

# DDUtil

DDUtil : Add normed scores.....	2
DDUtil : Column projections.....	7
DDUtil : Columns/Inertia analysis.....	10
DDUtil : Data modelling.....	13
DDUtil : Diagonalization.....	16
DDUtil : Residuals.....	18
DDUtil : Row projections.....	20
DDUtil : Rows/Inertia analysis.....	22
DDUtil : Supplementary columns.....	24
DDUtil : Supplementary rows.....	26
DDUtil : Triplet Transpose.....	29

## DDUtil : Add normed scores



Utilitaire d'interprétation d'analyse multivariée à un tableau.



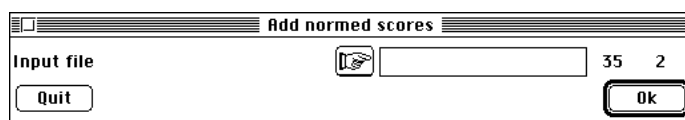
L'option sauvegarde les coordonnées dites "de variance unité" en plus des coordonnées dites "de variances égales aux valeurs propres". Cette opération a un sens qui peut varier suivant les analyses. Pour une analyse dont les colonnes sont centrées (ACP classiques, AFC, ACM) les coordonnées des lignes sont centrées et de variances égales aux valeurs propres. On peut avoir l'usage des coordonnées correspondantes de variance unité qui sont alors les composantes des composantes principales dans l'espace des colonnes.

Lorsque les coordonnées ne sont pas centrées, par exemple celles des colonnes dans une ACP normée, l'option donne les coordonnées standardisées (de norme unité) mais non centrées. Si la pondération associée est unitaire, comme en ACP normée, on obtient les coordonnées des projections des vecteurs de la base canonique sur la base des axes principaux, qui sont aussi les poids canoniques.

Les programmes anglo-saxons donnent souvent ces coordonnées de norme 1 en place des coordonnées de norme la valeur propre associée. On privilégie ainsi les vecteurs sur lesquels on projette par rapport aux coordonnées des projections.



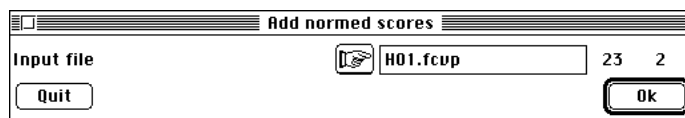
L'option utilise une seule fenêtre de dialogue :



Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.



L'opération a un sens spécifique en AFC. Exécuter l'AFC du fichier H01 (43-23) issu de la carte Forêts de la pile ADE•Data (Voir la fiche de COA : Reciprocal scaling).



File H01.fc11 contains the row scores with unit norm

It has 43 rows and 2 columns

File :H01.fc11

Col.	Mini	Maxi
1	-1.216e+00	2.439e+00
2	-1.457e+00	3.712e+00

File H01.fccl contains the column scores with unit norm

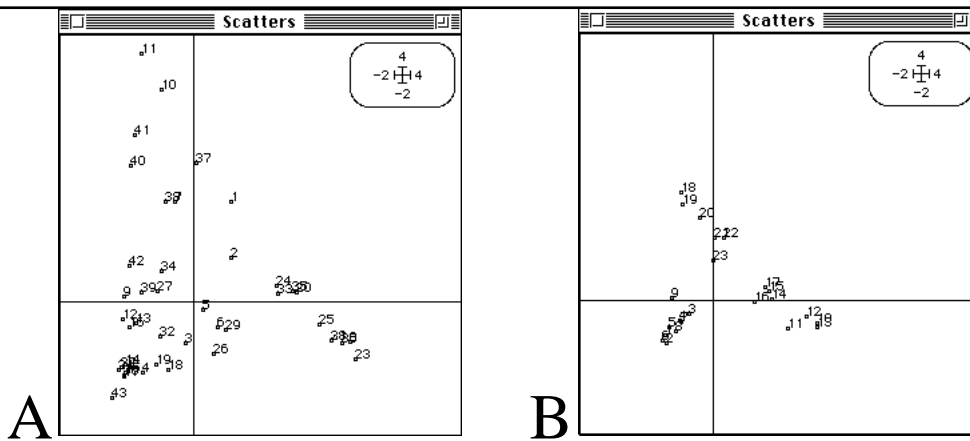
It has 23 rows and 2 columns

File :H01.fccl

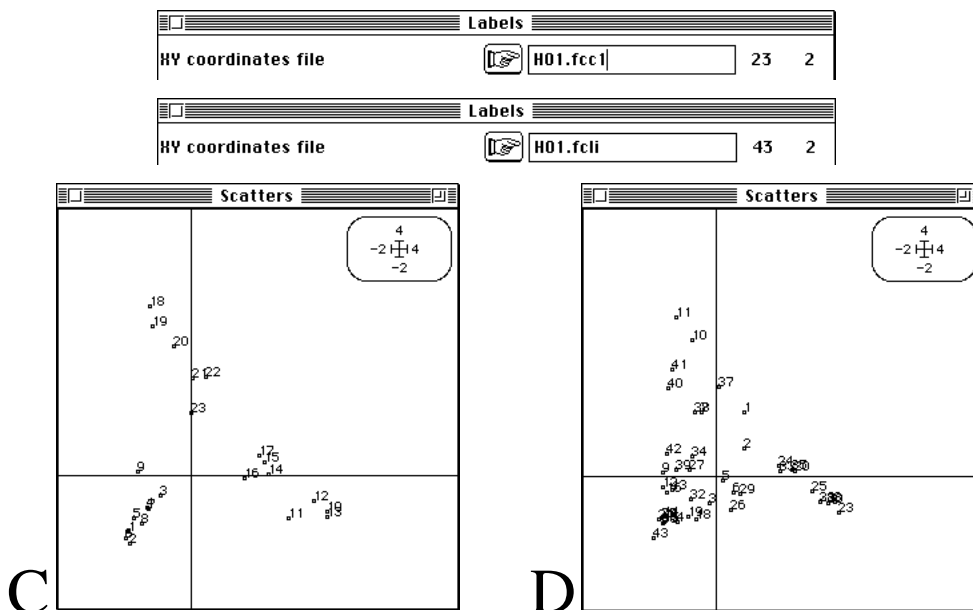
Col.	Mini	Maxi
1	-9.775e-01	2.054e+00
2	-1.023e+00	2.520e+00

Dans **Scatters : Labels** positionner les lignes avec les coordonnées de variance unité et les colonnes avec les coordonnées ordinaires :



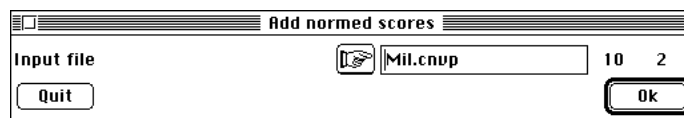


Inverser et utiliser les coordonnées des colonnes de variance unité et les coordonnées des lignes ordinaires :



Dans le premier cas on obtient une espèce à la moyenne des relevés qu'elle occupe. Dans le second cas, on obtient un relevé à la moyenne des espèces qu'il contient. On ne peut pas faire les deux à la fois. En utilisant les deux cartes ordinaires superposées, il faut donc penser à la dilatation implicite. Ces modes de représentation sont discutés à une dimension<sup>1</sup> et à deux dimensions<sup>2</sup>.

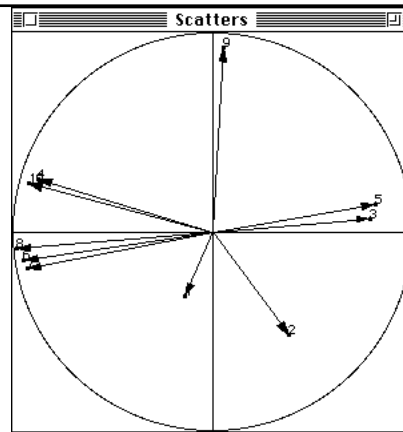
L'option a également un sens spécifique en ACP normée. Exécuter l'ACP du fichier Mil (24-10) issu de la carte Méaudret de la pile ADE•Data (Voir la fiche de PCA : Correlation matrix PCA).



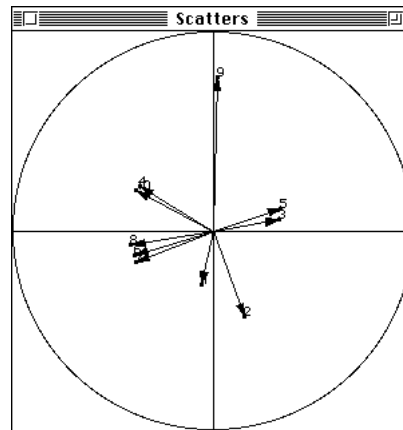
Les coordonnées ordinaires donnent un cercle de corrélation. Il s'agit de la projection des colonnes (vecteurs à 24 composantes de longueur unité) sur un plan défini par deux composantes principales, projection accompagné du cercle de rayon 1, intersection de la sphère unité avec ce plan.

Les coordonnées standardisées donnent une image voisine qui a un sens très différents. Il s'agit de la projection des vecteurs de la base canonique de l'espace des individus (vecteurs à 10 composantes de longueur unité) sur un plan défini par deux axes principaux, projection accompagné du cercle de rayon 1, intersection de la sphère unité avec ce plan.

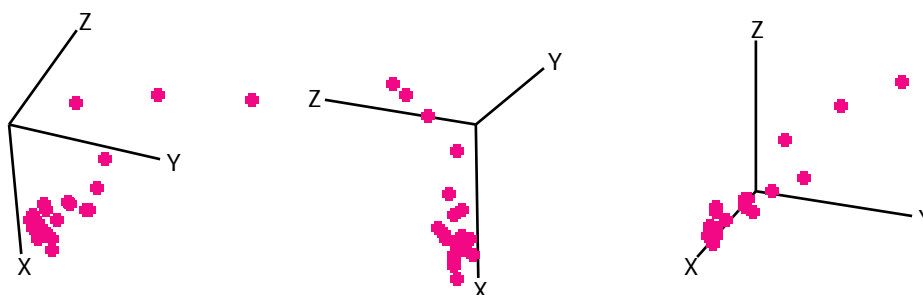
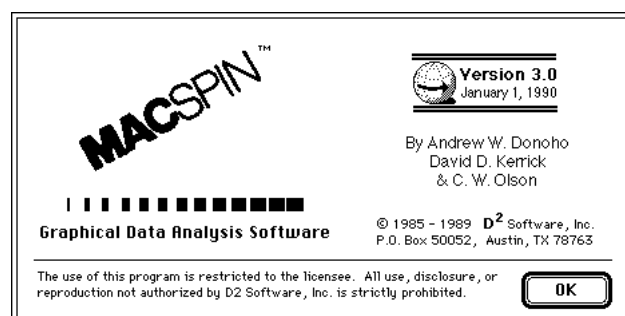
Labels	
HV coordinates file	<input type="text" value="Mil.cnco"/> 10 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="#"/>



Labels	
HV coordinates file	<input type="text" value="Mil.cnc1"/> 10 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="#"/>



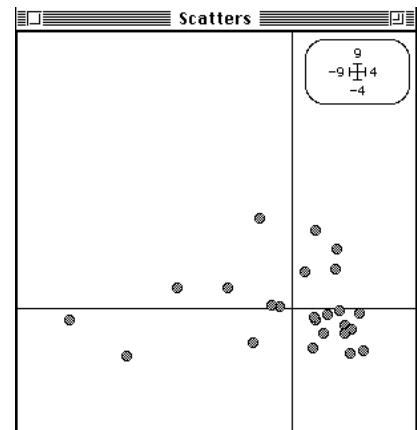
L'intérêt de la première représentation est de montrer les corrélations entre variables et coordonnées des lignes. L'intérêt de la seconde est d'être superposable avec la carte des individus dans une représentation Biplot<sup>3</sup>. On peut dans cette optique superposer les coordonnées ordinaires des lignes et les coordonnées normées des colonnes ou l'inverse. Cette opération est à concevoir comme l'obtention automatique d'une position qu'on obtient manuellement en faisant "tourner" des nuages de points à trois dimensions, comme dans MacSpin<sup>TM</sup> :



Il est clair que cette manière d'aborder les données perd tout son sens dès qu'il y a plus de trois variables. L'ACP propose une position optimale dans son plan 1-2.

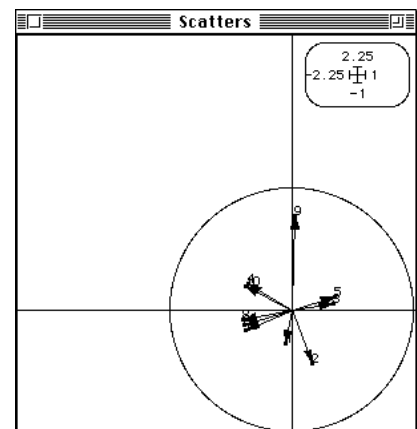
Placer les points par Scatters : Values :

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

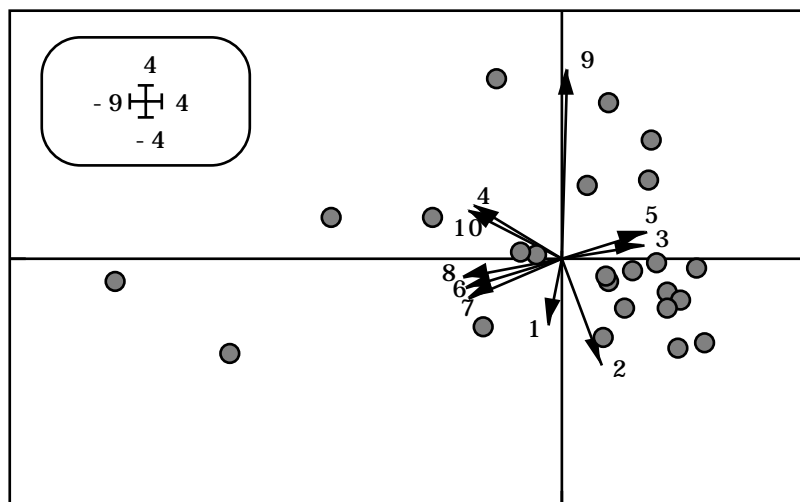


Placer les variables par Scatters : Labels :

Labels			
HV coordinates file		Mil.cnc1	10 2
H-axis column number (default = 1)			
V-axis column number (default = 2)			
Label file (or # for item numbers)		#	
Draw vectors from origin (yes = 1)		1	
Draw unit circle (yes = 1)		1	



Conserver les échelles relatives. Ici on décide de représenter les vecteurs de la base canonique multipliés par 4. Superposer dans MacDraw de Claris™ :



L'ACP peut être ainsi comprise comme un "MacSpin automatisé" en dimension quelconque.



Voir aussi l'usage de scores standardisés dans PCA : Decentring X[i,j] - Model[j]. Cette option est automatiquement mise en œuvre dans PCA : After row % transformation PCA.



<sup>1</sup> Chessel, D., Lebreton, J.D. & Prodon, R. (1982) Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés: cas d'un gradient simple. *Compte rendu hebdomadaire des séances de l'Académie des sciences. Paris, D : III, 295, 83-88.*

<sup>2</sup> Oksanen, J. (1987) Problems of joint display of species and site scores in correspondence analysis. *Vegetatio* : 72, 51-57.

<sup>3</sup> Gabriel, K.R. (1971) The biplot graphical display of matrices with application to principal component analysis. *Biometrika* : 58, 453-467.

Gabriel, K.R. (1972) Analysis of meteorological data by means of canonical decomposition and biplots. *Journal of applied meteorology* : 11, 1071-1077.

Bradu, D. & Gabriel, K.R. (1978) The biplot as a diagnostic tool for models of two-way tables. *Technometrics* : 20, 1, 47-68.

Gabriel, K.R. (1981) Biplot display of multivariate matrices for inspection of data and diagnosis. In : *Interpreting multivariate data*. Barnett, V. (Ed.) John Wiley and Sons, New York. 147-174.

Noter que dans ces travaux le biplot est considéré du point de vue reconstitution de données (voir DDUtil : Data Modelling) plutôt que du point de vue géométrique. Il s'applique à tous les types de centrage (voir HTA). Références connexes :

Ter Braak, C.J.F. (1983) Principal components biplots and alpha and beta diversity. *Ecology* : 64, 3, 454-462.

Ter Braak, C.J.F. (1990) Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika* : 55, 519-531.

Hermly, M. & Lewi, P.J. (1991) Multivariate ratio analysis, a graphical method for ecological ordination. *Ecology* : 72, 735-738.

Glasbey, C.A. (1992) A reduced rank regression model for local variation in solar radiation. *Applied Statistics* : 41, 381-387.

Rivas-Gonzalo, J.C. & Coll. (1993) Biplot analysis applied to enological parameters in the geographical classification of young red wines. *Am. J. Enol. Vitic* : 44, 302-308.

Génard, M., Souty, M., Holmes, S., Reich, M. & Breuils, L. (1994) Correlation among quality parameters of peach fruit. *Journal of the science of food and agriculture* : 66, 241-245.

## DDUtil : Column projections



Utilitaire d'interprétation d'analyse multivariée à un tableau.



La projection de lignes ou de colonnes supplémentaires est une pratique courante en analyse des données. Il ne faut pas confondre projection d'éléments supplémentaires (opération géométrique) et représentation sur un plan factoriel d'information supplémentaire (opération pratique qui peut prendre des formes extrêmement variées). Le module projette sur le sous-espace des composantes principales d'une analyse quelconque des vecteurs colonnes de même dimension que les colonnes du tableau traité **sans leur faire subir de transformation préalable**. La projection se fait au sens de la métrique (poids des lignes) de l'analyse de départ.



L'option utilise une seule fenêtre de dialogue :

Column projections			
Initial analysis (up)	<input type="text" value="DouMil.cnup"/>	11	2
Supplementary column file	<input type="text" value="DouPoi0"/>	30	27
Output file name	<input type="text" value="Sup"/>		
<input type="button" value="Quit"/>		<input type="button" value="Ok"/>	

Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.

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

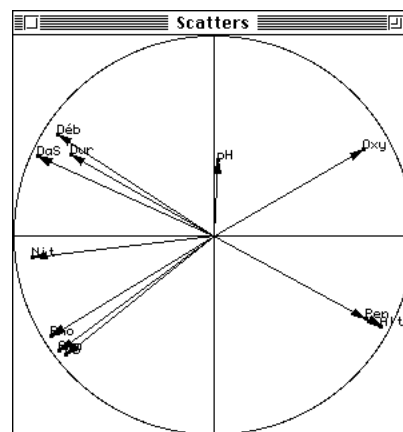
Nom du fichier de sortie. Il contient en lignes les colonnes du fichier précédent et en colonnes les facteurs de l'analyse de base.



Utiliser la carte Doubs de la pile ADE•Data pour obtenir le fichier DouMil (30-11) et le fichier VarMil\_label. Faire l'ACP normée de DouMil :

Correlation matrix PCA			
Matrix input file	<input type="text" value="DouMil"/>	30	11

Labels			
HV coordinates file	<input type="text" value="DouMil.cnco"/>	11	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="VarMil_label"/>		
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>		
Draw unit circle (yes = 1)	<input type="text" value="1"/>		



Utiliser la carte Doubs+1 de la pile ADE•Data pour obtenir le fichier DouPoi (30-27)

Char->Binary			
Text input file	<input type="text" value="DouPoi.car"/>		
Binary output file	<input type="text" value="DouPoi"/>		

et le fichier Poi\_label. On peut représenter en colonnes supplémentaires, au sens traditionnel du terme, les espèces par DDUtil : Supplementary columns :





Les colonnes supplémentaires ne sont jamais dans le tableau des colonnes actives.



En autorisant la projection de vecteurs quelconques, on permet l'usage des colonnes supplémentaires pour tous les types d'analyse (inter-intra, co-inertie, MFA, ...). Il convient de maîtriser la théorie pour s'en servir correctement.

Quand les colonnes supplémentaires doivent subir la même transformation préalable que les colonnes du tableau analysées, utiliser DDUtil : Supplementary columns. C'est la version ordinaire de l'emploi des colonnes supplémentaires.

## DDUtil : Columns/Inertia analysis



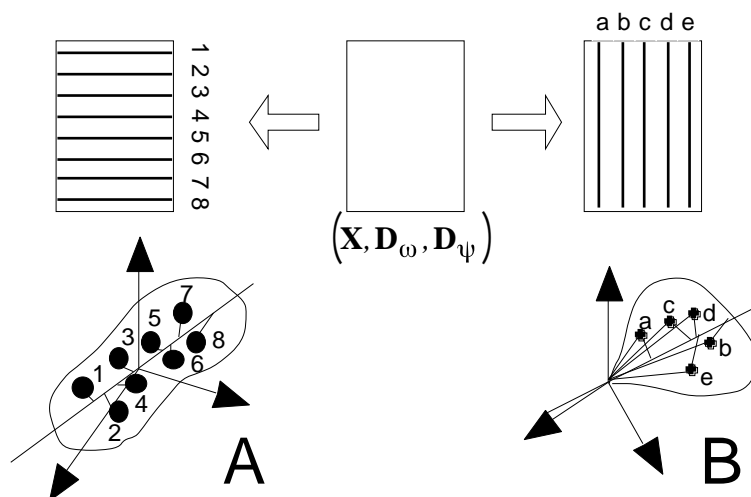
Utilitaire d'interprétation d'analyse multivariée à un tableau.



L'option édite le listing des statistiques d'inertie du nuage des colonnes.

Soient les deux nuages de points définis dualement par un seul triplet  $(\mathbf{X}, \mathbf{D}_\omega, \mathbf{D}_\psi)$ .  $\mathbf{X}$  est une matrice à  $n$  lignes et  $p$  colonnes et les produits scalaires associés aux pondérations définissent les matrices :

$$\mathbf{D}_\omega = \text{Diag}(\omega_1, \dots, \omega_p) \text{ et } \mathbf{D}_\psi = \text{Diag}(\psi_1, \dots, \psi_n).$$



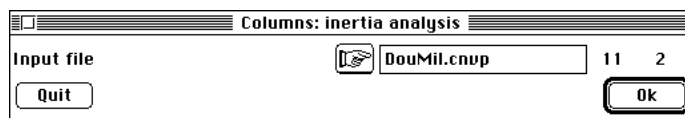
La somme des valeurs propres égale l'inertie totale  $I_T$  commune aux deux nuages. En ACP centrée cette inertie est la somme des variances. En ACP normé cette inertie vaut le nombre de variables. La quantité  $\lambda_k/I_T$  est l'inertie relative du vecteur principal de rang  $k$ .

L'inertie des projections sur la composante principale  $k$  se décompose en somme de contribution absolue (CA) de la variable  $j$  à la définition de l'axe  $k$ .

L'inertie totale se décompose en contributions à la trace des lignes et des colonnes. Le carré de la norme (variance en ACP centrée ou normée) d'une variable  $j$  se décompose entre contribution relative (CR) des composantes à la représentation de la ligne  $i$ . Les contributions relatives sont des carrés de cosinus. Les définitions utilisées sont celles qui ont été définies dans le cas particulier de l'analyse des correspondances par <sup>1</sup> (pp. 316-317).



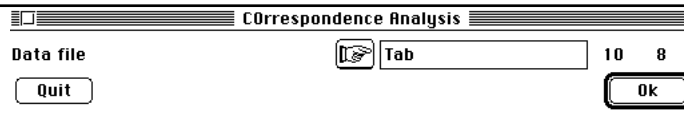
L'option utilise une seule fenêtre de dialogue :



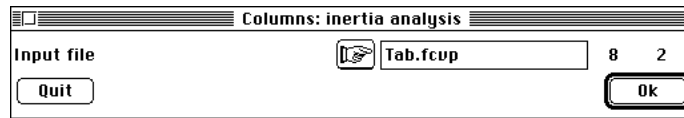
Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.



Utiliser la carte Hébergement de la pile ADE•Data pour obtenir le fichier Tab (10-8) reproduit dans <sup>1</sup> (p. 306). Exécuter l'AFC du tableau :



Utiliser le présent module :



Les résultats sont ceux de <sup>1</sup> (p. 319). Les contributions sont exprimées par des entiers en partie pour dix milles (3572 signifie 35.72 %).

Input file: Tab.fcta  
Number of rows: 10, columns: 8

Inertia: Two diagonal norm inertia analysis  
Total inertia: 0.0836022 - Number of axes: 3

File Tab.fccc contains the contribution of columns to the trace  
It has 8 rows and 1 column

Column inertia  
All contributions are in 1/10000

-----Absolute contributions-----

Num	Fac 1	Fac 2	Fac 3
1	3572	2197	1251
2	147	1024	439
3	2693	610	5626
4	375	4174	691
5	171	2	907
6	2543	1212	1028
7	440	561	11
8	54	217	45

-----Relative contributions-----

Num	Fac 1	Fac 2	Fac 3	Remains	Weight	Cont.
1	7290	2049	507	153	1459	2597
2	1162	3680	686	4471	1086	674
3	6547	678	2717	57	718	2180
4	1531	7780	560	128	4112	1299
5	2017	11	2117	5853	597	451
6	7238	1576	581	604	1240	1862
7	4713	2742	24	2518	429	495
8	661	1202	109	8026	355	438

Ces statistiques d'inertie s'étendent à tout type d'analyse à un tableau.

Utiliser la carte Doubs de la pile ADE•Data pour obtenir le fichier DouMil (30-11). Faire l'ACP normée de DouMil et utiliser la présente option :



Input file: DouMil.cnta  
Number of rows: 30, columns: 11

Inertia: Two diagonal norm inertia analysis  
Total inertia: 11 - Number of axes: 2

File DouMil.cncc contains the contribution of columns to the trace  
It has 11 rows and 1 column

Column inertia  
All contributions are in 1/10000

-----Absolute contributions-----

Num	Fac 1	Fac 2
1	1206	703
2	1113	922
3	914	770
4	956	1137
5	0	632
6	802	720
7	1040	1135
8	1285	51
9	930	1500
10	887	820
11	860	1605

-----Relative contributions-----

Num	Fac 1	Fac 2	Remains	Weight	Cont.
1	7627	1570	801	10000	909
2	7040	2059	900	10000	909
3	5779	1718	2502	10000	909
4	6049	2537	1413	10000	909
5	6	1410	8583	10000	909
6	5073	1608	3317	10000	909
7	6578	2533	887	10000	909
8	8127	113	1758	10000	909
9	5884	3347	767	10000	909
10	5609	1832	2557	10000	909
11	5439	3582	978	10000	909



Les statistiques d'inertie sont importantes dans les analyses à pondérations non uniformes (comme en AFC). Pour les analyses à pondérations uniformes, elles sont largement redondantes avec les cartes factorielles. En ACP, on peut y lire la décomposition des variances des variables transformées.



<sup>1</sup> Lebart, L., Morineau, A. & Fenelon, J.P. (1982) *Traitement des données statistiques. Méthodes et Programmes*. Dunod, 2<sup>o</sup> édition, Paris. 1-518.

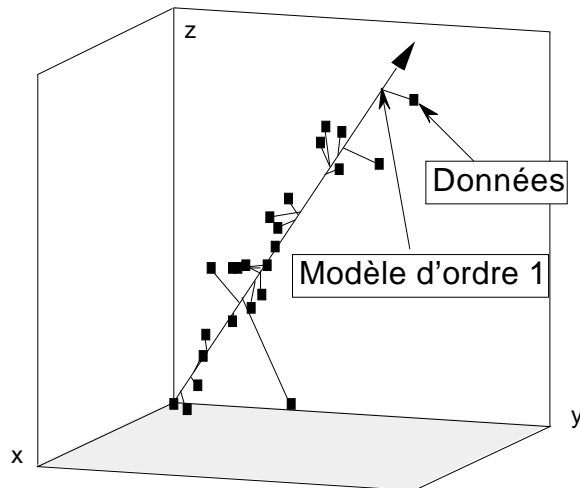
## DDUtil : Data modelling



Utilitaire d'interprétation d'analyse multivariée à un tableau.



L'option met en œuvre la formule de reconstitution des données après l'analyse d'un triplet statistique. On obtient le modèle de rang 1 en substituant au point de mesure sa projection sur le premier axe. On obtient le modèle d'ordre 2 en substituant au point de mesure sa projection sur le plan factoriel 1-2. On obtient le modèle d'ordre  $k$  en substituant au point de mesure sa projection sur le sous-espace défini par les  $k$  premiers axes.



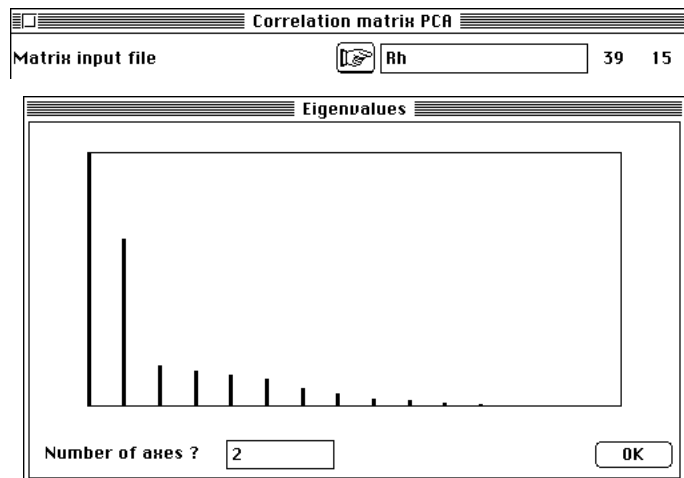
L'option utilise une seule fenêtre de dialogue :



Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.



Utiliser la carte Rhône de la pile ADE•Data pour obtenir le fichier Rh (39-15)<sup>1</sup>. Faire l'ACP normée de Rh :



Lancer la présente option :

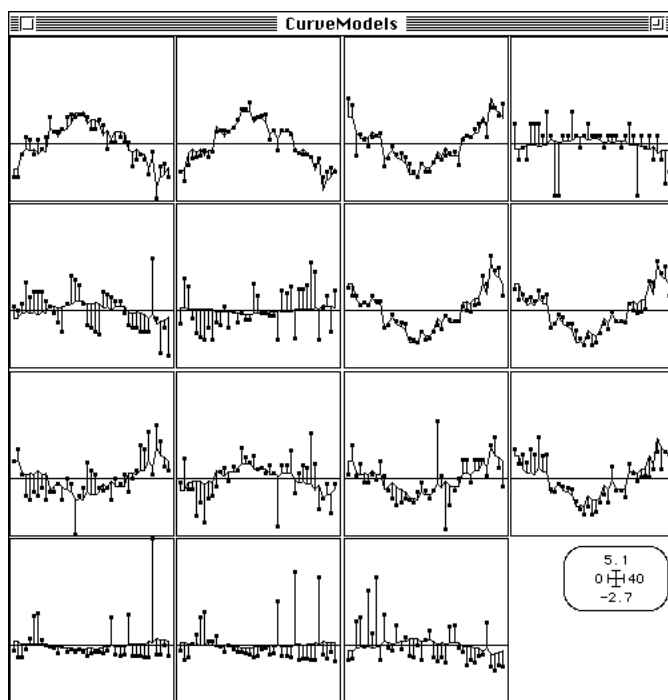
Modelling: Data reconstitution after PCA/CA  
Title of the analysis: Rh.cnta  
Number of rows: 39, columns: 15

File Rh.cnrA contains the model computed with 1 axe  
 It has 39 rows and 15 columns

File Rh.cnrB contains the model computed with 2 axes  
 It has 39 rows and 15 columns

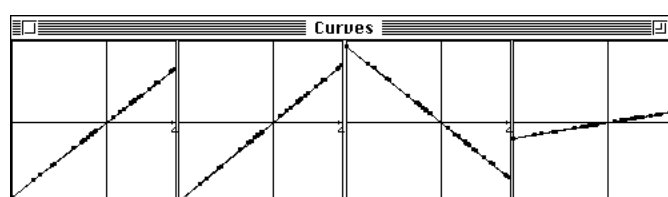
Représenter<sup>2</sup> les données (chronique multivariée) et le modèle de rang 1 par  
 CurveModels : Numerical :

Numerical	
H file (default = 1, 2, 3, ..., n)	<input type="text"/>
H file column number (default = 1)	<input type="text"/>
Model values file (no default)	<input type="text" value="Rh.cnrA"/> 39 15
Data values file (no default)	<input type="text" value="Rh.cnta"/> 39 15
Variable label file (optional)	<input type="text"/>
Draw model (1=yes, 2=no)	<input type="text" value="1"/>
Draw observed points (1=yes, 2=no)	<input type="text" value="1"/>
Draw residual sticks (1=yes, 2=no)	<input type="text" value="1"/>

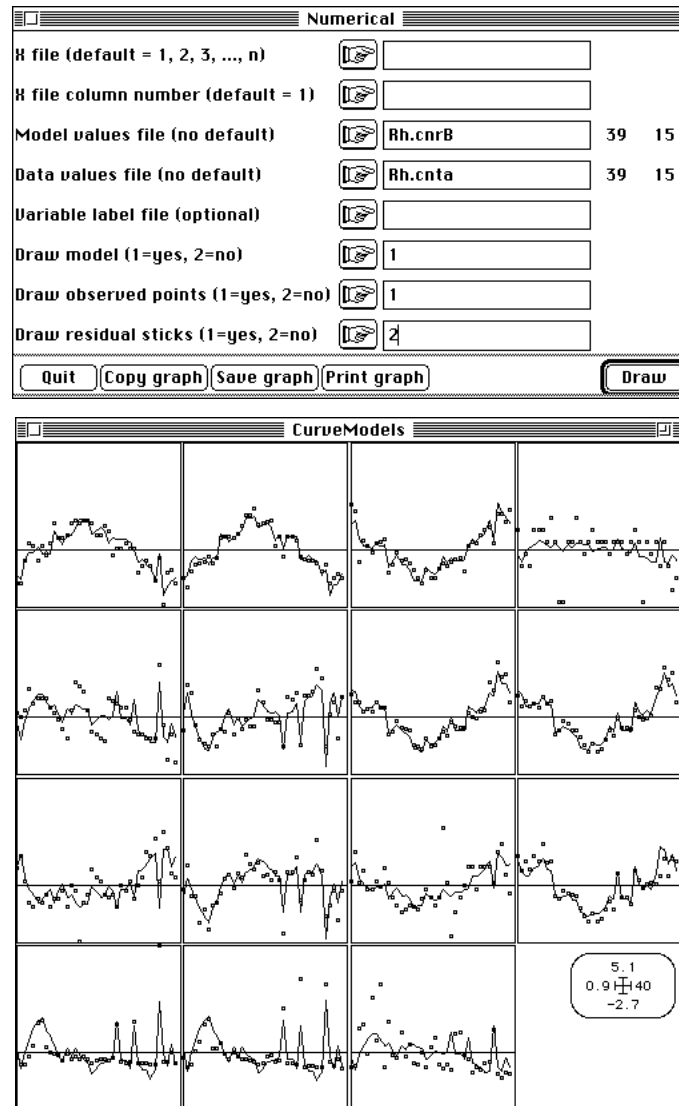


Toutes les courbes sont, à une constante multiplicative près par fenêtre, les mêmes et reproduisent la courbe fournie par la coordonnée sur l'axe 1 de l'ACP (effet saisonnier). Ceci est explicité dans la représentation coordonné (en abscisse) modèle (en ordonnée) dans Curves : Lines :

Lines	
H file (default = 1, 2, 3, ..., n)	<input type="text" value="Rh.cnli"/> 39 2
H file column number (default = 1)	<input type="text"/>
Y file (no default)	<input type="text" value="Rh.cnrA"/> 39 15



La reconstitution avec le modèle de rang 2 ajoute au modèle précédent une courbe unique (à une constante multiplicative près par variables) :



La formule de reconstitution des données permet de substituer aux observations un *simulateur*, c'est-à-dire un modèle<sup>3</sup>, utilisant les coordonnées des lignes, les coordonnées des colonnes et les valeurs propres (décomposition en valeurs singulières<sup>4</sup>).



Voir aussi les références de HTA : Edit inertia.



<sup>1</sup> Carrel, G. (1986) *Caractérisation physico-chimique du Haut-Rhône français et de ses annexes : incidences sur la croissance des populations d'alevins*. Thèse de doctorat. Université Lyon 1. 1-186.

<sup>2</sup> Carrel, G., Barthelemy, D., Auda, Y. & Chessel, D. (1986) Approche graphique de l'analyse en composantes principales normée : utilisation en hydrobiologie. *Acta Œcologica, Œcologia Generalis* : 7, 2, 189-203.

<sup>3</sup> Legay, J.M. (1973) *La méthode des modèles, état actuel de la méthode expérimentale*. Informatique et Biosphère, Paris. 11-69.

<sup>4</sup> Good, I.J. (1969) Some applications of the singular decomposition of a matrix. *Technometrics* : 11, 823-831.

# DDUtil : Diagonalization



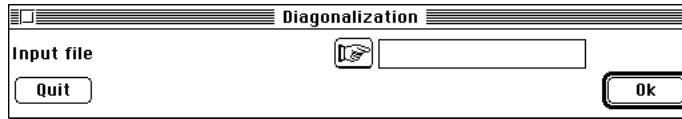
Outil d'analyse linéaire des données.



L'option est la partie commune de toutes les analyses à un tableau. Elle permet de diagonaliser un schéma de dualité entièrement préparé "à la main".



L'option utilise une seule fenêtre de dialogue :



Nom de fichier du type `---.##ta` comportant  $n$  lignes et  $p$  colonnes. Ces paramètres sont quelconques. Le dossier de travail doit contenir un fichier binaire de nom `---.##pl` ( $n$  lignes et 1 colonne de nombres positifs) et un fichier binaire de nom `---.##pc` ( $p$  lignes et une colonne de nombres positifs). Le tout forme un triplet statistique associant un tableau, une pondération des lignes et une pondération des colonnes. L'option fait l'analyse de ce schéma de dualité particulier. La chaîne `##` définit une nouvelle analyse.



Utiliser la carte Rhône de la pile ADE•Data pour obtenir le fichier Rh (39-15)

Faire l'ACP normée de Rh :

- Rh
- Rh.cnco
- Rh.cnli
- Rh.cnma
- Rh.cnpa
- Rh.cnpc
- Rh.cnpl
- Rh.cnta
- Rh.cnvp



Le dossier de travail contient les fichiers Rh.cnta (tableau centré et réduit par colonnes); Rh.cnpl (poids des lignes), Rh.cnpc (poids des colonnes). Détruire les autres et changer l'étiquette cn en XX. Il reste :

- Rh.XXpc
- Rh.XXpl
- Rh.XXta



```
DiagoRC: General program for two diagonal inner product analysis
Input file: Rh.XXta
--- Number of rows: 39, columns: 15
-----
Total inertia:      15
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.2743E+00 +0.4183 +0.4183 | 02  +4.1409E+00 +0.2761 +0.6943 |
03  +1.0082E+00 +0.0672 +0.7616 | 04  +8.5920E-01 +0.0573 +0.8188 |
...
File Rh.XXvp contains the eigenvalues and relative inertia for each axis
--- It has 15 rows and 2 columns

File Rh.XXco contains the column scores
--- It has 15 rows and 2 columns
File :Rh.XXco
| Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -9.673e-01 | 9.650e-01 |
| 2 | -8.543e-01 | 8.677e-01 |
|-----|-----|-----|
```

File Rh.XXli contains the row scores  
--- It has 39 rows and 2 columns  
File :Rh.XXli

Col.	Mini	Maxi
1	-5.760e+00	4.141e+00
2	-5.741e+00	3.354e+00

On retrouve les résultats de l'ACP.



Les modules PCA, COA, MCA et HTA utilisent cette procédure. On peut construire ainsi des “mélanges d’analyses” par reconstitution d’un triplet quelconque.



La diagonalisation est faite dans la plus petite des deux dimensions  $n$  ou  $p$ .



Le schéma de dualité a été inventé par P. CAZES<sup>1</sup> (1970), diffusé par CAILLIEZ & PAGES<sup>2</sup> (1976) et étendu par RAMSAY (1982)<sup>3</sup> :

*“It can be considered a fundamental algebraic advance over matrix analysis, and it is hoped that it will become a standard part of statistical language.”*

Il a été introduit en écologie statistique par ESCOUFIER (1987)<sup>4</sup> comme moyen d'adapter des abstractions mathématiques à des situations concrètes.

<sup>1</sup> Cazes, P. (1970) *Application de l'analyse des données au traitement de problèmes géologiques*. Thèse de 3<sup>o</sup> cycle, Faculté des Sciences de Paris. 1-132 + annexes.

<sup>2</sup> Cailliez, F. & Pages, J.P. (1976) *Introduction à l'analyse des données*. SMASH, 9 rue Duban, 75016 Paris. 1-616.

<sup>3</sup> Ramsay, J.O. (1982) When the data are functions. *Psychometrika* : 47, 4, 379-396.

<sup>4</sup> 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.

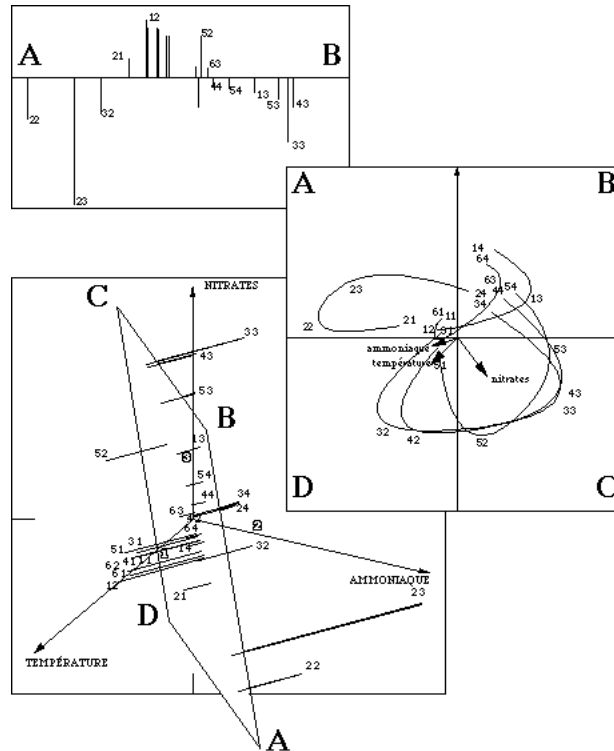
# DDUtil : Residuals



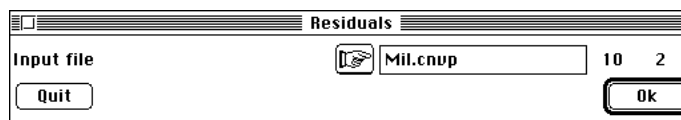
Utilitaire d'interprétation d'analyse multivariée à un tableau.



L'option met en œuvre la formule de reconstitution des données après l'analyse d'un triplet statistique mais édite les tableaux des résidus tandis que **DDUtil : Modelling** édite les tableaux des modèles. Par exemple, 6 stations et 4 dates donnent 24 points de mesure (11,12, ..., 64) où sont enregistrées 3 variables. ABCD est le plan d'inertie maximale. Vu "de face" il donne la carte factorielle. Vu "de profil" il donne les résidus de l'approximation. Les données sont des points, les modèles également, les erreurs sont des vecteurs liés (traits gras des projections). Maximiser la variabilité sur le plan ou minimiser l'erreur autour du plan sont équivalents.



L'option utilise une seule fenêtre de dialogue :



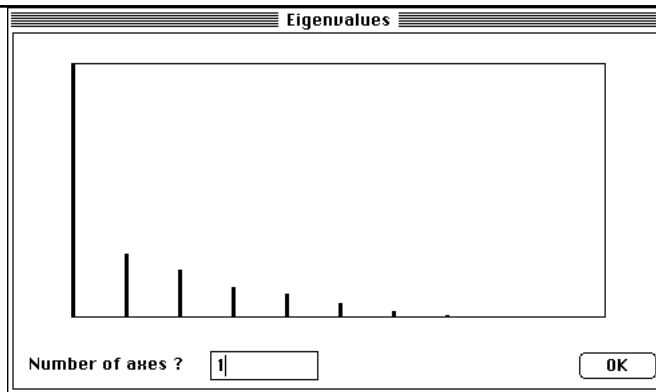
Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.



Utiliser la carte Méaudret de la pile ADE•Data pour obtenir le fichier Mil (24-10) et la carte Méaudret+1 pour obtenir le fichier Plan (24-2). Lire ce dernier avec **CategVar : Read Categ File** et faire l'ACP normée de Rh :



Le graphe des valeurs propres pourrait laisser croire qu'un seul facteur épuise l'information :

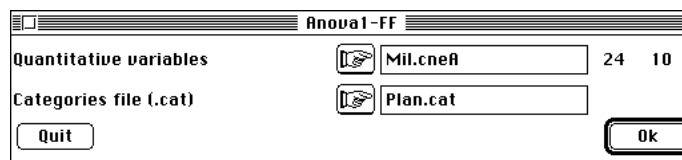


Utiliser la présente option qui enlève des données le modèle associé au premier facteur :

Residuals: computes the differences between the data reconstitution performed by the Modelling module and the data table (ta file).  
 Title of the analysis: Mil.cnta  
 Number of rows: 24, columns: 10

File Mil.cneA contains the residuals computed with 1 axe  
 It has 24 rows and 10 columns

Vérifier avec **Discrimin : Anova1-FF** que les composantes spatio-temporelles des variables résiduelles sont encore très significatives :



variable 1 from Mil.cneA versus variable 2 from Plan

Source	SS	d.f.	MS	F	Proba
Between	22.55	3	7.516	147.8	0
Within	1.017	20	0.05085		
Total	23.57	23			

...

variable 3 from Mil.cneA versus variable 2 from Plan

Source	SS	d.f.	MS	F	Proba
Between	3.76	3	1.253	4.717	0.012
Within	5.314	20	0.2657		
Total	9.073	23			

...

variable 4 from Mil.cneA versus variable 2 from Plan

Source	SS	d.f.	MS	F	Proba
Between	2.475	3	0.825	4.833	0.01092
Within	3.414	20	0.1707		
Total	5.889	23			

...

L'information, en dépit des apparences, ne se réduit pas à un facteur.



L'ACP peut être considérée comme pratique d'automodélisation. Elle a été inventée pour cela.

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* : 2, 559-572.

## DDUtil : Row projections



Utilitaire d'interprétation d'analyse multivariée à un tableau.



La projection de lignes ou de colonnes supplémentaires est une pratique courante en analyse des données. Il ne faut pas confondre projection d'éléments supplémentaires (opération géométrique) et représentation sur un plan factoriel d'information supplémentaire (opération pratique qui peut prendre des formes extrêmement variées). Le module projette sur le sous-espace des axes principaux d'une analyse quelconque des vecteurs lignes de même dimension que les lignes du tableau traité **sans leur faire subir de transformation préalable**. La projection se fait au sens de la métrique (poids des colonnes) de l'analyse de départ.



L'option utilise une seule fenêtre de dialogue :

Field	Value	Dimensions
Initial analysis (vp)	Rh.cnvp	15 2
Supplementary row file	Provi.MLRmod	39 15
Output file name	Sup	

Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.

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

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 Rhône de la pile ADE•Data pour obtenir le fichier Rh (39-15) et la carte Rhône+1 pour obtenir le fichier Deb (39-3). Faire l'ACP normée de Rh :

Field	Value	Dimensions
Matrix input file	Rh	39 15

On veut représenter sur la carte factorielle des lignes les modèles de la prédiction de Rh par Deb. Initialiser la régression multiple par LinearReg : Initialize :

Field	Value	Dimensions
Explanatory variables	Deb	39 3
Dependent variables	Rh.cnta	39 15
Output file name	Provi	

Exécuter la régression multiple par LinearReg : Modelling :

Field	Value
Input file	Provi.reg

Lancer la présente option :

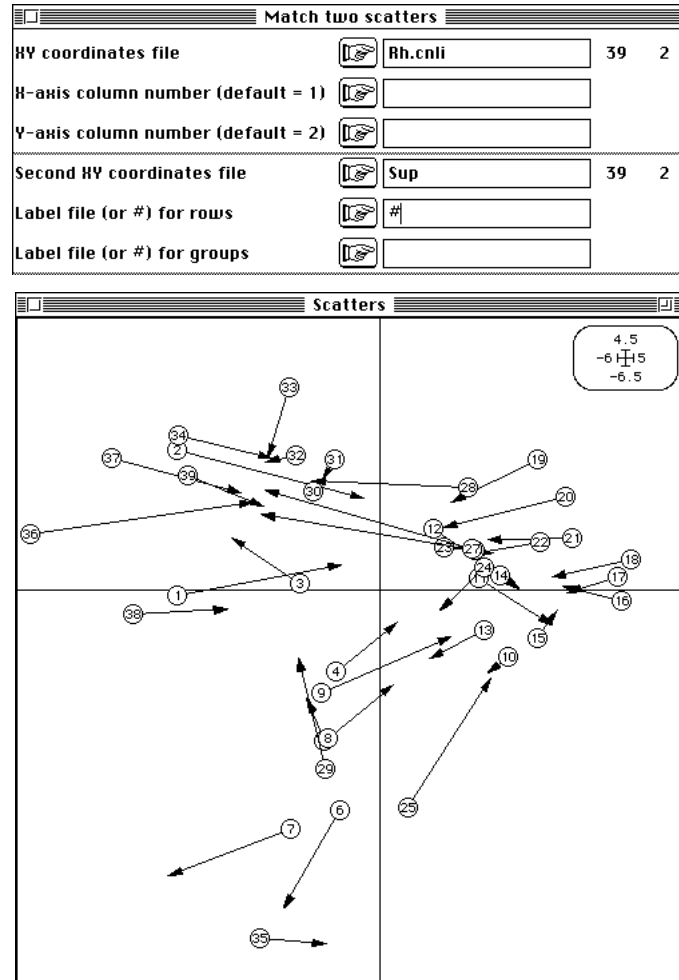
Title of the analysis: Rh.cnta  
Number of rows: 39, columns: 15

File Sup contains the coordinates of the projected rows  
It has 39 rows and 2 columns

File :Sup

Col.	Mini	Maxi
1	-3.493e+00	3.082e+00
2	-5.837e+00	2.188e+00

Représenter sur la carte factorielle les lignes du tableau de départ et les points estimés par la régression dans Scatters : Match two scatters :



Cet exemple introduit à l'ACP sur variables instrumentales (Projectors).

En autorisant la projection de vecteurs quelconques, on permet l'usage des lignes supplémentaires pour tous les types d'analyse (inter-intra, co-inertie, MFA, ...). Il convient de maîtriser la théorie pour s'en servir correctement.

Quand les lignes supplémentaires doivent subir la même transformation préalable que les lignes du tableau analysées, utiliser DDUtil : Supplementary rows. C'est la version ordinaire de l'emploi des lignes supplémentaires.

## DDUtil : Rows/Inertia analysis



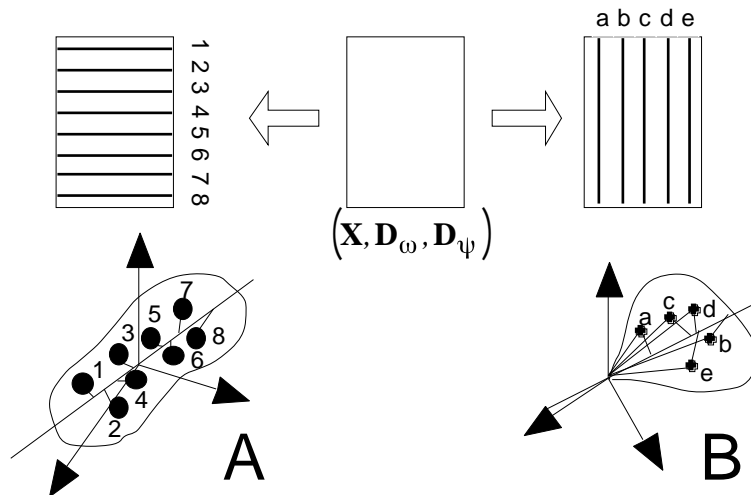
Utilitaire d'interprétation d'analyse multivariée à un tableau.



L'option édite le listing des statistiques d'inertie du nuage des lignes.

Soient les deux nuages de points définis dualement par un seul triplet  $(\mathbf{X}, \mathbf{D}_\omega, \mathbf{D}_\psi)$ .  $\mathbf{X}$  est une matrice à  $n$  lignes et  $p$  colonnes et les produits scalaires associés aux pondérations définissent les matrices :

$$\mathbf{D}_\omega = \text{Diag}(\omega_1, \dots, \omega_p) \text{ et } \mathbf{D}_\psi = \text{Diag}(\psi_1, \dots, \psi_n).$$



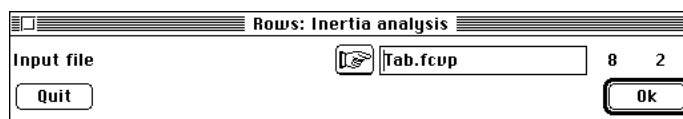
La somme des valeurs propres égale l'inertie totale  $I_T$  commune aux deux nuages. En ACP centrée cette inertie est la somme des variances. En ACP normé cette inertie vaut le nombre de variables. La quantité  $\lambda_k/I_T$  est l'inertie relative du vecteur principal de rang  $k$ .

L'inertie des projections sur l'axe principal  $k$  se décompose en somme de contribution absolue (CA) du point  $i$  à la définition de l'axe  $k$ .

L'inertie totale se décompose en contributions à la trace des lignes et des colonnes. Le carré de la distance d'une ligne  $i$  à l'origine se décompose entre contribution relative (CR) des axe à la représentation de la ligne  $i$ . Les contributions relatives sont des carrés de cosinus. Les définitions utilisées sont celles qui ont été définies dans le cas particulier de l'analyse des correspondances par <sup>1</sup> (pp. 316-317).



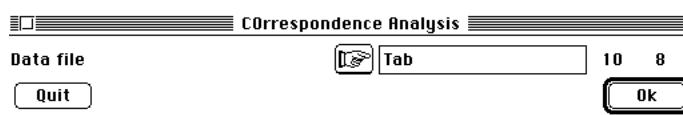
L'option utilise une seule fenêtre de dialogue :



Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cm pour ACM, fc pour AFC, cn pour ACP normée, ...) et vp est le fichier des valeurs propres.



Utiliser la carte Hébergement de la pile ADE•Data pour obtenir le fichier Tab (10-8) reproduit dans <sup>1</sup> (p. 306). Exécuter l'AFC du tableau :



Utiliser le présent module. Les résultats sont ceux de <sup>1</sup> (p. 319). Les contributions sont exprimées par des entiers en partie pour dix milles (111 signifie 1.11 %).

Input file: Tab.fcta  
 Number of rows: 10, columns: 8

Inertia: Two diagonal norm inertia analysis  
 Total inertia: 0.0836022 - Number of axes: 3

File Tab.fccl contains the contribution of rows to the trace  
 It has 10 rows and 1 column  
 Row inertia  
 All contributions are in 1/10000

-----Absolute contributions-----

Num	Fac 1	Fac 2	Fac 3
1	111	388	1725
2	2	613	1142
3	949	1604	346
4	3556	33	1667
5	318	242	831
6	103	800	22
7	3725	10	21
8	6	374	1215
9	0	5072	879
10	1224	859	2148

-----Relative contributions-----

Num	Fac 1	Fac 2	Fac 3	Remains	Weight	Cont.
1	829	1321	2548	5301	256	713
2	29	3353	2715	3901	83	443
3	5193	4008	376	421	958	969
4	9039	39	841	79	1486	2085
5	3280	1138	1698	3883	1382	515
6	1763	6227	76	1932	956	311
7	9720	12	10	255	2963	2031
8	79	2233	3149	4537	187	406
9	3	8633	650	712	457	1423
10	5895	1889	2055	159	1267	1101



Les statistiques d'inertie sont importantes dans les analyses à pondérations non uniformes (comme en AFC). Elles s'étendent à tout type d'analyse à un tableau, mais pour les analyses à pondérations uniformes, elles sont largement redondantes avec les cartes factorielles.



<sup>1</sup> Lebart, L., Morineau, A. & Fenelon, J.P. (1982) *Traitement des données statistiques. Méthodes et Programmes*. Dunod, 2<sup>o</sup> édition, Paris. 1-518.

## DDUtil : Supplementary columns



L'option projette sur le sous-espace des composantes principales d'une analyse des vecteurs colonnes de même dimension que les colonnes du tableau traité **en les transformant conformément au principe de l'analyse de départ**. La projection se fait au sens de la métrique (poids des lignes) de l'analyse de départ. Cette option est donc réservée aux analyses

- ACP normée (les colonnes supplémentaires sont normalisées) ;
- ACP centrée (les colonnes supplémentaires sont centrées) ;
- AFC (les colonnes supplémentaires sont considérées comme des distributions de fréquence sur l'ensemble des lignes).



L'option utilise une seule fenêtre de dialogue :

Field	Value	Rows	Columns
Initial analysis (vp)	Rh.cnvp	15	2
Supplementary column file	Deb	39	3
Output file name	DebColSup		

Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cp pour ACP centrée, fc pour AFC, cn pour ACP normée) et vp est le fichier des valeurs propres.

Fichier des colonnes supplémentaires. On doit y trouver autant de lignes 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 colonnes du fichier précédent et en colonnes les facteurs de l'analyse de base.



Utiliser la carte Rhône de la pile ADE•Data pour obtenir le fichier Rh (39-15) et la carte Rhône+1 pour obtenir le fichier Deb (39-3). Faire l'ACP normée de Rh :

Field	Value	Rows	Columns
Matrix input file	Rh	39	15

Add. col.	Mean	Var.
1	279.6	2.19E+04
2	104.6	6023
3	180.6	3.56E+04

Title of the analysis: Rh.cnta  
Number of rows: 39, columns: 15

File DebColSup contains the coordinates of the projected columns  
It has 3 rows and 2 columns

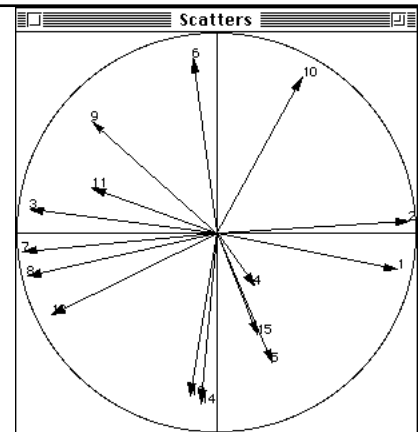
File :DebColSup

Col.	Mini	Maxi
1	-4.098e-01	6.409e-01
2	-8.796e-01	-2.810e-01

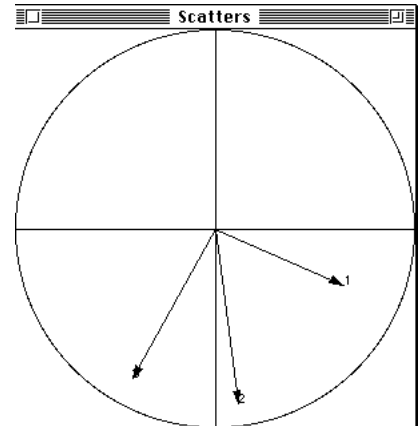
Tracer la carte des colonnes avec Scatters : Labels et sur la même fenêtre de dessin la carte des colonnes supplémentaires. Superposer éventuellement dans MacDraw de Claris™. En ACP normée les coordonnées des projections sont des coefficients de corrélation linéaire entre les variables supplémentaires et les coordonnées factorielles des lignes.


En AFC, les coordonnées des projections sont des pseudo-barycentres (centre de gravité déplacés par dilatation, voir <sup>1</sup> p.315-316).


Labels	
HV coordinates file	<input type="text" value="Rh.cnco"/> 15 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="#"/>
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>
Draw unit circle (yes = 1)	<input type="text" value="1"/>
Draw points (no = 2)	<input type="text" value="2"/>
Constrain H/V ratio (yes = 1)	<input type="text" value="1"/>
<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"/>	



Labels	
HV coordinates file	<input type="text" value="DebColSup"/> 3 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="#"/>
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>
Draw unit circle (yes = 1)	<input type="text" value="1"/>
Draw points (no = 2)	<input type="text" value="2"/>
Constrain H/V ratio (yes = 1)	<input type="text" value="1"/>
<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"/>	



 Voir aussi DDUtil : Supplementary rows.

 <sup>1</sup> Lebart, L., Morineau, A. & Fenelon, J.P. (1982) *Traitement des données statistiques. Méthodes et Programmes.* Dunod, 2<sup>o</sup> édition, Paris. 1-518.

## DDUtil : Supplementary rows



L'option projette sur le sous-espace des axes principaux d'une analyse des vecteurs lignes de même dimension que les lignes du tableau traité **en les transformant conformément au principe de l'analyse de départ**. La projection se fait au sens de la métrique (poids des colonnes) de l'analyse de départ. Cette option est donc réservée aux analyses

— ACP normée (les lignes supplémentaires sont modifiées comme les lignes de départ : soustraction de la moyenne par variable et division par l'écart-type correspondante) ;

— 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 et les coordonnées sont des pseudo-barycentres<sup>1</sup> p. 315-316).



L'option utilise une seule fenêtre de dialogue :

Field	Value	Dimensions
Initial analysis (vp)	Rh.cnvp	15 2
Supplementary column file	Deb	39 3
Output file name	DebColSup	

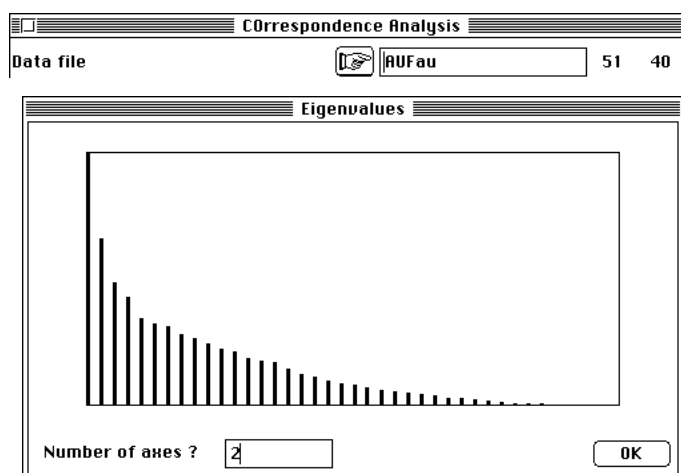
Nom de fichier du type ---.##vp issu d'une analyse à un tableau. ## est le type de l'analyse (cp pour ACP centrée, fc pour AFC, cn pour ACP normée) et vp est le fichier des valeurs propres.

Fichier des colonnes supplémentaires. On doit y trouver autant de lignes 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 colonnes du fichier précédent et en colonnes les facteurs de l'analyse de base.



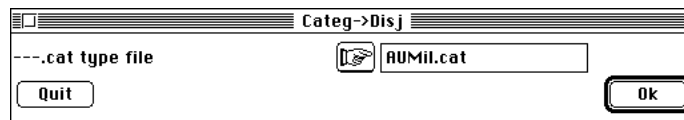
Utiliser la carte AviUrba de la pile ADE•Data pour obtenir le fichier AUFau (51-40) par TextToBin : Char->Binary. Faire l'AFC de ce tableau :



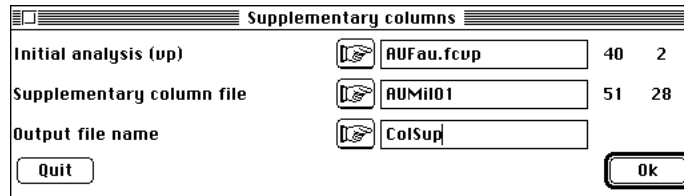
Utiliser la carte AviUrba+1 de la pile ADE•Data pour obtenir le fichier AUMil (51-11) par TextToBin : Char->Binary. Lire ce fichier par CategVar : Read Categ file :

Field	Value	Dimensions
Input file	AUMil	51 11

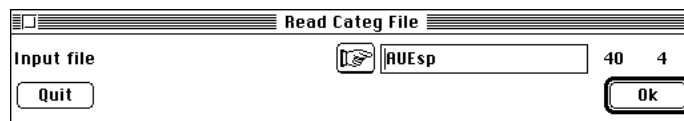
Obtenir le tableau disjonctif complet associé par CategVar : Categ->Disj :



Projeter en colonnes complémentaires les indicatrices des classes :



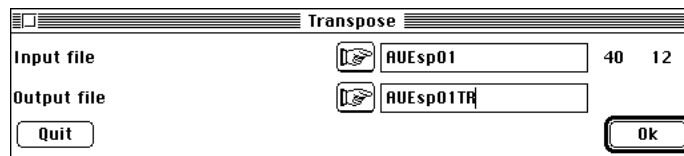
Utiliser la carte AviUrba+1 de la pile ADE•Data pour obtenir le fichier AUEsp (40-4) par TextToBin : Text->Binary. Lire ce fichier par CategVar : Read Categ file :



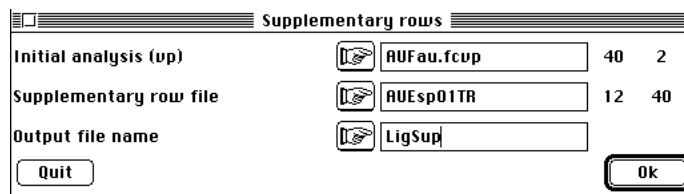
Obtenir le tableau disjonctif complet associé par CategVar : Categ->Disj :



Transposer ce fichier par FilesUtil : Transpose :



Projeter en lignes supplémentaires les indicatrices des classes :



Title of the analysis: AUFau.fcta  
Number of rows: 51, columns: 40

File LigSup contains the coordinates of the projected rows  
It has 12 rows and 2 columns

```
File :LigSup
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1 | -2.576e-01 | 1.806e+00 |
|  2 | -2.303e+00 | 2.489e+00 |
|----|-----|-----|
```

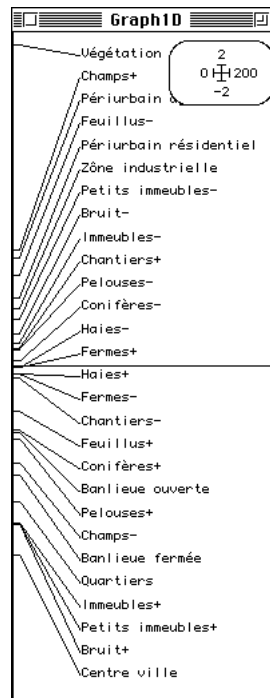
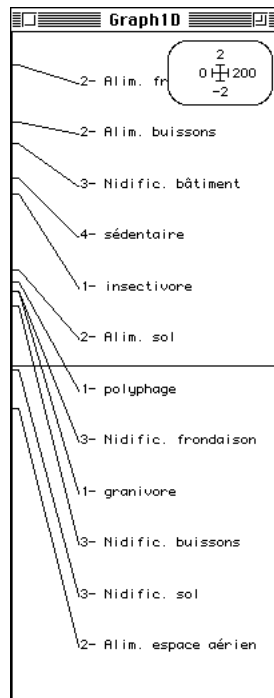
Récupérer les fichiers d'étiquettes des modalités des deux ensembles de variables qualitatives AUCode\_VarEsp et AUCode\_VarMil sur la carte précitée et mettre en forme dans BEdit™ :

1- insectivore  
 1- granivore  
 1- polyphage  
 2- Alim. sol  
 2- Alim. espace aérien  
 2- Alim. buissons  
 2- Alim. frondaisons  
 3- Nidific. sol  
 3- Nidific. bâtiment  
 3- Nidific. buissons  
 3- Nidific. frondaison  
 4- sédentaire  
 4- migrants

●●●  
 Haies+  
 Haies-  
 Feuillus+  
 Feuillus-  
 Conifères+  
 Conifères-  
 Bruit+  
 Bruit-  
 Végétation naturelle  
 Périurbain agricole  
 Périurbain résidentiel  
 Banlieue ouverte  
 Banlieue fermée  
 Quartiers  
 Centre ville  
 Zone industrielle

Représenter les lignes et les colonnes supplémentaires sur l'axe 1 par Graph1D : Labels :

Labels	
Data file (no default)	<input type="text" value="LigSup"/> 12 2
Rows label file (default = #)	<input type="text" value="AUCode_UarEsp"/>
Variable label file (or #)	<input type="text"/>
Vertical (1) or horizontal (2) graphs	<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"/>	



Labels	
Data file (no default)	<input type="text" value="ColSup"/> 28 2
Rows label file (default = #)	<input type="text" value="AUCode_UarMil"/>
Variable label file (or #)	<input type="text"/>
Vertical (1) or horizontal (2) graphs	<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"/>	



Pour relier directement les deux fichiers supplémentaires par le tableau central, utiliser le module RLQ.



<sup>1</sup> Lebart, L., Morineau, A. & Fenelon, J.P. (1982) *Traitement des données statistiques. Méthodes et Programmes*. Dunod, 2<sup>e</sup> édition, Paris. 1-518.

## DDUtil : Triplet Transpose



Utilitaire de gestion de données.



L'option transpose un tableau et permute les poids des lignes et les poids des colonnes.




L'option utilise une seule fenêtre de dialogue :


Triplet Transpose

Input file Mil.cmta 97 35

Generic output file name New

Quit Ok

 Nom de fichier du type `---.##ta` comportant  $n$  lignes et  $p$  colonnes. Ces paramètres sont quelconques. Le dossier de travail doit contenir un fichier binaire de nom `---.##pl` ( $n$  lignes et 1 colonne de nombres positifs) et un fichier binaire de nom `---.##pc` ( $p$  lignes et une colonne de nombres positifs). Le tout forme un triplet statistique associant un tableau, une pondération des lignes et une pondération des colonnes.

 Nom générique des fichiers de sortie.



Utiliser la carte Mafragh+2 de la pile ADE•Data pour obtenir par `TextToBin : Char->Binary` le fichier Mil (97-11). Le lire par `CategVar : Read Categ File`, en faire l'ACM par `MCA : Multiple Correspondence Analyse` et le transposer par la présente option.

```
Option: TripletTranspose
File New.trta contains the table (after transposition)
from the file: Mil.cmta
It has 35 rows and 97 columns
-----
File New.trpc contains the column weights
It has 97 rows and 1 column
File New.trpl contains the row weights
It has 35 rows and 1 column
-----
```

DDUtil:Diagonalize sur le nouveau triplet donne les résultats d'une ACM (variables en lignes, individus en colonnes).



Cette option intervient dans la gestion des K-tableaux (module `KTabUtil`), en particulier dans les assemblages d'analyses hétérogènes.

