif. Édition des résultats classique.



L'option utilise une seule fenêtre de dialogue :



 Fichier de variables quantitatives.

 Fichier de variables qualitatives.



Utiliser la carte Skulls (tableau 1.2 dans <sup>1</sup>) de la pile ADE-4•Data. Créer un fichier binaire Skulls (150-4) et un fichier Cla (150-1) lu par CategVar : Read Categ File.

variable 1 from Skulls versus variable 1 from Cla

Source	SS	d.f.	MS	F	Proba
Between	502.8	4	125.7	5.955	0.0002196
Within	3061	145	21.11		
Total	3564	149			

variable 2 from Skulls versus variable 1 from Cla

Source	SS	d.f.	MS	F	Proba
Between	229.9	4	57.48	2.447	0.04839
Within	3405	145	23.48		
Total	3635	149			

...

variable 4 from Skulls versus variable 1 from Cla

Source	SS	d.f.	MS	F	Proba
Between	61.2	4	15.3	1.507	0.2019
Within	1472	145	10.15		
Total	1533	149			



<sup>1</sup> Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215. Exemple traité pp. 6, 13, 51, 64, 72, 107, 112 et 117.

## Discrimin : Assignment\_New



Utilitaire d'affectation d'individus supplémentaires aux classes d'une analyse discriminante.



Une analyse discriminante définit  $m$  classes. Les lignes d'un tableau d'individus supplémentaires sont affectées à la classe dont le centre de gravité est le plus proche. Voir Manly (1994) <sup>1</sup>. La règle d'affectation est celle de la distance de Mahalanobis (W-) minimum.



L'option utilise une seule fenêtre de dialogue :

Assignment_New	
Data input file	<input type="text"/>
Supplementary row file	<input type="text"/>
Output file name	<input type="text"/>

Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.

Nom du fichier binaire des lignes supplémentaires.

Nom du fichier binaire de sortie (création).



L'option fonctionne pour les options de base cn (PCA : Correlation matrix PCA), cp (PCA : Covariance matrix PCA), nc (PCA : Non centred PCA) et rc (COA : Row % Triplet). Un inverse généralisé est utilisé pour le calcul de la métrique intra.



Utiliser la carte Skulls (tableau 1.2 dans <sup>1</sup>) de la pile ADE-4•Data. Créer un fichier binaire Skulls (150-4) et un fichier Cla (150-1) lu par CatgVar : Read Categ File. Effectuer l'ACP normée du tableau Skulls (PCA : Correlation matrix PCA) :

Correlation matrix PCA	
Matrix input file	<input type="text" value="Skulls"/> 150 4

Initialiser (Discrimin : Initialize/LinkPrep) :

Initialize: LinkPrep	
Statistical triplet	<input type="text" value="Skulls.cnta"/> 150 4
Categories file (.cat)	<input type="text" value="Cla.cat"/>
Selected variable (default=1)	<input type="text"/>
Output file name	<input type="text" value="A"/>

Utiliser la présente option :

Assignment_New	
Data input file	<input type="text" value="A.dis"/>
Supplementary row file	<input type="text" value="Skulls"/> 150 4
Output file name	<input type="text" value="W"/>

```
W- distance matrix
Groups are defined by column 1 of file Cla
Input statistical triplet: table Skulls.cnta
Number of rows: 150, columns: 4
Supplementary row: table Skulls
Number of rows: 150, columns: 4
Output file: W
Number of rows: 150, column: 1
Row = Supplementary row
Column = Assignment group
Rule: Minimum Mahalanobis distances
```

Le numéro de la classe à laquelle est affectée chaque ligne du tableau est dans le fichier W.  
Le lire avec CatgVar : Read Categ File :

Read Categ File	
Input file	W 150 1

Croiser ensuite l'affectation d'origine et l'affectation a posteriori (CategVar : CrossingCateg) :

CrossingCateg	
First ---.cat type file (Table X)	Cl.a.cat
Second ---.cat type file (table Y)	W.cat
Weight file (Default = none)	
Output file (RTDY)	W

Éditer la table de contingence croisée :

WBB					
	1	2	3	4	5
1	12.0000	8.0000	4.0000	4.0000	2.0000
2	10.0000	8.0000	5.0000	4.0000	3.0000
3	4.0000	4.0000	15.0000	2.0000	5.0000
4	3.0000	3.0000	7.0000	5.0000	12.0000
5	2.0000	4.0000	4.0000	9.0000	11.0000

On obtient exactement le tableau Table 8.3 de l'ouvrage de référence (p. 114).



<sup>1</sup>Manly, B.F. (1994) Multivariate Statistical Methods. A primer. Second edition. Chapman & Hall, London. 1-215. p. 114 & 118.

## Discrimin : Assignment\_Old



Calcul du pourcentage de bien classés dans une analyse discriminante.



Chaque ligne d'un tableau d'analyse discriminante est affecté à la classe du centre de gravité le plus proche au sens de la distance de Mahalanobis (métrique intra-classe  $W^{-1}$ ). Le tableau des nombres  $n(i,j)$  des individus de classe d'origine  $i$  affecté de cette manière à la classe  $j$  est édité et le pourcentage de bien classés est donné. Ce critère descriptif permet de savoir si une analyse discriminante est apte à affecter correctement un point dont on ne connaît pas la classe d'origine (Discrimin : Assignment\_New).



L'option utilise une seule fenêtre de dialogue :

Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



Utiliser la carte Skulls (tableau 1.2 dans <sup>1</sup>) de la pile ADE-4•Data. Créer un fichier binaire Skulls (150-4) et un fichier Cla (150-1) lu par CategVar : Read Categ File. Effectuer l'ACP normée du tableau Skulls (PCA : Correlation matrix PCA) :

Initialiser (Discrimin : Initialize/LinkPrep) :

Utiliser la présente option :

```
Groups are defined by column 1 of file Cla
Input statistical triplet: table Skulls.cnta
Number of rows: 150, columns: 4
Output file: A_Allo
Number of rows: 5, columns: 5
Row = Source group
Column = Assignment group
Rule: Minimum Mahalanobis distances
Allocation error N = 99 Frequency (*10000) = 6600
```

Éditer le tableau de sortie (Table 8.3 p. 114) :

A_Allo					
	1	2	3	4	5
1	12.0000	8.0000	4.0000	4.0000	2.0000
2	10.0000	8.0000	5.0000	4.0000	3.0000
3	4.0000	4.0000	15.0000	2.0000	5.0000
4	3.0000	3.0000	7.0000	5.0000	12.0000
5	2.0000	4.0000	4.0000	9.0000	11.0000



<sup>1</sup>Manly, B.F. (1994) Multivariate Statistical Methods. A primer. Second edition. Chapman & Hall, London. 1-215. p. 114.

## Discrimin : Between analysis/Run



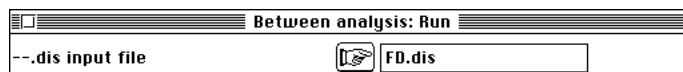
Méthode d'ordination linéaire : analyse interclasses.



Dans un triplet statistique  $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$   $\mathbf{X}$  est le tableau des données (après centrage initial),  $\mathbf{D}_p$  est la matrice des poids associés aux colonnes de  $\mathbf{X}$  et  $\mathbf{D}_n$  est la matrice des poids associés aux lignes de  $\mathbf{X}$ . Une variable qualitative répartit les  $n$  lignes en  $c$  classes.  $\mathbf{G}$  est le tableau des moyennes par classes et par variables ( $c$  lignes et  $p$  colonnes).  $\mathbf{D}_c$  est la matrice des poids associés aux lignes de  $\mathbf{G}$  (le poids d'une classe est la somme des poids des lignes de cette classe). L'analyse exécutée par cette option est celle du triplet  $(\mathbf{G}, \mathbf{D}_p, \mathbf{D}_c)$ . Les lignes de  $\mathbf{X}$  sont projetées en individus supplémentaires et donne des cartes factorielles de variance interclasses maximum. Les versions ACP et AFC sont décrites et illustrées dans <sup>1</sup>. Exemples d'utilisations dans <sup>2</sup>.



L'option utilise une seule fenêtre de dialogue :



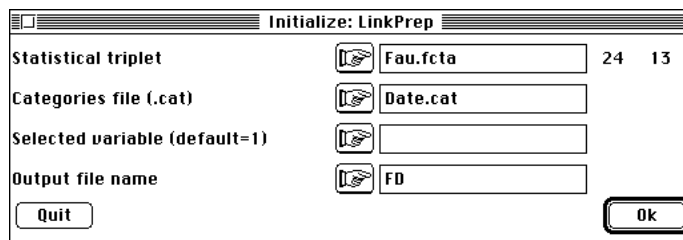
Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



Utiliser la carte Méaudret+2 pour obtenir le fichier Fau (24-13), Date (24-1) et Station (24-1). Lire les deux derniers avec CategVar : Read Categ File. Faire l'AFC du tableau Fau (COA : COrrespondence Analysis) :



Initialiser le lien entre cette analyse et la partition des relevés par dates (Discrimin : Initialize/LinkPrep) :



Exécuter l'analyse (fenêtre de dialogue ci-dessus).

```
-----
Between-class analysis
Groups are defined by column 1 of file Date
Input statistical triplet: table Fau.fcta
Number of rows: 24, columns: 13
total inertia: 7.963e-01
-----
```

```
File FD.beta contains means by groups of rows
It has 4 rows and 13 columns
File FD.bep1 contains the weight of each group
It has 4 rows and 1 column
File FD.bepc contains the weight of each column
It has 13 rows and 1 column
between-class inertia 2.169e-01 (ratio: 2.724e-01)
-----
```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.2981E-01	+0.5986	+0.5986	02	+7.0741E-02	+0.3262	+0.9247
03	+1.6323E-02	+0.0753	+1.0000	04	+0.0000E+00	+0.0000	+1.0000

```
File FD.bevp contains the eigenvalues and relative inertia for each axis
It has 4 rows and 2 columns
```



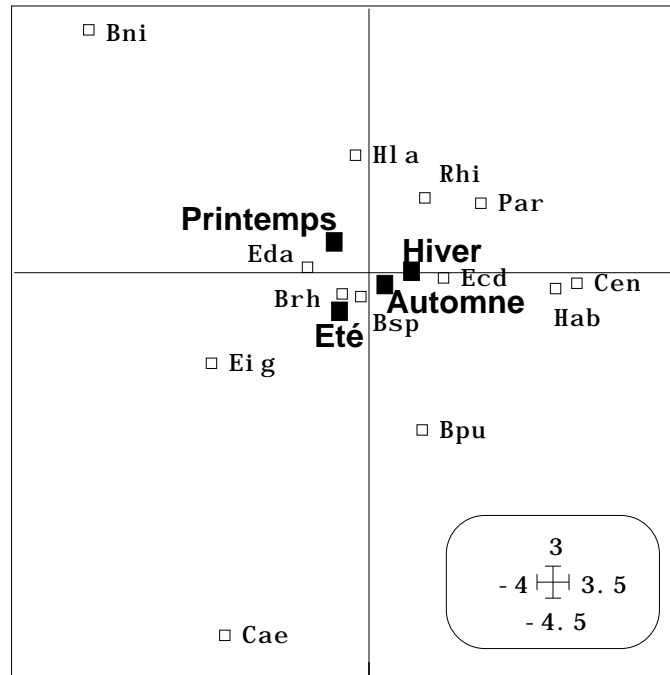
L'interclasse est une analyse d'inertie. Le tableau est celui des centres de gravité (moyennes par variable et par date). Les colonnes (espèces) sont celles de l'AFC initiale. Le poids des colonnes est celui de l'AFC initiale. Les lignes sont les classes (dates). Le poids d'une classe est la somme du poids des lignes (dans l'AFC initiale) de cette classe.

File FD.bec1 contains the column scores (unity norm)  
It has 13 rows and 2 columns

File :FD.bec1

Col.	Mini	Maxi
1	-3.140e+00	2.334e+00
2	-4.068e+00	2.757e+00

Le fichier ---.bec1 contient des scores colonnes de variance unité (Scatters : Labels) :



File FD.beli contains the standard scores of the centers of gravity (lambda norm)

It has 4 rows and 2 columns

File :FD.beli

Col.	Mini	Maxi
1	-4.244e-01	4.661e-01
2	-3.821e-01	4.170e-01

Le fichier ---.beli contient le score des classes obtenu par averaging sur le précédent. On maximise ainsi la variance des positions par dates et non la variance des positions par relevés. On obtient le même résultat en faisant l'AFC du tableau sommé par dates. On peut donc superposer scores de variance unitaire des espèces et scores par averaging des dates (ci-dessus). S'exprime ainsi la faiblesse des valeurs propres :

File FD.bels contains the standard row scores (lambda norm)

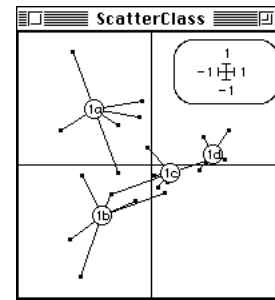
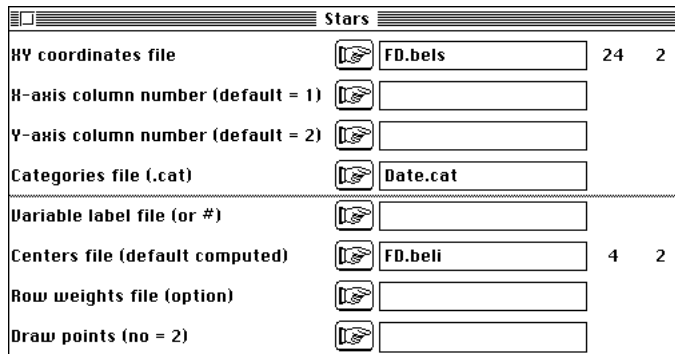
It has 24 rows and 2 columns

File :FD.bels

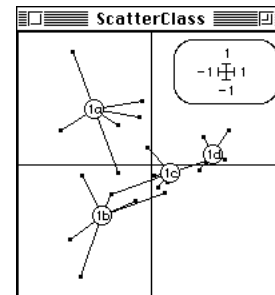
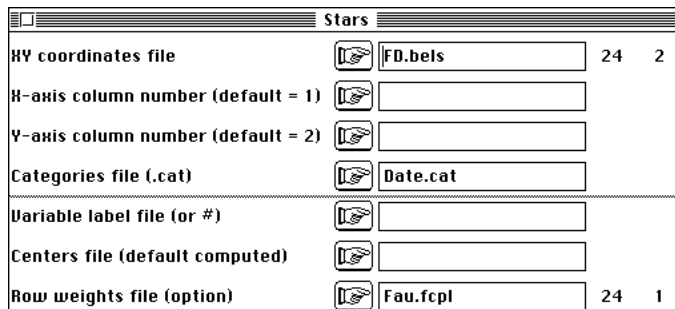
Col.	Mini	Maxi
1	-6.742e-01	5.876e-01
2	-8.458e-01	8.450e-01

Le fichier ---.bels contient le score des relevés obtenu par averaging sur le score des espèces. Les relevés sont ainsi à la moyenne des espèces et les dates à la moyenne des relevés du groupe correspondant.

Le module ScatterClass : Stars permet cette représentation, de deux manières. La première utilise les fichiers calculés :



La seconde refait les calculs avant le graphique :

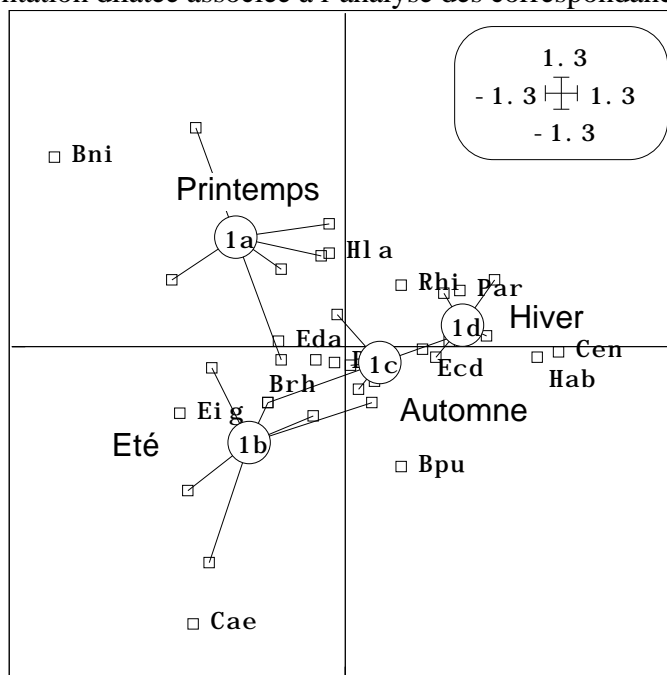


On notera l'utilisation impérative de la pondération AFC initiale pour calculer les centres de gravité.

File FD.beco contains the standard column scores (lambda norm)  
It has 13 rows and 2 columns  
File :FD.beco

Col.	Mini	Maxi
1	-1.131e+00	8.409e-01
2	-1.082e+00	7.332e-01

Le dernier fichier contient les scores de variances les valeurs propres, habituels dans une double représentation dilatée associée à l'analyse des correspondances<sup>3</sup> :





L'analyse interclasse s'emploie après toute analyse simple pour focaliser le point de vue sur les différences entre classes de lignes. Elle est toujours préférable à une analyse discriminante dès qu'on sort du cas classique de l'ACP normée et/ou quand le nombre de descripteurs devient important. Elle est toujours possible quelque soient les nombres de variables, de classes et d'individus par classe.



<sup>1</sup> Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Oecologica, Oecologia Generalis* : 8, 3, 403-426.

Dolédec, S. & Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique II- Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica, Oecologia Generalis* : 10, 3, 207-232.

<sup>2</sup> Beffy, J.L. & Dolédec, S. (1991) Mise en évidence d'une typologie spatiale dans le cas d'un fort effet temporel : un exemple en hydrobiologie. *Bulletin d'Ecologie* : 22, 3-11.

Lair, N. & Sargos, D. (1993) A 10 years study at four sites of the middle course of the River Loire. I - Patterns of change in hydrological, physical and chemical variables in relation to algal biomass. *Hydroécologie Appliquée* : 5, 1-27.

Degiori, F. & Grandmottet, J.P. (1993) Relations entre la topographie aquatique et l'organisation spatiale de l'ichtyofaune lacustre : définition des modalités spatiales d'une stratégie de prélèvements reproductible. *Bulletin Français de la Pêche et de la Pisciculture* : 329, 199-220.

Bornette, G., Amoros, C. & Collilieux, G. (1994) Role of seepage supply on the aquatic vegetation dynamics in former river channels: prediction testing using a hydroelectric construction. *Environmental Management* : 18, 223-234.

Bornette, G. & Heiler, G. (1994) Environmental and biological responses of former channels to river incision: a diachronic study on the upper Rhône river. *Regulated rivers: Research and Management* : 9, 72-92.

Bornette, G., Amoros, C., Castella, C. & Beffy, J.L. (1994) Succession and fluctuation in the aquatic vegetation of two former Rhône River channels. *Vegetatio* : 110, 171-184.

Dessaix, J., Fruget, J.F., Olivier, J.M. & Beffy, J.L. (1995) Changes of the macroinvertebrate communities in the dammed and by-passed sections of the French upper Rhône after regulation. *Regulated rivers: Research and Management* : 10, 265-279.

Franquet, E., Cellot, B., Pont, D. & Bournaud, M. (1995) Environmental and macroinvertebrate dynamics in the Lower Rhône River and a lateral dike field: a study matching two functioning descriptors. *Hydrobiologia* : 308, 207-217.

Bornette, G. & Large, A.R.G. (1995) Groundwater-surface water ecotones at the upstream part of confluences in former river channels. *Hydrobiologia* : 310, 123-137.

Dolédec, S., Dessaix, J. & Tachet, H. (1996) Changes within the Upper Rhône River macrobenthic communities after the completion of three hydroelectric schemes: anthropogenic effects or natural change?. *Archiv für Hydrobiologie* : 136, 19-40.

<sup>3</sup> Oksanen, J. (1987) Problems of joint display of species and site scores in correspondence analysis. *Vegetatio* : 72, 51-57.



## Discrimin : Discriminant analysis/Run



Méthode d'analyse multivariée classique (analyse discriminante linéaire) étendue à tout type de variables.



L'objectif habituel de l'analyse discriminante est de trouver des combinaisons de variables maximisant le rapport de la variance interclasse sur la variance totale, c'est à dire optimisant le pouvoir de discrimination entre groupes d'individus. L'option étend cet objectif à tout type de tableau soumis à une analyse de premier niveau. Pour la bibliographie, noter que celle de Cacoulios & Styán (1973) <sup>1</sup> cite 22 livres et 475 articles. D'après <sup>2</sup> (Section 3.3, Analyse factorielle discriminante, p. 251) celle de McLachlan <sup>3</sup> contient 1200 références



L'option utilise une seule fenêtre de dialogue :

Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



Utiliser la carte Skulls (tableau 1.2 dans <sup>4</sup>) de la pile ADE-4•Data. Créer un fichier binaire Skulls (150-4) et un fichier Cla (150-1) lu par CatgVar : Read Categ File. Effectuer l'ACP normée du tableau Skulls (PCA : Correlation matrix PCA) :

Noter la faiblesse des corrélations entre variables :

```
--- Correlation matrix ---
[ 1] 1000
[ 2] -62 1000
[ 3] -157 264 1000
[ 4] 183 147 -6 1000
-----
```

Initialiser (Discrimin : Initialize/LinkPrep) :

Tester la signification (Discrimin : Discriminant analysis/Test) :

```
number of random permutations: 1000 Observed: 3.533e-01
Histogram: minimum = 2.300e-02, maximum = 3.533e-01
number of simulations X<Obs: 1000 <frequency: 1.000e+00>
number of simulations X>=Obs: 0 <frequency: 0.000e+00>
```

```
***
*****
*****
*****
*****
*****
*****
*****
*****
****
****
```

Exécuter l'option.

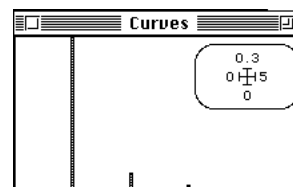
```
Discriminant analysis
Groups are defined by column 1 of file Cla
Input statistical triplet: table Skulls.cnta
Number of rows: 150, columns: 4
total inertia (norm C- generalised inverse) = matrix rank: 4.000e+00
```

L'analyse peut être considérée comme une analyse interclasse (Discrimin : Between analysis/Run) dans laquelle on utilise la norme de Mahalanobis en lieu et place de la métrique canonique ordinaire. L'inertie totale vaut 4 car avec la norme de Mahalanobis l'inertie projetée vaut 1 dans chacune des directions et l'inertie totale est le nombre de variables de départ.

```
between-class inertia (norm C-): 3.533e-01 (ratio: 8.833e-02)
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +2.9829E-01 +0.8443 +0.8443 | 02  +3.7535E-02 +0.1062 +0.9505 |
03  +1.5462E-02 +0.0438 +0.9943 | 04  +2.0163E-03 +0.0057 +1.0000 |
```

Il y a 4 variables et 5 classes. Le programme diagonalise dans la plus petite des deux dimensions et il ne peut y avoir que 4 valeurs propres non nulles. Noter que l'option prendra aussi bien 200 classes de 2 individus que 2 classes de 200 individus, avec le même nombre de variables.

File A.divp contains the eigenvalues and relative inertia for each axis  
It has 4 rows and 2 columns

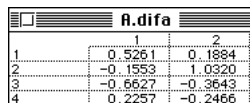


Le schéma utilisé par le programme est du type ACP ( $X$ ,  $C^{-}$ ,  $D$ ) ou  $X$  est le tableau des moyennes par classe et par variable,  $C^{-}$  une inverse généralisé de la matrice de corrélation (obtenu par diagonalisation :  $C$  n'est pas nécessairement inversible),  $D$  la pondération des classes (diagonale des poids des individus). L'inertie totale est la trace de  $BC^{-}$  dans la notation habituelle ( $B$  pour between,  $W$  pour within,  $C$  pour totale,  $C = B + W$ ). Les valeurs propres sont des rapports de corrélation (compris entre 0 et 1). Ils s'agit des pourcentages de variance expliquée optimaux qu'on puisse obtenir avec des combinaisons linéaires des variables de départ (sous contrainte de non corrélation progressive). Ici on sait qu'il y a des différences significatives entre populations. Les deux premières valeurs propres voisines (30% et 4%) sont conservées. Seule la première valeur propre donne lieu à interprétation.

File A.difa contains the coefficient of the discriminant scores  
It has 4 rows and 2 columns

File :A.difa

Col.	Mini	Maxi
1	-6.627e-01	5.261e-01
2	-3.643e-01	1.032e+00



	1	2
1	0.5261	0.1884
2	-0.1553	1.0320
3	-0.6627	-0.3643
4	0.2257	-0.2466

Ce fichier contient les poids canoniques (coefficients des combinaisons linéaires des variables normalisées de départ de variance totale unité et de variance interclasse maximale). Il permet de projeter éventuellement des individus supplémentaires. On peut utiliser aussi les corrélations entre variables canoniques et variables initiales pour repérer le rôle des variables dans la discrimination.

Des incohérences entre les deux représentations invitent à la prudence dans l'interprétation (voir documentation thématique fiche 5.3). La lecture du fichier permet d'écrire que l'équation de la première fonction discriminante est du type :

$$0.526 x_1 - 0.155 x_2 - 0.663 x_3 + 0.226 x_4$$

où  $x_j$  désigne la  $j^{\text{ème}}$  variable normalisée.

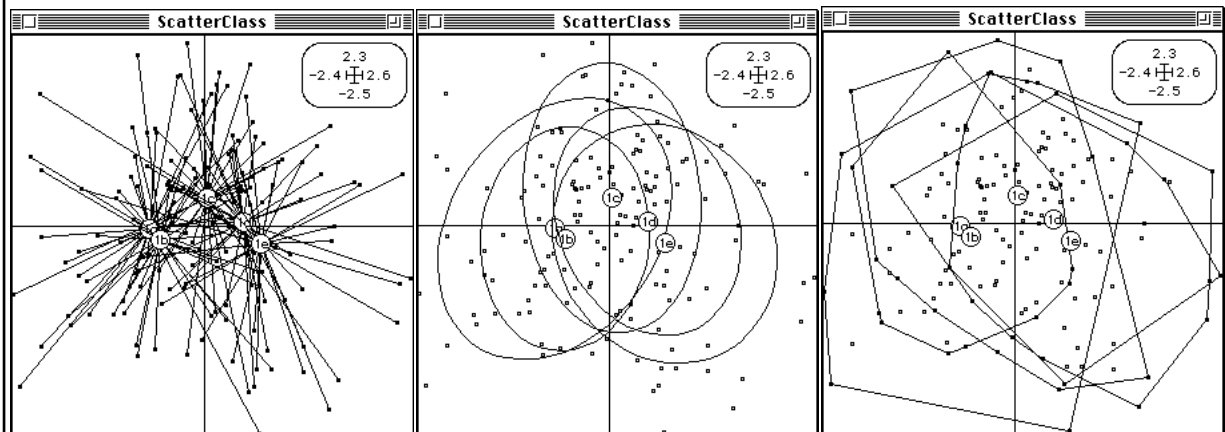
File A.dili contains the canonical scores of row (unit norm)

It has 150 rows and 2 columns

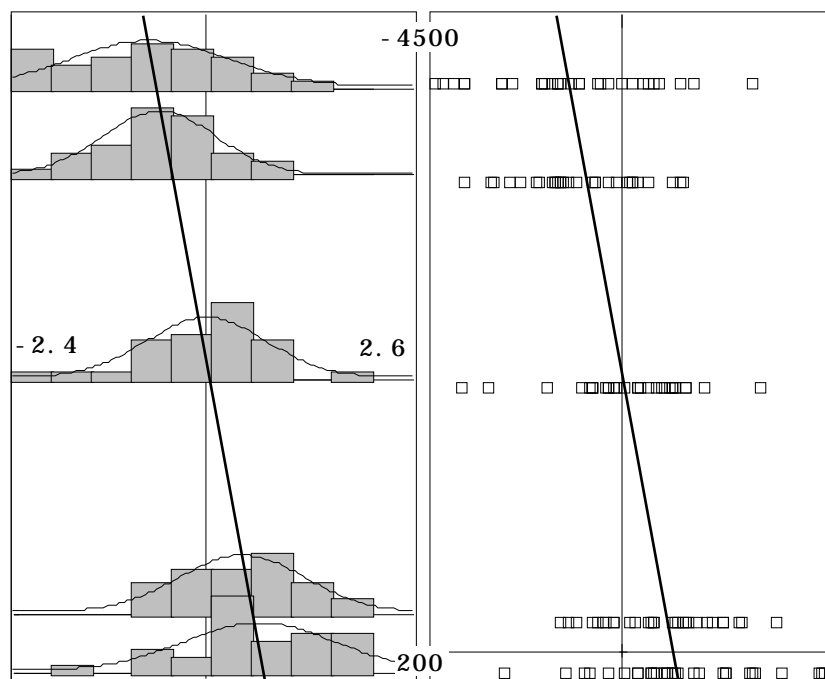
File :A.dili

Col.	Mini	Maxi
1	-2.382e+00	2.579e+00
2	-2.480e+00	2.203e+00

Ce fichier contient les valeurs des variables canoniques (combinaisons linéaires des variables normalisées de départ de variance totale unité et de variance interclasse maximale). Variables canoniques, fonctions discriminantes, scores de discrimination sont des termes équivalents. Ils servent à faire les cartes factorielles de discrimination avec ScatterClass : Stars par étoiles, ellipses ou polygones de contour :



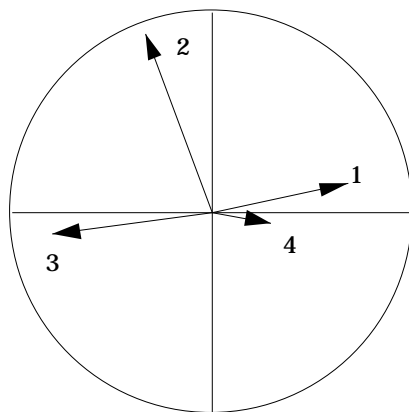
Ici, la représentation correcte est celle de la première variable canonique en fonction de la date des groupes (Graph1D : Histograms, CurveModels : Polynomials) :



File A.diax contains the principal axes  
It has 4 rows and 2 columns  
File :A.diax

Col.	Mini	Maxi
1	-7.878e-01	6.809e-01
2	-1.195e-01	8.879e-01

Ce fichier contient les coefficients de corrélation entre les variables canoniques (combinaisons linéaires des variables normalisées de départ de variance totale unité et de variance interclasse maximale) et les variables de départ .

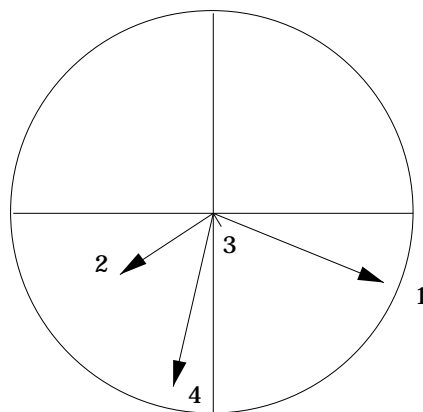


On retrouve le coefficient positif de x1 et le coefficient négatif de x3 dans l'équation de régression sur le score 1 de discrimination. Les variables 2 et 4 ne jouent pratiquement aucun rôle discriminant. On vérifiera par Discrimin : Anova1-FF.

File A.dicp contains the correlations between PCA scores  
and DA scores. It has 4 rows and 2 columns  
File :A.dicp

Col.	Mini	Maxi
1	-4.636e-01	8.645e-01
2	-8.791e-01	-6.908e-02

Ce fichier contient les coefficients de corrélation entre les variables canoniques (combinaisons linéaires des variables normalisées de départ de variance totale unité et de variance interclasse maximale) et les coordonnées des individus dans l'ACP normée de départ. On peut ainsi repérer éventuellement l'occurrence de facteurs (ACP) lointains et peu fiables (ci-dessous, le score 4 d'ACP est voisin du score 2 de la discriminante qui est sans signification) :







La principale difficulté de l'analyse discriminante tient aux équations des scores. Comme pour une équation de régression multiple ou d'ACP sur variables instrumentales, les poids des variables deviennent très rapidement instables numériquement dès que le nombre de mesures est insuffisant. Pour avoir une idée grossière, compter 20 points supplémentaires par variable supplémentaires. 200 lignes dans le tableau de départ supportent, dans les meilleurs cas, 10 colonnes. En deçà, les poids canoniques ne sont pas fiables.

Quand il y a moins de 5 individus supplémentaires par variable supplémentaire, l'analyse est douteuse. Quand il y a plus de variables que d'individus, elle est mathématiquement stupide. Compter une variable par modalité pour les variables qualitatives ou floues. En cas de doute, l'interclasse l'emporte. Après une ACM ou une AFC, l'interclasse est en général plus simple d'interprétation.

Après une ACM, l'option permet une analyse discriminante parce qu'elle utilise des inverses généralisés des matrices de covariances. Cette pratique est rarement utile vu le nombre des modalités. On préfère exécuter la discriminante sur les premières coordonnées factorielles de l'ACM (<sup>2</sup> Analyse factorielle discriminante qualitative, p. 275). Cette discrimination sur composantes est l'équivalent d'une régression sur composantes disponible dans le module OrthoVar. On lui préférera surtout l'analyse discriminante barycentrique <sup>5</sup> qui est exactement l'interclasse après une ACM. Elle est largement utilisée en raison de sa simplicité et de sa robustesse (<sup>2</sup> p. 276).

L'option ne permet pas la réaffectation des individus supplémentaires, car c'est rarement un problème écologique. On notera, par exemple, que dans l'illustration proposée par Manly, la discrimination est fortement significative et que les résultats sont probants, mais qu'il ne peut y avoir d'estimation possible de l'époque à partir des mesures. On trouvera dans la fiche thématique 5.3 des analyses discriminantes après des ACP doublement centrées.



<sup>1</sup> Cacoullos, T. & Styan, G.P.H. (1973) A bibliography of discriminant analysis. In : *Discriminant analysis and applications*. Cacoullos, T. (Ed.) Academic Press, New York, 1-434. 375-434.

<sup>2</sup> Lebart, L., Morineau, A. & Piron, M. (1995) *Statistique exploratoire multidimensionnelle*. Dunod, Paris. 1-439.

<sup>3</sup> McLachlan, G.J. (1992) *Discriminant analysis and Statistical pattern recognition*. J. Wiley, New York. -.

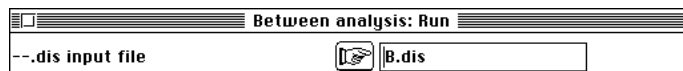
<sup>4</sup> Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215. Exemple traité pp. 6, 13, 51, 64, 72, 107, 112 et 117.

<sup>5</sup> Leclerc, A. (1976) Une étude de la relation entre une variable qualitative et un groupe de variables qualitatives. *International Statistical Review*: 44, 2, 241-248.

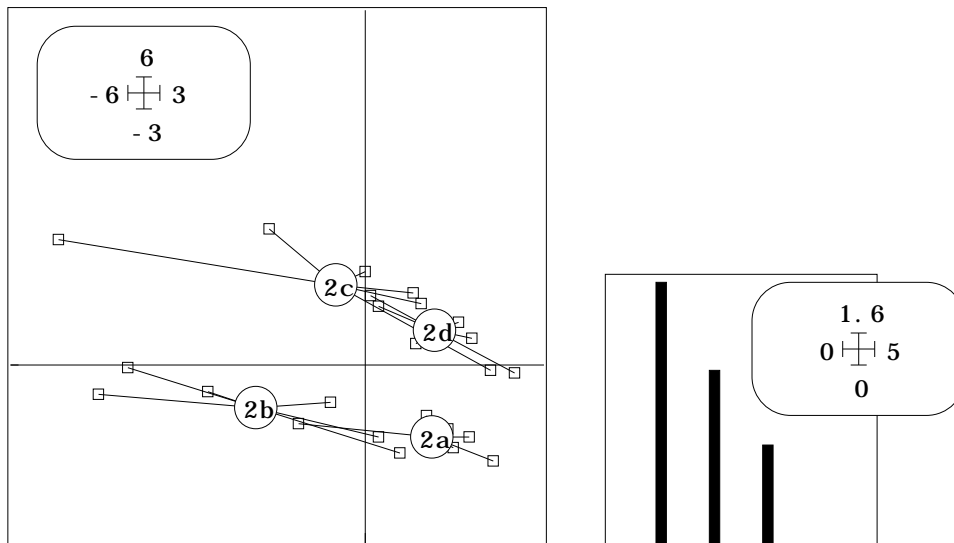




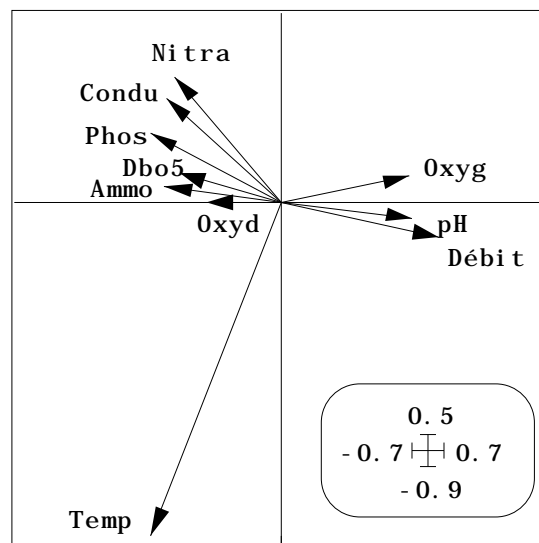
Le test apparaît comme sensiblement moins significatif. Exécuter l'analyse interclasse :



Dépouiller graphiquement (Scatters : Labels sur B.bels et Curves : Eigenvalues sur B.bevp) :



La discrimination entre saisons est sensiblement moins bonne. Par contre les deux axes sont interprétables (Scatters : Labels sur B.bec1) :



On retrouve l'identité de la variable température, mais sur l'axe 1 la limitation des indices de pollution avec les augmentations de débit. Le débat entre niveau de signification inférentielle et possibilité d'interpréter les résultats n'est pas clos. L'exemple montre que discriminante et inter-classe peuvent donner des résultats partiellement concordants et que les niveaux de signification sont des éléments dans un ensemble d'arguments complexes. La décorrélation des variables, vu l'équilibre entre nombre de points et nombre de variables semble ici à éviter.



<sup>1</sup> Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London. 1-281.

## Discrimin : Initialize/LinkPrep



Utilitaire de préparation de l'association d'un tableau et d'une partition.



Le module est consacré aux analyses d'un tableau dont les lignes sont répartis en groupes par une variable qualitative. L'option se contente d'enregistrer les paramètres du couplage dans un fichier texte qui servira d'entrée pour toutes les autres options à l'exception de Discrimin : Anova1-FF.



L'option utilise une seule fenêtre de dialogue :

Initialize: LinkPrep	
Statistical triplet	<input type="text"/>
Categories file (.cat)	<input type="text"/>
Selected variable (default=1)	<input type="text"/>
Output file name	<input type="text"/>
<input type="button" value="Quit"/>	<input type="button" value="Ok"/>

Nom d'un triplet statistique de nature quelconque. Les poids des lignes, les poids des colonnes et le tableau transformé seront utilisés dans les autres options.

Fichier descriptif du tableau de variables qualitatives dont l'une définit les classes de lignes du tableau précédent (après CategVar : Read Categ File ou un utilitaire créant des fichiers de variables qualitatives comme TextToBin : Create1Categ, TextToBin : Create2Categ, TextToBin : LabelToCateg).

Numéro de colonne utilisée dans le fichier précédent. Par défaut c'est la première.

Nom générique des fichiers créés.



Utiliser la carte Diversité+1 de la pile ADE-4•Data. Récupérer le fichier Mil.txt et transformer Mil.txt en binaire par TextToBin : Text->Binary. On y trouve 250 points et 11 variables quantitatives. Récupérer le fichier Mil\_Sta.label et le passer en binaire par TextToBin : LabelToCateg. On obtient un fichier Mil\_Sta avec une variable qualitative de 25 modalités dont les labels sont dans Mil\_Sta.123. Mil\_Sta répartit les 250 points de mesures entre 25 stations.

Faire l'ACP normée de Mil :

Correlation matrix PCA	
Matrix input file	<input type="text" value="Mil"/> 250 11

Associer cette analyse et la partition des relevés entre les stations :

Initialize: LinkPrep	
Statistical triplet	<input type="text" value="Mil.cnta"/> 250 11
Categories file (.cat)	<input type="text" value="Mil_Sta.cat"/>
Selected variable (default=1)	<input type="text"/>
Output file name	<input type="text" value="AM"/>

-----  
New TEXT file AM.dis contains the parameters:  
input file: Mil.cnta  
categorical variable file: Mil\_Sta.cat  
n° of categorical variable used: 1  
-----

Between and Within-class inertia  
Groups are defined by column 1 of file Mil\_Sta  
Input statistical triplet: table Mil.cnta  
total inertia: 1.100e+01

```
Between-class inertia 6.360e+00 (ratio: 5.781e-01)
Within-class inertia 4.640e+00 (ratio: 4.219e-01)
-----
```



L'option ne fait rien d'autre que de fabriquer un fichier texte qui enregistre les paramètres et de calculer les parts de l'inertie inter et intraclasse :

Les fichiers `--.dis` servent de points d'entrée dans toutes les autres options du module à l'exception de Discrimin : Anova1-FF.

## Discrimin : Supplementary rows



L'option calcule les scores canoniques des lignes d'un tableau d'individus supplémentaires **en les transformant conformément au principe de l'analyse de départ**. Cette option est donc réservée aux analyses discriminantes après les analyses de base :

- ACP normée (les lignes supplémentaires sont modifiées comme les lignes de départ : soustraction de la moyenne par variable et division par l'écart-type correspondant) ;
- ACP centrée (les lignes supplémentaires sont modifiées comme les lignes de départ : soustraction de la moyenne par variable) ;
- AFC (les lignes supplémentaires sont considérées comme des distributions de fréquence sur l'ensemble des colonnes).
- ACP non centrée (les lignes supplémentaires ne sont pas modifiées).



L'option utilise une seule fenêtre de dialogue :

Nom de fichier du type ---.dima issu d'une analyse discriminante (Discrimin : Discriminant analysis/Run).

Fichier des lignes supplémentaires. On doit y trouver autant de colonnes que dans l'analyse qui a fourni le fichier précédent. Ces colonnes ont une signification et une forme numérique compatible avec l'analyse de base.

Nom du fichier de sortie. Il contient en lignes les lignes du fichier précédent et en colonnes les facteurs de l'analyse de base.



Utiliser la carte Méaudret pour obtenir le fichier Mil (24-10) et Plan (24-2). Lire ce dernier avec CatgVar : Read Categ File. Faire l'ACP normée du tableau Mil (PCA : Correlation matrix PCA) :

Initialiser le lien entre cette analyse et la partition des relevés par dates (Discrimin : Initialize/LinkPrep) :

Exécuter l'analyse discriminante :

File B.dili contains the canonical scores of row (unity norm)  
It has 24 rows and 3 columns

File :B.dili

Col.	Mini	Maxi
1	-1.484e+00	1.219e+00
2	-1.891e+00	1.586e+00
3	-1.976e+00	1.881e+00

Exécuter l'option :

Supplementary rows	
Initial analysis (.dima)	B.dima
Supplementary row file	Mil 24 10
Output file name	MilSup

Title of the analysis: B.dima  
 Initial triplet : Mil.cnta  
 Rows: 24, Columns: 10  
 Supplementary rows: Mil  
 Rows: 24, Columns: 10

File MilSup contains the coordinates of rows projected as supplementary individuals  
 It has 24 rows and 3 columns

File :MilSup

Col.	Mini	Maxi
1	-1.484e+00	1.219e+00
2	-1.891e+00	1.586e+00
3	-1.976e+00	1.881e+00

Vus les fichiers choisis, on obtient à nouveau les scores canoniques de l'analyse initiale.



## Discrimin : Within Analysis/Run



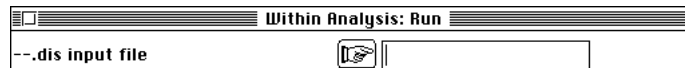
Analyse multivariée avec élimination de l'effet d'une variable qualitative.



L'option exécute l'analyse intraclasse complémentaire de l'analyse interclasse. Cette méthode cherche la structure interne des groupes de lignes tandis que l'interclasse cherche ce qui sépare les groupes de lignes. Les deux objectifs sont complémentaires. L'interclasse est une ACPVI et introduit aux méthodes à projection de variables (Projectors). L'intraclasse introduit aux méthodes à  $K$  tableaux (Partie 5).



L'option utilise une seule fenêtre de dialogue :



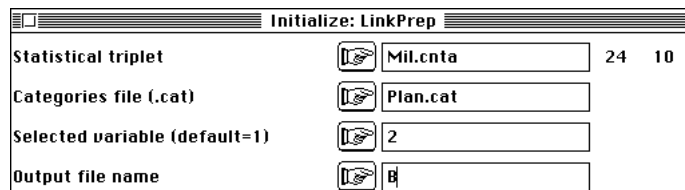
Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



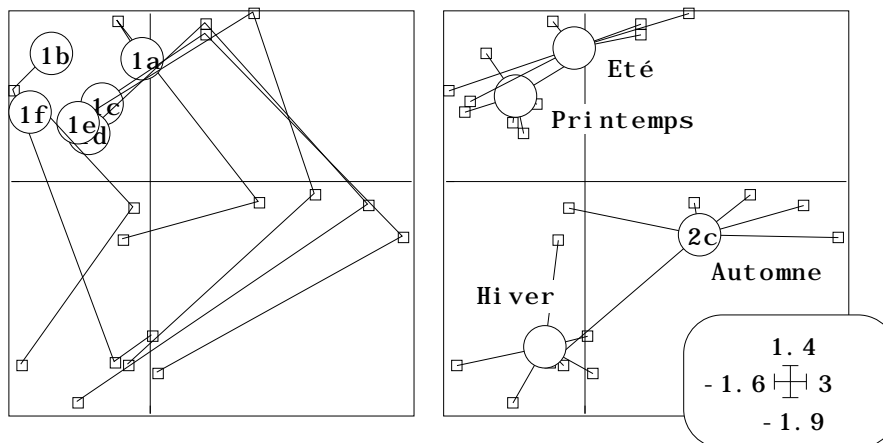
Utiliser la carte Méaudret pour obtenir le fichier Mil (24-10) et Plan (24-2). Lire ce dernier avec CatgVar : Read Categ File. Faire l'ACP normée du tableau Mil (PCA : Correlation matrix PCA) :



Initialiser le lien entre cette analyse et la partition des relevés par dates (Discrimin : Initialize/LinkPrep) :



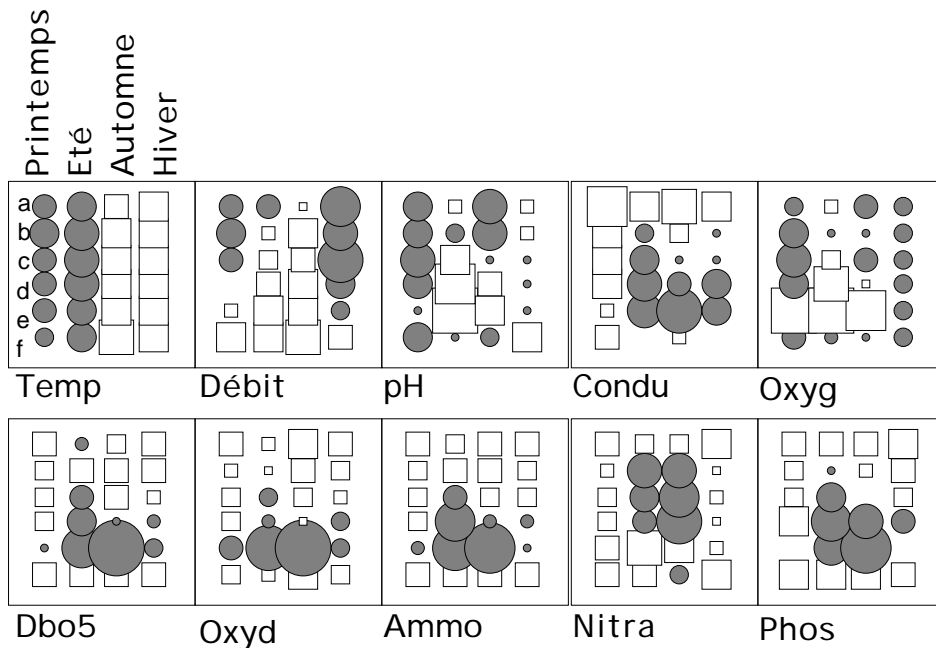
Observer que (ScatterClass : Trajectories et ScatterClass : Stars) que le plan 2-3 se l'ACP normée décrit des variations temporelles :



L'analyse sans a priori décrit simultanément rythme saisonnier et typologie spatiale intervenant dans le processus spatio-temporel généré par le déplacement de l'eau et l'évolution de la température externe. On peut examiner ce processus sous plusieurs angles au sens géométrique du terme c'est-à-dire choisir de privilégier ou d'éliminer une des composantes du plan d'expérience. Représenter les données par Scatters : Values :

Values	
HV coordinates file	Plan 24 2
H-axis column number (default = 1)	2
V-axis column number (default = 2)	1
G values file	Mil.cnta 24 10

Min/Max					
Min. abscissa:	0	<input checked="" type="checkbox"/>	Horiz. graphs:	5	<input checked="" type="checkbox"/>
Max. abscissa:	5	<input checked="" type="checkbox"/>	Vert. graphs:	5	<input checked="" type="checkbox"/>
Min. ordinate:	0	<input checked="" type="checkbox"/>	Nb. grad. H:	1	
Max. ordinate:	7	<input checked="" type="checkbox"/>	Nb. grad. V:	1	
Window height:	350		G factor:	0.8	
Window width:	350				



Exécuter l'analyse intra-classe :

Within Analysis: Run	
--.dis input file	B.dis

Le listing donne les éléments de départ :

```
-----
Within-class analysis
Groups are defined by column 2 of file Plan
Input statistical triplet: table Mil.cnta
Number of rows: 24, columns: 10
total inertia: 1.000e+01
-----
```

L'option prépare un nouveau triplet statistique. Le tableau a autant de lignes et de colonnes que le tableau de départ. Les poids des lignes sont ceux du tableau de départ. Les poids des colonnes sont ceux du tableau de départ. Le tableau est recentré par classes. Son inertie est dite intra-classe. Elle représente ici 68 % de l'inertie totale.

```
File B.whta contains the block-centered array
It has 24 rows and 10 columns
File B.whpc contains the column weight
It has 10 rows and 1 column
File B.whpl contains the row weight
It has 24 rows and 1 column
```

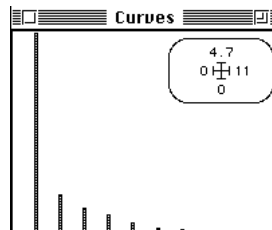
within-class inertia 6.814e+00 (ratio: 6.814e-01)

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+4.6505E+00	+0.6825	+0.6825	02	+8.7006E-01	+0.1277	+0.8102
03	+5.5652E-01	+0.0817	+0.8918	04	+3.9004E-01	+0.0572	+0.9491
05	+2.0546E-01	+0.0302	+0.9792	06	+6.5492E-02	+0.0096	+0.9888
07	+3.1483E-02	+0.0046	+0.9935	08	+2.2419E-02	+0.0033	+0.9968
09	+1.2484E-02	+0.0018	+0.9986	10	+9.6367E-03	+0.0014	+1.0000

Le nouveau triplet donne une analyse d'inertie standard.

File B.whvp contains the eigenvalues and relative inertia for each axis  
It has 10 rows and 2 columns

On consulte ses valeurs propres (Curves : Eigenvalues) :



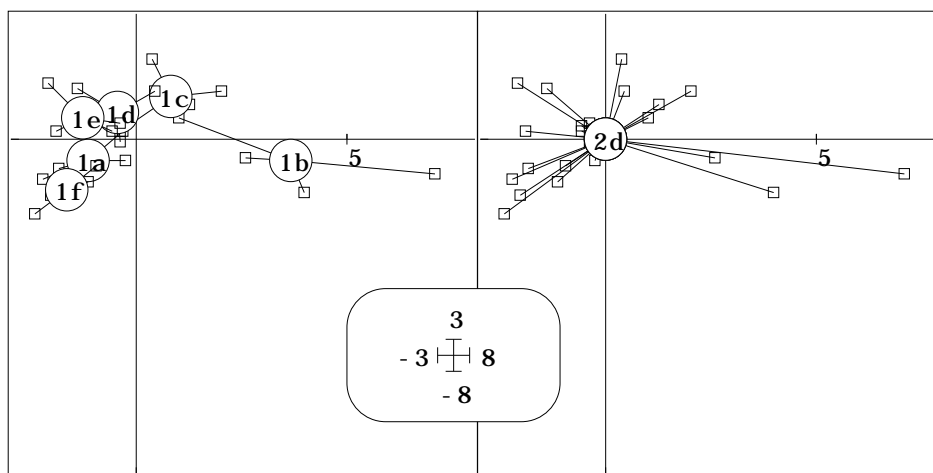
File B.whli contains the standard scores of the rows of the centered table (lambda norm)

It has 24 rows and 2 columns

File :B.whli

Col.	Mini	Maxi
1	-2.410e+00	7.108e+00
2	-1.759e+00	1.872e+00

L'essentiel tient sur un axe. La diagonalisation donne des axes d'inertie intraclasse sur lesquels sont projetés les lignes du nuage recentré (ScatterClass : Stars) :



On vérifie que les nuages par dates ont été ramené à l'origine et on trouve une structure interstation par élimination de l'effet temporel.

File B.whls contains the scores of the rows of the initial table (lambda norm)

It has 24 rows and 2 columns

File :B.whls

Col.	Mini	Maxi
1	-2.175e+00	7.433e+00
2	-1.457e+00	2.745e+00

Les lignes du tableau avant le recentrage par classe ont été projeté en individus supplémentaires sur les axes d'inertie intra-classe :

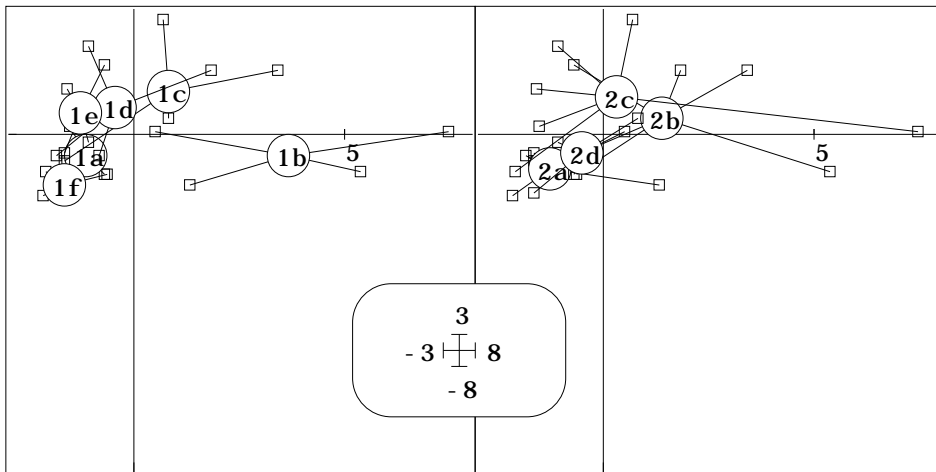
**Stars**

HV coordinates file  24 2

H-axis column number (default = 1)

V-axis column number (default = 2)

Categories file (.cat)



Les fichiers ---.whli (données recentrées) et ---.whls (données d'origine) sont appariées et les points d'une même classe sont déplacés de la même façon :

**Match two scatters**

HV coordinates file

H-axis column number (default = 1)

V-axis column number (default = 2)

Second HV coordinates file

Label file (or #) for rows

Label file (or #) for groups

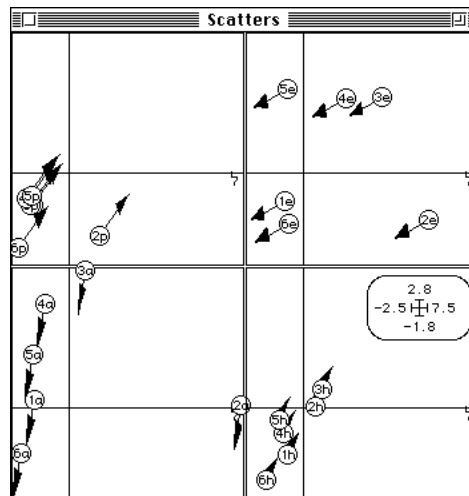
**Row & col. selection**

Col. selection:

Row selection method:  File  Keyboard

Row selection file (.cat):

Selection col. number:



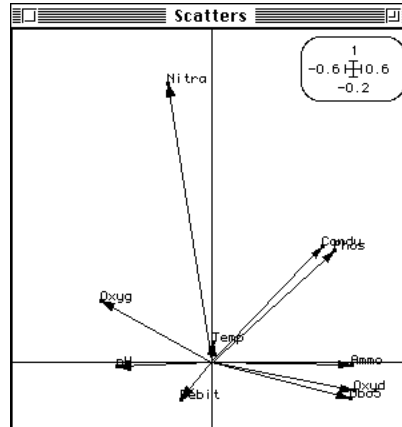
File B.whcl contains the scores of the column (unity norm)  
It has 10 rows and 2 columns

File :B.whc1

Col.	Mini	Maxi
1	-3.296e-01	4.288e-01
2	-1.140e-01	8.359e-01

Ce fichier contient les poids des variables dans les scores des relevés :

Labels	
HV coordinates file	B.whc1
H-axis column number (default = 1)	
V-axis column number (default = 2)	
Label file (or # for item numbers)	Label_Var
Draw vectors from origin (yes = 1)	1



Les variables Débit et Température à variations temporelles sont éliminées de la discussion.

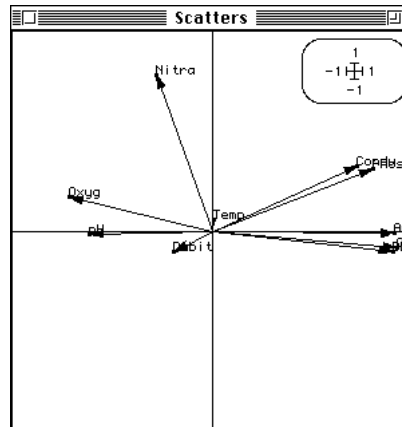
File B.whco contains the standard column scores (lambda norm)  
It has 10 rows and 2 columns

File :B.whco

Col.	Mini	Maxi
1	-7.108e-01	9.248e-01
2	-1.064e-01	7.797e-01

Le fichier contient les covariances entre scores de l'analyse et variables recentrées c'est-à-dire les covariances partielles entre scores et variables sachant l'effet temporel :

Labels	
HV coordinates file	B.whco
H-axis column number (default = 1)	
V-axis column number (default = 2)	
Label file (or # for item numbers)	Label_Var
Draw vectors from origin (yes = 1)	1



De manière générale, ce module s'utilise pour étudier la complémentarité entre l'analyse inter-classe et l'analyse intra-classe définie par une même partition. Les calculs sont les mêmes dans COA : Internal COA (mais la partition peut porter sur les lignes ou les colonnes ou les deux) et les résultats sont voisins dans les analyses PCA : Partial normed PCA et PCA : Within group normalized PCA. Dans ces options on élimine l'effet parasite et on s'en désintéresse totalement alors que dans la présente option la discussion sur l'effet éliminé est encore possible.



Bouroche, J.M. (1975) Analyse des données ternaires: la double analyse en composantes principales. Thèse de 3<sup>o</sup> cycle, Université de Paris VI. 1-57 + annexes.

Benzecri, J.P. (1983) Analyse de l'inertie intra-classe par l'analyse d'un tableau de correspondances. *Les Cahiers de l'Analyse des données* : 8, 3, 351-358.

Carlier, A. . (1985) Applications de l'analyse factorielle des évolutions et de l'analyse intra-périodes. *Statistique et Analyse des Données* : 10, 27-53.

Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Oecologica, Oecologia Generalis* : 8, 3, 403-426.

Dolédec, S. & Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique II- Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica, Oecologia Generalis* : 10, 3, 207-232.

Dolédec, S. & Chessel, D. (1991) Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India* : 1, 133-155.

Des exemples de telles analyses sont disponibles dans les références ci-dessous.



Castella, E., Ohlenforst, H., Neuhaus, K. & Obrdlik, P. (1989) Fluctuations of physico-chemical variables and macroinvertebrate populations within a floodplain of the Upper Rhine. Methodological approach. *Verhandlungen der Internationalen Vereinigung für theoretische und angewandte Limnologie* : 24, 1812-1816.

Crowl, T.A. & Schnell, G.D. (1990) Factors determining population density and size distribution of a freshwater snail in streams: effects of spatial scale. *Oikos* : 59, 359-367.

Castella, E., Richardot-Coulet, M., Roux, C. & Richoux, P. (1991) Aquatic macroinvertebrate assemblages of two contrasting floodplain: the Rhône and Ain rivers, France. *Regulated rivers: Research and Management* : 6, 289-300.

Castella, E., Speight, M.C.D., Obrdlik, P., Schneider, E. & Lavery, T. (1994) A methodological approach to the use of terrestrial invertebrates for the assessment of alluvial wetlands. *Wetlands Ecology and Management* : 3, 17-36.

Lamouroux, N., Pellegrin, F., Nandris, D. & Kohler, F. (1995) The *Coffea arabica* fungal pathosystem in New Caledonia: interactions at two different spatial scales. *Journal of Phytopathology* : 143, 403-413.

Bornette, G. & Large, A.R.G. (1995) Groundwater-surface water ecotones at the upstream part of confluences in former river channels. *Hydrobiologia* : 310, 123-137.

## Discrimin : Within Parameters



Utilitaire de calcul d'indices de diversité.



Calcul de l'inertie par classe comme indice de diversité sur tous les types de variables. L'option est basée sur une observation simple. L'indice de Simpson dans la version 1-D (voir 1) s'écrit :

$$I_S = 1 - D = 1 - \sum_{i=1}^p \frac{2}{i}$$

où  $i$  est l'abondance relative du taxon  $i$  dans le relevé et  $p$  le nombre total de taxons. C'est l'inertie totale de l'ACP centrée du tableau associant à chaque ligne-individu et à chaque colonne-espèce la valeur 1 si l'individu appartient à l'espèce et 0 sinon. C'est la variabilité totale de la variable qualitative "Nom d'espèce". Pour des individus répartis en classe, donc en relevés, la diversité d'un relevé est la variance intra-classe du même tableau. Pour tout type de variables, l'inertie intra-classe caractérise la variabilité totale de la classe et est donc une mesure de diversité interne à la classe.



L'option utilise une seule fenêtre de dialogue :



Nom de fichier de type ---.dis créé par l'option Discrimin : Initialize/LinkPrep.



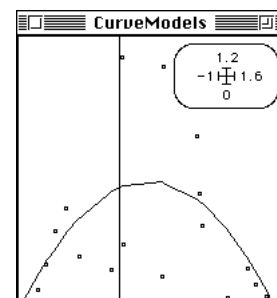
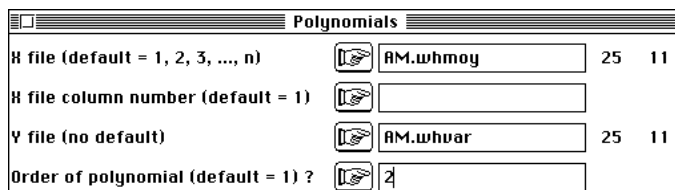
Installer les données conformément à la fiche de l'option Discrimin : Initialize/LinkPrep. Lancer l'option :

```
-----
Within-class analysis
Groups are defined by column 1 of file Mil_Sta
Input statistical triplet: table Mil.cnta
Number of rows: 250, columns: 11
total inertia: 11.000001
-----
```

```
File AM.whmoy contains the array of means
It has 25 rows and 11 columns
File AM.whvar contains the array of variances
It has 25 rows and 11 columns
File AM.whdiv contains the within-class inertia
It has 25 rows and 1 column
```

Le fichier ---.whmoy contient les moyennes par classes (en lignes) et par variables (en colonnes) ; le fichier --.whvar contient les variances par classes (en lignes) et par variables (en colonnes) ; le fichier ---.whdiv contient les combinaison des variances utilisant les poids des variables de l'analyse de départ. Ici c'est simplement la somme des variances ou inertie intra-classe.

On peut étudier la relation moyenne-variance. Par exemple, visualiser la relation moyenne (abscisse) contre variance (ordonnée) pour la variable galets avec CurveModels : Polynomials :



Quand on atteint 100 %, il est clair que la diversité de l'habitat redevient nulle. On a la courbe typique de la valeur indicatrice d'un taxon qui est nulle pour une espèce rare, optimale pour une espèce présente une fois sur deux et à nouveau nulle pour une espèce très fréquente. La diversité est maximum pour les valeurs intermédiaires. Plusieurs variables montre ce type de relations moyenne-variance.



<sup>1</sup> Magurran, A.E. (1988) *Ecological diversity and its measurement*. Croom Helm Limited, London. 1-179.