

COA

COA : COrrrespondence Analysis.....	2
COA : Decentred COA.....	8
COA : Internal COA.....	11
COA : NSCA_Col_Profiles.....	14
COA : NSCA_Row_Profiles.....	18
COA : Reciprocal scaling.....	19
COA : Row weighted COA.....	23
COA : Row_%_Triplet.....	25
COA : Total inertia test.....	28

COA : COrréspondence Analysis



Méthode d'analyse multivariée à un tableau.



L'option exécute l'analyse des correspondances d'un tableau de nombres positifs ou nuls. L'analyse des correspondances ou Analyse factorielle des correspondances est encore connue et utilisée sous les noms de — *method of reciprocal averages* — *reciprocal averaging* — *additive scoring* — *appropriatescoring* — *canonicalscoring* — *Guttman weighting* — *principal component analysis of qualitative data* — *optimal scaling* — *Hayashi's theory of quantification* — *simultaneous linear regression* — *correspondence factor analysis*¹.

\mathbf{T} est la matrice de départ, n_{ij} la valeur du terme de la ligne i et de la colonne j , n la somme totale de tous les éléments du tableau, I le nombre de lignes et J le nombre de colonnes. On note :

$$n_{i.} = \sum_{j=1}^J n_{ij} \quad n_{.j} = \sum_{i=1}^I n_{ij} \quad n = \sum_{i=1}^I n_{i.} = \sum_{j=1}^J n_{.j}$$

$$p_{ij} = \frac{n_{ij}}{n} \quad p_{i.} = \frac{n_{i.}}{n} \quad p_{.j} = \frac{n_{.j}}{n}$$

L'AFC d'un tableau de nombre positifs ou nul \mathbf{T} est l'analyse d'un schéma de dualité avec le jeu de paramètres² :

1 — Transformation initiale : double centrage multiplicatif. Le tableau traité est \mathbf{Z} de terme général :

$$z_{ij} = \frac{p_{ij}}{p_{i.} p_{.j}} - 1 = \frac{p_{ij} - p_{i.} p_{.j}}{p_{i.} p_{.j}}$$

2 — Pondération des lignes : chaque ligne a le poids $p_{i.}$ avec $\sum_{i=1, I} p_{i.} = 1$;

3 — Pondération des colonnes : chaque colonne a le poids $p_{.j}$ avec $\sum_{j=1, J} p_{.j} = 1$.



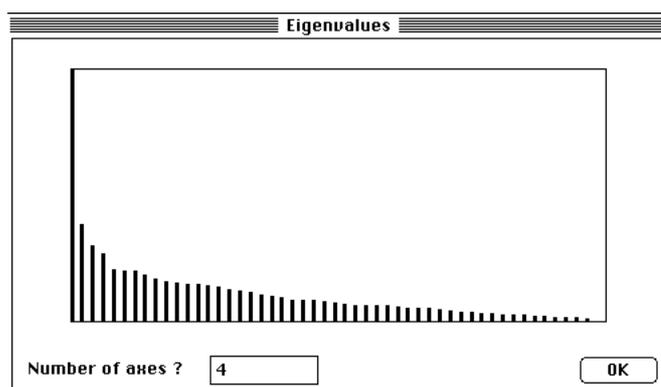
L'option utilise une seule fenêtre de dialogue :



Nom du fichier d'entrée binaire (valeurs positives ou nulles).



Utiliser la carte Avi-Veg de la pile ADE•Data pour obtenir par TextToBin : Char->Binary le fichier AVFau (182-51). Lancer le module :



Sélectionner le nombre de facteurs à conserver après affichage des valeurs propres. L'essentiel est dans le premier facteur.

```
fc/COA: Correspondance analysis
Input file: AVFau
Number of rows: 182, columns: 51
File AVFau.fcpl contains the edge distribution of rows
It has 182 rows and 1 column
File AVFau.fcpc contains the edge distribution of columns
It has 51 rows and 1 column
File AVFau.fcta contains the doubly centred table DI-1*P*DJ-1 -II*1J'
It has 182 rows and 51 columns
File AVFau.fcma contains:
    the number of rows: 182
    the number of columns: 51
    the total number: 1639
```

```
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: AVFau.fcta
--- Number of rows: 182, columns: 51
```

```
-----
Total inertia: 4.46833
-----
```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+7.5325E-01	+0.1686	+0.1686	02	+2.9291E-01	+0.0656	+0.2341
03	+2.2934E-01	+0.0513	+0.2855	04	+2.0467E-01	+0.0458	+0.3313
05	+1.5729E-01	+0.0352	+0.3665	06	+1.5144E-01	+0.0339	+0.4003
07	+1.5077E-01	+0.0337	+0.4341	08	+1.3927E-01	+0.0312	+0.4653
09	+1.2810E-01	+0.0287	+0.4939	10	+1.2161E-01	+0.0272	+0.5211
11	+1.1768E-01	+0.0263	+0.5475	12	+1.1435E-01	+0.0256	+0.5731
13	+1.1113E-01	+0.0249	+0.5979	14	+1.0869E-01	+0.0243	+0.6223
15	+1.0457E-01	+0.0234	+0.6457	16	+9.8787E-02	+0.0221	+0.6678
17	+9.3491E-02	+0.0209	+0.6887	18	+8.9476E-02	+0.0200	+0.7087
19	+8.3038E-02	+0.0186	+0.7273	20	+7.8628E-02	+0.0176	+0.7449
21	+7.1855E-02	+0.0161	+0.7610	22	+6.6172E-02	+0.0148	+0.7758
23	+6.4570E-02	+0.0145	+0.7902	24	+6.3748E-02	+0.0143	+0.8045
...							
39	+2.9257E-02	+0.0065	+0.9549	40	+2.7162E-02	+0.0061	+0.9610

```
File AVFau.fcvp contains the eigenvalues and relative inertia for each
axis
--- It has 51 rows and 2 columns
```

```
File AVFau.fcco contains the column scores
--- It has 51 rows and 4 columns
```

```
File :AVFau.fcco
```

Col.	Mini	Maxi
1	-1.701e+00	9.726e-01
2	-2.165e+00	2.070e+00
3	-1.995e+00	1.351e+00
4	-2.646e+00	1.054e+00

```
File AVFau.fcli contains the row scores
--- It has 182 rows and 4 columns
```

```
File :AVFau.fcli
```

Col.	Mini	Maxi
1	-1.642e+00	9.361e-01
2	-1.680e+00	1.439e+00
3	-1.291e+00	1.002e+00
4	-2.184e+00	1.160e+00



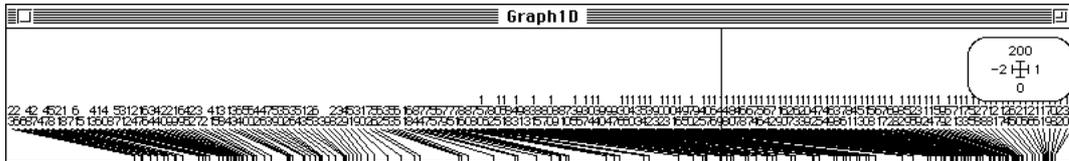
La procédure donne le même résultat sur un tableau et son transposé. Les calculs sont optimisés en fonction du mode de présentation du tableau.

Les cartes factorielles sont centrées pour les pondérations marginales. La manière la plus simple de comprendre comment fonctionne l'analyse est de représenter l'abondance entre colonnes par ligne ou entre lignes par colonne en utilisant les coordonnées factorielles.

La coordonnée des lignes de rang 1 est un code numérique qui positionne les 182 relevés sur un axe. Utiliser Graph1D : Labels :

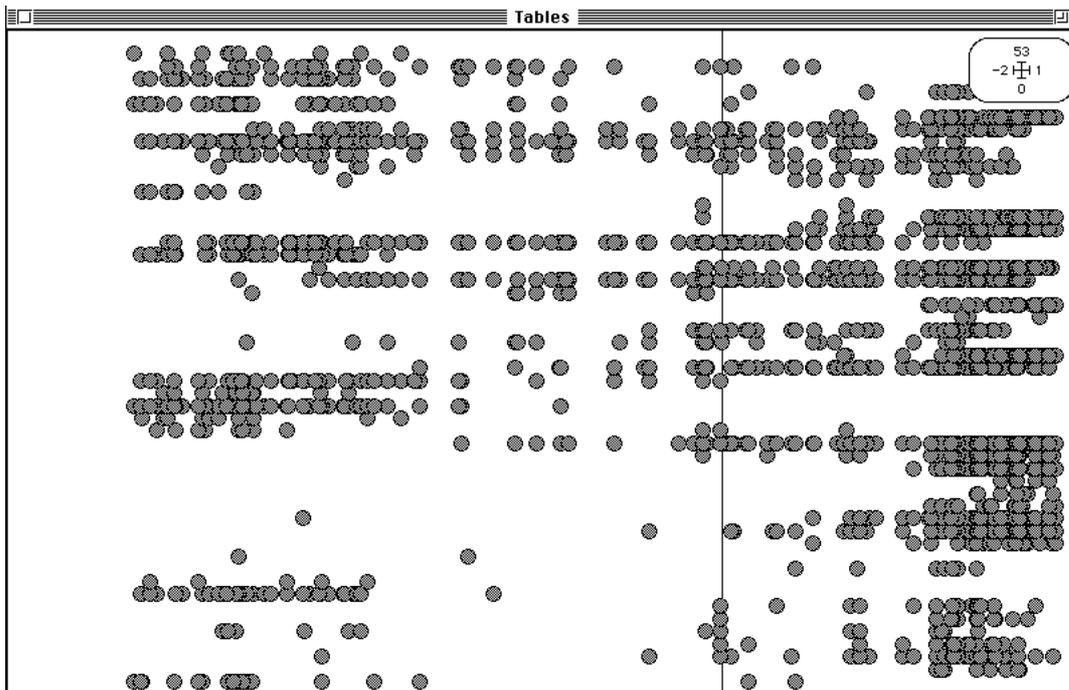
Labels			
Data file (no default)	<input type="button" value="Browse"/>	AVFau.fcli	182 4
Rows label file (default = #)	<input type="button" value="Browse"/>		
Variable label file (or #)	<input type="button" value="Browse"/>		
Vertical (1) or horizontal (2) graphs	<input type="button" value="Browse"/>	2	
<input type="button" value="Quit"/> <input type="button" value="Copy graph"/> <input type="button" value="Save graph"/> <input type="button" value="Print graph"/> <input type="button" value="Draw"/>			

Row & col. selection	
Col. selection:	<input type="text" value="1"/>
Row selection method:	<input type="radio"/> File <input checked="" type="radio"/> Keyboard
Graph number: 1	<input type="button" value="Validate & next graph"/>
Row selection:	<input type="text"/>
<input type="button" value="Draw"/>	



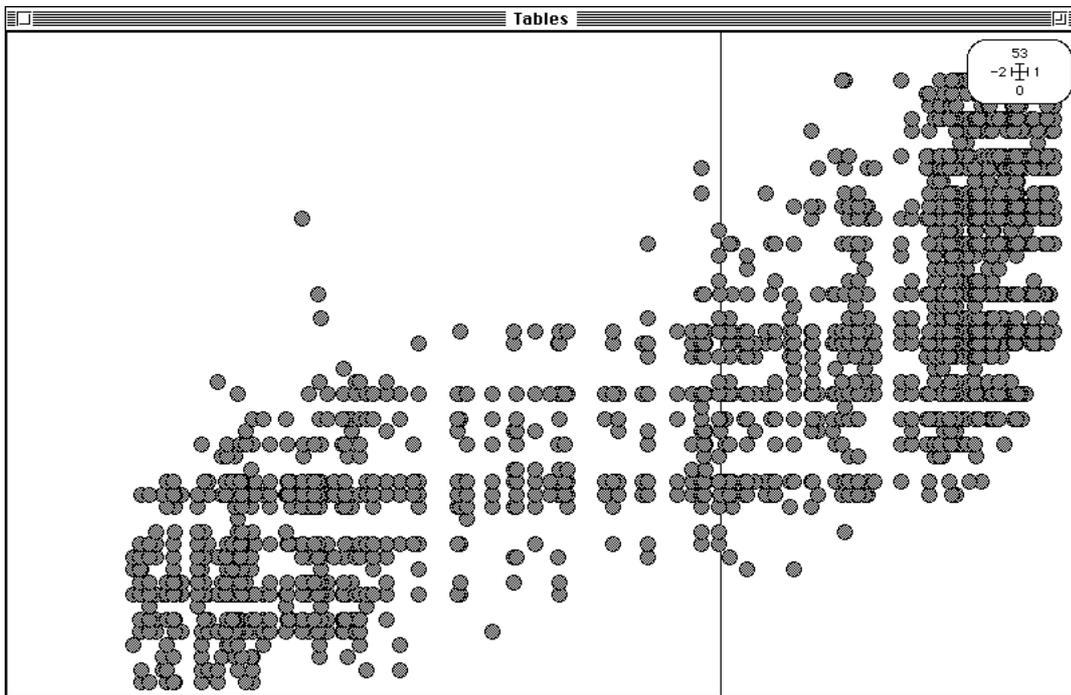
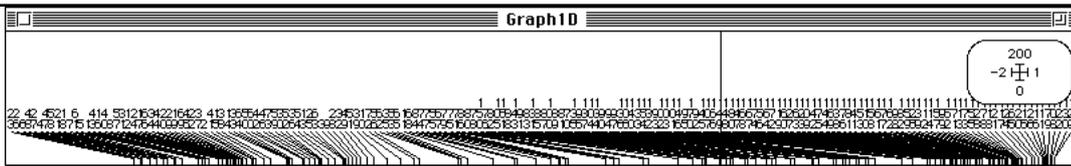
Chaque colonne est une espèce qui est présente dans certains des relevés. Transposer le tableau dans AVFau_TR (FilesUtil : Transpose) et utiliser Tables : Values :

Values			
Input table file	<input type="button" value="Browse"/>	AVFau_TR	51 182
H-axis position file	<input type="button" value="Browse"/>	AVFau.fcli	182 4

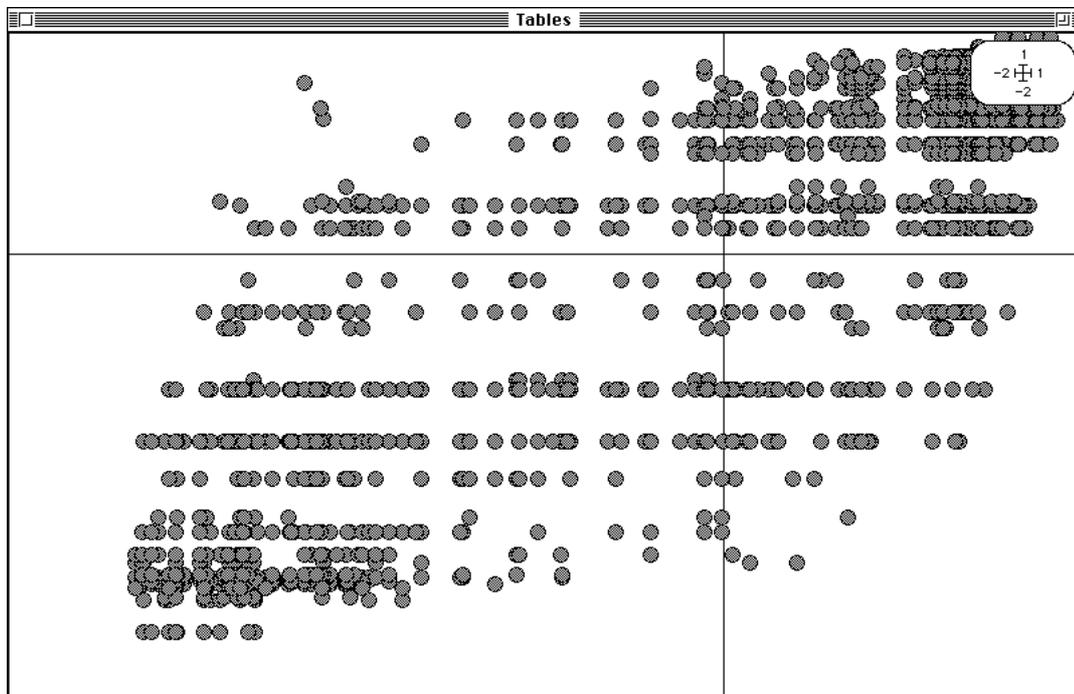


On voit la répartition de chacune des espèces le long de l'axe sur lequel sont ordonnés les 182 relevés. Chacune des espèces a une position moyenne sur cet axe. Ces positions moyennes varient d'une espèce à l'autre. Cela est plus visible si on ordonne le graphique en fonction de ces positions. La coordonnée des espèces est proportionnelle à cette moyenne et permet de faire cette mise en ordre :

Values			
Input table file	<input type="button" value="Browse"/>	AVFau_TR	51 182
H-axis position file	<input type="button" value="Browse"/>	AVFau.fcli	182 4
Column number (default = 1)	<input type="button" value="Browse"/>		
Y-axis position file	<input type="button" value="Browse"/>	AVFau.fcco	51 4
Column number (default = 1)	<input type="button" value="Browse"/>		
H-axis: Ordination (1) or Ranking (2)	<input type="button" value="Browse"/>		
Y-axis: Ordination (1) or Ranking (2)	<input type="button" value="Browse"/>	2	



En utilisant l'ordination au lieu du rangement, on obtient :



Ainsi l'AFC positionne les relevés sur un axe pour que la variance des moyennes par espèce soit maximale, positionne les espèces sur un axe pour que la variance des moyennes par relevé soit maximale, positionne les espèces et les relevés sur un axe pour que la corrélation soit maximale. La solution commune est la première coordonnée de l'AFC.



La procédure de calcul qui donne essentiellement une liste de valeurs propres et des coordonnées (scores) aux lignes et aux colonnes est la partie commune d'un grand nombre de propriétés, ce qui explique que l'AFC a été redécouverte plusieurs fois et est pratiquée de diverses manières³. Pour des raisons théoriques, c'est une des principales méthodes de la statistique descriptive multivariée.

En écologie, on s'en sert pour représenter les résultats d'une classification⁴, en particulier en phyto-sociologie⁵. Elle a été introduite pour analyser les tableaux faunistiques en hydrobiologie⁶, en ornithologie⁷, en planctonologie⁸. La représentation des coordonnées factorielles en fonction du temps ou de l'espace⁹ introduit en écologie la notion de discrimination par l'AFC. Le modèle d'ordination réciproque est repéré par Hill¹⁰ et correspond au modèle général des gradients environnementaux¹¹.

Indépendamment, Feoli & Orlóci (1979)¹² s'attribuent la procédure sous le nom de *analysis of concentration*, en partant de l'article de¹³ qui parle de *analysis of association*, alors qu'on peut y voir¹⁴ une analyse en composantes principales doublement standardisée, en partant de¹⁵. Des dizaines d'articles utilisent, précisent et commentent la méthode.

La procédure est clairement décrite par Hirschfeld (1935)¹⁶ comme base théorique des codages, Williams (1952)¹³ pour mesurer la corrélation entre variables qualitatives, Kendall & Stuart (1961)¹⁷ comme méthode d'estimation fishérienne. L'approche géométrique¹⁸ enrichit l'interprétation et est popularisée dans¹⁹.

Après cette option, le module DDUtil est disponible.



1 Nishisato, S. (1980) *Analysis of categorical data : dual scaling and its applications*. University of Toronto Press, London. 1-276.

2 Escoufier, Y. (1982) L'analyse des tableaux de contingence simples et multiples. *Metron* : 40, 53-77.

Escoufier, Y. (1985) L'analyse des correspondances : ses propriétés et ses extensions. In : *Proceedings 45th session. Institut International de la Statistique*. Amsterdam. 28.2.1-28.2.16.

3 Buyse, M. (1983) Les différentes approches conduisant à l'analyse des correspondances. *Biométrie-praximétrie* : 23, 1, 1-26.

Rijkevorsel, J. van. (1987) *The application of fuzzy coding and hoerseshoes in multiple correspondence analysis*. DSWO Press, Leiden. 1-272.

4 Roux, G. & Roux, M. (1967) A propos de quelques méthodes de classification en phytosociologie. *Revue de Statistique Appliquée* : 25, 59-72.

5 Guinochet, M. (1973) *Phytosociologie*. Masson, Paris. 1-228.

6 Lévêque, C. & Gaborit, M. (1972) Utilisation de l'analyse factorielle des correspondances pour l'étude des peuplements en Mollusques benthiques du lac Tchad. *Cahiers ORSTOM, Série Hydrobiologie* : 4, 1, 47-66.

7 Lebreton, J.D. (1973) Etude des déplacements saisonniers des Sarcelles d'hiver, *Anas c. crecca* L., hivernant en Camargue à l'aide de l'analyse factorielle des correspondances. *Compte rendu hebdomadaire des séances de l'Académie des sciences. Paris, D* : III, 277, 2417-2420.

8 Ibanez, F. & Seguin, G. (1972) Etude du cycle annuel du zooplancton d'Abidjan. Comparaison de plusieurs méthodes d'analyse multivariable. *Investigación pesquera* : 36, 81-108.

9 Hatheway, W.H. (1971) Contingency table analysis of rain forest vegetation. In : *Statistical Ecology. III Many species populations ecosystems and systems analysis*. Patil, G.P., Pielou, E.C. & Waters, W.E. (Eds.) Pennsylvania State University Press. 271-314.

Dessier, A. & Laurec, A. (1978) Le cycle annuel du zooplancton à Pointe-Noire (RP Congo). Description mathématique. *Oceanologica acta* : 1, 3, 285-304.

Esteve, J. (1978) Les méthodes d'ordination : éléments pour une discussion. In : *Biométrie et Ecologie*. Legay, J.M. & Tomassone, R. (Eds.) Société Française de Biométrie, Paris. 223-250.

¹⁰ Hill, M.O. (1973) Reciprocal averaging : an eigenvector method of ordination. *Journal of Ecology* : 61, 237-249.

Hill, M.O. (1974) Correspondence analysis : A neglected multivariate method. *Journal of the Royal Statistical Society, C* : 23, 340-354.

¹¹ Bates, J.W. & Brown, D.H. (1981) Epiphyte differentiation between *Quercus petraea* and *Fraxinus excelsior* trees in a maritime area of South West England. *Vegetatio* : 48, 61-70.

Prodon, R. & Lebreton, J.D. (1981) Breeding avifauna of a Mediterranean succession : the holm oak and cork oak series in the eastern Pyrénées. 1 : Analysis and modelling of the structure gradient. *Oikos* : 37, 21-38.

¹² Feoli, E. & Orloci, L. (1979) Analysis of concentration and detection of underlying factors in structured tables. *Vegetatio* : 40, 49-54.

¹³ Williams, E.J. (1952) Use of scores for the analysis of association in contingency tables. *Biometrika* : 39, 274-289.

¹⁴ Noy-Meir, I. (1973) Data transformations in ecological ordination. I. Some advantages of non-centering. *Journal of Ecology* : 61, 329-341.

¹⁵ Benzecri, J.P. (1969) Statistical analysis as a tool to make patterns emerge from data. In : *Methodologies of pattern recognition*. Watanabe, S. (Ed.) Academic Press, New-York. 35-60.

¹⁶ Hirschfeld, H.O. (1935) A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences* : 31, 520-524.

¹⁷ Kendall, D.G. & Stuart, A. (1961) *The advanced theory of statistics. Vol 2: Inference and relationships*. Cha. 33 : Categorized data. Griffin, London. 536-591.

¹⁸ Benzécri, J.P. & Coll. (1973) *L'analyse des données. II L'analyse des correspondances*. Bordas, Paris. 1-620.

¹⁹ Greenacre, M. (1984) *Theory and applications of correspondence analysis*. Academic Press, London. 1-364.

Greenacre, M. & Hastie, T. (1987) The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association* : 82, 398, 437-447.



L'illustration ornithologique ci-dessus, publiée par Prodon & Lebreton (1981)¹¹, utilise la vision d'écologie statistique de Hill (1974)¹⁰ qui s'appuie sur le théorème de Williams (1952)¹³. Réserver l'assertion selon laquelle l'AFC est une méthode "française" d'analyse des données pour les moments de détente.

COA : Decentred COA



Variante originale de l'analyse des correspondances¹, qui fait partie des analyses des correspondances sur modèles².



L'analyse des correspondances est doublement centrée et doublement pondérée. Ceci implique qu'elle regroupe deux analyses sur profils (profils par lignes et profils par colonnes) décentrées sur les distributions marginales³ (Ch. 4). Ce n'est pas forcément une bonne chose quand on doit comparer chacune des distributions colonnes, non pas à la moyenne des distributions colonnes, mais à une distribution de référence extérieure au tableau de données. Cette option résout cette question.

\mathbf{T} est la matrice de départ, n_{ij} la valeur du terme de la ligne i et de la colonne j , n la somme totale de tous les éléments du tableau, I le nombre de lignes et J le nombre de colonnes. Soient \mathbf{r} le vecteur de décentrage à I composantes, r_i la valeur de l'élément de rang i , R la somme totale de tous les éléments du vecteur. On considère, par exemple, que n_{ij} est le nombre de relevés de la campagne i qui contient l'espèce j et r_i le nombre de relevés de cette campagne. On note la fréquence d'apparition de l'espèce j dans les relevés de la campagne i :

$$f_{j/i} = \frac{n_{ij}}{r_i}$$

La fréquence totale d'apparition de l'espèce j est d'autre part :

$$f_j = \frac{n_{.j}}{R} \text{ avec } n_{.j} = \sum_{i=1, I} n_{ij}$$

L'AFC d'un tableau de nombre positifs ou nul \mathbf{T} décentrée par \mathbf{r} est l'analyse d'un schéma de dualité⁴ avec le jeu de paramètres :

1 — Transformation initiale : décentrage colonne. Le tableau traité est \mathbf{Z} de terme général :

$$z_{ij} = \frac{f_{j/i}}{f_j} - 1$$

2 — Pondération des lignes : chaque ligne a le poids r_i/R avec $\sum_{i=1, I} r_i/R = 1$;

3 — Pondération des colonnes : chaque colonne a le poids f_j .



L'option utilise une seule fenêtre de dialogue :

Decentred COA			
Data file	Fau	35	9
Column reference file	NumPoint/li	35	1
Generic output file	New		
l = Save new frequency table	1		
<input type="button" value="Quit"/>		<input type="button" value="Ok"/>	

Fichier du tableau d'entrée (nombres positifs ou nuls).

Fichier du vecteur de décentrage.

Nom générique des fichiers de sortie.

Option de conservation de a table des fréquences par lignes.



Utiliser la carte Poissons de la pile ADE•Data. Obtenir les fichiers Fau (35-9), NumPoint/li (35-1) et Label_Esp (étiquettes).

```

fc/COA: Decentred correspondance analysis
Input file: Fau
Number of rows: 35, columns: 9
Column reference: NumPoint/li
Number of rows: 35, column: 1
Total:          7009
File New.frpl contains the edge distribution of rows
It has 35 rows and 1 column
Frequency distribution from column reference file NumPoint/li
File New.frpc contains the edge distribution of columns
It has 9 rows and 1 column
File New contains the table [P(j/i)]
It has 35 rows and 9 columns
File New.frta contains the table [P(j/i)/F(j)-1]
It has 35 rows and 9 column
Col:   1 Khi2 =   148.5 DDL =   34 Proba =           0 Trace Cont. =   0.0493
Col:   2 Khi2 =   274.7 DDL =   34 Proba =           0 Trace Cont. =   0.0912
Col:   3 Khi2 =   449.3 DDL =   34 Proba =           0 Trace Cont. =   0.149
Col:   4 Khi2 =   103.5 DDL =   34 Proba =   8.67E-09 Trace Cont. =   0.0344
Col:   5 Khi2 =   495.5 DDL =   34 Proba =           0 Trace Cont. =   0.165
Col:   6 Khi2 =   275.7 DDL =   34 Proba =           0 Trace Cont. =   0.0915
Col:   7 Khi2 =     555 DDL =   34 Proba =           0 Trace Cont. =   0.184
Col:   8 Khi2 =   426.1 DDL =   34 Proba =           0 Trace Cont. =   0.141
Col:   9 Khi2 =   283.2 DDL =   34 Proba =           0 Trace Cont. =   0.094

```

DiagoRC: General program for two diagonal inner product analysis

Input file: New.frta
--- Number of rows: 35, columns: 9

Total inertia: 0.429685

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.3945E-01	+0.5573	+0.5573	02	+7.4861E-02	+0.1742	+0.7315
03	+3.6563E-02	+0.0851	+0.8166	04	+2.6486E-02	+0.0616	+0.8782
05	+1.7233E-02	+0.0401	+0.9183	06	+1.0756E-02	+0.0250	+0.9434
07	+8.9793E-03	+0.0209	+0.9643	08	+8.4688E-03	+0.0197	+0.9840
09	+6.8872E-03	+0.0160	+1.0000				

File New.frvp contains the eigenvalues and relative inertia for each axis
--- It has 9 rows and 2 columns

File New.frco contains the column scores
--- It has 9 rows and 2 columns

File :New.frco

Col.	Mini	Maxi
1	2.576e-02	8.960e-01
2	-2.567e-01	4.870e-01

File New.frli contains the row scores
--- It has 35 rows and 2 columns

File :New.frli

Col.	Mini	Maxi
1	-6.710e-01	1.851e+00
2	-4.343e-01	7.516e-01



Les tests Khi2 du listing n'ont de sens que dans la situation expérimentale citée. Ce sont les Khi2 d'ajustement entre la distribution observée sur une colonne et la distribution de référence (Cf. la théorie des profils écologique [EcolTools : Ecological Profiles](#)). Noter encore que le tableau traité peut être compris de deux manières :

$$\frac{f_{j/i}}{f_j} = \frac{\frac{n_{ij}}{r_i}}{\frac{n_{.j}}{R}} = \frac{n_{ij}}{r_i} \frac{R}{n_{.j}} = \frac{n_{ij}}{n_{.j}} \frac{R}{r_i} = \frac{p_{i/j}}{r_i/R}$$



Les coordonnées factorielles des lignes (relevés) issues de l'analyse présentent des propriétés remarquables. Elles sont centrées pour la pondération externe $\frac{r_1}{R}, \dots, \frac{r_I}{R}$.

Elles ont des variances λ_k (k-ième valeur propre de l'analyse) et non corrélées pour la pondération des colonnes.

Les colonnes (coordonnées factorielles des colonnes) sont positionnées (à un coefficient $1/\sqrt{\lambda_k}$ près) à la moyenne des lignes pour leur distribution de fréquences de points.

L'analyse positionne les lignes pour maximiser la moyenne pondérée des carrés des écarts entre la position occupée par une distribution colonne donnée et la position de la distribution de référence qui est à l'origine.

Après cette option, le module DDUtil est disponible.



¹ Dolédec, S., Chessel, D. & Olivier, J.M. (1995) L'analyse des correspondances décentrée: application aux peuplements ichtyologiques du haut-Rhône. *Bulletin français de pisciculture* (sous presse).

² Escoufier, B. (1984) Analyse factorielle en référence à un modèle. Applications à l'analyse d'un tableau d'échanges. *Revue de Statistique Appliquée* : 32, 4, 25-36.

Escoufier, Y. (1985) L'analyse des correspondances : ses propriétés et ses extensions. In : *Proceedings 45th session. Institut International de la Statistique*. Amsterdam. 28.2.1-28.2.16.

³ Greenacre, M. (1984) *Theory and applications of correspondence analysis*. Academic Press, London. 1-364.

⁴ Escoufier, Y. (1987) The duality diagram : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute, Serie G. Springer Verlag, Berlin. 139-156.

COA : Internal COA



Variante de l'analyse des correspondances. Analyse des correspondances intra-classes simple¹ ou double, dite encore analyse des correspondances internes².



Un tableau d'analyse des correspondances à I lignes et J colonnes est partitionné en L blocs lignes et K blocs colonnes. On ne veut tenir compte dans sa description que des structures internes des sous-tableaux définis par les deux partitions. L'option introduit dans le schéma de dualité de l'analyse des correspondances deux projecteurs et ajoute les aides à l'interprétation qui en découlent. La situation type est celle de l'analyse d'un tableau faunistique décomposé par groupes taxonomiques et par saison d'échantillonnage. Lorsque les structures observées du point de vue choisi sont faibles, par rapport aux structures globales définies par l'analyse des correspondances, l'analyse des correspondances internes donne des résultats radicalement différents.

Quand L et K valent 1, on a une analyse des correspondances simple. Quand un des deux paramètres seulement vaut 1, on a une analyse des correspondances intra-classes.



L'option utilise une seule fenêtre de dialogue :

Input file (---)	ardfau	43	35
Row indicator (Default = 1 variable)	ardblolig	4	1
Col indicator (Default = 1 variable)	ardblocol	6	1
Output file (default = ---)			

Nom du fichier d'entrée binaire (valeurs positives ou nulles).

Fichier indicateur des blocs de lignes.

Fichier indicateur des blocs de colonnes.

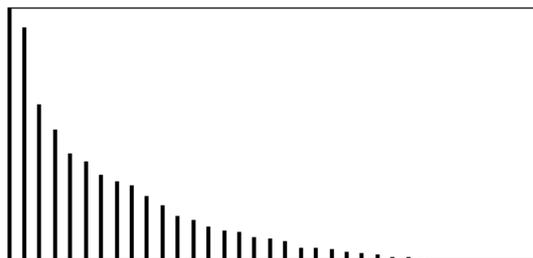
Nom générique des fichiers de sortie (par défaut, le nom du fichier d'entrée).



Particularités



Utiliser la carte Ardèche de la pile ADE•Data pour obtenir le fichier ardfau_(43-35) par TextToBin : Char->Binary et par TextToBin : BlocTo1Categ les fichiers ardblolig (4-1) et ardblocol (6-1) et les fichiers d'indicatrice de classes associés. Exécuter la présente option :



```
Internal correspondence analysis
File ardfau.wwpl contains the row weights
It has 43 rows and 1 column
File ardfau.wwpc contains the column weights
It has 35 rows and 1 column
Row bloc number : 4
--> Row number in a bloc in file : ardblolig
Col bloc number : 6
--> Row number in a bloc in file : ardblocol
File ardfau.wwil contains class indicator
It has 43 rows and 1 column
File ardfau.wwic contains class indicator
```

It has 35 rows and 1 column
 File ardfau.wwta contains the doubly centered matrix
 It has 43 rows and 35 column

 DiagoRC: General program for two diagonal inner product analysis
 Input file: ardfau.wwta
 --- Number of rows: 43, columns: 35

 Total inertia: 0.448464

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+6.8581E-02	+0.1529	+0.1529	02	+6.3251E-02	+0.1410	+0.2940
03	+4.2535E-02	+0.0948	+0.3888	04	+3.5656E-02	+0.0795	+0.4683
05	+2.9109E-02	+0.0649	+0.5332	06	+2.7068E-02	+0.0604	+0.5936
07	+2.3163E-02	+0.0516	+0.6452	08	+2.1602E-02	+0.0482	+0.6934
09	+2.0494E-02	+0.0457	+0.7391	10	+1.7512E-02	+0.0390	+0.7781
11	+1.5024E-02	+0.0335	+0.8116	12	+1.2167E-02	+0.0271	+0.8388
13	+1.1056E-02	+0.0247	+0.8634	14	+9.0633E-03	+0.0202	+0.8836
15	+8.2919E-03	+0.0185	+0.9021	16	+7.8503E-03	+0.0175	+0.9196
17	+6.3848E-03	+0.0142	+0.9339	18	+5.8888E-03	+0.0131	+0.9470
19	+5.2470E-03	+0.0117	+0.9587	20	+3.5722E-03	+0.0080	+0.9667
21	+3.4801E-03	+0.0078	+0.9744	22	+2.9082E-03	+0.0065	+0.9809
23	+2.2384E-03	+0.0050	+0.9859	24	+1.8397E-03	+0.0041	+0.9900
25	+1.6856E-03	+0.0038	+0.9938	26	+1.0721E-03	+0.0024	+0.9962
27	+8.8481E-04	+0.0020	+0.9981	28	+5.0391E-04	+0.0011	+0.9993
29	+3.3459E-04	+0.0007	+1.0000	30	+0.0000E+00	+0.0000	+1.0000
31	+0.0000E+00	+0.0000	+1.0000	32	+0.0000E+00	+0.0000	+1.0000
33	+0.0000E+00	+0.0000	+1.0000	34	+0.0000E+00	+0.0000	+1.0000
35	+0.0000E+00	+0.0000	+1.0000				

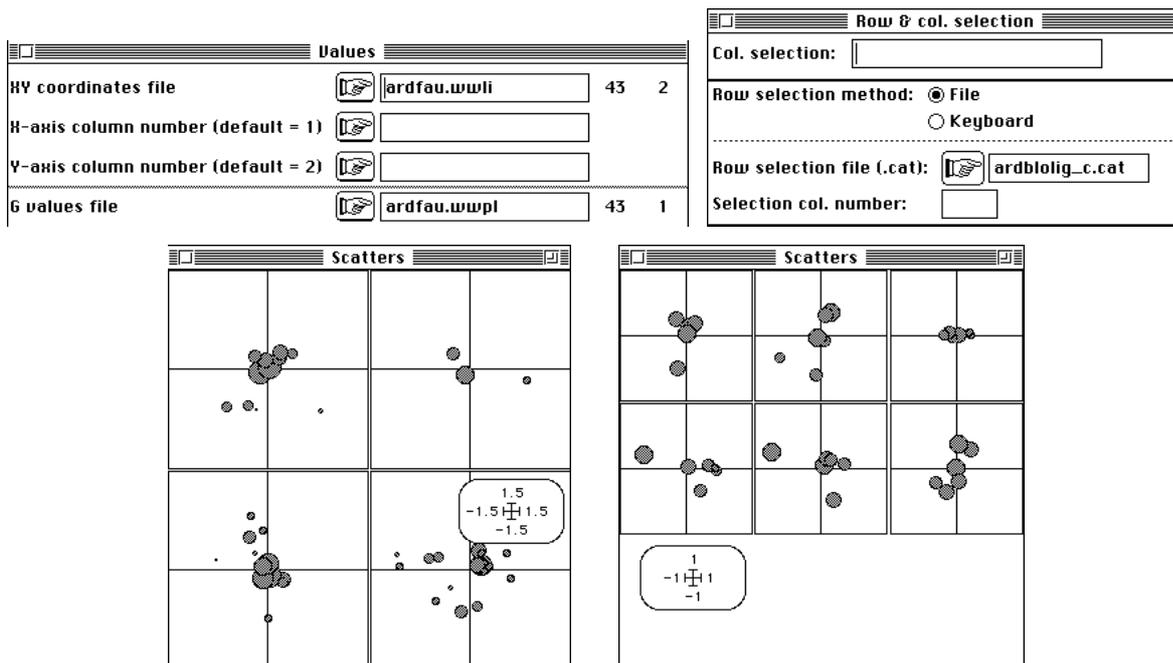
File ardfau.wwvp contains the eigenvalues and relative inertia for each axis
 --- It has 35 rows and 2 columns

File ardfau.wwco contains the column scores
 --- It has 35 rows and 2 columns

File :ardfau.wwco

Col.	Mini	Maxi
1	-7.202e-01	4.860e-01
2	-6.261e-01	3.558e-01

Les coordonnées des lignes (ci-dessous à gauche) et des colonnes (ci-dessous à droite) sont centrées par bloc pour la pondération correspondante :



File ardfau.wvli contains the row scores
 --- It has 43 rows and 2 columns
 File : ardfau.wvli

Col.	Mini	Maxi
1	-1.080e+00	8.699e-01
2	-7.461e-01	8.054e-01

Les valeurs propres qui sont des moyennes d'inertie projetée par bloc (ou inertie intra-classe¹) se décomposent. On trouve le numéro du bloc, son poids, sa part de valeur propre pour chaque axe conservé.

Bloc:	1	0.288	0.0511	0.0563
Bloc:	2	0.0653	0.115	0.0263
Bloc:	3	0.37	0.0204	0.0709
Bloc:	4	0.277	0.14	0.069
-----			0.0686	0.0633

Cette part s'exprime en pourcentage de participation :

Eigenvalue		percentage	
Bloc:	1	0.214	0.256
Bloc:	2	0.11	0.0271
Bloc:	3	0.11	0.415
Bloc:	4	0.566	0.302

La même édition est faite sur l'ensemble des blocs de colonnes :

Bloc:	1	0.186	0.0109	0.0706
Bloc:	2	0.18	0.0413	0.106
Bloc:	3	0.105	0.017	0.00068
Bloc:	4	0.136	0.192	0.032
Bloc:	5	0.19	0.138	0.0613
Bloc:	6	0.203	0.0243	0.0736
-----			0.0686	0.0633

Eigenvalue		percentage	
Bloc:	1	0.0295	0.207
Bloc:	2	0.108	0.302
Bloc:	3	0.0262	0.00113
Bloc:	4	0.381	0.0691
Bloc:	5	0.383	0.184
Bloc:	6	0.0721	0.236

Cette analyse est la plus simple des analyses à deux projecteurs². Pour les analyses intra-classes, voir aussi Discrimin : Within analysis: Run.



¹ 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.

² Cazes, P., Chessel, D. & Dolédec, S. (1988) L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie. *Revue de Statistique Appliquée* : 36, 39-54.



Après cette option, le module DDUtil est disponible.

COA : NSCA_Col_Profiles



Méthode d'ordination d'un tableau due à N. Lauro & L. D'Ambra ¹.



L'Analyse non symétrique des correspondances ordonne un tableau **T** de valeurs positives ou nulles comme l'analyse des correspondances. On note :

$$\mathbf{T} = [n_{ij}]_{\substack{1 \leq i \leq I, \\ 1 \leq j \leq J}} \quad n_{i.} = \sum_{j=1}^J n_{ij} \quad n_{.j} = \sum_{i=1}^I n_{ij} \quad n_{..} = \sum_{j=1}^J n_{.j}$$

$$f_{i.} = \frac{n_{i.}}{n_{..}} \quad \sum_{i=1}^I f_{i.} = 1 \quad f_{.j} = \frac{n_{.j}}{n_{..}} \quad \sum_{j=1}^J f_{.j} = 1$$

$$f_{ij|j} = \frac{n_{ij}}{n_{.j}} \quad \sum_{j=1}^J f_{ij|j} = 1 \quad f_{j|i} = \frac{n_{ij}}{n_{i.}} \quad \sum_{i=1}^I f_{j|i} = 1$$

$$\mathbf{D}_J = \text{Diag}(f_{.1}, \dots, f_{.j}, \dots, f_{.J}) \quad \mathbf{D}_I = \text{Diag}(f_{1.}, \dots, f_{i.}, \dots, f_{I.})$$

L'analyse proposée par l'option est celle du triplet :

$$(\mathbf{K}_0, \mathbf{D}_J, \mathbf{D}_I) \quad \mathbf{K}_0 = [f_{i|j} - f_{i.}]_{\substack{1 \leq i \leq I, \\ 1 \leq j \leq J}}$$



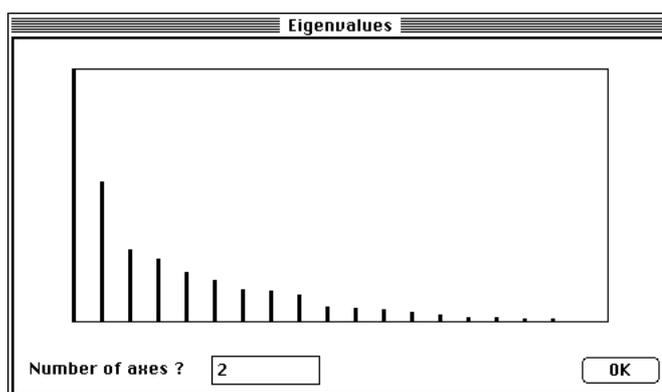
L'option utilise une seule fenêtre de dialogue :



Nom du fichier binaire d'entrée.



Utiliser la carte JV73_Poissons ² de la pile de données. Dans le fichier Poi 19 colonnes (espèces) décrivent la répartition de l'ichtyofaune sur 92 stations (relevés).



```
zc/NSCA: Column Profiles
Input file: Poi
Number of rows: 92, columns: 19
File Poi.zcta contains the centred table P*DJ-1 -VI*1J'
It has 92 rows and 19 columns
File Poi.zcpc contains the margin distribution of columns
It has 19 rows and 1 column
File Poi.zcpl contains the row weights (1)
It has 92 rows and 1 column
```

Le triplet étudié est décrit.

```
File Poi.zcma contains:
the number of rows: 92
the number of columns: 19
```

the total number: 1363

L'inertie totale de cette analyse est une partie de la quantité (notée ci-dessous comme un indice de Simpson, terme correct quand les colonnes sont des espèces) :

$$H = \sum_{i=1}^I f_i (1 - f_i) = 1 - \sum_{i=1}^I f_i^2$$

Total Simpson Diversity = 9.8337e-01

Inertia = 2.2831e-02

Le rapport de l'inertie de l'analyse sur H est un pourcentage de variance expliquée (Voir Light & Margolin ³).

Light & Margolin 1971 R2 = 2.3217e-02

Quand le tableau étudié est une vraie table de contingence (ce qui n'est pas le cas ici), la signification statistique se fait par :

CATANOVA Index = C = 2.8776e+03

Approximated as Khi2 1638 DF

Proba Khi2 > Obs = 0.0000e+00

Le triplet est diagonalisé :

DiagoRC: General program for two diagonal inner product analysis

Input file: Poi.zcta

--- Number of rows: 92, columns: 19

Total inertia: 0.0228309

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+7.3279E-03	+0.3210	+0.3210	02	+4.0768E-03	+0.1786	+0.4995
03	+2.0929E-03	+0.0917	+0.5912	04	+1.8508E-03	+0.0811	+0.6723
05	+1.4555E-03	+0.0638	+0.7360	06	+1.2181E-03	+0.0534	+0.7894
07	+9.3588E-04	+0.0410	+0.8304	08	+9.1208E-04	+0.0399	+0.8703
09	+7.7500E-04	+0.0339	+0.9043	10	+4.4945E-04	+0.0197	+0.9239
11	+4.1656E-04	+0.0182	+0.9422	12	+3.7022E-04	+0.0162	+0.9584
13	+2.7697E-04	+0.0121	+0.9705	14	+2.1873E-04	+0.0096	+0.9801
15	+1.4857E-04	+0.0065	+0.9866	16	+1.1964E-04	+0.0052	+0.9919
17	+1.0022E-04	+0.0044	+0.9962	18	+8.5644E-05	+0.0038	+1.0000
19	+0.0000E+00	+0.0000	+1.0000				

File Poi.zcvp contains the eigenvalues and relative inertia for each axis

--- It has 19 rows and 2 columns

La procédure générale de diagonalisation des triplets donne un fichier des coordonnées des colonnes et des lignes. Il n'est pas conseillé d'utiliser le second pour l'interprétation :

File Poi.zcco contains the column scores

--- It has 19 rows and 2 columns

File :Poi.zcco

Col.	Mini	Maxi
1	-9.742e-02	1.687e-01
2	-1.406e-01	2.336e-01

File Poi.zcli contains the row scores

--- It has 92 rows and 2 columns

File :Poi.zcli

Col.	Mini	Maxi
1	-9.083e-03	3.267e-02
2	-1.778e-02	2.357e-02

On débutera le dépouillement en utilisant les coordonnées normées :

File Poi.zcl1 contains the row scores with unit norm

It has 92 rows and 2 columns

File :Poi.zcl1

Col.	Mini	Maxi
1	-1.061e-01	3.817e-01
2	-2.784e-01	3.692e-01

Trajectories	
HY coordinates file	Poi.zc11
H-axis column number (default = 1)	
Y-axis column number (default = 2)	
Categories file (.cat)	Riv_c.cat

Ces scores des relevés sont centrés pour la pondération uniforme et la somme des carrés des valeurs vaut l'unité. Les colonnes (espèces) sont des distributions de fréquence qui fournissent des positions moyennes par averaging :

File Poi.zccs contains the column scores from averaging
It has 19 rows and 2 columns
File :Poi.zccs

Col.	Mini	Maxi
1	-4.319e-02	2.230e-01
2	-1.522e-01	2.220e-01

Labels	
HY coordinates file	Poi.zccs
H-axis column number (default = 1)	
Y-axis column number (default = 2)	
Label file (or # for item numbers)	Label_Poi

Les poids observés des relevés $(f_1, \dots, f_i, \dots, f_I)$ donnent leur position moyenne qui est aussi la moyenne pondérée des scores des espèces pour la pondération $(f_1, \dots, f_j, \dots, f_J)$:

File Poi.zccg contains the gravity center scores from averaging
It has 1 rows and 2 columns
File :Poi.zccg

Col.	Mini	Maxi
1	5.423e-02	5.423e-02
2	-1.165e-02	-1.165e-02

Labels	
HY coordinates file	Poi.zccg

Les trois figures sont superposées pour l'interprétation (ci-dessous). Les relevés sont positionnés sur un ou plusieurs axes par des codes numériques de somme de carrés unités. Chaque espèce prend une position moyenne et la variance de ces positions moyennes est optimisée. L'écart entre la position de l'espèce et la position moyenne des espèces définit la coordonnée du fichier ---.zcco.

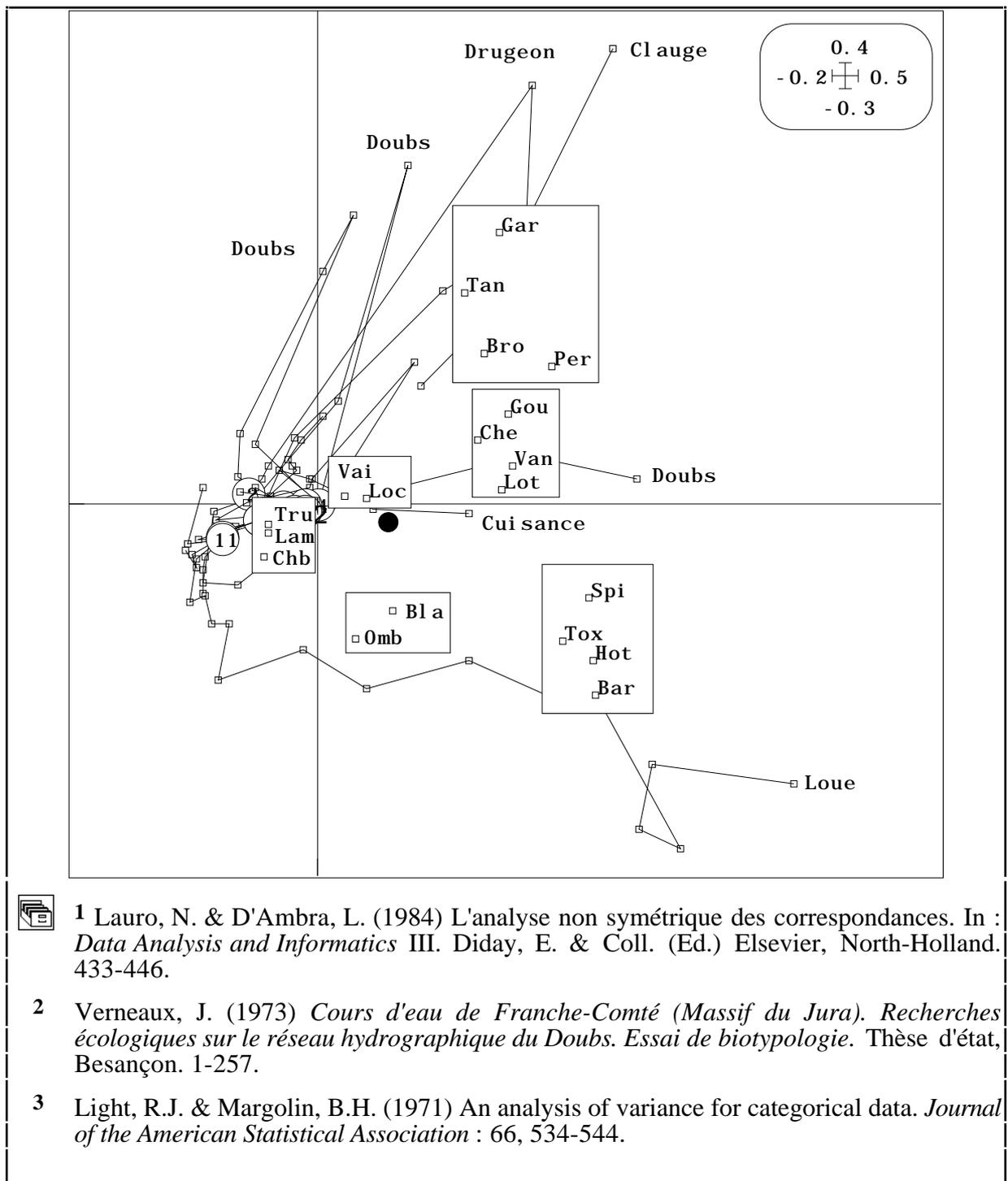


Les options COA : NSCA Row Profiles et COA : NSCA Col Profiles sont symétriques. Si l'une est appliquée à un tableau et l'autre est appliquée au tableau transposé, on obtient le même résultat. Mais si elles sont appliquées au même tableau, on obtient des résultats très différents. Si on n'a pas de raisons de choisir l'une ou l'autre, l'AFC (COA : COrrrespondence Analysis) s'impose.

Le test CATANOVA n'a de sens que pour les tableaux d'entier. C'est une alternative au Khi2 d'une table de contingence.



La nature et le rôle de cette analyse est discutée dans la fiche thématique 2.8. Voir le module Niche et la fiche thématique 4.7 pour une version proche de l'ANSC.



1 Lauro, N. & D'Ambra, L. (1984) L'analyse non symétrique des correspondances. In : *Data Analysis and Informatics III*. Diday, E. & Coll. (Ed.) Elsevier, North-Holland. 433-446.

2 Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse d'état, Besançon. 1-257.

3 Light, R.J. & Margolin, B.H. (1971) An analysis of variance for categorical data. *Journal of the American Statistical Association* : 66, 534-544.

COA : NSCA_Row_Profiles



Méthode d'ordination d'un tableau due à N. Lauro & L. D'Ambra ¹.



L'Analyse non symétrique des correspondances ordonne un tableau **T** de valeurs positives ou nulles comme l'analyse des correspondances. On note :

$$\mathbf{T} = [n_{ij}]_{\substack{i \in I, 1 \\ j \in J}} \quad n_{i.} = \sum_{j=1}^J n_{ij} \quad n_{.j} = \sum_{i=1}^I n_{ij} \quad n_{..} = \sum_{j=1}^J n_{.j}$$

$$f_{i.} = \frac{n_{i.}}{n_{..}} \quad \sum_{i=1}^I f_{i.} = 1 \quad f_{.j} = \frac{n_{.j}}{n_{..}} \quad \sum_{j=1}^J f_{.j} = 1$$

$$f_{ij|j} = \frac{n_{ij}}{n_{.j}} \quad \sum_{j=1}^J f_{ij|j} = 1 \quad f_{j|i} = \frac{n_{ij}}{n_{i.}} \quad \sum_{i=1}^I f_{j|i} = 1$$

$$\mathbf{D}_J = \text{Diag}(f_{.1}, \dots, f_{.J}) \quad \mathbf{D}_I = \text{Diag}(f_{1.}, \dots, f_{I.})$$

L'analyse proposée par l'option est celle du triplet :

$$(\mathbf{L}_0, \mathbf{I}d_J, \mathbf{D}_I) \quad \mathbf{L}_0 = [f_{j|i} - f_{.j}]_{\substack{i \in I, 1 \\ j \in J}}$$

Elle est strictement équivalente à COA : NSCA_Col_Profiles sur le tableau transposé.



L'option utilise une seule fenêtre de dialogue :



Nom du fichier binaire d'entrée.



Utiliser le fichier DanishEdu de la carte de données Medica. Les données sont proposées par Light & Margolin 2 pour illustrer leur test CATANOVA. On obtient les résultats détaillés par les auteurs :

```
Total Simpson Diversity = 7.2118e-01
Inertia = 1.5129e-02
Light & Margolin 1971 R2 = 2.0978e-02
CATANOVA Index = C = 5.1691e+01
Approximated as Khi2 4 DF
Proba Khi2 > Obs = 1.3526e-09
```



Utiliser la carte JV73_Poissons de la pile de données. Transposer le fichier Poi (92-19) en PoiTR (19-92) et retrouver les résultats de l'option COA : NSCA_Col_Profiles appliquée à Poi.



¹ Lauro, N. & D'Ambra, L. (1984) L'analyse non symétrique des correspondances. In : *Data Analysis and Informatics III*. Diday, E. & Coll. (Ed.) Elsevier, North-Holland. 433-446.

² Light, R.J. & Margolin, B.H. (1971) An analysis of variance for categorical data. *Journal of the American Statistical Association* : 66, 534-544.

COA : Reciprocal scaling



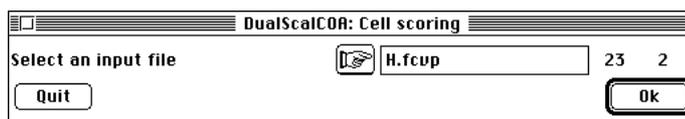
Aide à l'interprétation d'une analyse des correspondances comme analyse canonique¹.



On sait que l'AFC d'un tableau positionne les lignes à la moyenne de leur distribution sur les colonnes et les colonnes à la moyenne de leur distribution sur les lignes. On connaît moins la notion d'AFC comme analyse canonique. Ceci permet d'attribuer à chaque cellule du tableau traité des coordonnées factorielles et de voir les lignes et les colonnes du tableau comme des distributions de fréquences sur le même nuage des cellules.



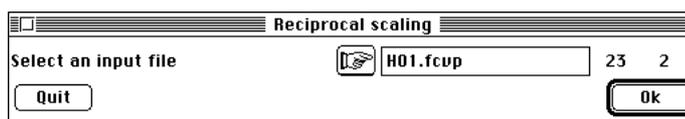
L'option utilise une seule fenêtre de dialogue :



 Nom de fichier du type --.fcvp créé par COA : COrespondence Analysis.



Utiliser la carte Forêts de la pile ADE•Data pour obtenir, par TextToBin : Char->Binary, le fichier H (43-23). Le passer en présence-absence par Bin->Bin : 0 or 1 Reducing (ce cas particulier donne à chaque correspondance espèce-relevé le même poids et facilite l'illustration, mais l'opération n'est évidemment pas obligatoire). Le fichier H01 (43-23) est soumis à l'AFC :



```
DualScalCOA: Reciprocal scaling after a CA
Title of the analysis: H01.
Number of rows: 43, columns: 23
```

On attribue à chaque cellule du tableau comportant une note non nulle (correspondance réelle) un score sur chaque axe factoriel. Ce score est de moyenne nulle et de variance unité (pour la pondération issue directement du tableau). Sur deux axes différents, ces scores sont non corrélés. Il y a dans le tableau 295 correspondances :

```
File H01_score contains the coordinates of non empty table cells
It has 295 rows and 2 columns
```

Une correspondance est une case non nulle du tableau. La valeur correspondante est le poids de cette correspondance. Ces poids sont rangés dans un vecteur (vérifier ici qu'ils sont tous égaux à l'unité). Ce vecteur est donc une réécriture du tableau, ligne par ligne, en omettant tous les zéros :

```
File H01_weigh contains the value, in the initial table, of these cells
It has 295 rows and 1 columns
```

Une correspondance, une case non nulle du tableau, est caractérisée par son numéro de ligne et son numéro de colonne. Ces valeurs sont rangées dans un fichier de deux variables qualitatives qui subit le traitement habituel :

```
File H01_class contains the row and column number of these cells
It has 295 rows and 2 columns
```

```
-----
| Description of a coding matrix |
-----
```

```
Qualitative variables file: H01_class
Number of rows: 295, variables: 2, categories: 66
```

```
Description of categories:
```

```
-----
Variable number 1 has 43 categories
```

```

-----
[ 1]Category: 1 Num: 11 Freq.: 0.03729
[ 2]Category: 2 Num: 18 Freq.: 0.06102
[ 3]Category: 3 Num: 8 Freq.: 0.02712
...
[ 41]Category: 41 Num: 3 Freq.: 0.01017
[ 42]Category: 42 Num: 3 Freq.: 0.01017
[ 43]Category: 43 Num: 2 Freq.: 0.00678

```

Variable number 2 has 23 categories

```

-----
[ 44]Category: 1 Num: 18 Freq.: 0.06102
[ 45]Category: 2 Num: 18 Freq.: 0.06102
[ 46]Category: 3 Num: 17 Freq.: 0.05763
...
[ 64]Category: 21 Num: 12 Freq.: 0.04068
[ 65]Category: 22 Num: 7 Freq.: 0.02373
[ 66]Category: 23 Num: 12 Freq.: 0.04068

```

Auxiliary binary output file H01_classModa: Indicator vector of modalities
It contains variable number for each modality
It has 66 rows (modalities) and one column

Auxiliary ASCII output file H01_class.123: labels (two characters) for 66 modalities
It contains one label for each modality
It has 66 rows (modalities) and labels 1,2, ..., 43, 1, 2, ...23

Chaque ligne du tableau est une distribution de poids sur l'ensemble des correspondances. Elle définit donc sur chaque score des correspondances une moyenne (proportionnelle à la coordonnée classique de cette ligne) et une variance qui sont éditées et conservées dans un fichier :

Row	1	Mean	0.541	1.356	Var	0.167	0.285
Row	2	Mean	0.540	0.582	Var	0.274	0.323
Row	3	Mean	-0.104	-0.568	Var	0.222	0.049
...							
Row	41	Mean	-0.810	2.260	Var	0.002	0.373
Row	42	Mean	-0.892	0.473	Var	0.005	0.743
Row	43	Mean	-1.141	-1.320	Var	0.000	0.002

File H01_mvli contains the within-rows mean-var.-covar.
It has 43 rows and 5 columns

Chaque colonne du tableau est une distribution de poids sur l'ensemble des correspondances. Elle définit donc sur chaque score des correspondances une moyenne (proportionnelle à la coordonnée classique de cette colonne) et une variance qui sont éditées et conservées dans un fichier :

Col	1	Mean	-0.880	-0.762	Var	0.042	0.196
Col	2	Mean	-0.860	-0.926	Var	0.067	0.092
Col	3	Mean	-0.421	-0.290	Var	0.117	0.210
...							
Col	21	Mean	0.025	1.312	Var	0.163	0.294
Col	22	Mean	0.209	1.322	Var	0.105	0.146
Col	23	Mean	0.017	0.833	Var	0.096	0.203

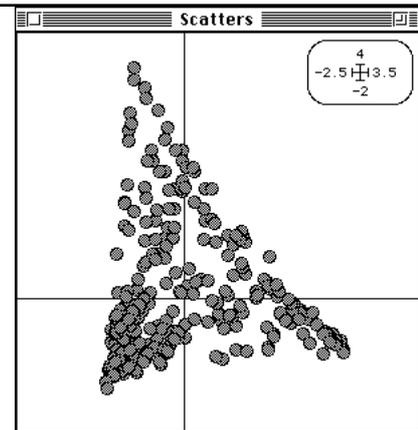
File H01_mvco contains the within-columns mean-var.-covar.
It has 23 rows and 5 columns



Cette option étend la notion de positions duales (moyennes) à celles de variabilités duales (variances)¹. Les variances associées aux espèces sont des amplitudes d'habitat et les variances associées aux relevés sont des indices de diversité. On obtient des aides à l'interprétation graphique très particulières.

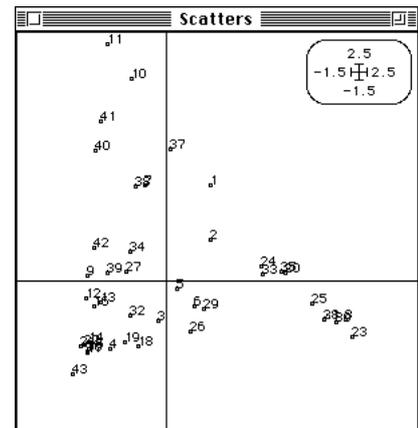
Les correspondances forment des cartes factorielles pondérées (Scatters : Values) :

Values	
HV coordinates file	<input type="text" value="H01_score"/> 295 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
G values file	<input type="text" value="H01_weigh"/> 295 1
Dot if G = 0 (yes = 1)	<input type="text"/>
Constrain H/U ratio (yes = 1)	<input type="text"/>
<input type="button" value="Quit"/> <input type="button" value="Copy graph"/> <input type="button" value="Save graph"/> <input type="button" value="Print graph"/> <input type="button" value="Draw"/>	

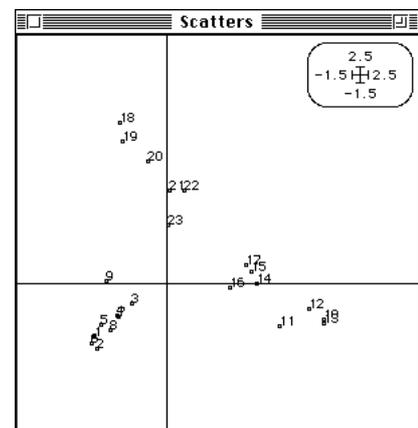


Représenter les cartes factorielles classiques :

Labels	
HV coordinates file	<input type="text" value="H01.fcli"/> 43 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="#"/>



Labels	
HV coordinates file	<input type="text" value="H01.fcco"/> 23 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="#"/>

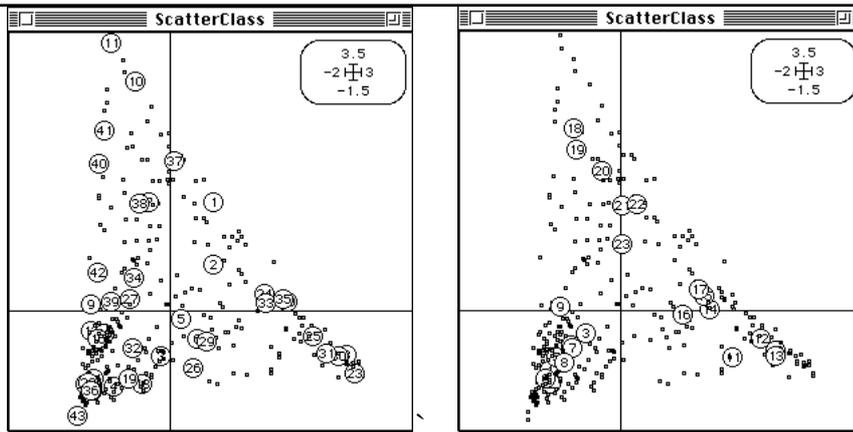


ScatterClass permet d'identifier les lignes et les colonnes comme ensemble de correspondances :

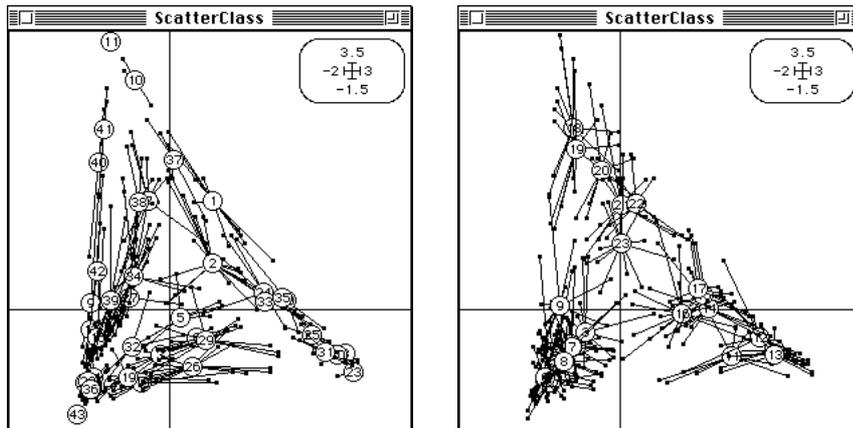
Labels	
HV coordinates file	<input type="text" value="H01_score"/> 295 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Categories file (.cat)	<input type="text" value="H01_class.cat"/>

Row & col. selection	
Col. selection:	<input type="text" value="1"/>
Row selection method:	<input type="radio"/> File <input checked="" type="radio"/> Keyboard
Graph number: 1	<input type="button" value="Validate & next graph"/>
Row selection:	<input type="text"/>
<input type="button" value="Draw"/>	

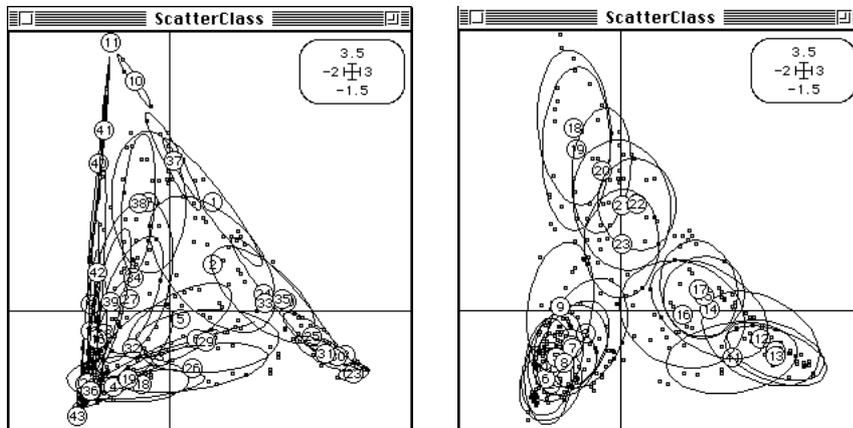
La sélection de colonnes dans ce module permet d'obtenir la carte des lignes (1) et la carte des colonnes (2) comme moyenne sur l'ensemble des correspondances :



L'option ScatterClass : Stars donne la représentation :



On peut utiliser encore l'option ScatterClass : Ellipses :



Noter les grandes différences entre les structures internes des nuages lignes et colonnes.



1 Thioulouse, J. & Chessel, D. (1992) A method for reciprocal scaling of species tolerance and sample diversity. *Ecology* : 73, 670-680.

2 Lebart, L., Morineau, L. & Warwick, K.M. (1984) *Multivariate descriptive analysis: correspondence and related techniques for large matrices*. John Wiley and Sons, New York. 1-231.

COA : Row weighted COA



Variante de l'analyse des correspondances.



L'analyse des correspondances est une analyse canonique¹ (voir [COA : Reciprocal scaling](#)). C'est son originalité radicale. Cela n'a pas que des avantages. Sur un tableau relevés-espèces, les lignes (relevés) porte un poids proportionnel à la somme des abondances des espèces qu'on y trouve. Cela impose aux méthodes de couplage qui utilise une AFC, en particulier l'analyse canonique des correspondances² (voir [Projectors : PCA on Instrumental Variables](#)), de calculer les moyennes et les variances des variables environnementales avec ces poids et donc de faire varier les paramètres environnementaux avec la faune qu'on y trouve. C'est une bonne chose si on pense environnement = environnement de la faune et une mauvaise si on pense environnement tout cours (en particulier pollution). Dans ce cas on peut utiliser une AFC à poids imposés qui conserve les propriétés de l'AFC et donne au poids des lignes une valeur désirée.

\mathbf{T} est la matrice de départ, n_{ij} la valeur du terme de la ligne i et de la colonne j , n la somme totale de tous les éléments du tableau, I le nombre de lignes et J le nombre de colonnes. Soient \mathbf{r} le vecteur de pondération imposée à I composantes, r_i la valeur de l'élément de rang i , R la somme totale de tous les éléments du vecteur. On note :

$$n_i = \sum_{j=1}^J n_{ij} \quad p_{j/i} = \frac{n_{ij}}{n_i} \quad q_i = \frac{r_i}{R} \quad t_j = \sum_{i=1}^I q_i p_{j/i}$$

L'AFC pondérée de \mathbf{T} est l'AFC ordinaire du tableau de terme général $r_i p_{j/i}$, qui redonne aux lignes les poids q_i et aux colonnes les poids t_j , c'est à dire l'analyse du schéma de dualité³ de paramètres :

1 — Tableau traité est \mathbf{Z} de terme général :

$$z_{ij} = \frac{p_{j/i}}{t_j} - 1 = \frac{p_{j/i} - t_j}{t_j}$$

2 — Pondération des lignes : chaque ligne a le poids q_i avec $\sum_{i=1, I} q_i = 1$;

3 — Pondération des colonnes : chaque colonne a le poids t_j avec $\sum_{j=1, J} t_j = 1$.



L'option utilise une seule fenêtre de dialogue :

Nom de fichier du tableau d'entrée.

Nom du fichier du tableau de pondération des lignes (par défaut, on introduit la pondération uniforme).

Nom générique des fichiers de sortie.



Utiliser la carte Elections de la pile ADE•Data pour obtenir ([TextToBin : Text->Binary](#)) les fichiers Res (40-9) et Vot (40-1).

fc/COA: Modified Correspondance analysis
Input file: Res

Number of rows: 40, columns: 9
 File New.fcpl contains the edge distribution of rows
 It has 40 rows and 1 column
 Weighting file: Vot
 File New.fcpc contains the edge distribution of columns
 It has 9 rows and 1 column
 File New.fcta contains the doubly centred table DI-1*P*DJ-1 -1I*1J'
 It has 40 rows and 9 columns
 File New contains the new table
 It has 40 rows and 9 columns
 File New.fcpa contains:
 the number of rows: 40
 the number of columns: 9
 the total number: 3.9982E+05

 DiagoRC: General program for two diagonal inner product analysis
 Input file: New.fcta
 --- Number of rows: 40, columns: 9

Total inertia: 0.0700367

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+3.5689E-02	+0.5096	+0.5096	02	+1.8317E-02	+0.2615	+0.7711
03	+9.7664E-03	+0.1394	+0.9106	04	+3.6985E-03	+0.0528	+0.9634
05	+1.4751E-03	+0.0211	+0.9844	06	+9.0337E-04	+0.0129	+0.9973
07	+1.2972E-04	+0.0019	+0.9992	08	+5.6897E-05	+0.0008	+1.0000
09	+0.0000E+00	+0.0000	+1.0000				

File New.fcvp contains the eigenvalues and relative inertia for each axis
 --- It has 9 rows and 2 columns

File New.fcco contains the column scores
 --- It has 9 rows and 2 columns

File :New.fcco

Col.	Mini	Maxi
1	-2.130e-01	3.963e-01
2	-1.912e-01	1.599e-01

File New.fcli contains the row scores
 --- It has 40 rows and 2 columns

File :New.fcli

Col.	Mini	Maxi
1	-2.203e-01	4.127e-01
2	-2.092e-01	2.592e-01



Cette analyse est équivalente à une AFC floue (MCA : Fuzzy Correspondance Analysis) quand on considère que les colonnes du tableau de départ sont les modalités d'une unique variable floue. Elle peut être considérée comme la version AFC de l'ACP sur profils lignes (PCA : After row % transformation PCA). Si on utilise une pondération uniforme, elle équivaut à l'AFC du tableau ramené en pourcentage par lignes (Bin->Bin : Frequencies).



¹ Lebart, L., Morineau, L. & Warwick, K.M. (1984) *Multivariate descriptive analysis: correspondence and related techniques for large matrices*. John Wiley and Sons, New York. 1-231.

² Ter Braak, C.J.F. (1986) Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 69, 69-77.

³ Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.

File Color.rcvp contains the eigenvalues and relative inertia for each axis

--- It has 5 rows and 2 columns

Les coordonnées des colonnes ne sont pas utiles dans cette option :

File Color.rcco contains the column scores

--- It has 5 rows and 2 columns

File :Color.rcco

Col.	Mini	Maxi
1	-1.468e-01	1.521e-01
2	-8.273e-02	4.719e-02

Les coordonnées des lignes sont exactement celles de l'AFC ordinaire (COA : CO Correspondence Analysis).

File Color.rcli contains the row scores

--- It has 4 rows and 2 columns

File :Color.rcli

Col.	Mini	Maxi
1	-4.403e-01	7.037e-01
2	-2.459e-01	1.658e-01

On interprète en utilisant les coordonnées normalisées de variance unité qui sont exactement celle qu'on obtiendrait avec COA : CO Correspondence Analysis puis DDUtil : Add normed scores :

File Color.rcc1 contains the column scores with unit norm (DJ scalar product)

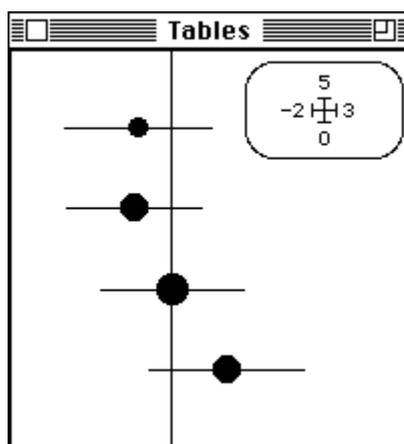
It has 5 rows and 2 columns

File :Color.rcc1

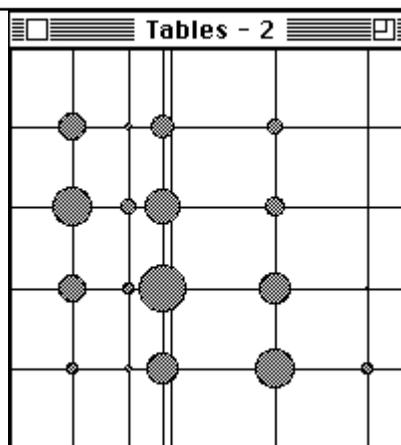
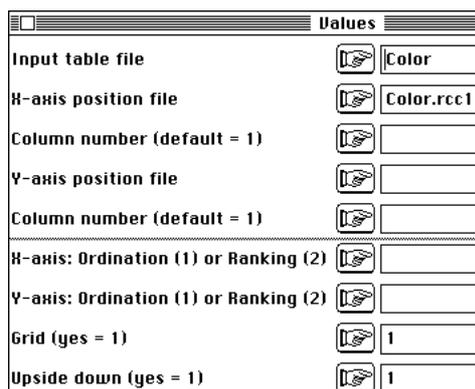
Col.	Mini	Maxi
1	-1.217e+00	2.471e+00
2	-1.198e+00	1.716e+00

On obtient un code des colonnes de variance unité qui maximise la variance des moyennes conditionnelles par lignes, ce qui est le modèle de l'analyse discriminante de Fisher¹ auquel Gower² attribue la naissance de l'analyse des correspondances :

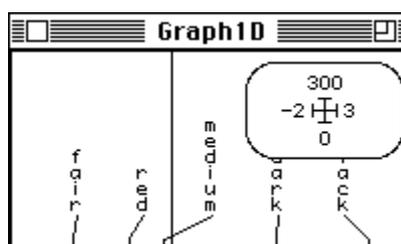
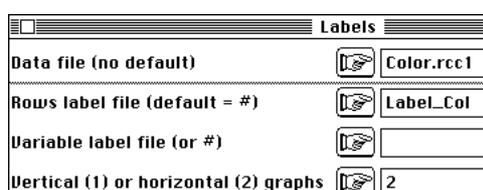
TabMeanVar	
Input table file	<input type="button" value="..."/> Color
H-axis position file	<input type="button" value="..."/> Color.rcc1
Column number (default = 1)	<input type="button" value="..."/> 1
Y-axis position file	<input type="button" value="..."/>
Column number (default = 1)	<input type="button" value="..."/>
H-axis: Ordination (1) or Ranking (2)	<input type="button" value="..."/>
Y-axis: Ordination (1) or Ranking (2)	<input type="button" value="..."/>
1 = Row distribution	<input type="button" value="..."/> 1
Upside down (yes = 1)	<input type="button" value="..."/> 1



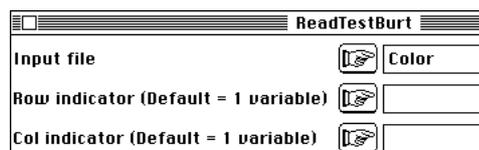
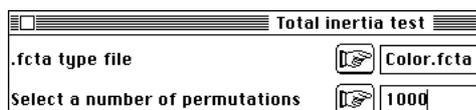
Assembler avec :



Labéliser avec Graph1D : Labels :



Voir cette pratique dans la fiche thématique 2-7 “Structures à une dimension”. Le résumé graphique est légitime car le test d’association (COA : Total inertia test) est très significatif comme le khi2 (CategVar : ReadTestBurt) :



Cette option a été implantée pour être utilisée avec le module Discrimin pour exécuter l’analyse discriminante des correspondances, méthode originale de ADE-4 décrite dans la fiche thématique 5-6 et utilisée dans ³.



L’analyse d’inertie des lignes de cette variante (DDUtil : Rows/Inertia analysis) est exactement la même que l’analyse d’inertie des lignes (DDUtil : Rows/Inertia analysis) de l’AFC ordinaire (COA : CORrespondence Analysis). L’analyse d’inertie des colonnes de cette variante n’a pas de sens. L’analyse d’inertie des colonnes de l’AFC ordinaire (COA : CORrespondence Analysis) est celle de l’analyse par cette variante du tableau transposé. (FilesUtil : Transpose). Le programme ordinaire effectue donc une double analyse d’inertie des profils lignes centrés ou des profils colonnes centrés. Voir la meilleure explication dans ⁴. Voir encore dans les messages sur le forum Adelist la question de Gilles CRAGUE (crague@enpc.enpc.fr, 28/02/97) et les réponses qu’elle a suscitées.



¹ Fisher, R.A. (1940) The precision of discriminant functions. *Annals of Eugenics* : 10, 422-438.

² Gower, J.C. (1990) Fisher's optimal scores and multiple correspondence analysis. *Biometrics* : 46, 947-961.

³ Perrière, G., Lobry, J.R. & Thioulouse, J. (1996) Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acid sequences. *CABIOS* : 12, 519-524.

⁴ Lebart, L., Morineau, A. & Piron, M. (1995) *Statistique exploratoire multidimensionnelle*. Dunod, Paris. 1-439.

COA : Total inertia test



Test de permutation sur une table de contingence permettant un test du type Khi2 dans toutes les circonstances numériques.



Une table de contingence est la réécriture compacte d'un couple de variables qualitatives :

	A	B	C
1	3	1	0
2	0	1	0
3	0	1	2

1	1
1	1
1	1
2	2
3	2
3	3
3	3

En permutant au hasard les valeurs de la seconde colonne, on engendre une table de contingence ayant les mêmes marges que la première. On peut tester l'indépendance des deux variables qualitatives en utilisant l'inertie totale (proportionnelle au Khi2) de l'AFC comme mesure de liaison entre les deux variables.



L'option utilise une seule fenêtre de dialogue :

Nom du fichier ---.fcta qui renvoie à l'AFC d'un tableau.

Nombre de permutations.



Dialogues intermédiaires



Particularités



Utiliser la carte Chats de la pile ADE•Data pour obtenir le fichier ChaTab (7-8). Exécuter l'AFC de cette table de contingence :

Utiliser la présente option qui affiche :

```

number of random matching: 1000   Observed: 0.192598
Histogram:  minimum = 0.054612, maximum = 0.259525
number of simulation X<Obs: 988 (frequency: 0.988000)
number of simulation X>=Obs: 12 (frequency: 0.012000)

*
****
*****
*****
*****
*****
*****
*****
*****
*****
*****
***
***
●-> *
*

```

Environ 1% des permutations aléatoires fournissent une inertie totale supérieure à l'observation. Comparer avec le test Khi2 (CategVar : ReadTestBurt) :

ReadTestBurt

Input file 7 8

Row indicator (Default = 1 variable)

Col indicator (Default = 1 variable)

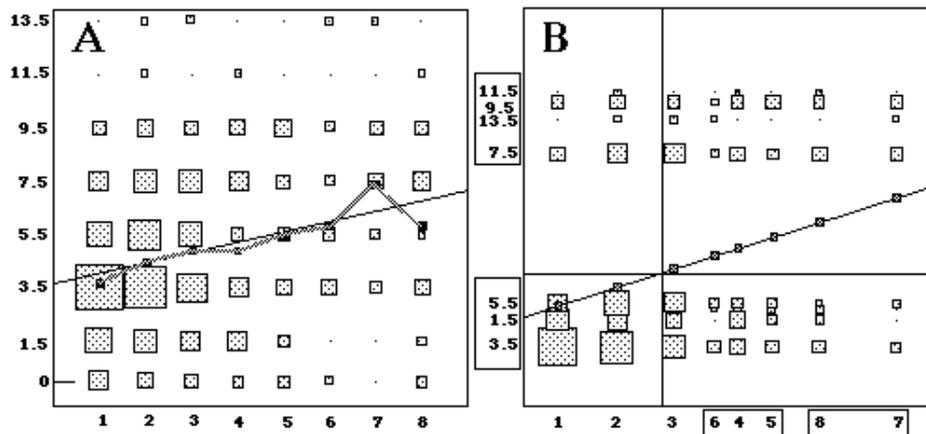
 Khi2 tests on Burt's crossed table ChaTab

 Var 1 versus var 1:Khi2 = 6.7409e+01 ddl = 42 Proba = 0.007736

On obtient un seuil de signification voisin, bien qu'en principe les conditions d'utilisation du test ne soient pas remplies (robustesse).



L'AFC de cette table de contingence est légitime et son dépouillement est un bon exercice :



Pour 350 chattes¹ on connaît l'âge (1 an à 8 ans et plus) et le nombre de chatons produits dans l'année 1 ou 2 (1.5), 3 ou 4 (3.5), ..., 13 ou 14 (13.5). La liaison entre les 2 variables est exprimée par le nuage bivarié élémentaire (A : surface des carrés proportionnelle au nombre d'observations), la courbe de régression (rapport de corrélation de 0.089) et la droite de régression (carré de corrélation de 0.073). On y "voit" une liaison linéaire comme dans un cas d'école de statistique descriptive. L'AFC de la table de contingence qui ne tient aucun compte de l'origine quantitative des données recode les lignes et les colonnes du tableau en maximisant la corrélation observée et, pour ce faire, rend linéaire la liaison entre les codes lignes et colonnes² : Les modalités associées à l'âge sont ordonnées sur 5 niveaux (1, 2, 3, 4 à 6, 7 et plus) et celles qui concernent la fécondité sont groupées en deux classes (1 à 6, 7 à 14). Le rapport de corrélation, le carré de corrélation et la première valeur propre valent 0.114. La liaison se fait d'abord par le nombre de portées. (extrait de ³).



¹Legay, J.M. & Pontier, D. (1985) Relation âge-fécondité dans les populations de Chats domestiques, *Felis catus*. *Mammalia* : 49, 395-402.

²Hirschfeld, H.O. (1935) A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences* : 31, 520-524.

³Chessel, D. & Thioulouse, J. (1990) Auto-modélisation en analyse des données. In : *Modélisation confluent des sciences*. Brissaud, M., Forsé, M. & Zighed, A. (Eds.) Editions du CNRS. 71-86.

