

HTA

HTA : Column centring.....	2
HTA : Double centring additive.....	5
HTA : Double centring multiplicative.....	8
HTA : Edit inertia.....	11
HTA : No centring.....	14
HTA : Overall centring.....	16
HTA : Row centring.....	19

HTA : Column centring



Méthode d'analyse multivariée à un tableau.



L'analyse d'un tableau homogène (HTA) à centrage par colonne est l'ACP centrée classique (PCA : Covariance matrix PCA) Elle est implantée dans ce module, pour des raisons pédagogiques, avec les ACP à centrages par lignes, simple et doubles, beaucoup moins connues d'un public large. Cette ACP est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : centrage par colonne. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, \tilde{p}_i le poids relatif associé à la ligne i ($\sum_{i=1,n} \tilde{p}_i = 1$). Le centrage par colonne est le changement de variable $x_{ij} \mapsto x_{ij} - m_{.j}$, avec $m_{.j} = \sum_{i=1,n} \tilde{p}_i x_{ij}$;

2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = \tilde{p}_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ $\tilde{p}_i = 1/n$ (pondération unitaire), c) $\tilde{p}_i = p_i / \sum_{i=1,n} p_i$ (pondération quelconque à lire dans un fichier à une colonne) ;

3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = 1/p$ (pondération uniforme), b) $q_j = 1$ (pondération unitaire, utilisée par défaut), c) pondération quelconque à lire dans un fichier à une colonne.

L'option donne les mêmes résultats que PCA : Covariance matrix PCA mais conserve un fichier supplémentaire.



L'option utilise une seule fenêtre de dialogue :

Fichier du tableau d'entrée.

Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicité de la pile ADE-4•Data pour obtenir le fichier Toxi (11-15) et sa signification (toxicité de 11 produits sur 15 organismes cibles en log $\mu\text{mol/l}$).

Homogeneous Table Analysis: option .cp

Each row has 0.066667 weight (Sum = 1)
File T.cppl contains the row weights
It has 15 rows and 1 column

Each column has 0.090909 weight (Sum = 1)
File T.cppc contains the column weights
It has 15 rows and 1 column

Input file: T
Number of rows: 15, columns: 11
File T.cpta contains the (column) centred table
It has 15 rows and 11 columns
File T.cpmo contains the model (Model = mean by column)
It has 15 rows and 11 columns

DiagoRC: General program for two diagonal inner product analysis

Input file: T.cpta
--- Number of rows: 15, columns: 11

Total inertia: 0.243892

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.3691E-01	+0.5614	+0.5614	02	+4.5760E-02	+0.1876	+0.7490
03	+2.1896E-02	+0.0898	+0.8388	04	+1.3675E-02	+0.0561	+0.8948
05	+1.0287E-02	+0.0422	+0.9370	06	+7.5118E-03	+0.0308	+0.9678
07	+4.4306E-03	+0.0182	+0.9860	08	+2.1671E-03	+0.0089	+0.9949
09	+5.8307E-04	+0.0024	+0.9972	10	+3.8943E-04	+0.0016	+0.9988
11	+2.8205E-04	+0.0012	+1.0000				

File T.cpvv contains the eigenvalues and relative inertia for each axis
--- It has 11 rows and 2 columns

File T.cpcoc contains the column scores
--- It has 11 rows and 3 columns

File :T.cpcoc

Col.	Mini	Maxi
1	-2.104e-01	1.159e+00
2	-2.919e-01	4.447e-01
3	-3.391e-01	9.483e-02

File T.cpli contains the row scores
--- It has 15 rows and 3 columns

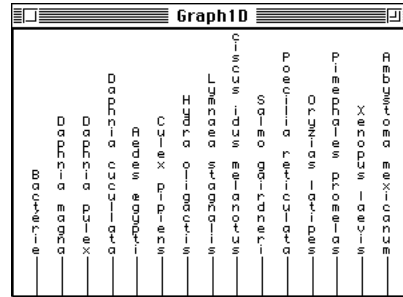
File :T.cpli

Col.	Mini	Maxi
1	-8.436e-01	4.202e-01
2	-2.984e-01	5.887e-01
3	-4.150e-01	2.458e-01

On insiste dans ce module sur la présentation du centrage comme celle d'un choix de modèle évident, puis d'élimination de ce modèle.

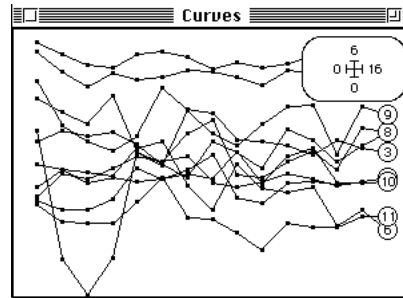
Préparer la légende (code des lignes représenté par Graph1D : Labels) :

Labels	
Data file (no default)	Numero 15 1
Rows label file (default = #)	Code_cible
Variable label file (or #)	
Vertical (1) or horizontal (2) graphs	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"/>	



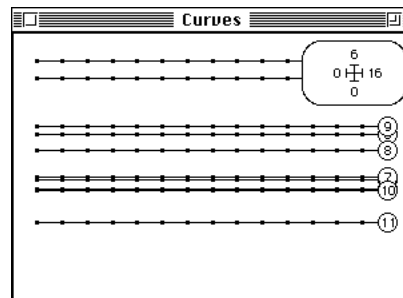
Représenter les données (Curves : Lines). Une courbe est une colonne du tableau :

Lines	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	T 15 11



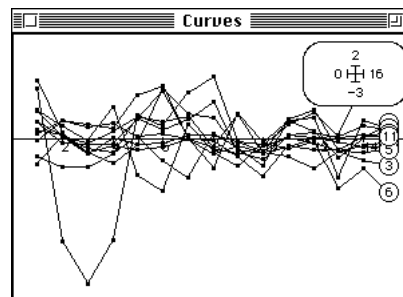
Représenter le modèle (la toxicité ne dépend que du produit) :

Lines	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	T.cpmo 15 11



Représenter le tableau traité (tableau de données - tableau modèle) :

Lines	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	T.cpta 15 11



L'analyse fonctionne ensuite comme une ACP ordinaire. Le centrage, comme élimination de modèle, est totalement explicité, et non pas caché comme dans un programme classique.



Thioulouse, J., Devillers, J., Chessel, D. & Auda, Y. (1991) Graphical techniques for multidimensional data analysis. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W. (Eds.) Kluwer Academic Publishers. 153-205.

HTA : Double centring additive



Méthode d'analyse multivariée à un tableau.



L'analyse d'un tableau homogène (HTA) à double centrage additif est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : double centrage additif. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, \tilde{p}_i le poids relatif associé à la ligne i ($\sum_{i=1,n} \tilde{p}_i = 1$), \tilde{q}_j le poids relatif associé à la colonne j ($\sum_{j=1,p} \tilde{q}_j = 1$). Le double centrage additif est le changement de variable $x_{ij} \mapsto x_{ij} - m_{i.} - m_{.j} + m_{..}$, avec :

$$m_{i.} = \sum_j \tilde{q}_j x_{ij} \quad m_{.j} = \sum_i \tilde{p}_i x_{ij}$$
$$m_{..} = \sum_{ij} \tilde{p}_i \tilde{q}_j x_{ij} = \sum_j \tilde{q}_j m_{.j} = \sum_i \tilde{p}_i m_{i.}$$

2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = \tilde{p}_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ $\tilde{p}_i = 1/n$ (pondération unitaire), c) $\tilde{p}_i = p_i / \sum_{i=1,n} p_i$ (pondération quelconque à lire dans un fichier à une colonne) ;

3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = \tilde{q}_j = 1/p$ (pondération uniforme), b) $q_j = 1$ $\tilde{q}_j = 1/p$ (pondération unitaire, utilisée par défaut), c) $\tilde{q}_j = q_j / \sum_{j=1,p} q_j$ (pondération quelconque à lire dans un fichier à une colonne).



L'option utilise une seule fenêtre de dialogue :

Fichier du tableau d'entrée.

Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicité de la pile ADE-4•Data pour obtenir le fichier Toxi (11-15) et sa signification (toxicité de 11 produits sur 15 organismes cibles en log $\mu\text{mol/l}$).

Homogeneous Table Analysis: option .cc

Each row has 0.066667 weight (Sum = 1)
File T.ccpl contains the row weights
It has 15 rows and 1 column

Each column has 0.090909 weight (Sum = 1)
File T.ccpc contains the column weights
It has 15 rows and 1 column

Input file: T
Number of rows: 15, columns: 11
File T.ccta contains the doubly centred table
It has 15 rows and 11 columns
File T.ccmo contains the model
Model = mean by row + mean by column - overall mean
It has 15 rows and 11 columns

DiagoRC: General program for two diagonal inner product analysis

Input file: T.ccta
--- Number of rows: 15, columns: 11

Total inertia: 0.210923

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.2329E-01	+0.5845	+0.5845	02	+4.1942E-02	+0.1988	+0.7834
03	+1.6788E-02	+0.0796	+0.8630	04	+1.1309E-02	+0.0536	+0.9166
05	+7.5144E-03	+0.0356	+0.9522	06	+4.9620E-03	+0.0235	+0.9757
07	+3.2095E-03	+0.0152	+0.9910	08	+1.1363E-03	+0.0054	+0.9963
09	+4.2834E-04	+0.0020	+0.9984	10	+3.4226E-04	+0.0016	+1.0000
11	+0.0000E+00	+0.0000	+1.0000				

File T.ccvp contains the eigenvalues and relative inertia for each axis
--- It has 11 rows and 2 columns

File T.ccco contains the column scores
--- It has 11 rows and 2 columns

File :T.ccco

Col.	Mini	Maxi
1	-1.040e+00	3.700e-01
2	-4.223e-01	3.387e-01

File T.ccli contains the row scores

--- It has 15 rows and 2 columns

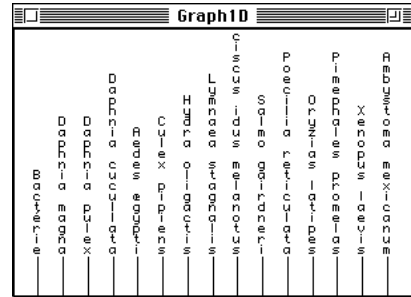
File :T.ccli

Col.	Mini	Maxi
1	-4.090e-01	7.739e-01
2	-2.588e-01	5.647e-01

Préparer la légende (code des lignes représenté par Graph1D : Labels) :

The dialog box titled 'Labels' contains the following fields and buttons:

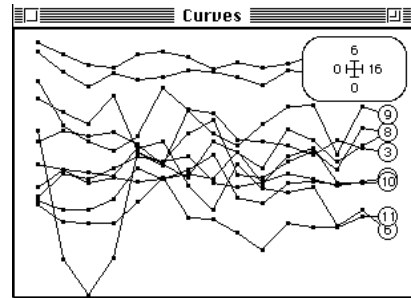
- Data file (no default):** 15 1
- Rows label file (default = #):**
- Variable label file (or #):**
- Vertical (1) or horizontal (2) graphs:** 1 2
- Buttons:** Quit, Copy graph, Save graph, Print graph, Draw



Représenter les données (Curves : Lines). Une courbe est une colonne du tableau :

The dialog box titled 'Lines' contains the following fields and buttons:

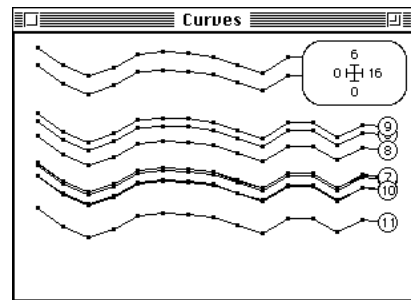
- X file (default = 1, 2, 3, ..., n):**
- X file column number (default = 1):**
- Y file (no default):** 15 11
- Buttons:**



Représenter le modèle (la toxicité dépend de manière additive de la cible et du produit) :

The dialog box titled 'Lines' contains the following fields and buttons:

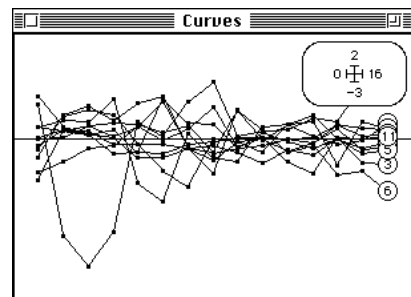
- X file (default = 1, 2, 3, ..., n):**
- X file column number (default = 1):**
- Y file (no default):** 15 11
- Buttons:**



Représenter le tableau traité (tableau de données - tableau modèle) :

The dialog box titled 'Lines' contains the following fields and buttons:

- X file (default = 1, 2, 3, ..., n):**
- X file column number (default = 1):**
- Y file (no default):** 15 11
- Buttons:**



L'analyse fonctionne ensuite comme une ACP ordinaire.



Après cette option, le module DDUtil est disponible.

HTA : Double centring multiplicative



Méthode d'analyse multivariée à un tableau.



L'analyse d'un tableau homogène (HTA) à double centrage multiplicatif est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : double centrage multiplicatif. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, p_i le poids associé à la ligne i , q_j le poids associé à la colonne j . Le double centrage multiplicatif est le changement de variable $x_{ij} \mapsto x_{ij} - \alpha_i \beta_j$, avec la contrainte :

$$\sum_{ij} p_i q_j (x_{ij} - \alpha_i \beta_j)^2 \text{ minimum}$$

Le modèle est estimé par la reconstitution au rang 1 de la matrice par décomposition en valeurs singulières (reconstitution de données par le premier facteur de l'ACP non centrée, dite encore décomposition biplot de rang 1, ou encore approximation d'Eckart-Young¹ d'ordre 1).


2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ (pondération unitaire), c) pondération quelconque à lire dans un fichier à une colonne ;


3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = 1/p$ (pondération uniforme), b) $q_j = 1$ (pondération unitaire, utilisée par défaut), c) pondération quelconque à lire dans un fichier à une colonne.



L'option utilise une seule fenêtre de dialogue :

 Fichier du tableau d'entrée.

 Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

 Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

 Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

 Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicornée de la pile ADE-4•Data pour obtenir le fichier vivo (34-4) et sa signification (toxicité de 34 produits pour 4 durée d'exposition). Utiliser les paramètres implicite :

Homogeneous Table Analysis: option .dm

Each row has 0.029412 weight (Sum = 1)
File vivo.dmpl contains the row weights
It has 34 rows and 1 column

Each column has unit weight (Sum = 4)
File vivo.dmpc contains the column weights
It has 34 rows and 1 column

Input file: vivo
Number of rows: 34, columns: 4
File vivo.dmta contains the doubly centred table
It has 34 rows and 4 columns
File vivo.dmmo contains the model
Model = A(i)*B(j) optimal from the first axe of non centered PCA
It has 34 rows and 4 columns

DiagoRC: General program for two diagonal inner product analysis
Input file: vivo.dmta
--- Number of rows: 34, columns: 4

Total inertia: 687.313

Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum
01	+5.3807E+02	+0.7829	02	+1.1266E+02	+0.1639
03	+3.6583E+01	+0.0532	04	+0.0000E+00	+0.0000

File vivo.dmvp contains the eigenvalues and relative inertia for each axis
--- It has 4 rows and 2 columns

File vivo.dmco contains the column scores
--- It has 4 rows and 2 columns

File :vivo.dmco

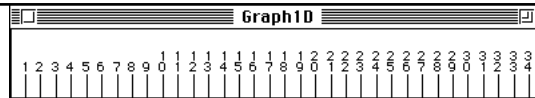
Col.	Mini	Maxi
1	-1.311e+01	1.745e+01
2	-8.430e+00	4.641e+00

File vivo.dmlj contains the row scores
--- It has 34 rows and 2 columns

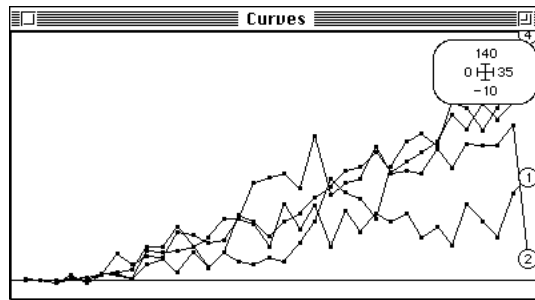
File :vivo.dmlj

Col.	Mini	Maxi
1	-8.204e+01	4.379e+01
2	-3.742e+01	1.856e+01

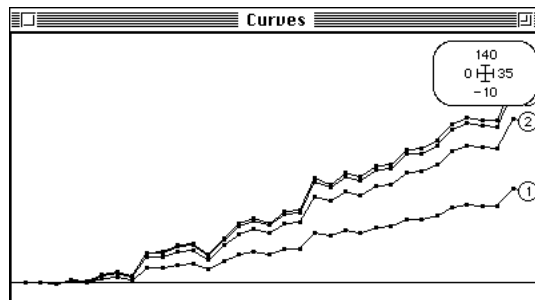
Mettre en place la légende :



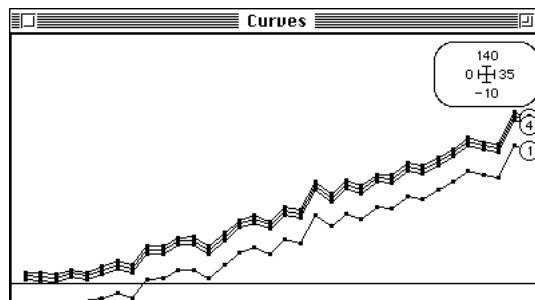
Représenter les données :



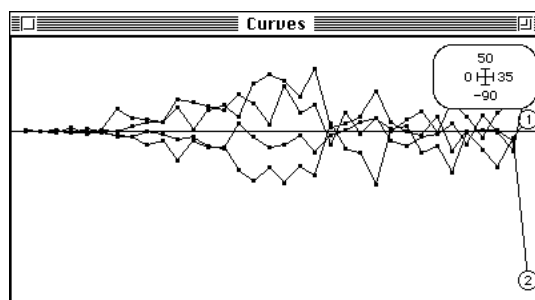
Visualiser le modèle multiplicatif :



Comparer avec le modèle additif :



Représenter les données traitées :



L'analyse fonctionne ensuite comme une ACP ordinaire. Elle a une propriété de double centrage du tableau concurrente de celle HTA : Double centrage additive et de celle de COA : Correspondence Analysis.



Après cette option, le module DDUtil est disponible.



¹ Eckart, C. & Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika* : 1, 211-218.

HTA : Edit inertia



Utilitaire pour le choix de l'analyse d'un tableau homogène.



L'objet de l'option est de donner une mesure de la pertinence de divers modèles simples, dits modèles évidents, du tableau. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, p_i le poids associé à la ligne i , q_j le poids associé à la colonne j . Soit \mathbf{Z} un modèle du tableau, z_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes. L'erreur globale commise avec ce modèle est l'inertie de l'ACP décentrée du tableau $\mathbf{X} - \mathbf{Z}$, soit :

$$E(\mathbf{Z}) = \sum_{ij} p_i q_j (x_{ij} - z_{ij})^2$$

Cette valeur est éditée pour les modèles suivants :

$$z_{ij} = 0 \quad (\text{modèle nul})$$

$$z_{ij} = m_{..} \quad (\text{modèle constant, estimation LS})$$

$$z_{ij} = m_{i.} \quad (\text{effet ligne simple, estimation LS})$$

$$z_{ij} = m_{.j} \quad (\text{effet colonne simple, estimation LS})$$


$$z_{ij} = m_{i.} + m_{.j} - m_{..} \quad (\text{effet additif, estimation LS})$$


$$z_{ij} = \alpha_i \beta_j \quad (\text{effet multiplicatif, estimation LS})$$


La notation LS est celle des moindres carrés (Least Squares). L'analyse du tableau $\mathbf{X} - \mathbf{Z}$, pour chacune des possibilités donne chacune des options du module.




L'option utilise une seule fenêtre de dialogue :

 Fichier du tableau d'entrée.

 Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

 Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

 Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

 Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicité de la pile ADE-4•Data pour obtenir le fichier Toxi (11-15) et sa signification (toxicité de 11 produits sur 15 organismes cibles en log $\mu\text{mol/l}$).

```

input file: T
Option 1 = No action (non centred PCA)          Inertia = 12.123
Option 2 = Centred table (overall centred PCA)  Inertia = 1.35621
Option 3 = Centred (zero mean) columns         Inertia = 0.243892
Option 4 = Centred (zero mean) rows           Inertia = 1.32324
Option 5 = Additive model                      Inertia = 0.210923
Option 6 = Multiplicative model               Inertia = 0.217928

```

Le modèle correspondant au centrage colonne (effet colonne simple) est le meilleur compromis entre simplicité et efficacité. L'introduction d'un effet ligne ou d'un effet double ne présente que peu d'intérêt.

Utiliser la carte Toxicornée de la pile ADE-4•Data pour obtenir le fichier vivo (34-4) et sa signification (toxicité de 34 produits pour 4 durée d'exposition). Utiliser le paramétrage implicite :



```

input file: vivo
Option 1 = No action (non centred PCA)          Inertia = 9551.68
Option 2 = Centred table (overall centred PCA)  Inertia = 4220.53
Option 3 = Centred (zero mean) columns         Inertia = 4014.18
Option 4 = Centred (zero mean) rows           Inertia = 1212.48
Option 5 = Additive model                      Inertia = 1006.13
Option 6 = Multiplicative model               Inertia = 687.313

```

Le gain du modèle multiplicatif est manifeste et caractérise les données (voir [HTA : double centrage multiplicatif](#))



Il existe une littérature ancienne et considérable sur les ACP à centrage non classique, les modèles additif et multiplicatif, la reconstitution des tableaux. Le poids des options imposées par les logiciels de grande diffusion, même les plus récents, en réserve l'usage aux statisticiens professionnels. Il y a plusieurs filières bibliographiques parallèles. L'ACP non centrée est décrite dans Whittle (1952)¹ qui introduit clairement au double centrage multiplicatif. Le double centrage additif est introduit par Mandel (1961)² qui renvoie à ³. Les mêmes voies ont été ouverte par Gollob (1968)⁴ qui renvoie à Eckart et Young (1936)⁵. La question des mélanges des modèles additifs et multiplicatifs, qui n'est pas simple, a été abordée en ⁶. Les techniques parallèles de centrage, référées au mode R et Q maintenant oubliés, sont discutés par Okamoto (1972)⁷. L'usage des outils graphiques et des centrages est abordé dans ⁸. La logique de cette option est explicitée dans ⁹.



¹ Whittle, P. (1952) On principal components and least square methods of factor analysis. *Skandinavisk aktuarietidskrift* : 35, 223-239.

² Mandel, J. (1961) Non additivity in two-way analysis of variance. *Journal of the American Statistical Association* : 65, 878-888.

Mandel, J. & McCrackin, F.L. (1963) Analysis of families of curves. *Journal of Research — National Bureau of Standards, A, Physics and Chemistry*: 67, 259-267.

Mandel, J. (1969) A method for fitting empirical surfaces to physical or chemical data. *Technometrics* : 11, 411-429.

Mandel, J. (1971) A new analysis of variance model for non-additive data. *Technometrics* : 13, 1-18.

Mandel, J. (1972) Principal components, analysis of variance and data structure. *Statistica Neerlandica* : 26, 119-129.

³ Fisher, R.A. & MacKenzie, W.A. (1923) Studies in crop variation. II The manurial response of different potato varieties. *Journal of Agricultural Science*: 13, 311.

⁴ Gollob, H.F. (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* : 33, 73-115.

⁵ Eckart, C. & Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika* : 1, 211-218.

⁶ Johnson D.E. & Graybill F.A. (1972) An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* : 67, 862-868.

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

⁷ Okamoto, M. (1972) Four techniques of principal component analysis. *Journal of the Japanese Statistical Society* : 2, 63-69.

⁸ Devillers, J., Thioulouse, J. & Karcher, W. (1993) Chemometrical Evaluation of Multispecies-Multichemical Data by Means of Graphical Techniques Combined with Multivariate Analyses. *Ecotoxicology and Environmental Safety* : 26. 333-345.

⁹ Devillers, J. & Chessel, D. (1995b) Comparaison of in vivo and in vitro toxicity tests from co-inertia analysis. In : *Computer-Aided Molecular Design. Applications in Agrochemicals, Materials and Pharmaceuticals*. Reynolds, C.H., Holloway, M.K. & Cox, H.K. (Eds.) ACS Symposium Series 589. American Chemical Society, Washington. 250-266.

Devillers, J. & Chessel, D. (1995a) Can the enucleated rabbit eye test be a suitable alternative for the in vivo eye test? A chemometrical response. *Toxicology Modelling* : 1, 21-34.

HTA : No centring



Méthode d'analyse multivariée à un tableau.



L'analyse d'un tableau homogène (HTA) non centrée est l'ACP (PCA : Non centered PCA) Elle est réimplantée dans ce module pour des raisons d'homogénéité des ensembles de fonctions disponibles. Cette ACP est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : aucune.

2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ (pondération unitaire), c) pondération quelconque à lire dans un fichier à une colonne ;

3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = 1/p$ (pondération uniforme), b) $q_j = 1$ (pondération unitaire, utilisée par défaut), c) pondération quelconque à lire dans un fichier à une colonne.



L'option utilise une seule fenêtre de dialogue :

The dialog box titled "No centring" has the following fields:

- Matrix input file
- Row weighting option
- Column weighting option
- Option: file for row weighting
- Option: file for column weighting

Buttons: Quit, Ok

Fichier du tableau d'entrée.

Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Le centrage est considéré dans ce module comme l'élimination d'un modèle évident. Il est éliminé de l'analyse mais garde un intérêt considérable par lui-même. Ceci évite à l'analyse de retrouver ce modèle ou d'être très influencé par lui. On utilise une ACP non centrée chaque fois qu'aucun modèle n'est en jeu.

Considérons les exemples des autres options du module : La toxicité est-elle nulle pour chaque produit et chaque cible ?

La réponse est évidemment non, l'expérience serait stupide. La toxicité est-elle constante pour chaque produit et chaque cible ? Sans doute non, le tableau serait vraiment sans signification ?

Dépend-elle du produit, de la cible, des deux à la fois ? La réponse est positive, un graphique suffit à le montrer. Ceci est évident, c'est-à-dire attendu.

Y-a-t'il encore une autre information, plusieurs types de comportement des produits ou de cible, des associations cibles-produits ? Les analyses du tableau des résidus sont là pour répondre.

La stratégie est du type $D = ME + MC + E$ (donnée = modèle évident + modèle caché + erreur). Si on fait systématiquement une ACP classique, systématiquement on postule $ME =$ effet colonne et systématiquement on l'enlève. Ceci peut être nuisible (et l'est souvent) de deux manières. L'effet colonne n'existe pas ou partiellement et on l'enlève à tort. L'effet ligne est évident et on le laisse encore à tort (l'analyse le retrouvera et on dira encore que la méthode est triviale).

Il existe des situations du type $D = PE + MC + E$ (donnée = prévision externe + modèle caché + erreur). L'analyse du tableau D-PE (Données - prévisions) utilise cette option pour définir le modèle caché.



L'option s'emploie comme PCA : Non centered PCA.

HTA : Overall centring



Méthode d'analyse multivariée à un tableau homogène.



L'analyse d'un tableau homogène (HTA) à centrage unique est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : centrage unique. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, \tilde{p}_i le poids relatif associé à la ligne i ($\sum_{i=1,n} \tilde{p}_i = 1$), \tilde{q}_j le poids relatif associé à la colonne j ($\sum_{j=1,p} \tilde{q}_j = 1$). Le centrage unique est le changement de variable $x_{ij} \mapsto x_{ij} - m_{..}$, avec :

$$m_{..} = \sum_{ij} \tilde{p}_i \tilde{q}_j x_{ij} \text{ (moyenne générale) } ;$$

2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = \tilde{p}_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ $\tilde{p}_i = 1/n$ (pondération unitaire), c) $p_i = 1$ $\tilde{p}_i = p_i / \sum_{i=1,n} p_i$ (pondération quelconque à lire dans un fichier à une colonne) ;

3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = \tilde{q}_j = 1/p$ (pondération uniforme), b) $q_j = 1$ $\tilde{q}_j = 1/p$ (pondération unitaire, utilisée par défaut), c) $q_j = 1$ $\tilde{q}_j = q_j / \sum_{j=1,p} q_j$ (pondération quelconque à lire dans un fichier à une colonne).



L'option utilise une seule fenêtre de dialogue :

Overall centring	
Matrix input file	T
Row weighting option	1
Column weighting option	2
Option: file for row weighting	
Option: file for column weighting	
Quit	Ok

Fichier du tableau d'entrée.

Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicité de la pile ADE-4•Data pour obtenir le fichier Toxi (11-15) et sa signification (toxicité de 11 produits sur 15 organismes cibles en log $\mu\text{mol/l}$).

Homogeneous Table Analysis: option .cu

Each row has 0.066667 weight (Sum = 1)
File T.cupl contains the row weights
It has 15 rows and 1 column

Each column has 0.090909 weight (Sum = 1)
File T.cupc contains the column weights
It has 15 rows and 1 column

Input file: T
Number of rows: 15, columns: 11
File T.cuta contains the centred (overall mean) table
It has 15 rows and 11 columns
File T.cumo contains the model
Model = overall mean
It has 15 rows and 11 columns

DiagoRC: General program for two diagonal inner product analysis

Input file: T.cuta
--- Number of rows: 15, columns: 11

Total inertia: 1.35621

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.1248E+00	+0.8294	+0.8294	02	+1.3480E-01	+0.0994	+0.9288
03	+4.1838E-02	+0.0308	+0.9596	04	+1.9748E-02	+0.0146	+0.9742
05	+1.1633E-02	+0.0086	+0.9827	06	+9.0660E-03	+0.0067	+0.9894
07	+7.3928E-03	+0.0055	+0.9949	08	+4.0247E-03	+0.0030	+0.9978
09	+2.1372E-03	+0.0016	+0.9994	10	+4.7790E-04	+0.0004	+0.9998
11	+3.0302E-04	+0.0002	+1.0000				

File T.cuwp contains the eigenvalues and relative inertia for each axis
--- It has 11 rows and 2 columns

File T.cuco contains the column scores
--- It has 11 rows and 2 columns

File :T.cuco

Col.	Mini	Maxi
1	-1.553e+00	2.060e+00
2	-1.133e+00	2.633e-01

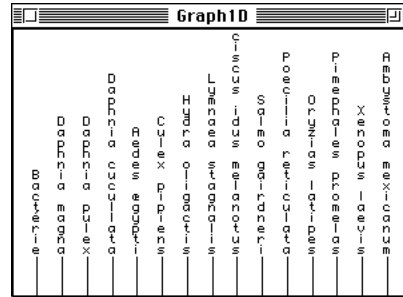
File T.culi contains the row scores
--- It has 15 rows and 2 columns

File :T.culi

Col.	Mini	Maxi
1	8.175e-01	1.208e+00
2	-4.280e-01	8.208e-01

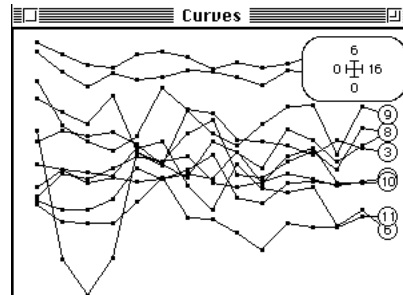
Préparer la légende (code des lignes représenté par Graph1D : Labels) :

Labels	
Data file (no default)	<input type="text" value="Numero"/> 15 1
Rows label file (default = #)	<input type="text" value="Code_cible"/>
Variable label file (or #)	<input type="text"/>
Vertical (1) or horizontal (2) graphs	<input type="text" value="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"/>	



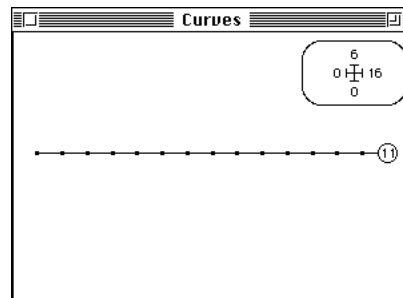
Représenter les données (Curves : Lines). Une courbe est une colonne du tableau :

Lines	
X file (default = 1, 2, 3, ..., n)	<input type="text"/>
X file column number (default = 1)	<input type="text"/>
Y file (no default)	<input type="text" value="T"/> 15 11



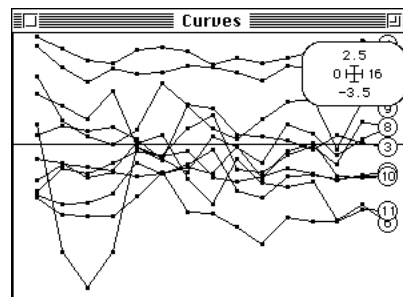
Représenter le modèle (la toxicité ne dépend ni du produit ni de la cible) :

Lines	
X file (default = 1, 2, 3, ..., n)	<input type="text"/>
X file column number (default = 1)	<input type="text"/>
Y file (no default)	<input type="text" value="T.cumo"/> 15 11



Représenter le tableau traité (tableau de données - tableau modèle) :

Lines	
X file (default = 1, 2, 3, ..., n)	<input type="text"/>
X file column number (default = 1)	<input type="text"/>
Y file (no default)	<input type="text" value="T.cutaj"/> 15 11



L'analyse fonctionne ensuite comme une ACP ordinaire.



Après cette option, le module DDUtil est disponible.

HTA : Row centring



Méthode d'analyse multivariée à un tableau.



L'analyse d'un tableau homogène (HTA) à centrage par ligne est l'analyse d'un schéma de dualité avec le jeu de paramètres :

1 — Transformation initiale : centrage par ligne. \mathbf{X} est la matrice de départ, x_{ij} la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, \tilde{q}_j le poids relatif associé à la colonne j $\sum_{j=1,p} \tilde{q}_j = 1$. Le centrage par ligne est le changement de variable $x_{ij} \mapsto x_{ij} - m_i$, avec $m_i = \sum_{j=1,p} \tilde{q}_j x_{ij}$;

2 — Pondération des lignes (3 options) : a) chaque ligne a le poids $p_i = 1/n$ (pondération uniforme, utilisée par défaut), b) $p_i = 1$ (pondération unitaire), c) pondération quelconque à lire dans un fichier à une colonne ;

3 — Pondération des colonnes (3 options) : a) chaque colonne a le poids $q_j = \tilde{q}_j = 1/p$ (pondération uniforme), b) $q_j = 1$ $\tilde{q}_j = 1/p$ (pondération unitaire, utilisée par défaut), c) $q_j = 1$ $\tilde{q}_j = q_j / \sum_{j=1,p} q_j$ (pondération quelconque à lire dans un fichier à une colonne).

L'option donne les mêmes résultats que PCA : Covariance matrix PCA exécutée sur le tableau d'entrée transposé, mais conserve un fichier supplémentaire.



L'option utilise une seule fenêtre de dialogue :

Matrix input file		T
Row weighting option		1
Column weighting option		2
Option: file for row weighting		
Option: file for column weighting		

Fichier du tableau d'entrée.

Option de pondération des lignes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Option de pondération des colonnes (1 — uniforme, 2 — unitaire, 3 — à lire dans un fichier).

Nom du fichier de pondération des lignes (accepté si l'option correspondante a été sélectionnée).

Nom du fichier de pondération des colonnes (accepté si l'option correspondante a été sélectionnée).



Utiliser la carte Toxicité de la pile ADE-4•Data pour obtenir le fichier Toxi (11-15) et sa signification (toxicité de 11 produits sur 15 organismes cibles en log $\mu\text{mol/l}$).

Homogeneous Table Analysis: option .cl

Each row has 0.066667 weight (Sum = 1)
File T.clpl contains the row weights

It has 15 rows and 1 column

Each column has 0.090909 weight (Sum = 1)

File T.clpc contains the column weights

It has 15 rows and 1 column

Input file: T

Number of rows: 15, columns: 11

File T.clta contains the (column) centred table

It has 15 rows and 11 columns

File T.clmo contains the model (Model = mean by row)

It has 15 rows and 11 columns

DiagoRC: General program for two diagonal inner product analysis

Input file: T.clta

--- Number of rows: 15, columns: 11

Total inertia: 1.32324

Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum
01 +1.1247E+00	+0.8500	+0.8500	02 +1.2167E-01	+0.0920	+0.9419
03 +3.7091E-02	+0.0280	+0.9699	04 +1.2702E-02	+0.0096	+0.9795
05 +1.1308E-02	+0.0085	+0.9881	06 +7.4684E-03	+0.0056	+0.9937
07 +4.9553E-03	+0.0037	+0.9975	08 +2.4503E-03	+0.0019	+0.9993
09 +5.1492E-04	+0.0004	+0.9997	10 +3.6433E-04	+0.0003	+1.0000
11 +0.0000E+00	+0.0000	+1.0000			

File T.clvp contains the eigenvalues and relative inertia for each axis

--- It has 11 rows and 2 columns

File T.clco contains the column scores

--- It has 11 rows and 2 columns

File :T.clco

Col.	Mini	Maxi
1	-2.069e+00	1.544e+00
2	-4.152e-01	1.017e+00

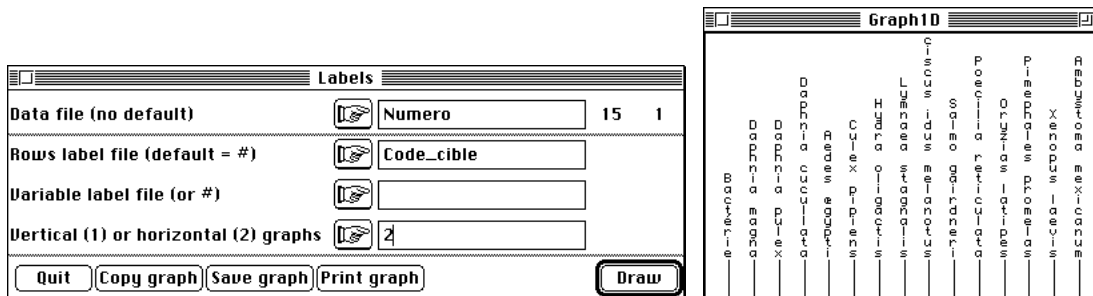
File T.clli contains the row scores

--- It has 15 rows and 2 columns

File :T.clli

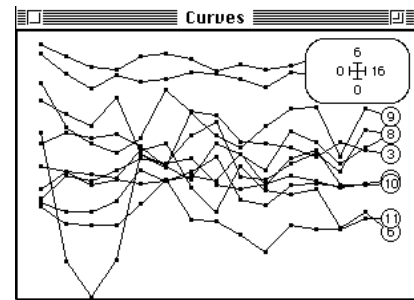
Col.	Mini	Maxi
1	-1.209e+00	-8.193e-01
2	-7.537e-01	4.139e-01

Préparer la légende (code des lignes représenté par Graph1D : Labels) :



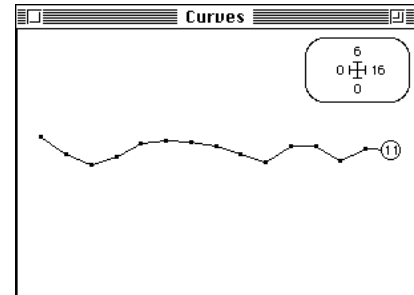
Représenter les données (Curves : Lines). Une courbe est une colonne du tableau :

Lines	
H file (default = 1, 2, 3, ..., n)	<input type="text"/>
H file column number (default = 1)	<input type="text"/>
Y file (no default)	T <input type="text"/> 15 11



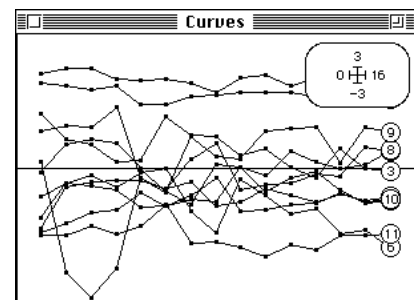
Représenter le modèle (la toxicité ne dépend que de la cible) :

Lines	
H file (default = 1, 2, 3, ..., n)	<input type="text"/>
H file column number (default = 1)	<input type="text"/>
Y file (no default)	T.cimo <input type="text"/> 15 11



Représenter le tableau traité (tableau de données - tableau modèle) :

Lines	
H file (default = 1, 2, 3, ..., n)	<input type="text"/>
H file column number (default = 1)	<input type="text"/>
Y file (no default)	T.citq <input type="text"/> 15 11



L'analyse fonctionne ensuite comme une ACP ordinaire.



Après cette option, le module DDUtil est disponible.



Le centrage par lignes d'un tableau est le centrage par colonne de son transposé. On appelle une des deux analyses en mode Q et l'autre en mode R, avec aucune autre définition que le mode Q n'est pas le mode R et réciproquement. Comme chacune des deux comporte deux analyses duales (nuage des lignes et nuage des colonnes), il y a ambiguïté.

En ACP centrée il y a donc implicitement 4 analyses d'inertie sous-jacentes. Les questions de centrage et de transposition des tableaux conduisent alors à la définition des analyses O/P/Q/R/S/T.

L'essentiel est alors dans le principe¹ :

"It is little more trouble for the author, and much more informative to the reader, simply to state precisely what was done."



¹ Ivimey Cook, R.B., Proctor, M.C.F. & Wigston, D.L. (1969) On the problem of the "R/Q" terminology in multivariate analysis of biological data. *Journal of Ecology* : 57, 673-675.

