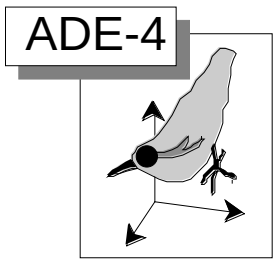


LinearReg

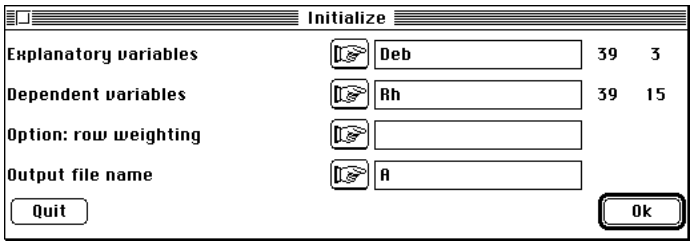
LinearReg : Initialize.....2
LinearReg : MLR -> Modelling.....3
LinearReg : MLR -> MultCorCoeff.....7
LinearReg : MLR -> New Data.....9
LinearReg : PLS -> Modelling.....10
LinearReg : PLS -> New Data.....15
LinearReg : PLS -> Randomization Test.....17

LinearReg : Initialize

Utilitaire de contrôle de données : préparation de régressions multiples.

L'option s'emploie pour définir un jeu de paramètres qui sera utilisé par toutes les autres options de ce module. On s'intéresse ici au cas d'un **groupe de *p* variables explicatives quelconques**, de *k* variables à expliquer et de modèles de prédiction multivariés indépendants pour chacune des variables à prédire.

L'option utilise une seule fenêtre de dialogue :



Nom du fichier binaire contenant les variables explicatives.

Nom du fichier binaire contenant les variables à prédire.

Fichier de pondération des lignes (par défaut, on utilise la pondération uniforme).

Nom générique des fichiers de sortie.

Utiliser la carte Rhône de la pile ADE-4•Data pour obtenir les fichiers Rh (39-15), Date (39-1) et Code_VarX. Utiliser la carte Rhône+1 pour obtenir les fichiers Deb (39-3) et Code_VarY. Associer le fichier des variables à prédire (Rh) et des variables prédictives (Deb) par la présente option.

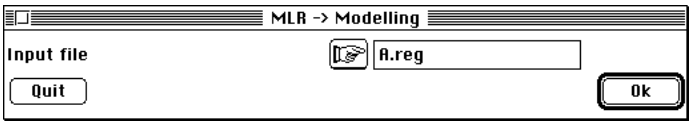
```
-------------------------------------------------------
New TEXT file A.reg contains the parameters:
----> Explanatory variables: Deb [39][3]
----> Dependent variable file: Rh [39][15]
----> Row weight file: Uniform weight
-------------------------------------------------------
```

L'option enregistre ces paramètres dans un fichier texte. On entre dans les autres options du module par le fichier ---.reg ainsi créé.

LinearReg : MLR -> Modelling

Régression multiple classique à coefficient constant : calcul des valeurs prédites.

L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par LinearReg : Initialize.

Utiliser l'exemple[1] introduit dans la fiche de LinearReg : Initialize :

```
Multiple Linear Regression
------------------------------------------
```

Les moyennes et variances des variables explicatives sont éditées :

```
Multiple Linear Regression
------------------------------------------
Explanatory variable file: Deb
It has 39 rows and 3 columns
Var.|  Mean    | Variance |
   1| 2.796e+02| 2.190e+04|
   2| 1.046e+02| 6.023e+03|
   3| 1.806e+02| 3.560e+04|
------------------------------------------
```

Les moyennes, variances et pourcentages de variance expliquée des variables à expliquer (régression multiple indépendante pour chacune d'entre elles sur le même ensemble d'explicatives) sont éditées :

```
Dependent variable file: Rh
It has 39 rows and 15 columns
R2 = Squared multiple correlation coefficient
Var.|  Mean    | Variance |    R2    |
   1| 1.377e+01| 5.310e+01| 5.837e-01|
   2| 1.310e+01| 3.016e+01| 5.292e-01|
•••
  14| 3.582e+00| 1.572e+01| 6.652e-01|
  15| 3.087e+00| 5.424e+00| 2.986e-01|
------------------------------------------
```

Les valeurs prédites par des modèles du type $y = a_1x_1 + ... + a_px_p + b$ sont conservées :

```
File A.MLRmod has 39 rows and 15 columns
It contains the linear models resulting
from separate multiple linear regression of each dependent variable
upon the set of explanatory variables
File :A.MLRmod
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 6.140e+00| 2.250e+01|
|   2| 5.019e+00| 1.963e+01|
•••
|  14| 6.267e-01| 1.580e+01|
|  15| 1.576e+00| 7.106e+00|
|----|----------|----------|
```

Les résidus des prévisions (données - modèles) sont conservées :

```
File A.MLRres has 39 rows and 15 columns
File :A.MLRres
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.154e+01| 9.008e+00|
|   2|-9.298e+00| 7.477e+00|
•••
```

```
|  14|-4.189e+00| 1.113e+01|
|  15|-2.706e+00| 5.955e+00|
|----|----------|----------|
```

Les coefficients des variables dans les modèles sont conservés :

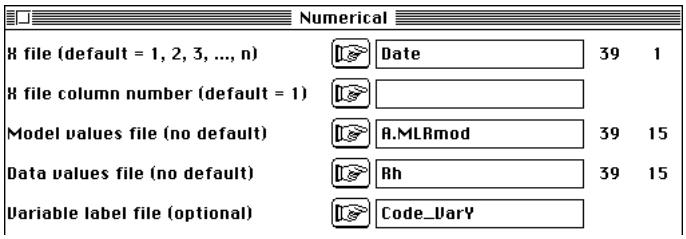
```
File A.MLRw1 has 3 rows and 15 columns
It contains the regression coefficients
Rows : explanatory variables / Columns : dependent variables
Models for normalized (mean = 0 / variance =1) variables
File :A.MLRw1
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-3.686e-01| 5.709e-01|
|   2|-5.751e-01| 4.258e-01|
•••
|  14|-9.725e-02| 6.995e-01|
|  15| 1.664e-01| 3.923e-01|
|----|----------|----------|
```

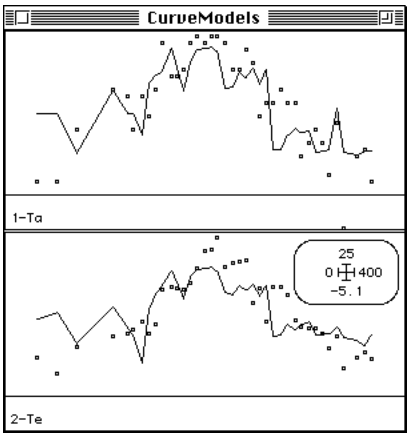
Pour faciliter la lecture, ces coefficients sont ceux des modèles écrits sous la forme :

$$\frac{\mathbf{y}-m(\mathbf{y})}{\sqrt{var(\mathbf{y})}} = \tilde{a}_1 \frac{\mathbf{x}_1 - m(\mathbf{x}_1)}{\sqrt{var(\mathbf{x}_1)}} + \ldots + \tilde{a}_p \frac{\mathbf{x}_p - m(\mathbf{x}_p)}{\sqrt{var(\mathbf{x}_p)}}$$

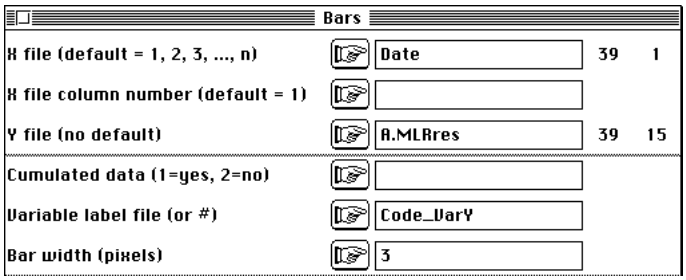
où *m* et *var* sont les moyennes et les variances.

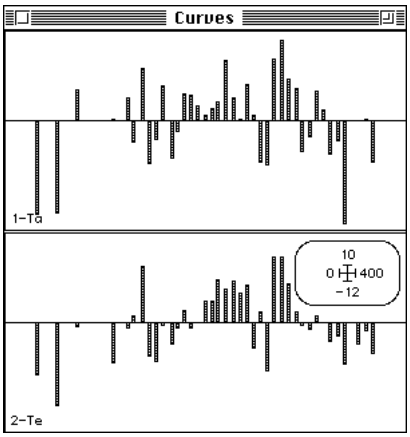
Représenter (CurveModels : Numerical) les modèles et les données pour les deux premières variables :



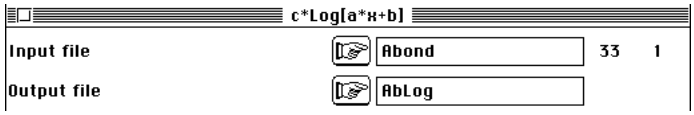


Représenter (Curves : Bars) les résidus :

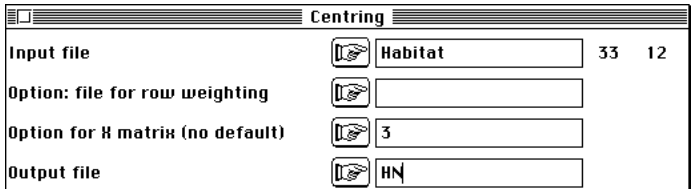


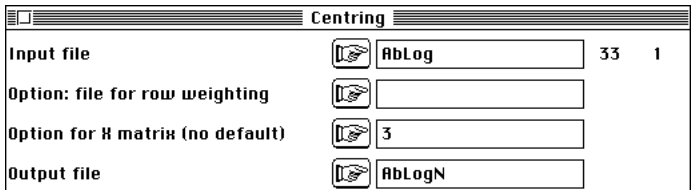


Les difficultés d'utilisation de la régression multiple ne doivent pas être sous-estimées. Elles proviennent en général du rapport entre le nombre de lignes et le nombre de variable du tableau des explicatives. Utiliser la carte Truite[2] pour obtenir les fichiers Habitat (33-12), Abond (33-1) et label_Var. Transformer la variable à prédire (Bin->Bin) :

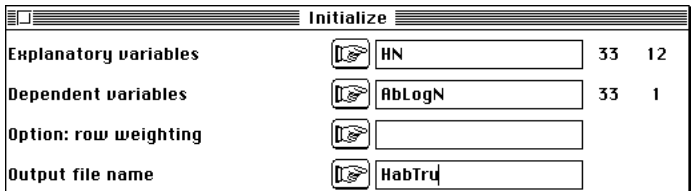


Normaliser les variables explicatives et la variable à expliquer (Bin->Bin) :

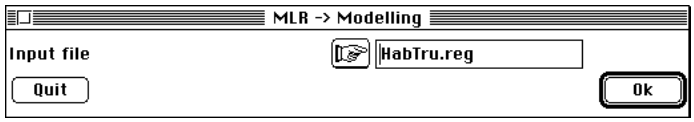




Associer par LinearReg : Initialize :



Utiliser la présente option :




```
Multiple Linear Regression
------------------------------------------
```

Les variables sont normalisées :

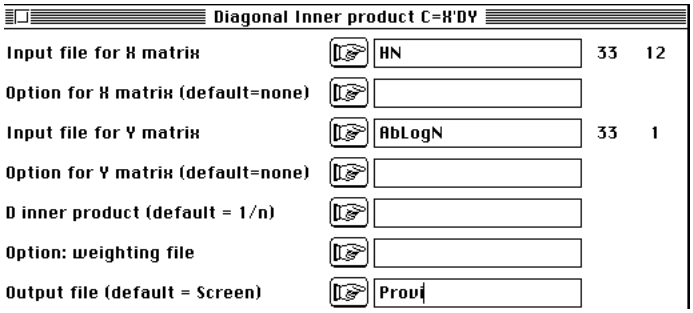
```
Explanatory variable file: HN
It has 33 rows and 12 columns
Var.|  Mean    | Variance |
   1|-2.709e-09| 1.000e+00|
•••
------------------------------------------
Dependent variable file: AbLogN
It has 33 rows and 1 columns
R2 = Squared multiple correlation coefficient
Var.|  Mean    | Variance |    R2    |
   1| 2.484e-09| 1.000e+00| 8.387e-01|
```

La régression donne 84% de variance expliquée.

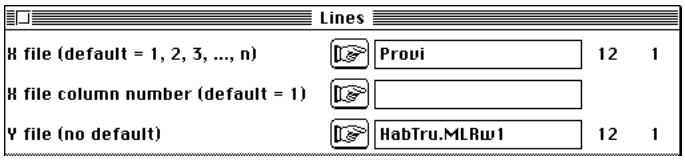
```
------------------------------------------
File HabTru.MLRmod has 33 rows and 1 columns
•••
File HabTru.MLRres has 33 rows and 1 columns
•••
------------------------------------------
File HabTru.MLRw1 has 12 rows and 1 columns
It contains regression coefficients
Rows : explanatory variables / Columns : dependent variables
Models for normalized (mean = 0 / variance =1) variables
```

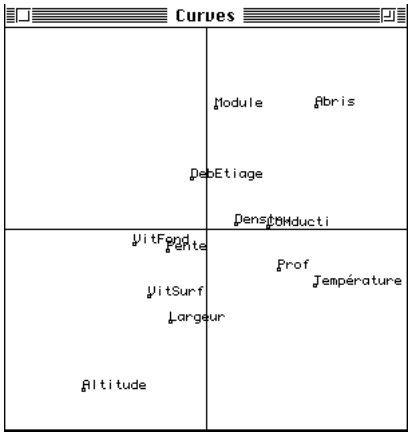
```
File :HabTru.MLRw1
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-8.012e-01| 6.070e-01|
|----|----------|----------|
```

Calculer les corrélations explicatives-expliquées (MatAlg : Diagonal Inner Product) :



Comparer les deux familles de paramètres par Curves : Lines :





L'incohérence entre coefficients de régression (rôle des variables dans le modèle de prévision global) et coefficients de corrélation (rôle des variables dans les prévisions univariées séparées) est l'indication qu'une limitation du nombre d'explicatives s'impose.

⚠ Les coefficients de régression sont disponibles si la matrice des covariances entre explicatives est inversible (variables non redondantes). Dans le cas contraire, les explicatives ne sont pas indépendantes et le modèle n'est pas identifiable. Il existe cependant (solution unique obtenue par projection) et il est calculé avec un inverse généralisé.

Voir aussi la fiche 2 du fascicule 5 de la documentation thématique.

[1] 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.

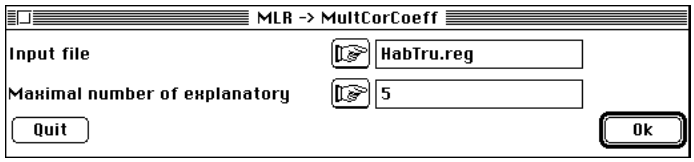
[2] Baran, P., Delacoste, M., Lascaux, J.M. & Belaud, A. (1993) Relations entre les caractéristiques de l'habitat et les populations de truites communes (*Salmo trutta* L.) de la vallée de la Neste d'Aure. *Bull. Fr. Pêche Piscic* : 331, 321-340.

LinearReg : MLR -> MultCorCoeff

Utilitaire pour la sélection de variables en régression multiple.

L'objectif est de calculer le pourcentage de variance expliquée pour toutes les combinaisons de 1 variable explicative, de 2 variables explicatives, ..., de *k* variables explicatives.

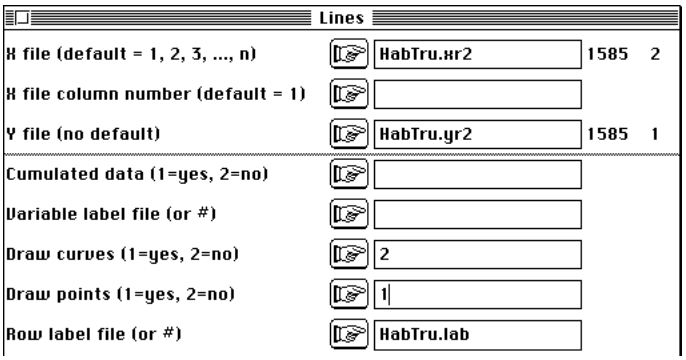
L'option utilise une seule fenêtre de dialogue :

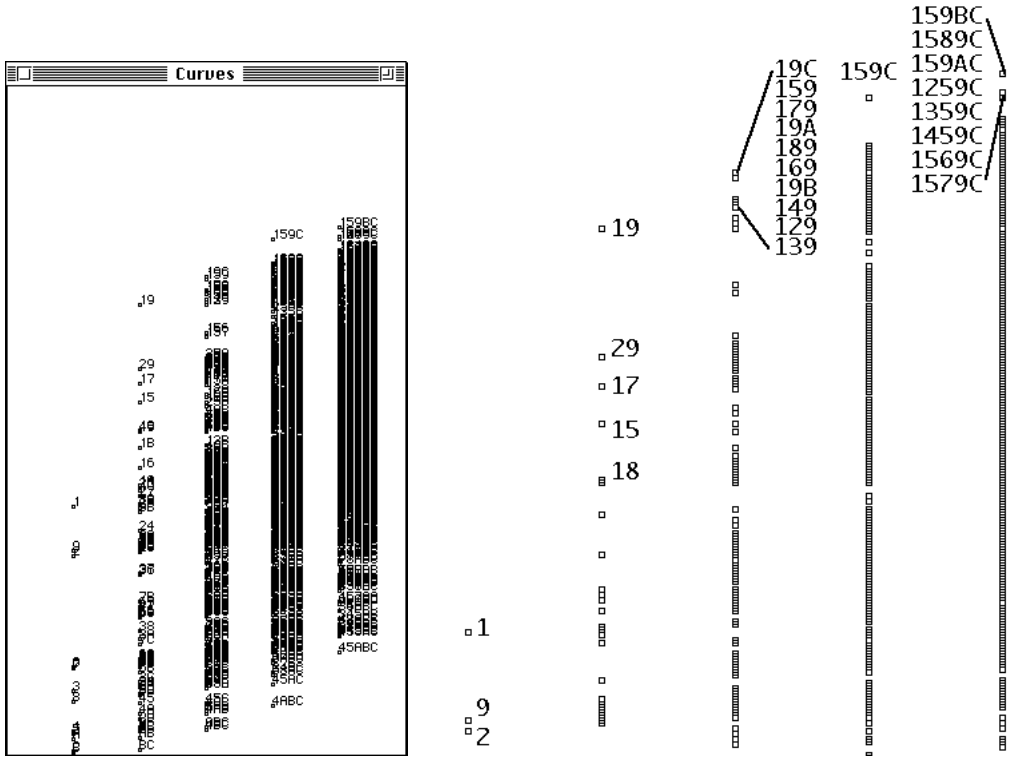


Fichier des paramètres créé par LinearReg : Initialize.

Nombre maximum d'explicatives à prendre en compte (*k*). Se méfier de l'explosion combinatoire du nombre de combinaisons.

Utiliser l'exemple mis en place dans la fiche de LinearReg : MLR -> Modelling :






```
Multiple correlation on all possible combinations of variables
--------------------------------------------
Explanatory variable file: HN
```

```
It has 33 rows and 12 columns
----------------------------------------
Dependent variable file: AbLogN
It has 33 rows and 1 columns
Maximum number for explanatory variables: 5
----------------------------------------
Multiple correlation coefficient (R2): HabTru.yr2
It has 1585 rows and 1 columns
On one row, R2 of each dependent variable for a given set of explanatory variables
Label of the combinations in file HabTru.lab with 1585 rows
Number of explanatory variables used in column 1 of file HabTru.xr2 with 1585 rows
Rank of the projection subspace in column 2 of file HabTru.xr2 with 1585 rows
----------------------------------------
```

Chaque combinaison de variables donne une valeur du carré de corrélation multiple et une étiquette identifiant cette combinaison. On peut représenter, à l'aide des fichiers créés par cette option, l'évolution du la variance expliquée en fonctions du nombre de variables utilisées. Employer Curves : Lines (ci-dessus). La paire d'explicatives 1-9 combine l'économie de variables et la précision du résultat.

La régression multiple n'a pas de sens dès que le nombre de variables explicatives dépassent le nombre d'individus. Dans ce cas, la version PLS s'impose[1]. La régression multiple pose des problèmes considérables si les explicatives sont fortement corrélées et on lui préfère alors la régression sur composantes[2] (voir OrthoVar). Les deux stratégies entretiennent des relations étroites.

[1] Cramer, R.D. III, Bunce, J.D., Patterson, D.E. & Frank, I.E. (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships* : 7, 18-25.

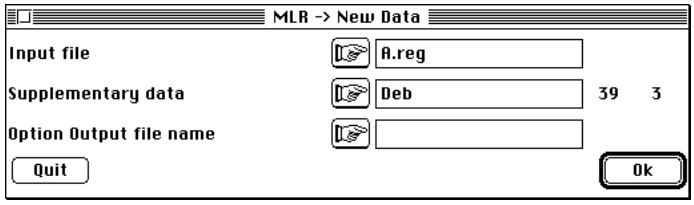
[2] Morzuch, B. J. & Ruark, G. A. (1991) Principal components regression to mitigate the effects of multicollinearity. *Forest Science* : 37, 191-199.

LinearReg : MLR -> New Data

Utilitaire en régression multiple.

L'option permet l'extension de la prédiction à des valeurs supplémentaires des variable explicatives.

L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par LinearReg : Initialize.

Fichier des valeurs supplémentaires des variables explicatives.

Nom du fichier de sortie (par défaut, il utilise le nom du fichier d'entrée).

Utiliser l'exemple introduit dans la fiche de LinearReg : Initialize :

```
Multiple Linear Regression
------------------------------------------
Explanatory variable file: Deb
It has 39 rows and 3 columns
------------------------------------------
Dependent variable file: Rh
It has 39 rows and 15 columns
------------------------------------------
New data file: Deb
It has 39 rows and 15 columns
------------------------------------------
File Deb.MLRSup has 39 rows and 15 columns
It contains the linear models resulting
from separate multiple linear regression of each dependent variable
upon the set of explanatory variables
File :Deb.MLRSup
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 6.140e+00| 2.250e+01|
|   2| 5.019e+00| 1.963e+01|
•••
|  14| 6.267e-01| 1.580e+01|
|  15| 1.576e+00| 7.106e+00|
|----|----------|----------|
```

Vu le choix du fichier des valeurs supplémentaires, le contenu du ficher Deb.MLRSup est évidemment identique au contenu du fichier A.MLRmod créé par LinearReg : MLR -> Modelling.

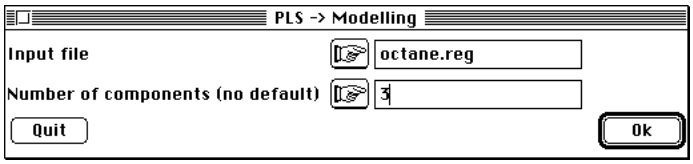
L'exécution du présent module suppose l'existence du fichier ---.MLRw1 (analyse de titre ---.reg) créé par LinearReg : MLR -> Modelling. Ceci implique qu'on a utilisé au préalable cette option et que la matrice des corrélations des explicatives est de plein rang. Si ce n'est pas le cas, ce fichier n'est pas disponible et l'opération demandée est impossible.

LinearReg : PLS -> Modelling

Régression PLS : méthode alternative à la régression multiple classique dès que le nombre des explicatives est grand, en particulier supérieur au nombre d'individus. La méthode est fortement conseillée pour des explicatives fortement corrélées.

Le module exécute la régression PLS ou régression partiellement aux moindres carrés (Partial Least Squares) de première génération. Inventée en chimiométrie, dont elle est un standard méthodologique (synthèse complète dans Lindgren[1]), la régression PLS gagne à être connue en écologie. L'algorithme utilisé est décrit par Ter Braak & Juggins (1993, p. 487)[2].

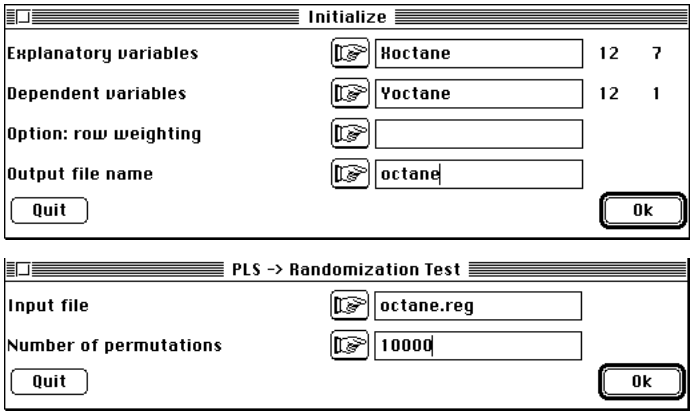
L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par LinearReg : Initialize.

Nombre de composantes utilisées (voir LinearReg : PLS -> Randomization test).

Utiliser la carte Octane de la pile ADE-4•Data pour obtenir les fichiers Xoctane (12-7), Yoctane (12-1) et Code_Var_X. Initialiser et tester le nombre de composantes :




```
PLS1 - Permutation test
----------------------------------------
Explanatory variable file: Xoctane
It has 12 rows and 7 columns
----------------------------------------
Dependent variable file: Yoctane
It has 12 rows and 1 columns
------------------------------
------ Vari number:    1 ------
------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     0| 0.000e+00|
|   2| 10000|    36| 3.600e-03|
|   3| 10000|   114| 1.140e-02|
|   4| 10000|  7103| 7.103e-01|
|   5| 10000|  1726| 1.726e-01|
|   6| 10000|  7915| 7.915e-01|
|   7| 10000|  9615| 9.615e-01|
```

On doit donc garder trois composantes dans l'exécution de la présente option :

```
PLS1 - Modelling
----------------------------------------
Explanatory variable file: Xoctane
It has 12 rows and 7 columns
```

```
----------------------------------------
Dependent variable file: Yoctane
It has 12 rows and 1 columns
----------------------------------------

|--------------| Col:   1 |--------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 9.236e-01| 9.236e-01| 9.236e-01|
|   2| 7.641e-02| 5.275e-02| 6.904e-01| 9.763e-01|
|   3| 2.366e-02| 1.421e-02| 6.008e-01| 9.906e-01|
```

Ce tableau se lit progressivement. La variance initiale de la variable à prédire vaut 1. La première composante en explique 92%. Il reste un résidu de variance 0.0764. 69% de ce résidu sont expliqués par une seconde combinaison des explicatives. Ceci porte le total de variance expliquée à 97.6 % de la variance initiale. Il reste un résidu de prédiction de variance 0.0237. 60% de cette quantité sont expliqués par une troisième composante. Ceci porte le pourcentage de variance expliquée à 99.06 %.

Les coefficients du modèle global (réunissant les trois composantes) sont conservés :

```
File octane.PLSw1 has 7 rows and 1 columns
It contains coefficients
File :octane.PLSw1
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-2.932e-01| 4.564e-01|
|----|----------|----------|
```

Editer ces coefficients et noter que le modèle s'écrit sur les variables normalisées :

$$\frac{\mathbf{y}-m(\mathbf{y})}{\sqrt{var(\mathbf{y})}} = \tilde{a}_1 \frac{\mathbf{x}_1 - m(\mathbf{x}_1)}{\sqrt{var(\mathbf{x}_1)}} + \ldots + \tilde{a}_p \frac{\mathbf{x}_p - m(\mathbf{x}_p)}{\sqrt{var(\mathbf{x}_p)}}$$

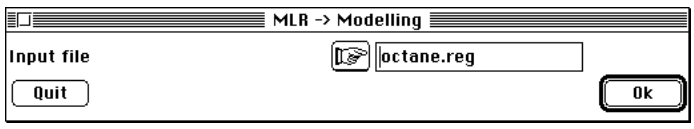
$$\tilde{a}_1 = -0.139 \quad \tilde{a}_2 = -0.2087 \quad \tilde{a}_3 = -0.1376 \quad \tilde{a}_4 = -0.2932$$

$$\tilde{a}_5 = -0.0384 \quad \tilde{a}_6 = 0.4564 \quad \tilde{a}_7 = -0.1434$$

Ces résultats sont ceux de [3] (p. 29) obtenu sur ces données[4] avec le logiciel SIMCA[5]. Le modèle et le résidu sont conservés :

```
----------------------------------------
File octane.PLSmod has 12 rows and 1 columns
It contains components models
File :octane.PLSmod
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 8.150e+01| 9.759e+01|
|----|----------|----------|
----------------------------------------
File octane.PLSres has 12 rows and 1 columns
It contains residuals
File :octane.PLSres
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-9.414e-01| 1.141e+00|
|----|----------|----------|
```

La comparaison avec la régression ordinaire (MLR) et la régression sur composantes (PCR) est intéressante. Pour la régression ordinaire, on a :

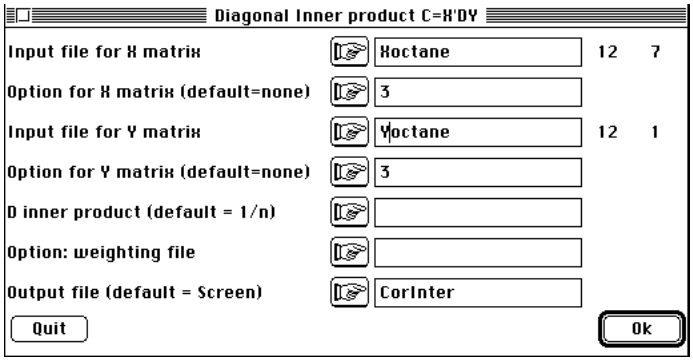


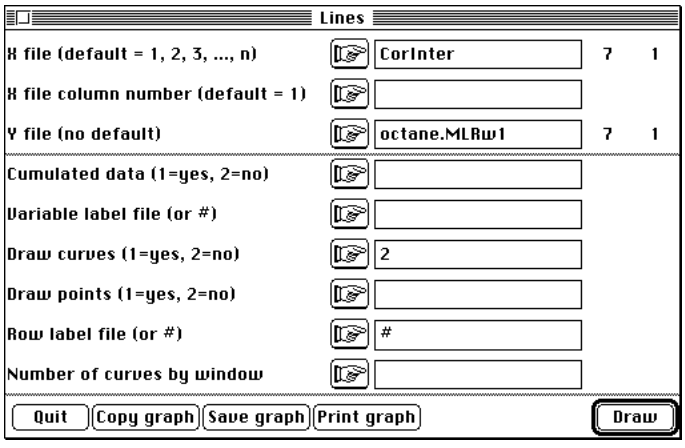

```
R2 = Squared multiple correlation coefficient
Var.|  Mean    | Variance |    R2    |
   1| 8.858e+01| 3.898e+01| 9.925e-01|
----------------------------------------
```

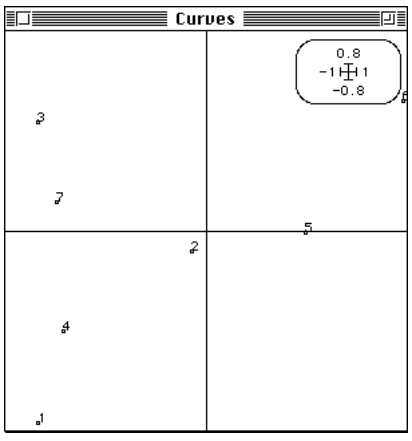
```
File octane.MLRmod has 12 rows and 1 columns
It contains the linear models resulting
from the separate multiple linear regression of each dependent variable
upon the set of explanatory variables
File :octane.MLRmod
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 8.140e+01| 9.802e+01|
|----|----------|----------|

File octane.MLRres has 12 rows and 1 columns
It contains (data - model) matrix
File :octane.MLRres
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-8.141e-01| 1.207e+00|
|----|----------|----------|
File octane.MLRw1 has 7 rows and 1 columns
It contains regression coefficients
Rows : explanatory variables / Columns : dependent variables
Models for normalized (mean = 0 / variance =1) variables
File :octane.MLRw1
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-7.697e-01| 5.206e-01|
|----|----------|----------|
```

99.2 % de variance sont expliqués (c'est l'optimum puisque la régression multiple maximise ce critère). On compare les coefficients des deux modèles à travers les corrélations explicatives/expliquée (MatAlg) :

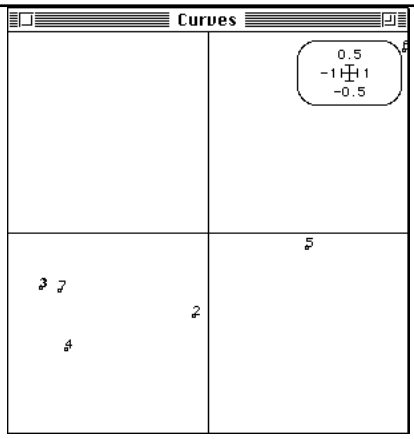


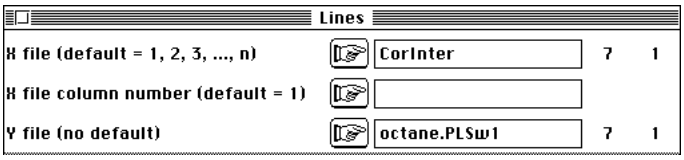




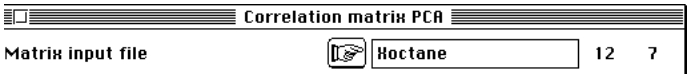
Une corrélation négative peut s'accompagner d'un poids dans le modèle positif. L'incohérence impose la limitation des variables dans le modèle de régression classique. La régression PLS permet d'échapper à ce phénomène.

Elle assure la cohérence entre coefficient de régression et coefficient de corrélation au prix d'une perte très faible de qualité de prédiction :





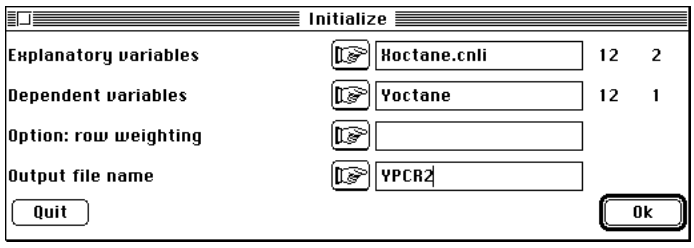
Pour la régression sur composantes, faire l'ACP du tableau des explicatives (PCA) :



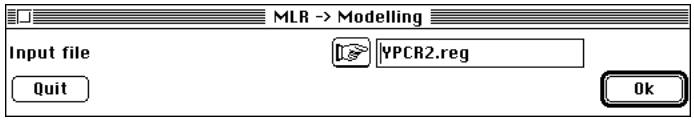
On garde deux axes et on note les corrélations fortes entre explicatives :

```
---------------------- Correlation matrix -------------------
[  1] 1000
[  2]  104 1000
[  3] 1000  101 1000
[  4]  371 -537  374 1000
[  5] -548 -293 -548 -211 1000
[  6] -805 -191 -805 -646  463 1000
[  7]  603 -590  607  916 -274 -656 1000
------------------------------------------------------------
Total inertia:        7
-----------------------
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +4.0255E+00 +0.5751 +0.5751   |02   +1.9104E+00 +0.2729 +0.8480 |
03   +5.7263E-01 +0.0818 +0.9298   |04   +4.7818E-01 +0.0683 +0.9981 |
05   +1.3237E-02 +0.0019 +1.0000   |06   +8.1139E-05 +0.0000 +1.0000 |
07   +0.0000E+00 +0.0000 +1.0000
```

Les deux premiers axes conservent 85 % d'inertie. On utilise les coordonnées factorielles comme prédicteurs :




```
-----------------------------------------------------
New TEXT file YPCR2.reg contains the parameters:
----> Explanatory variables: Xoctane.cnli [12][2]
----> Dependant variable file: Yoctane [12][1]
----> Row weight file: Uniform weight
-----------------------------------------------------
```




```
Multiple Linear Regression
-----------------------------------------
Explanatory variable file: Xoctane.cnli
It has 12 rows and 2 columns
Var.|  Mean     | Variance |
   1|-2.484e-09| 4.026e+00|
   2|-3.849e-08| 1.910e+00|
-----------------------------------------
```

```
Dependent variable file: Yoctane
It has 12 rows and 1 columns
R2 = Squared multiple correlation coefficient
Var.|  Mean    | Variance |    R2    |
   1| 8.858e+01| 3.898e+01| 9.071e-01|
----------------------------------------
File YPCR2.MLRmod has 12 rows and 1 columns
It contains the linear models resulting
from the separate multiple linear regression of each dependent variable
upon the set of explanatory variables
File :YPCR2.MLRmod
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 8.164e+01| 9.629e+01|
|----|----------|----------|
----------------------------------------
File YPCR2.MLRres has 12 rows and 1 columns
It contains (data - model) matrix
File :YPCR2.MLRres
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-3.036e+00| 4.286e+00|
|----|----------|----------|
```

91 % de variance sont expliquée au lieu de 99 %. On ne peut pas retrouver le rôle des variables dans ce modèle.

⚠ On s'en tient ici à la PLS de première génération (une seule variable expliquée). Ceci recouvre la quasi totalité des cas pratiques. La PLS de deuxième génération est à l'ACPVI (Module Projectors) ce que la régression PLS de première génération est à la régression multiple. On préfèrera l'analyse de co-inertie pour étudier les liens entre deux paquets de variables (Module CoInertia).

L'option limite les sorties du programme à sa fonction principale de construction d'un prédicteur.

[1] Lindgren, F. (1994) *Third generation PLS. Some elements and applications*. Research Group for Chemometrics. Department of Organic Chemistry. Umeå University. S-901 87 Umeå, Sweden. ISBN 91-7174-911-X, 1-57 & 5 papers.

[2] Ter_Braak, C.J.T. & Juggins, S. (1993) Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* : 269/270, 485-502.

[3] Tenenhaus, M., Gauchi, J.P. & Ménardo, C. (1995) Régression PLS et applications. *Revue de Statistique Appliquée* : 43, 7-63.

[4] Kettaneh-Wold, N. (1992) Analysis of mixture data with partial least squares. *Chemometrics and Intelligent Laboratory Systems* : 14, 57-69.

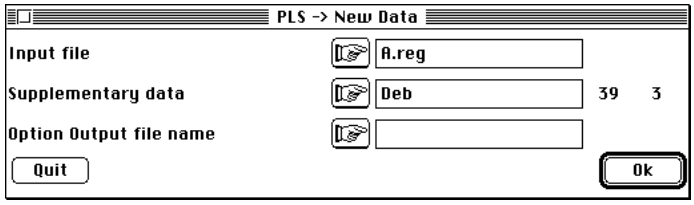
[5] SIMCA (1991) "Soft Independant Modeling of Class Analogy", Version 4.3R, Umetri AB Box, 1456, S-901 24 Umea.

LinearReg : PLS -> New Data

Utilitaire en régression PLS.

L'option permet l'extension de la prédiction à des valeurs supplémentaires des variable explicatives.

L'option utilise une seule fenêtre de dialogue :

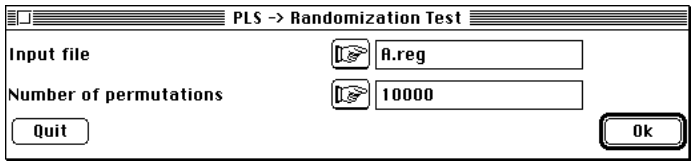


Fichier des paramètres créé par LinearReg : Initialize.

Fichier des valeurs supplémentaires des variables explicatives.

Nom du fichier de sortie (par défaut, il utilise le nom du fichier d'entrée).

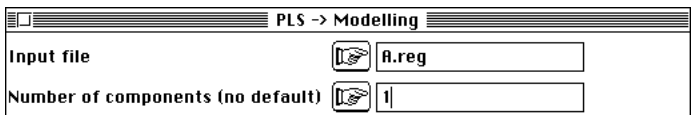
Utiliser l'exemple introduit dans la fiche de LinearReg : Initialize. Chercher le nombre de composantes à utiliser dans la régression :




```
PLS1 - Permutation test
-----------------------------------------
Explanatory variable file: Deb
It has 39 rows and 3 columns
-----------------------------------------
Dependent variable file: Rh
It has 39 rows and 15 columns
------------------------------
------ Vari number:    1 ------
------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     0| 0.000e+00|
|   2| 10000|  7251| 7.251e-01|
|   3| 10000|  4691| 4.691e-01|
------------------------------
------ Vari number:    2 ------
------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     0| 0.000e+00|
|   2| 10000|  4769| 4.769e-01|
|   3| 10000|  2601| 2.601e-01|
------------------------------
```

•••

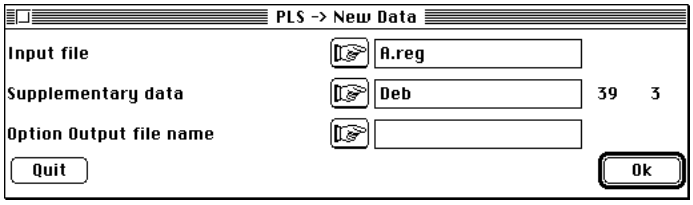
Exécuter la régression :




```
PLS1 - Modelling
-----------------------------------------
Explanatory variable file: Deb
It has 39 rows and 3 columns
-----------------------------------------
```

```
Dependent variable file: Rh
It has 39 rows and 15 columns
----------------------------------------
|--------------| Col:   1 |--------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 5.724e-01| 5.724e-01| 5.724e-01|
|--------------| Col:   2 |--------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 4.982e-01| 4.982e-01| 4.982e-01|
|--------------| Col:   3 |--------------------|
•••
```

Utiliser des individus supplémentaires :




```
PLS Regression
----------------------------------------
Explanatory variable file: Deb
It has 39 rows and 3 columns
----------------------------------------
Dependent variable file: Rh
It has 39 rows and 15 columns
----------------------------------------
New data file: Deb
It has 39 rows and 15 columns
----------------------------------------
File Deb.PLSSup has 39 rows and 15 columns
It contains the linear models resulting
from the separate PLS regression of each dependant variable
upon the set of explanatory variables
File :Deb.PLSSup
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 6.148e+00| 2.202e+01|
|   2| 7.213e+00| 1.918e+01|
•••
|  14| 5.600e-01| 1.413e+01|
|  15| 1.681e+00| 6.891e+00|
|----|----------|----------|
```

Vu le choix du fichier des valeurs supplémentaires, le contenu du ficher Deb.PLSSup est évidemment identique au contenu du fichier A.PLSmod créé par LinearReg : PLS -> Modelling.

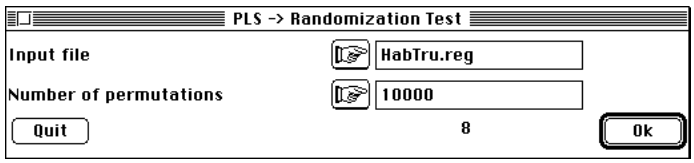
⚠ L'exécution du présent module suppose l'existence du fichier ---.PLSw1 (analyse de titre ---.reg) créé par LinearReg : PLS -> Modelling.

LinearReg : PLS -> Randomization Test

Utilitaire pour le choix du nombre d'itérations dans une régression PLS.

La régression PLS est une méthode itérative. A chaque pas on cherche une combinaison linéaire explicative, on fait la prévision linéaire, on enlève la prévision de la variable à expliquer et la combinaison prédictive des variables explicatives. On obtient une nouvelle variable à prédire, résidu du tour précédent et de nouvelles variables prédictrices indépendantes de la combinaison déjà utilisée. La question porte donc sur le nombre d'itérations à utiliser. Par tests de permutation réinitialisés à chaque tour, on compte la fréquence des permutations aléatoires qui donnerait un pourcentage d'explication aussi bon. On ne retiendra pour LinearReg : PLS -> Modelling que le nombre d'itérations franchement significatives.

L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par LinearReg : Initialize.

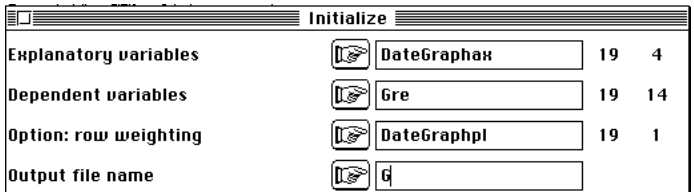
Nombre de simulations utilisées (*n*). Le compteur affiche le numéro de l'itération qui est fixé à 8 maximum. A chaque itération on fait *n* permutations.

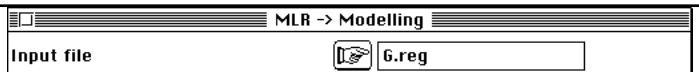
Utiliser l'exemple mis en place dans la fiche de LinearReg : MLR -> Modelling :

```
PLS1 - Permutation test
----------------------------------------
Explanatory variable file: HN
It has 33 rows and 12 columns
----------------------------------------
Dependent variable file: AbLogN
It has 33 rows and 1 columns
------------------------------
------ Vari number:    1 ------
------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     0| 0.000e+00|
|   2| 10000|  1572| 1.572e-01|
|   3| 10000|  4362| 4.362e-01|
|   4| 10000|  7130| 7.130e-01|
|   5| 10000|  3277| 3.277e-01|
|   6| 10000|  5763| 5.763e-01|
•••
```

Au premier tour, on ne trouve aucune permutation aléatoire qui dépasse l'observation, signe que la prédiction a un sens. Au second tour on en trouve 16%, ce qui implique qu'il n'est pas justifié de tenir compte de plus d'une seule itération. Le nombre de pas de cette procédure est limitée à 8 s'il y a plus de 8 variables explicatives. Il est limité au nombre d'explicatives dans le cas contraire. En utilisant toutes les composantes possibles on obtient la régression multiple ordinaire.

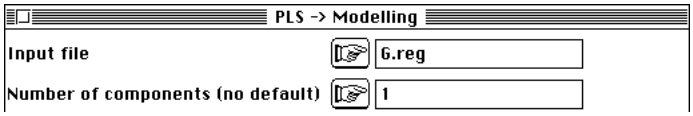
Si les variables explicatives sont indépendantes régression PLS et régression MLR donnent **exactement** les mêmes résultats. Utiliser l'exemple mis en place dans la fiche de OrthoVar : Initialize. Exécuter LinearReg : Initialize :






```
Multiple Linear Regression
-----------------------------------------
Explanatory variable file: DateGraphax
It has 19 rows and 4 columns
Var.|  Mean    | Variance |
   1|-3.488e-09| 1.000e+00|
   2|-3.060e-09| 1.000e+00|
   3| 1.642e-08| 1.000e+00|
   4|-8.027e-09| 1.000e+00|
-----------------------------------------
Dependent variable file: Gre
It has 19 rows and 14 columns
R2 = Squared multiple correlation coefficient
Var.|  Mean    | Variance |    R2    |
   1| 4.596e+01| 9.940e+02| 7.932e-01|
   2| 1.812e+01| 2.401e+02| 9.071e-01|
•••
  13| 1.183e+01| 1.249e+02| 9.677e-01|
  14| 1.257e+01| 8.986e+01| 9.659e-01|
-----------------------------------------
```


```
-------------------------------
------ Vari number:    1 ------
-------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     3| 3.000e-04|
|   2| 10000| 10000| 1.000e+00|
|   3| 10000| 10000| 1.000e+00|
|   4| 10000| 10000| 1.000e+00|
-------------------------------
------ Vari number:    2 ------
-------------------------------
|Step|Nrepet|X>Xobs| Frequency|
|   1| 10000|     0| 0.000e+00|
|   2| 10000| 10000| 1.000e+00|
|   3| 10000| 10000| 1.000e+00|
|   4| 10000| 10000| 1.000e+00|
•••
```




```
PLS1 - Modelling
-----------------------------------------
Explanatory variable file: DateGraphax
It has 19 rows and 4 columns
Dependent variable file: Gre
It has 19 rows and 14 columns
-----------------------------------------
|---------------| Col:   1 |---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 7.932e-01| 7.932e-01| 7.932e-01|
|---------------| Col:   2 |---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 9.071e-01| 9.071e-01| 9.071e-01|
•••
|---------------| Col:  13 |---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 9.677e-01| 9.677e-01| 9.677e-01|
|---------------| Col:  14 |---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 9.659e-01| 9.659e-01| 9.659e-01|
```

Plus les explicatives sont corrélées et plus l'usage de la PLS s'impose devant celui de la MLR.

PLS et MLR sont complètement associées dans la constitution de modèles linéaires. Les liaisons non linéaires sont ignorées par l'une et l'autre des deux méthodes. Il en est de même pour PCR (régression sur composantes dans OrthoVar).

L'option utilise un test de permutation[1] dans la logique des couplages de tableaux[2] (exemple dans [3]) proche de la validation croisée[4] utilisée dans SIMCA d'après [5].

La régression PLS est très étudiée et utilisée en chimiométrie [6]. Les écologues trouveront dans cette discipline des dizaines de référence.

La première composante de la PLS est très liée avec la première coordonnée de l'analyse de co-inertie [7], mais la logique de l'extension de la PLS à deux tableaux diffère sensiblement de l'analyse de co-inertie de deux tableaux.

[1] Good, P. (1994) Permutation tests. Springer-Verlag, New-York. 1-228.

[2] Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1994) Refined approximations to permutation tests for multivariate inference. *ComputationalStatistics and Data Analysis*, 20, 643-656.

[3] Fraile, L., Escoufier, Y. & Raibaut, A. (1993) Analyse des correspondances de données planifiées : Etude de la chémotaxie de la larve infestante d'un parasite. *Biometrics* : 49, 1142-1153.

[4] Wold , S. (1978) Cross-validation estimation of the number of components in factor and principal components models. *Technometrics* : 20, 397-405.

Cramer, R.D. III, Bunce, J.D., Patterson, D.E. & Frank, I.E. (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships* : 7, 18-25.

[5] Gauchi, J.P. (1995) Utilisation de la régression PLS pour l'analyse des plans d'expériences en chimie de formulation. *Revue de Statistique Appliquée* : 43, 65-89.

[6] Geladi, P. & Kowalski, B.R. (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta* : 1, 185, 19-32.

Geladi, P. (1988) Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics* : 2, 231-246.

Höskuldsson, A. (1988) PLS regression methods. *Journal of Chemometrics* : 2, 211-228.

Wold, S. (1995) PLS for multivariate linear modeling. In : *Chemometric Methods in Molecular Design*. van der Waterbeemd. (Ed.) VCH, Weinhein, Germany. 195-218.

[7] Devillers, J. & Chessel, D. (1994) Graphical Analysis as an Aid in Medicinal Chemistry. In : *Chemometric Methods in Molecular Design*. van der Waterbeemd. (Ed.) VCH, Weinhein, Germany. 165-176.