UNIWIN VERSION 10.7.0

ANALYSE NIPALS

Révision : 26/08/2026

Définition 1
Entrée des données 2
Données manquantes 3
Exemple 1 : Fichier VEHICULE sans validation croisée 3
L'option Rapports 6
L'option Graphiques 8
Exemple 2 : Fichier VEHICULE avec validation croisée 13
Exemple 3 : Fichier VEHICULE2 sans validation croisée 16
Exemple 4 : Fichier VEHICULE2 avec validation croisée 20
Les variables internes créées par la procédure 21
Références 22

Définition

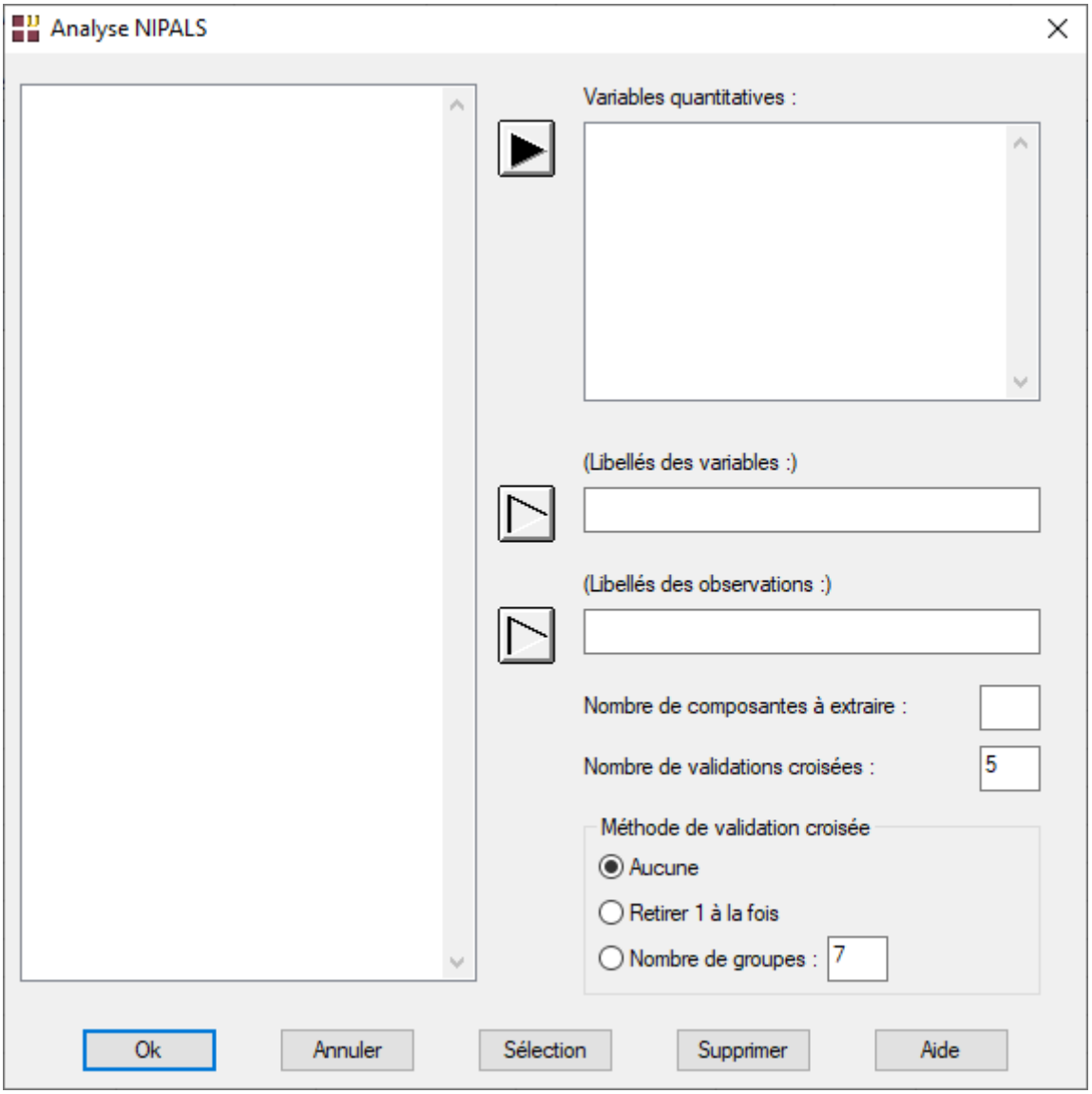
L'analyse NIPALS (Nonlinear Iterative Partial Least Squares) est une méthode séquentielle d'analyse en composantes principales autorisant la présence de données manquantes dans les données. Cette méthode a été initialement présentée par Herman Wold en 1966. Par la validation croisée (méthode de Krzanowski), elle permet également de définir le nombre de composantes à retenir.

Le mont NIPALS

Elle affiche un rapport présentant les diverses statistiques calculées ainsi que les graphiques des R2 cumulés, des Q2 cumulés, des plans factoriels des observations et des poids des variables, des cercles des corrélations ainsi que des diagrammes des contributions, des cosinus carrés cumulés, des distances carrées à l'origine et des distances normalisées au modèle.

Entrée des données

Cliquons sur l'icône NIPALS dans le ruban Décrire. La boîte de dialogue montrée ci-dessous s'affiche :



Cette boîte de dialogue permet de définir les variables quantitatives utilisées dans l'analyse.

Note : Les données sont automatiquement centrées et réduites.

Elle permet, en option, d'indiquer les noms des variables contenant les libellés des variables quantitatives et les libellés des observations.

Le nombre de composantes à extraire peut être précisé. Il est au plus égal au minimum des nombres d'observations et de variables. S'il n'est pas précisé, toutes les composantes sont automatiquement extraites.

Si la validation croisée est mise en œuvre pour déterminer le nombre de composantes à retenir, il est possible d'indiquer le nombre de validations croisées à réaliser et de préciser la méthode utilisée : « Retirer 1 à la fois » ou « Nombre de groupes ».

Données manquantes

Dans cette procédure les données manquantes sont permises et la procédure NIPALS les estimera en se basant sur le modèle développé.

Exemple 1 : Fichier VEHICULE sans validation croisée

Nous utiliserons le fichier VEHICULE pour illustrer ce premier exemple. Ce fichier sans données manquantes contient les informations suivantes concernant 24 modèles de voitures :

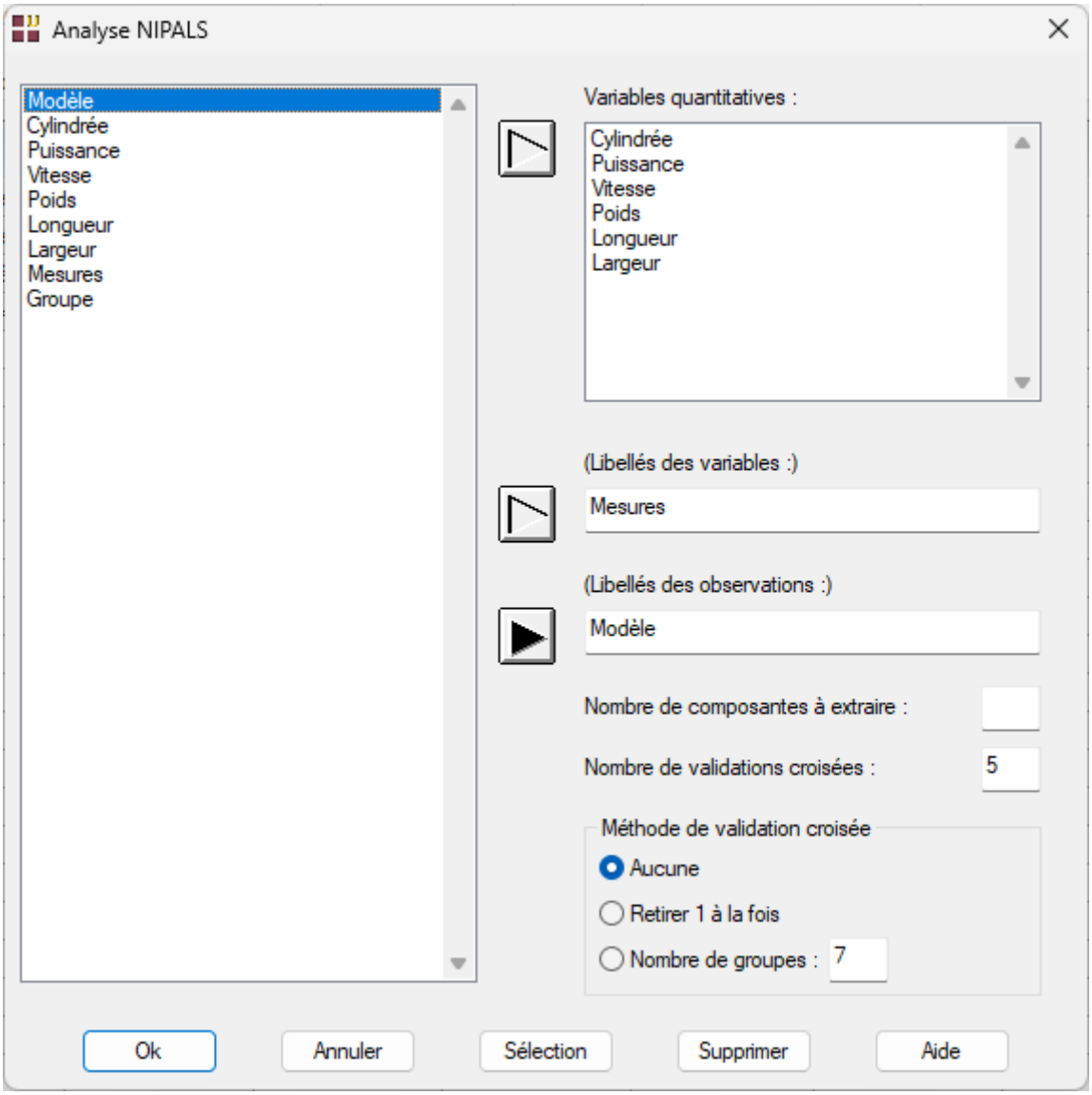
Données

	Modèle	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur	Mesures	Groupe
	Modèles	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur		
	Type = C	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = C	Type = N
	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 6	Longueur = 24
1	Honda-Civic	1396	90	174	850	369	166	Cylindrée	1
2	R19	1721	92	180	965	415	169	Puissance	1
3	Fiat-Tipo	1580	83	170	970	395	170	Vitesse	1
4	405	1769	90	180	1080	440	169	Poids	1
5	R21	2068	88	180	1135	446	170	Longueur	1
6	BX	1769	90	182	1060	424	168	Largeur	1
7	BMW-530i	2986	188	226	1510	472	175		1
8	Rover-827i	2675	177	222	1365	469	175		1
9	R25	2548	182	226	1350	471	180		1
10	Opel-Omega	1998	122	190	1255	473	177		1
11	405-Break	1905	125	194	1120	439	171		1
12	Ford-Sierra	1993	115	185	1190	451	172		1
13	BMW-325ix	2494	171	208	1300	432	164		2
14	Audi-90-Quattro	1994	160	214	1220	439	169		2
15	Ford-Scorpio	2933	150	200	1345	466	176		2
16	Espace	1995	120	177	1265	436	177		2
17	Nissan-Vanette	1952	87	144	1430	436	169		2
18	VW-Caravelle	2109	112	149	1320	457	184		2
19	Ford-Fiesta	1117	50	135	810	371	162		2
20	Fiat-Uno	1116	58	145	780	364	155		2
21	205	1580	80	159	880	370	156		2
22	205-Rallye	1294	103	189	805	370	157		2
23	Seat-Ibiza-SXI	1461	100	181	925	363	161		2
24	AX-Sport	1294	95	184	730	350	160		2

Cliquons sur l'icône NIPALS dans le ruban Décrire.

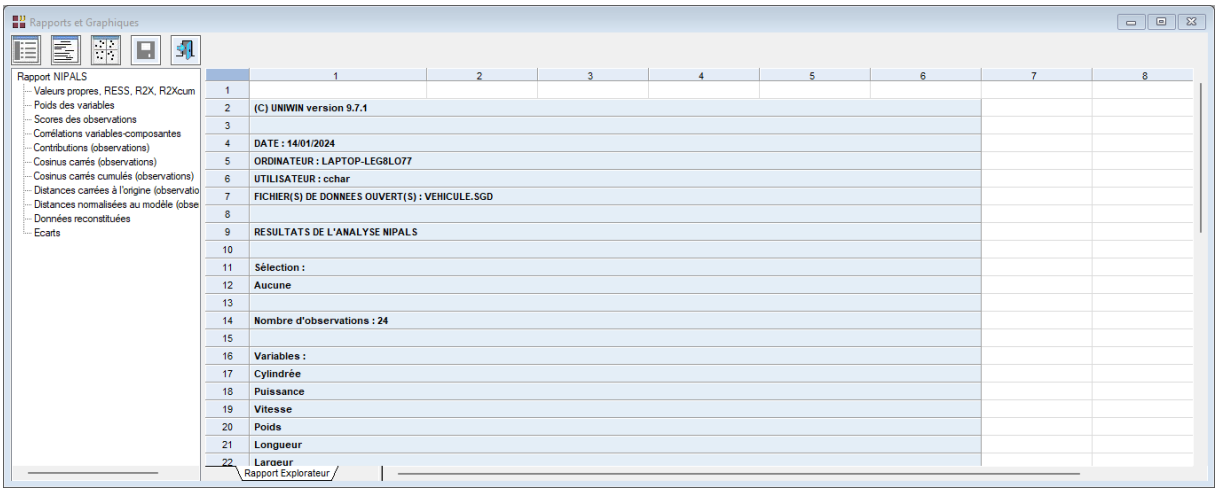
Sélectionnons les variables 'Cylindrée', 'Puissance', 'Vitesse', 'Poids', 'Longueur' et 'Largeur' comme variables quantitatives, la variable 'Mesures' pour les libellés des variables et la variable 'Modèles' pour les libellés des observations.

Nous ne renseignons pas le champ concernant le nombre de composantes à extraire (ainsi, 6 composantes seront extraites) et ne demandons pas la validation croisée.



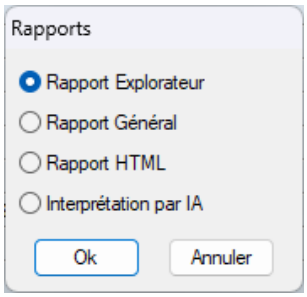
Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul de l'Analyse NIPALS.

Après quelques instants, l'écran suivant s'affiche :

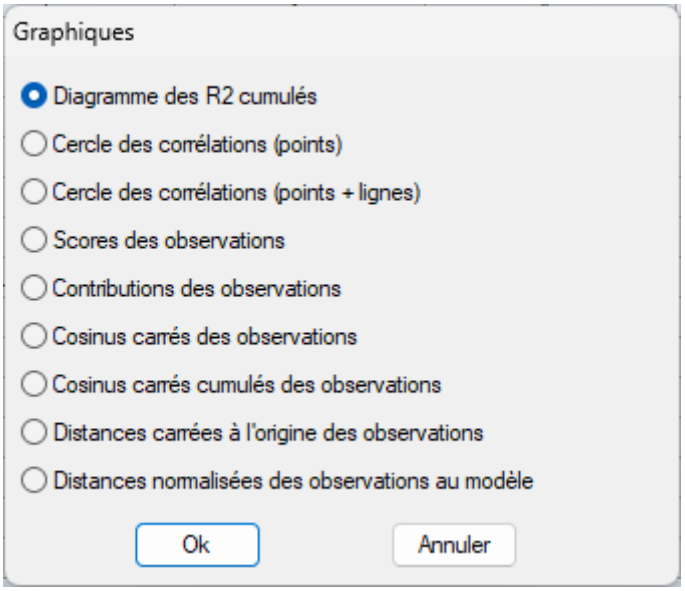


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données' de rappeler la boîte de dialogue d'entrée des données.

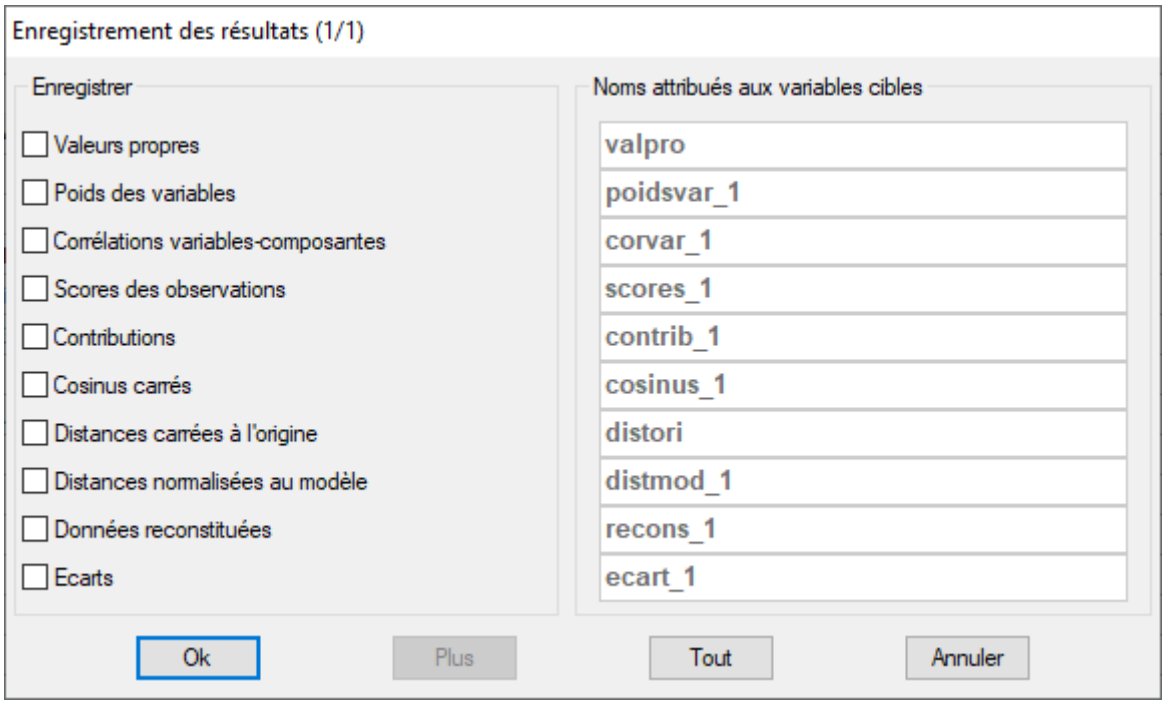
L'icône 'Rapports' affiche la boîte de dialogue des options pour les rapports :



et l'icône 'Graphiques' affiche la boîte de dialogue des options pour les graphiques :



L'icône 'Enregistrer' permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

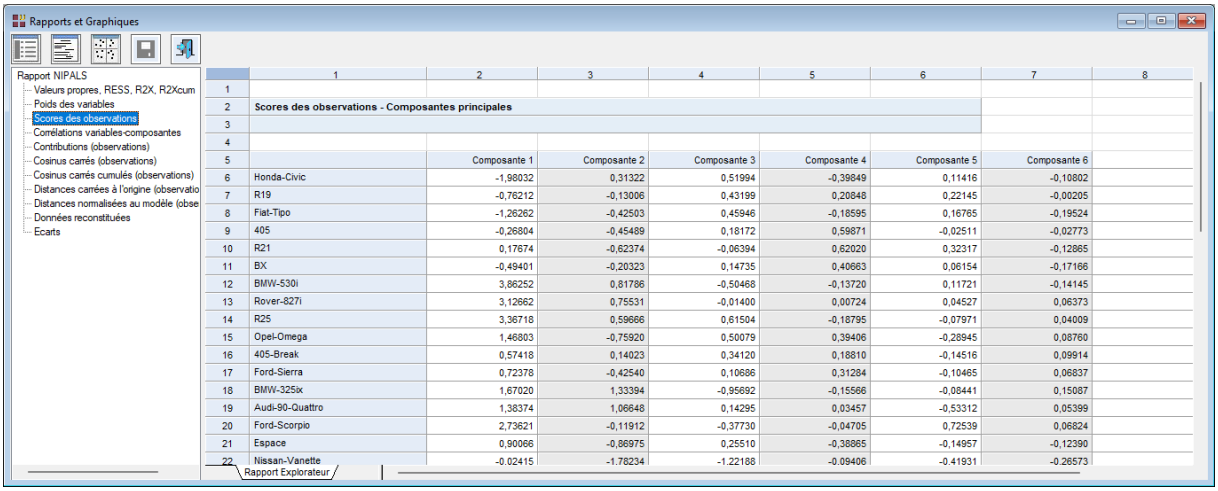


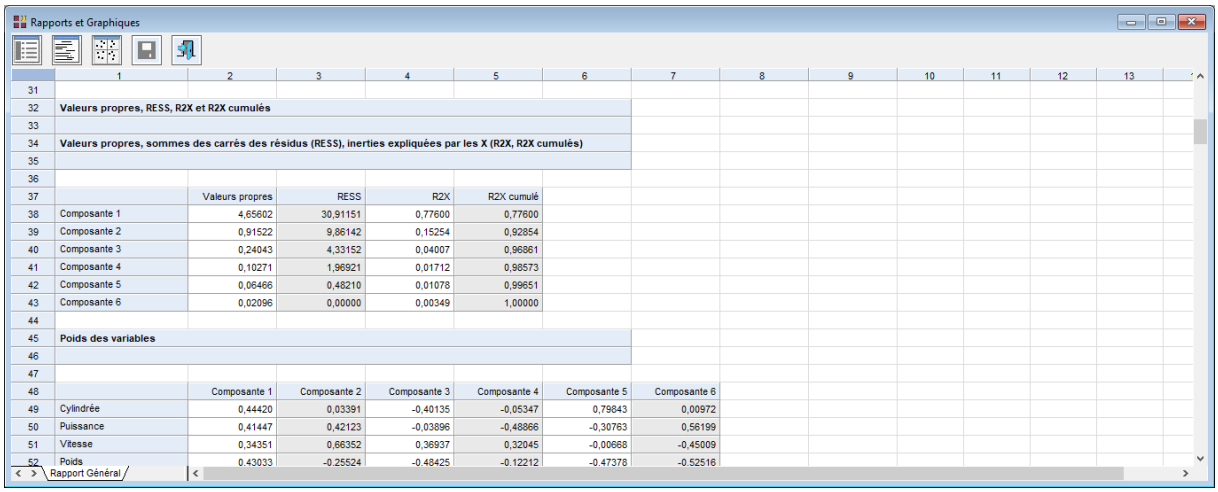
L'icône 'Quitter' permet de quitter l'analyse.

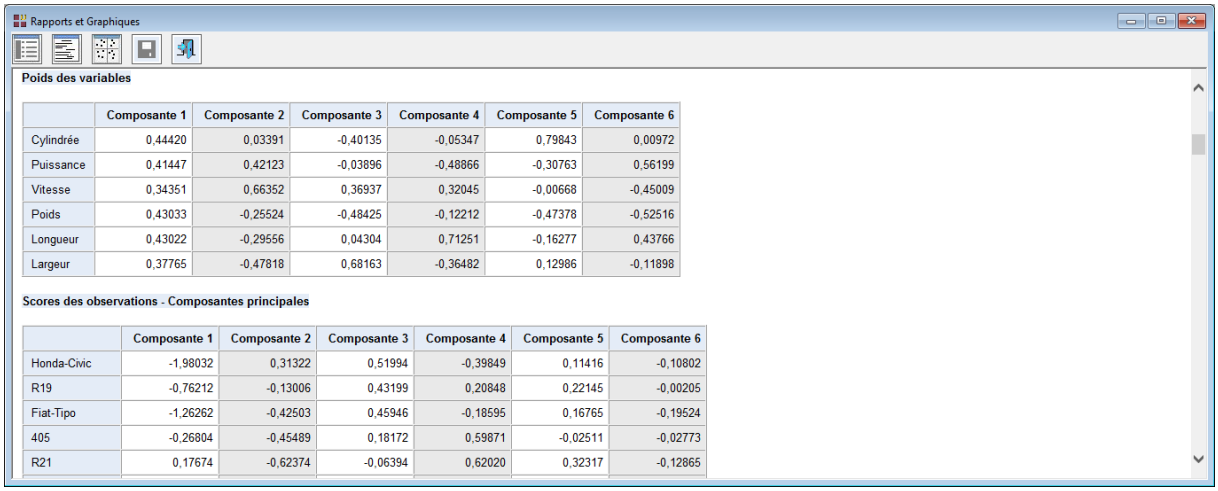
L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme Explorateur, Général ou HTML et d'accéder à l'aide à l'interprétation par Intelligence Artificielle.

Voici trois exemples du rapport pour notre analyse : Explorateur, Général, HTML.







Ce rapport affiche les informations suivantes :

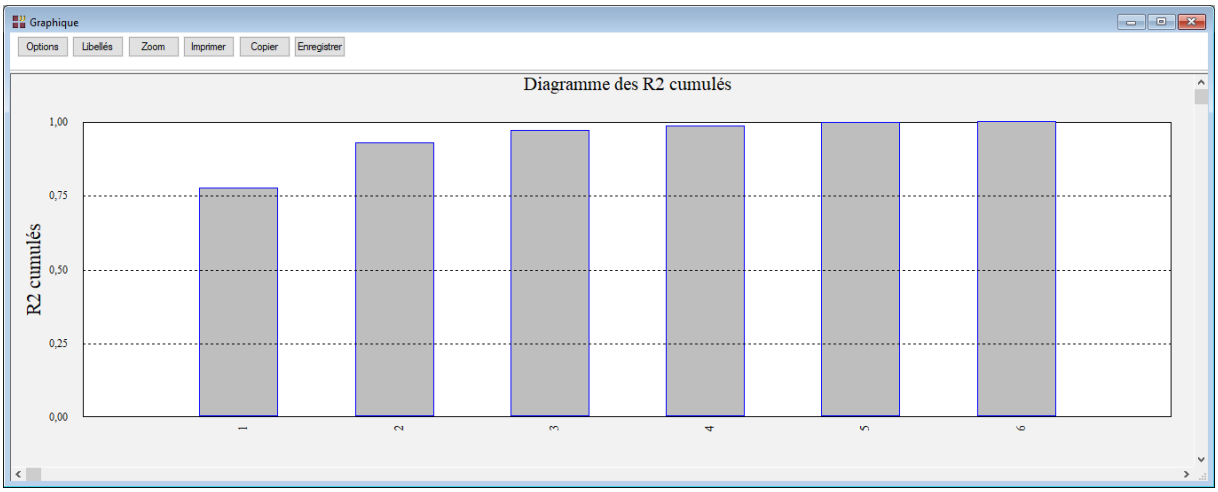
- Valeurs propres, RESS, R2X et R2X cumulés
- Poids des variables
- Scores des observations
- Corrélations entre les variables et les composantes principales
- Contributions des observations
- Cosinus carrés des observations
- Cosinus carrés cumulés des observations
- Distances carrées à l'origine des observations
- Distances normalisées des observations au modèle
- Données reconstituées
- Ecarts entre les données d'origine et les données reconstituées

L'option Graphiques

Cette option permet d'obtenir divers graphiques pour l'analyse NIPALS.

- L'option Diagramme des R2 cumulés

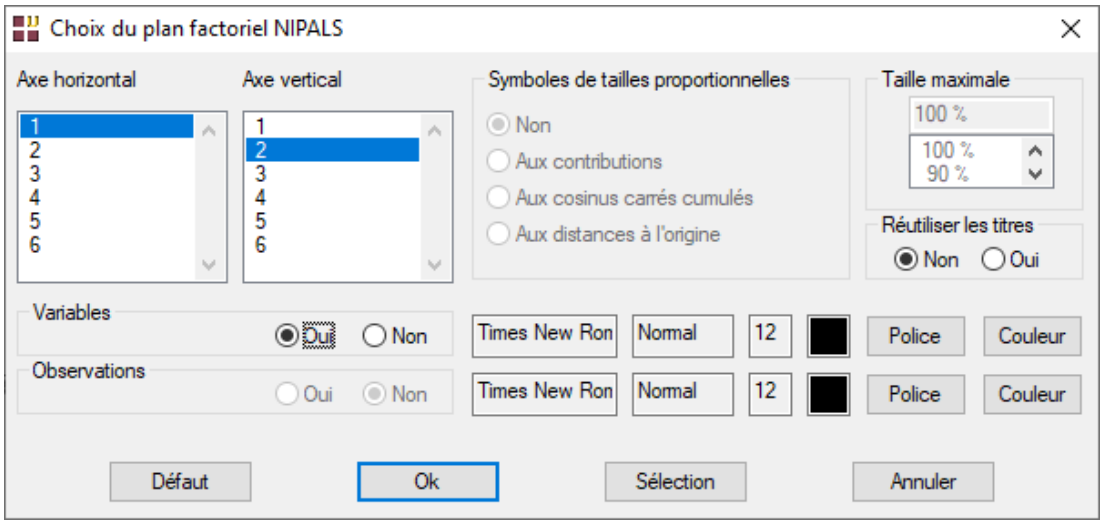
Cette option affiche les valeurs des R2 cumulés pour les axes extraits.



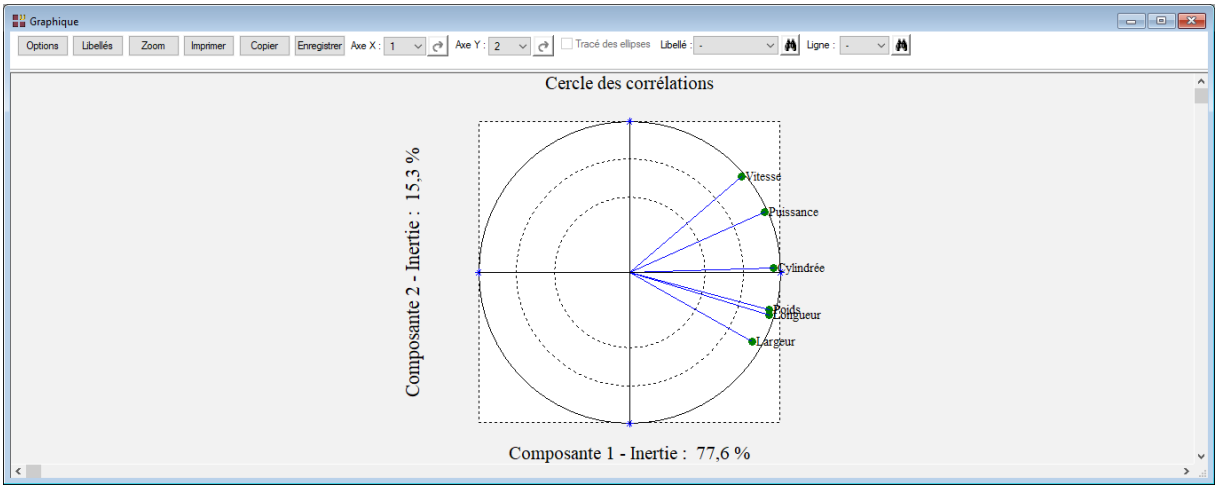
- Les options Cercle des corrélations

Ces options permettent d'afficher le cercle de corrélations des variables et de choisir si on désire tracer les lignes reliant les points à l'origine du cercle. L'option sans ces lignes (points) est utile lorsqu'il y a un grand nombre de variables représentées. Choisissons les variables avec lignes.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.



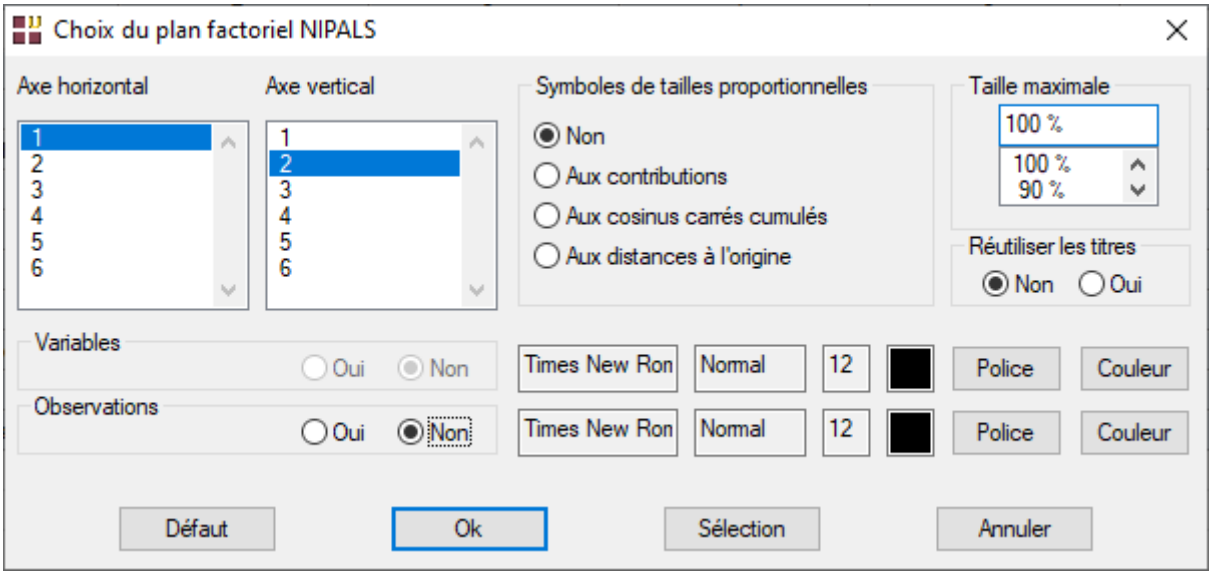
Un exemple du cercle factoriel obtenu est montré ci-après.

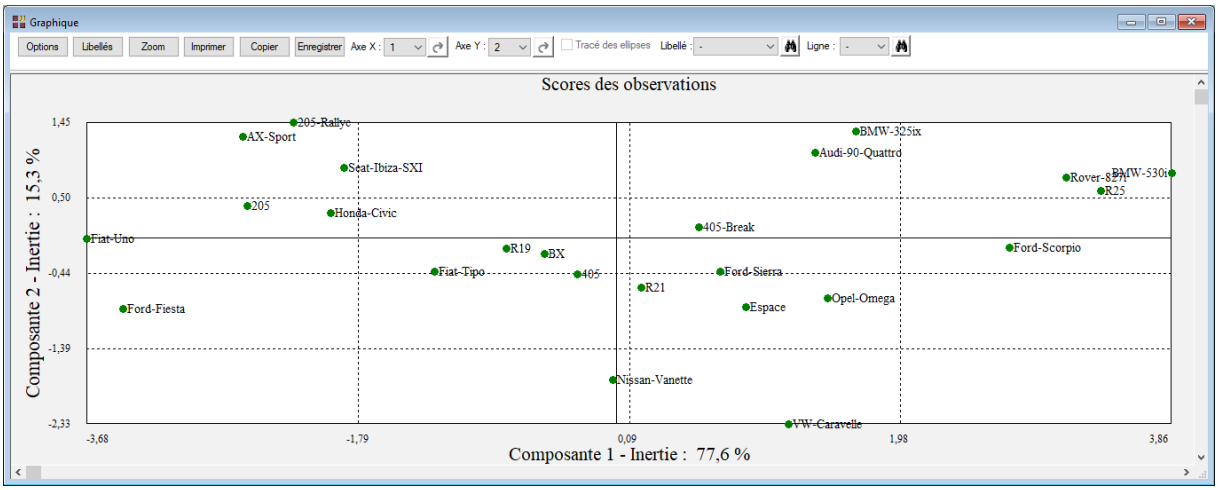


- L'option Scores des observations

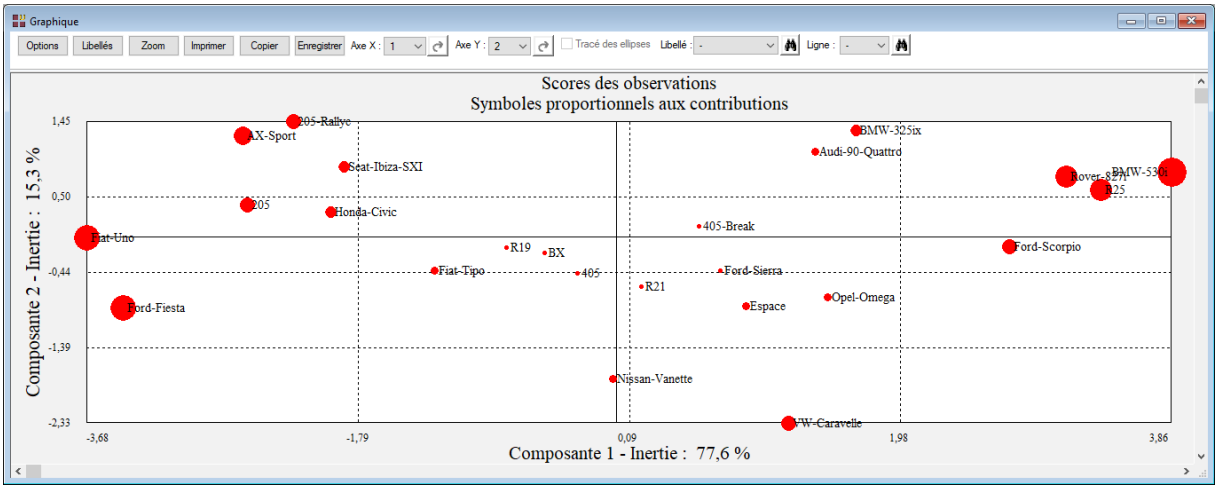
Cette option permet d'afficher des plans factoriels des observations.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des observations, de choisir la couleur et la police, d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail et de préciser si les symboles des points doivent être proportionnels aux contributions, aux cosinus carrés, aux distances à l'origine ou tous de la même taille. L'option 'Taille maximale' permet de contrôler les tailles des bulles en fonction d'un pourcentage de la taille de la bulle la plus grande.





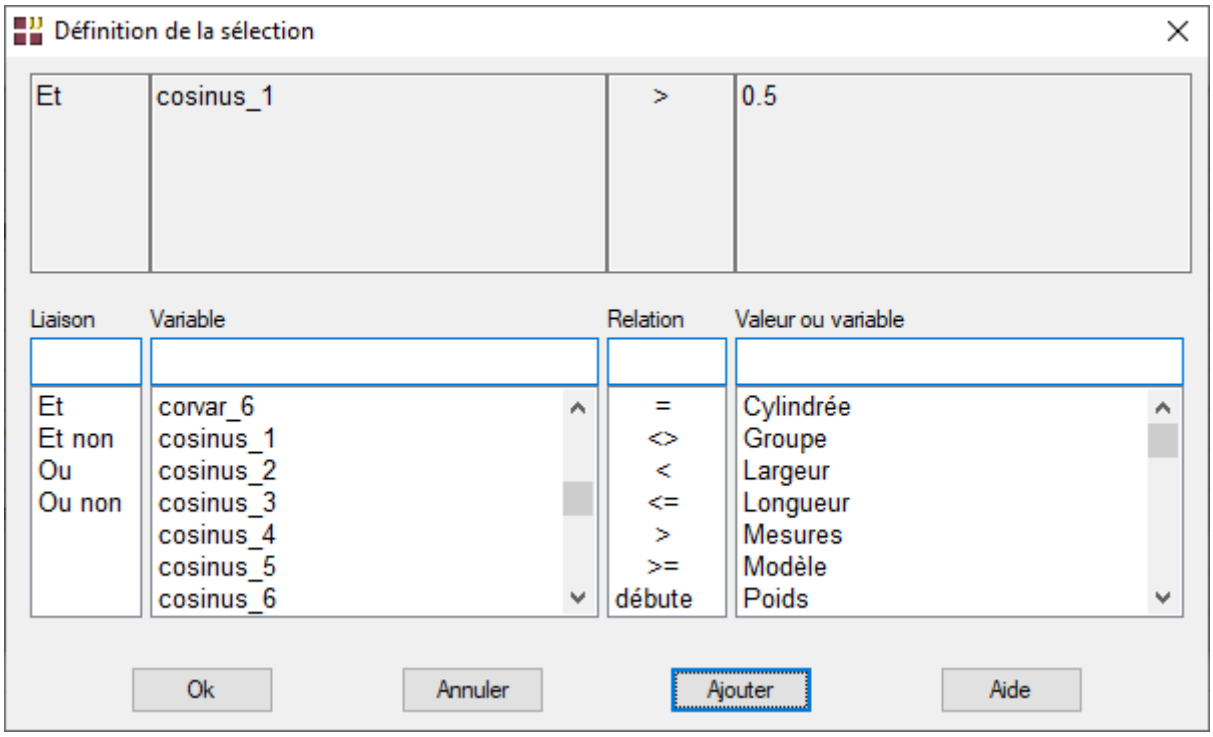
Affichons ce même graphique avec des symboles de points proportionnels aux contributions.



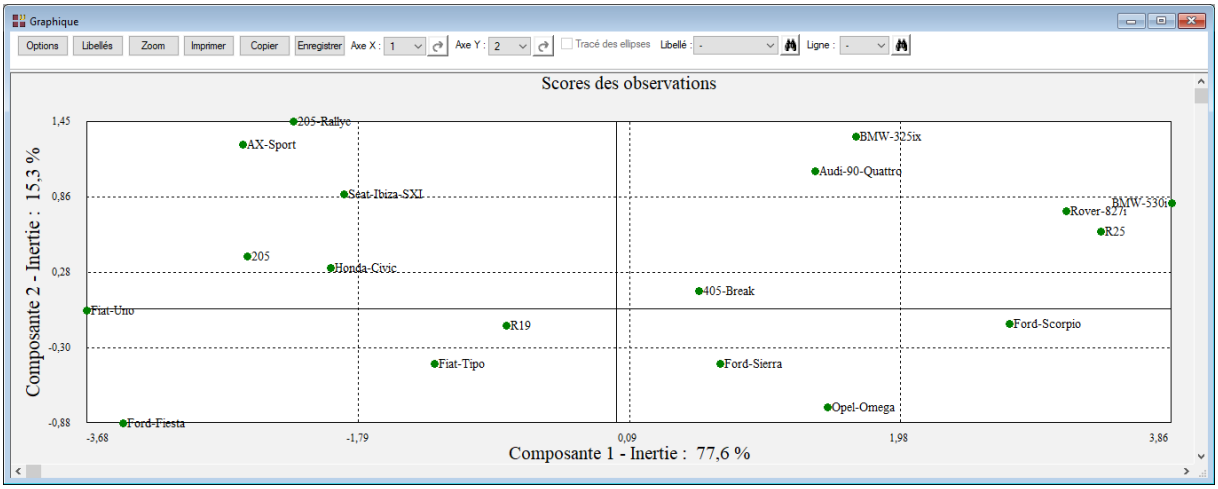
A noter le bouton ‘Sélection’ qui permet de n’afficher que les points qui satisfont une condition logique, par exemple une condition basée sur les contributions ou les cosinus carrés. Ceci est très pratique dans le cas d’un nuage comportant beaucoup d’observations : seuls les points significatifs d’un point de vue statistique sont affichés.

Les principaux résultats de l’analyse sont accessibles dans la boîte de dialogue ‘Sélection’ en plus des données du fichier VEHICULE.

Comme exemple, mettons en œuvre une sélection basée sur la variable ‘cosinus_1’ qui contient les cosinus carrés sur le premier axe factoriel.



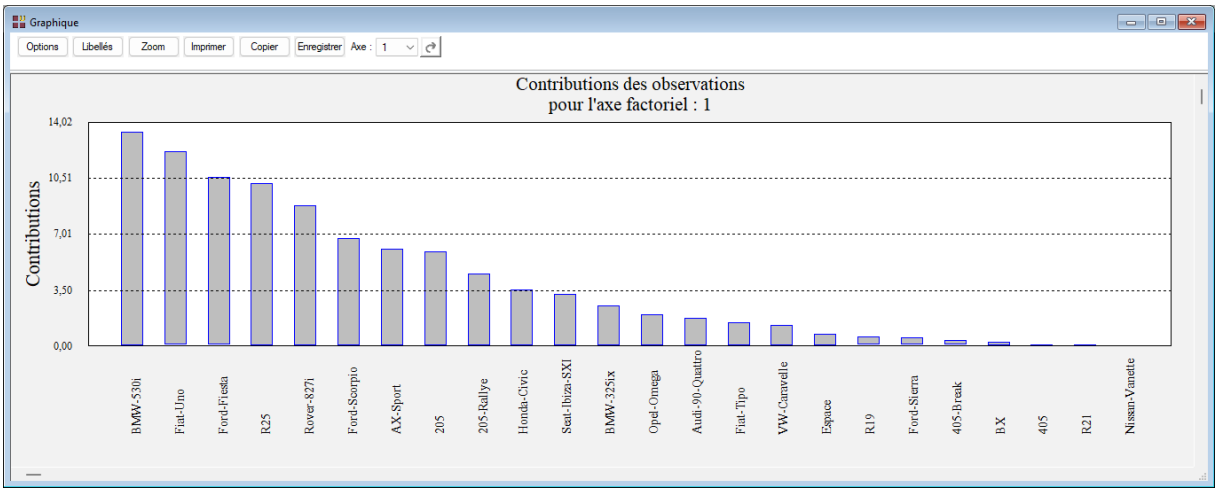
La sélection définie comporte 18 éléments.

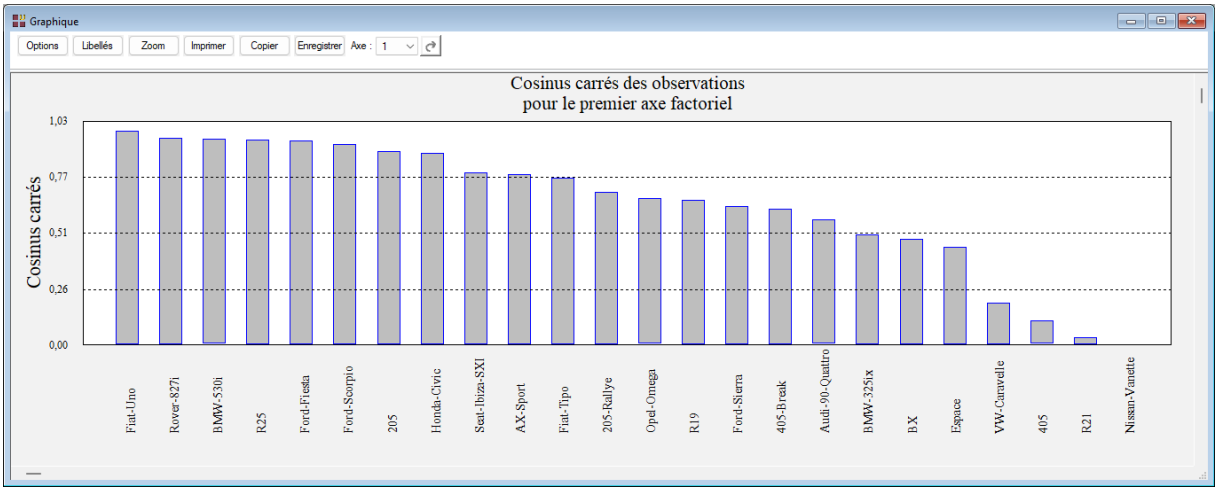


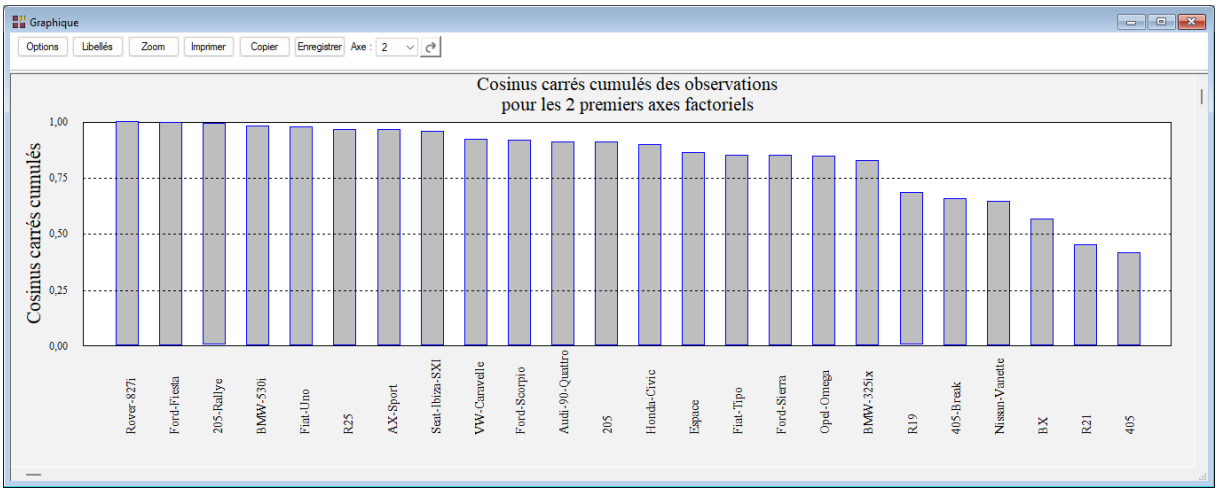
- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés, Distances carrées à l'origine, Distances normalisées au modèle

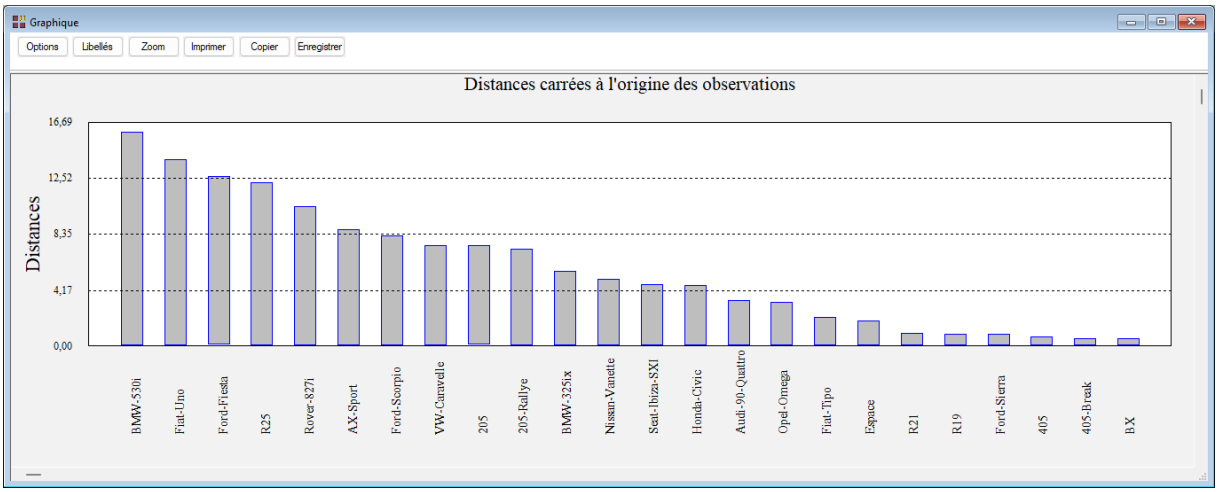
Ces options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse.

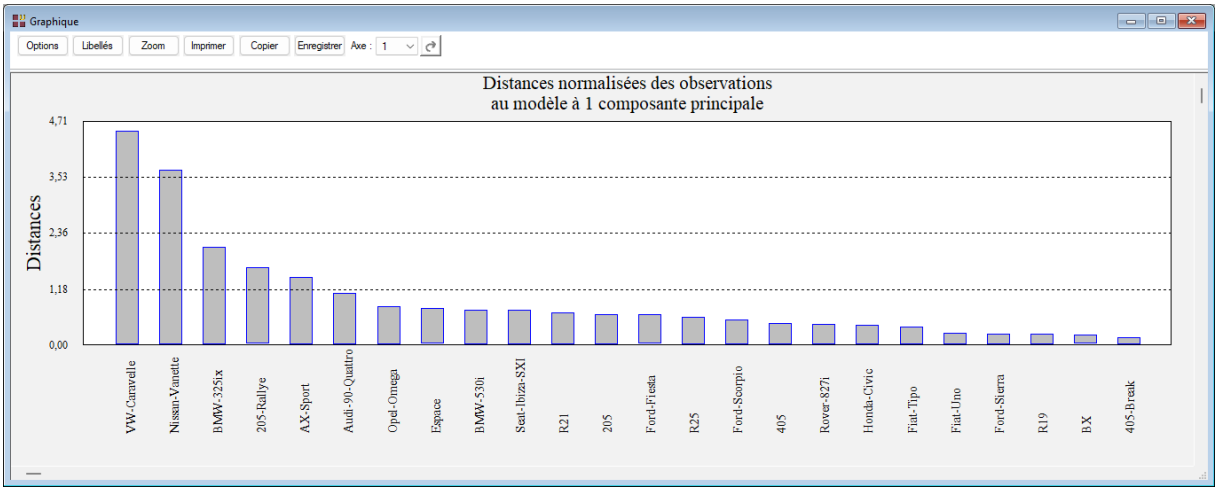
Elles mettent en évidence les observations qui contribuent fortement à la définition des axes factoriels, qui sont bien projetés dans les plans factoriels et qui sont éloignés de l'origine ou du modèle.


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe : 1
Contributions des observations
pour l'axe factoriel : 1
Contributions
14,02
10,51
7,01
3,50
0,00
BMW-530i
Fiat-Uno
Ford-Fiesta
R25
Rover-827i
Ford-Scorpio
AX-Sport
205
205-Rallye
Honda-Civic
Seat-Ibiza-SXI
BMW-325ix
Opel-Omega
Audi-90-Quattro
Fiat-Tipo
VW-Caravelle
Espace
R19
Ford-Sierra
405-Break
BX
405
R21
Nissan-Vanette


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe : 1
Cosinus carrés des observations
pour le premier axe factoriel
Cosinus carrés
1,03
0,77
0,51
0,26
0,00
Fiat-Uno
Rover-827i
BMW-530i
R25
Ford-Fiesta
Ford-Scorpio
205
Honda-Civic
Seat-Ibiza-SXI
AX-Sport
Fiat-Tipo
205-Rallye
Opel-Omega
R19
Ford-Sierra
405-Break
Audi-90-Quattro
BMW-325ix
BX
Espace
VW-Caravelle
405
R21
Nissan-Vanette


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe : 2
Cosinus carrés cumulés des observations
pour les 2 premiers axes factoriels
Cosinus carrés cumulés
1,00
0,75
0,50
0,25
0,00
Rover-827i
Ford-Fiesta
205-Rallye
BMW-530i
Fiat-Uno
R25
AX-Sport
Seat-Ibiza-SXI
VW-Caravelle
Ford-Scorpio
Audi-90-Quattro
205
Honda-Civic
Espace
Fiat-Tipo
Ford-Sierra
Opel-Omega
BMW-325ix
R19
405-Break
Nissan-Vanette
BX
R21
405





Exemple 2 : Fichier VEHICULE avec validation croisée

Reprenons cet exemple avec le fichier VEHICULE et demandons maintenant la mise en œuvre de la validation croisée pour définir le nombre de composantes à retenir.

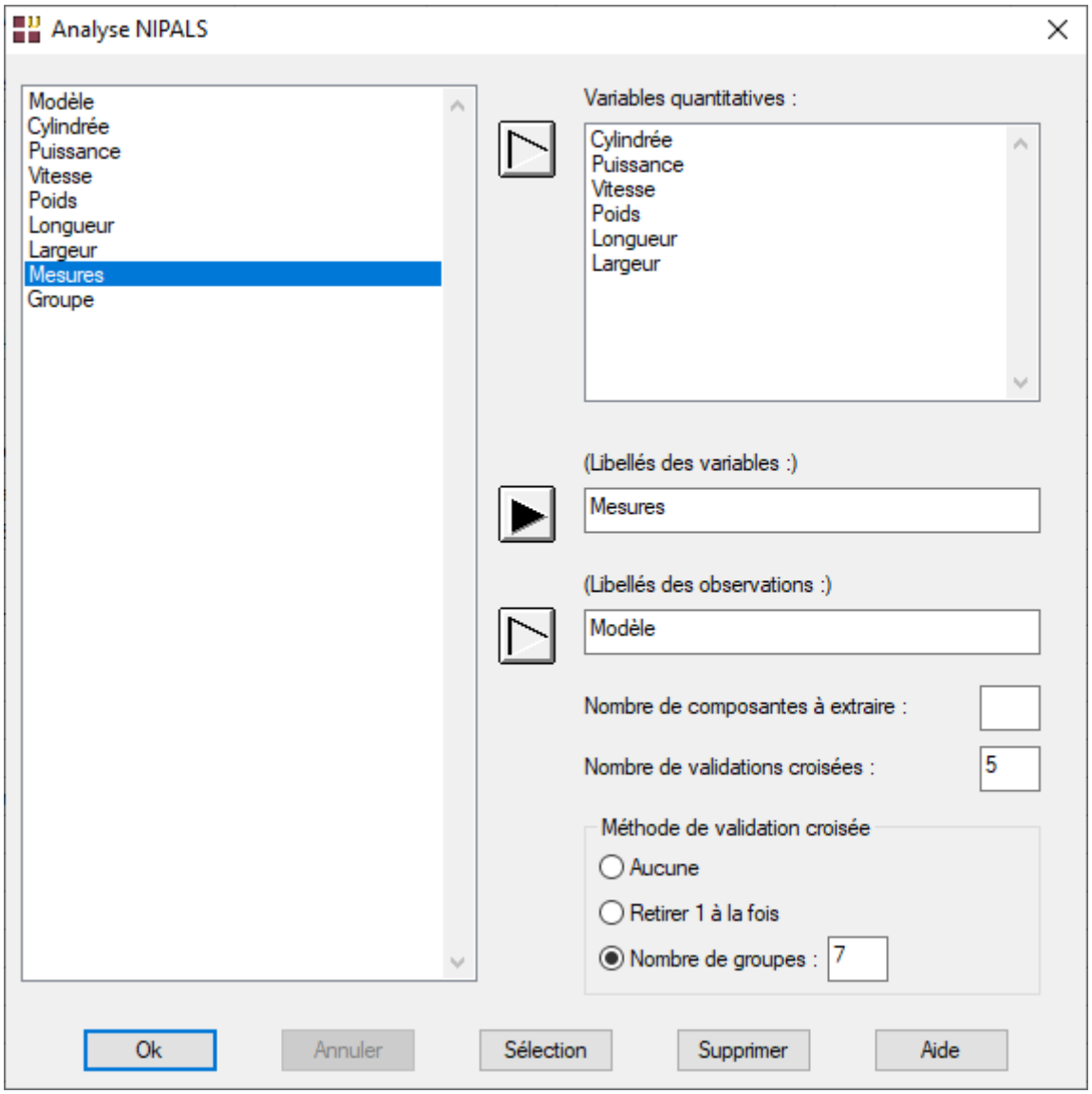
Choisissons la méthode basée sur un nombre de groupes égal à 7 et demandons la réalisation de 5 validations croisées.

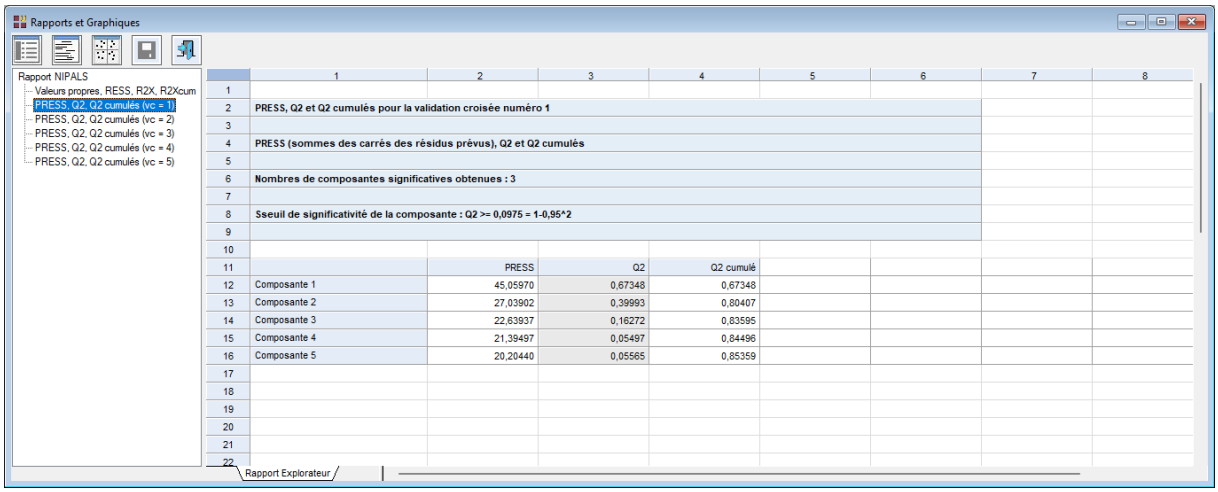
Après quelques instants, le rapport affiche :

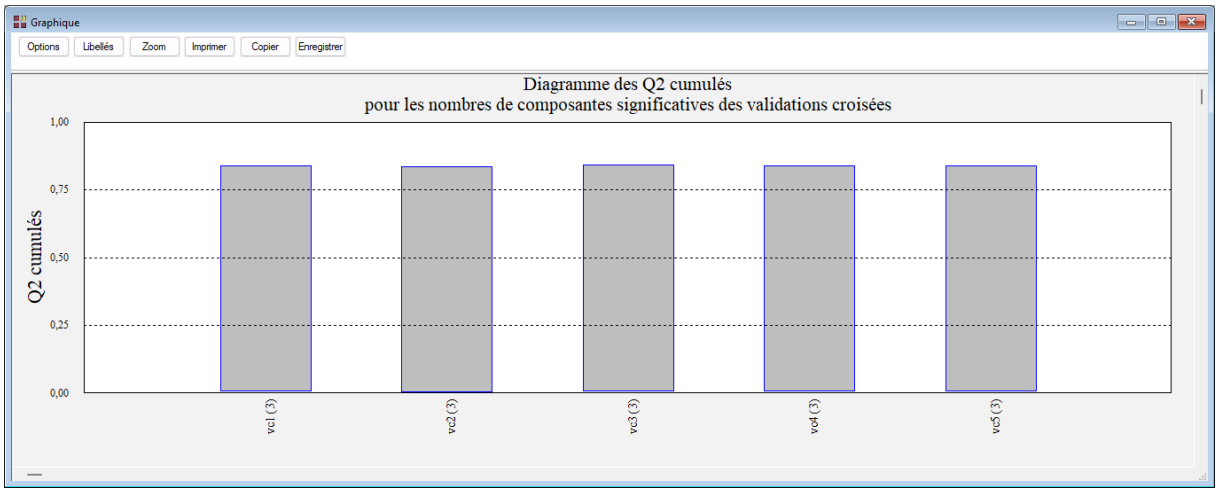
- Valeurs propres, RESS, R2X et R2X cumulés
- PRESS, Q2 et Q2 cumulés pour chacune des validations croisées

Les tableaux de Q2 précisent également les nombres de composantes à retenir.

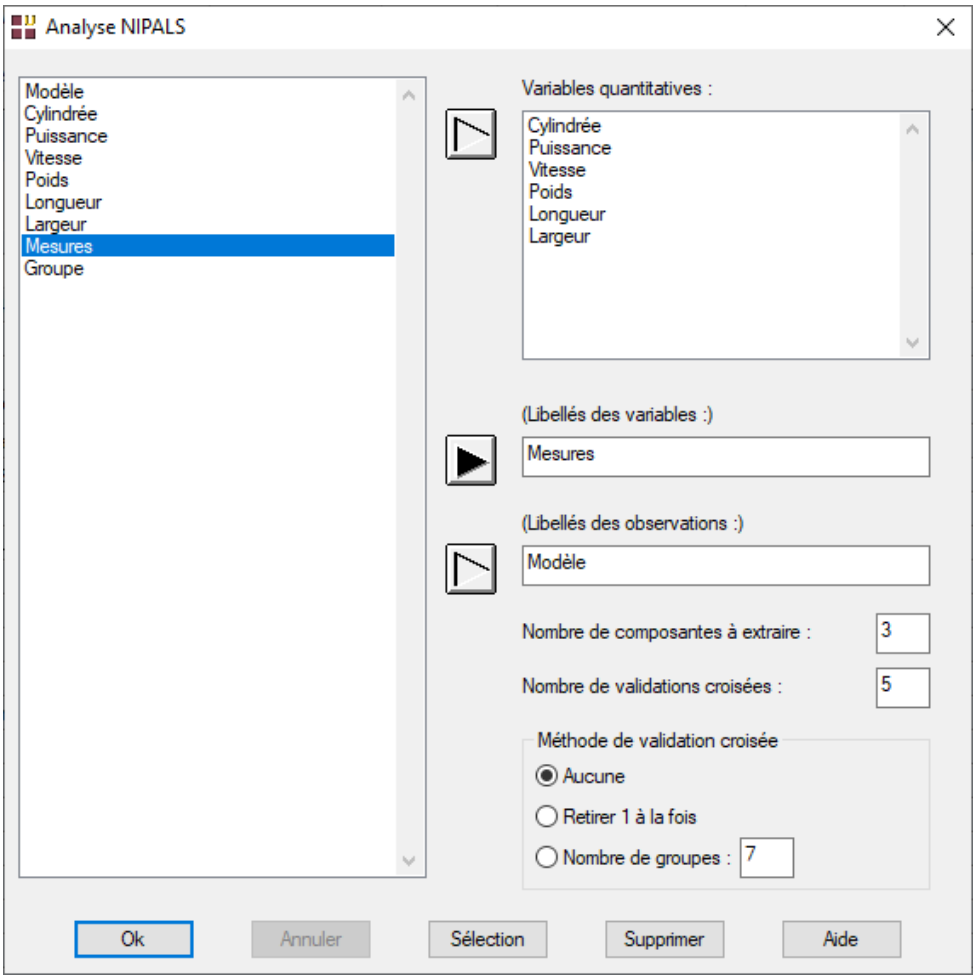
Une synthèse graphique est disponible via l'option 'Graphique des Q2 cumulés'. Les libellés de l'axe horizontal indiquent le numéro de la validation croisée et le nombre de composantes à retenir. Par exemple 'vc2 (3)' indique la validation croisée n°2 pour laquelle 3 composantes sont à retenir.


Analyse NIPALS
Modèle
Cylindrée
Puissance
Vitesse
Poids
Longueur
Largeur
Mesures
Groupe
Variables quantitatives :
Cylindrée
Puissance
Vitesse
Poids
Longueur
Largeur
(Libellés des variables :)
Mesures
(Libellés des observations :)
Modèle
Nombre de composantes à extraire :
Nombre de validations croisées : 5
Méthode de validation croisée
Aucune
Retirer 1 à la fois
Nombre de groupes : 7
Ok
Annuler
Sélection
Supprimer
Aide


Rapports et Graphiques
Rapport NIPALS
Valeurs propres, RESS, R2X, R2Xcum
PRESS, Q2, Q2 cumulés (vc = 1)
PRESS, Q2, Q2 cumulés (vc = 2)
PRESS, Q2, Q2 cumulés (vc = 3)
PRESS, Q2, Q2 cumulés (vc = 4)
PRESS, Q2, Q2 cumulés (vc = 5)
PRESS, Q2 et Q2 cumulés pour la validation croisée numéro 1
PRESS (sommes des carrés des résidus prévus), Q2 et Q2 cumulés
Nombres de composantes significatives obtenues : 3
Sseuil de significativité de la composante : Q2 >= 0,0975 = 1-0,95^2
PRESS Q2 Q2 cumulé
Composante 1 45,05970 0,67348 0,67348
Composante 2 27,03902 0,39993 0,80407
Composante 3 22,63937 0,16272 0,83595
Composante 4 21,39497 0,05497 0,84496
Composante 5 20,20440 0,05565 0,85359
Rapport Explorateur



Dans cet exemple, toutes les validations croisées indiquent que 3 composantes doivent être retenues. Il suffit alors de relancer la procédure NIPALS en précisant ce nombre dans la zone 'Nombre de composantes à extraire' et d'indiquer qu'aucune validation croisée n'est à effectuer.



Un rapport similaire à celui de l'exemple 1 est alors obtenu pour les 3 composantes significatives extraites.

Exemple 3 : Fichier VEHICULE2 sans validation croisée

Nous utiliserons le fichier VEHICULE2 pour illustrer ce troisième exemple.

Ce fichier avec données manquantes contient les informations suivantes concernant 24 modèles de voitures :

Données

	Modèle	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur	Mesures	Groupe
	Type = C	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = C	Type = N
	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 24	Longueur = 23	Longueur = 6	Longueur = 24
1	Honda Civic		90	174	850	369	166	Cylindrée	1
2	Renault 19	1721		180	965	415	169	Puissance	1
3	Fiat Tipo	1580	83		970	395	170	Vitesse	1
4	Peugeot 405	1769	90	180		440	169	Poids	1
5	Renault 21	2068	88	180	1135		170	Longueur	1
6	Citroen BX	1769	90	182	1060	424		Largeur	1
7	BMW 530i		188	226	1510	472	175		1
8	Rover-827i	2675		222	1365	469	175		1
9	Renault 25	2548	182		1350	471	180		1
10	Opel Omega	1998	122	190		473	177		1
11	Peugeot 405 Break	1905	125	194	1120		171		1
12	Ford Sierra	1993	115	185	1190	451			1
13	BMW 325ix		171	208	1300	432	164		2
14	Audi 90 Quattro	1994		214	1220	439	169		2
15	Ford Scorpio	2933	150		1345	466	176		2
16	Renault Espace	1995	120	177		436	177		2
17	Nissan Vanette	1952	87	144	1430		169		2
18	VW Caravelle	2109	112	149	1320	457			2
19	Ford Fiesta		50	135	810	371	162		2
20	Fiat Uno	1116		145	780	364	155		2
21	Peugeot 205	1580	80		880	370	156		2
22	Peugeot 205 Rallye	1294	103	189		370	157		2
23	Seat Ibiza SX I	1461	100	181	925		161		2
24	Citroen AX Sport	1294	95	184	730	350			2
25									

Note : Dans le cadre de l'analyse d'un fichier comportant des données manquantes, certaines des options pour les rapports et pour les graphiques ne sont pas proposées.

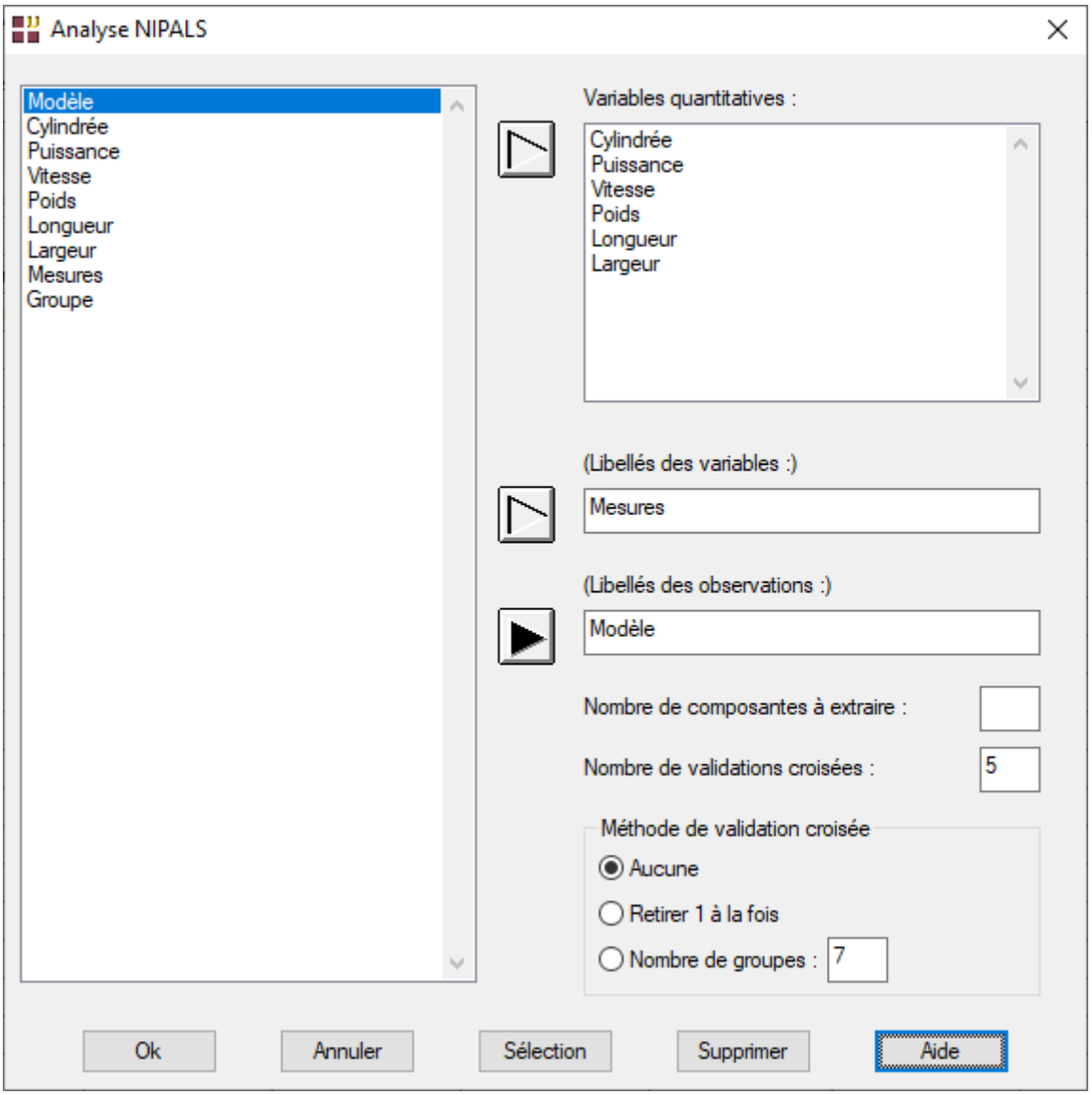
Cliquons sur l'icône NIPALS dans le ruban Décrire.

Sélectionnons les variables 'Cylindrée', 'Puissance', 'Vitesse', 'Poids', 'Longueur' et 'Largeur' comme variables quantitatives, la variable 'Mesures' pour les libellés des variables et la variable 'Modèles' pour les libellés des observations.

Nous ne renseignons pas le champ concernant le nombre de composantes à extraire (ainsi, 6 composantes seront extraites) et ne demandons pas la validation croisée.

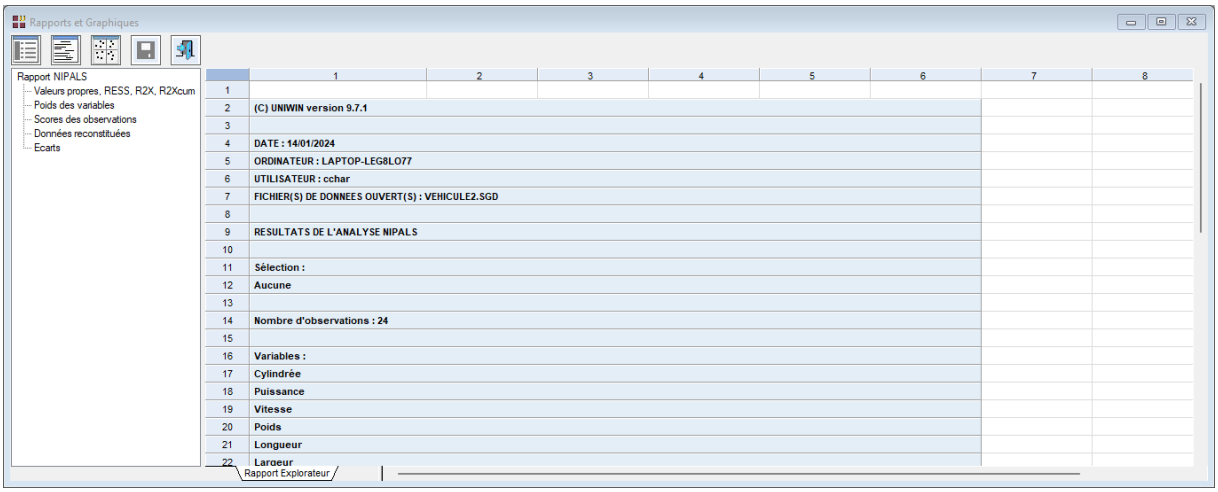
Les données sont automatiquement centrées et réduites.

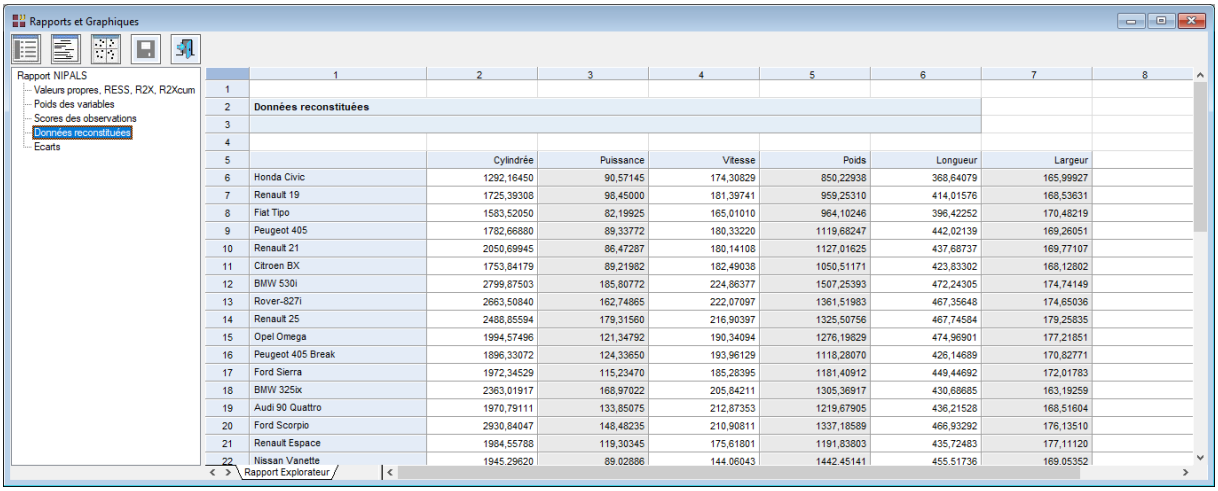
Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul de l'Analyse NIPALS.



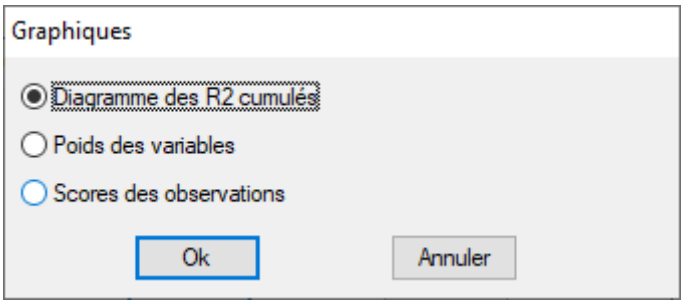
Le rapport affiche :

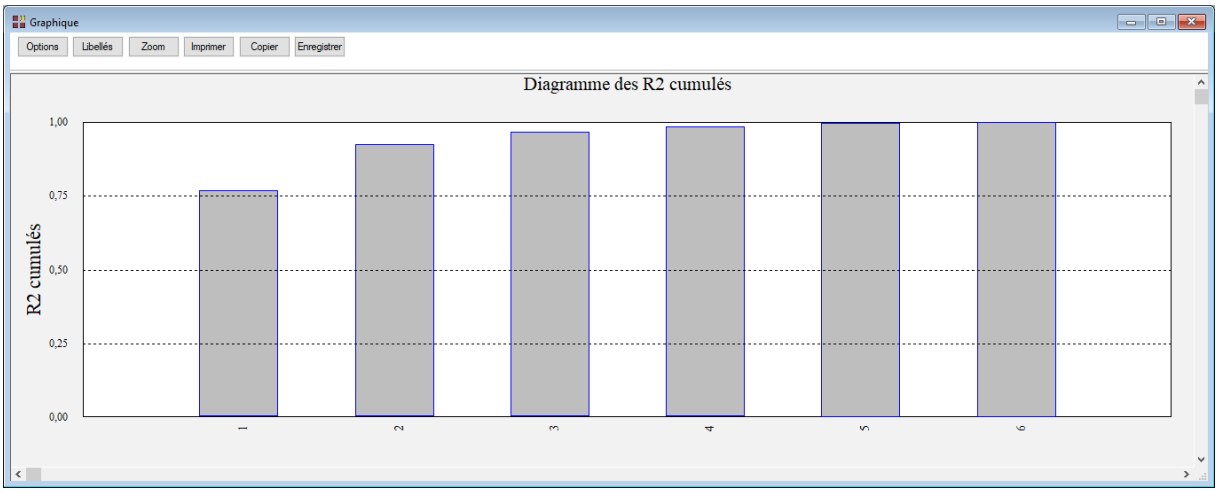
- Les valeurs propres
- Les sommes des carrés des résidus (RESS)
- Les R-carrés et R-carrés cumulés
- Les poids des variables
- Les scores des observations
- Les données reconstituées et donc les estimations des données manquantes
- Les écarts entre les données observées et les données reconstituées

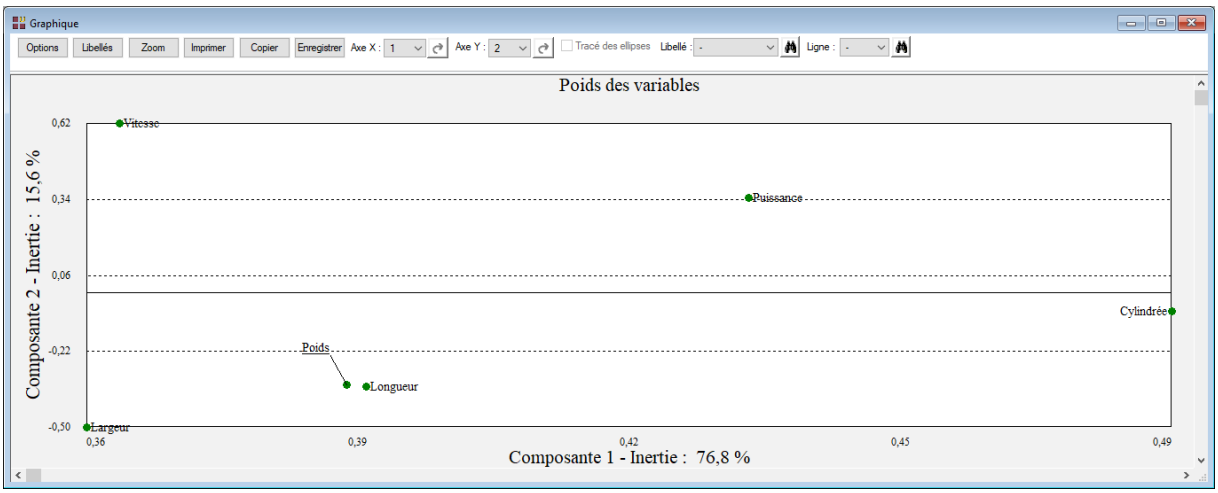


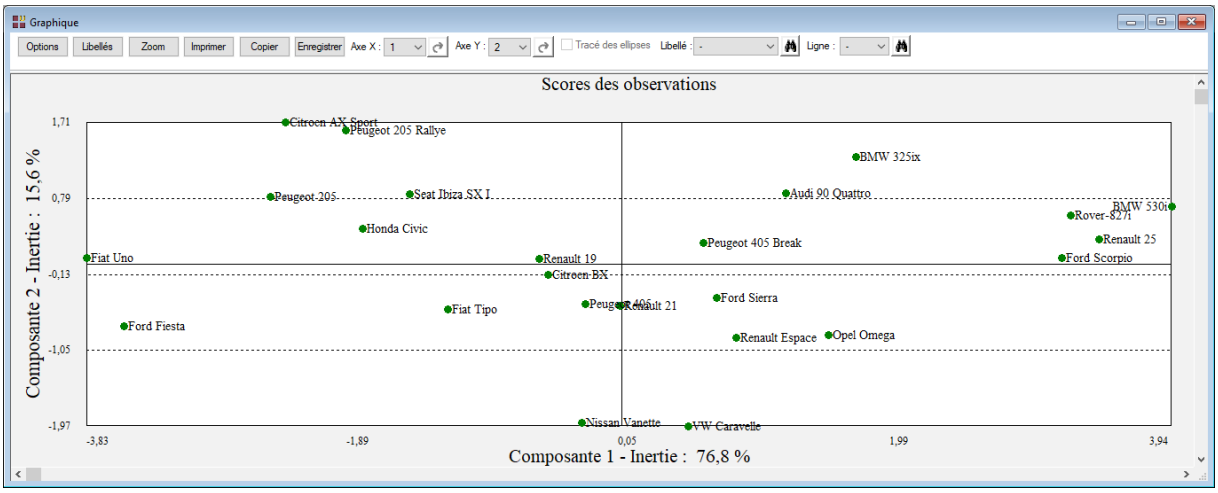


Les graphiques proposés dans le cas d'une analyse avec données manquantes sont :




Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Diagramme des R2 cumulés
R2 cumulés
1,00
0,75
0,50
0,25
0,00
1
2
3
4
5
6

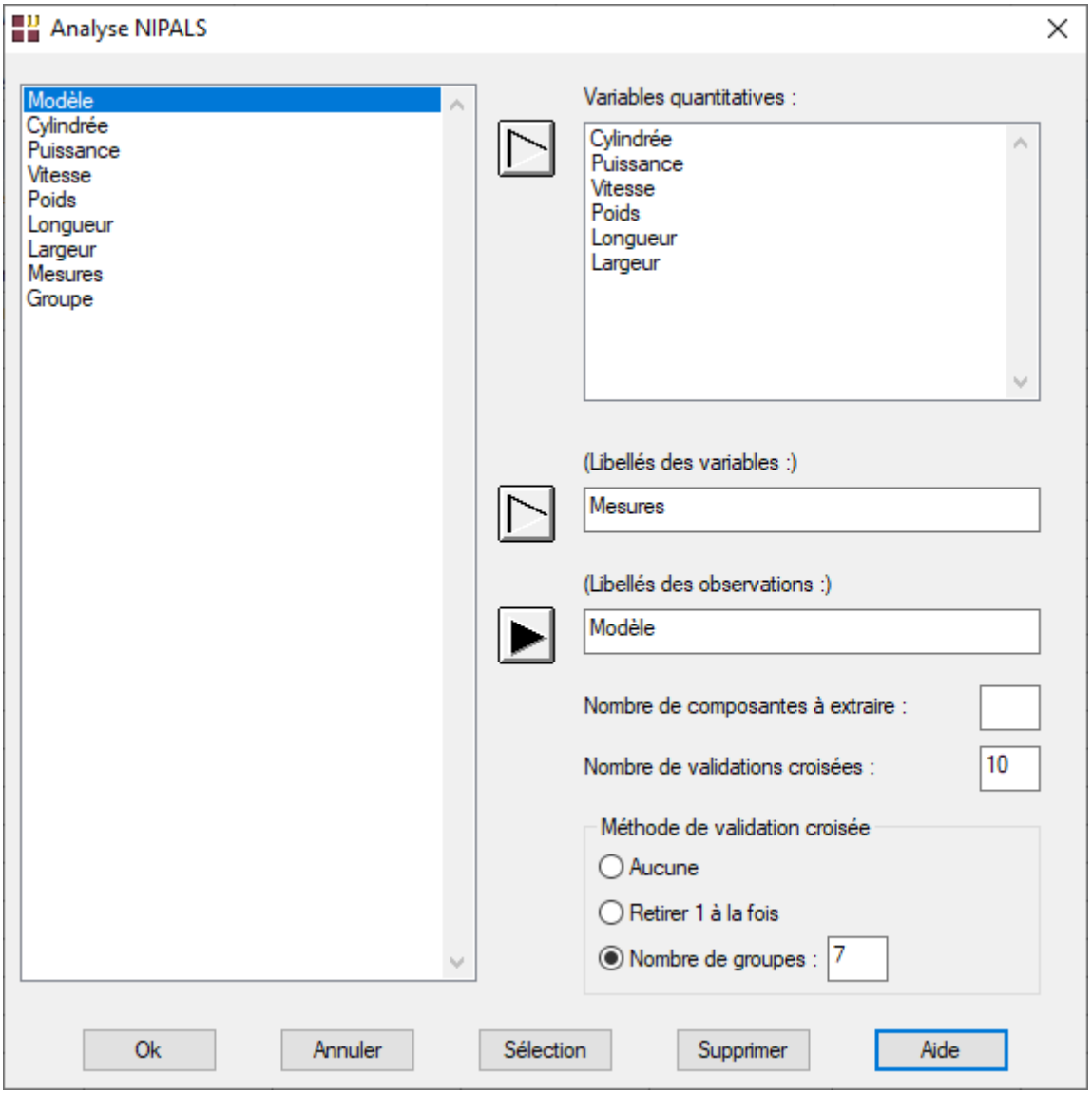

Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe X : 1
Axe Y : 2
Tracé des ellipses
Libellé :
Ligne :
Poids des variables
Composante 2 - Inertie : 15,6 %
0,62
0,34
0,06
-0,22
-0,50
Vitesse
Puissance
Cylindrée
Poids
Longueur
Largeur
0,36
0,39
0,42
0,45
0,49
Composante 1 - Inertie : 76,8 %


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe X : 1
Axe Y : 2
Tracé des ellipses
Libellé :
Ligne :
Scores des observations
Composante 2 - Inertie : 15,6 %
1,71
0,79
-0,13
-1,05
-1,97
Citroen AX Sport
Peugeot 205 Rallye
BMW 325ix
Peugeot 205
Seat Ibiza SX I
Audi 90 Quattro
BMW 530i
Rover-827i
Honda Civic
Peugeot 405 Break
Renault 25
Fiat Uno
Renault 19
Ford Scorpio
Citroen BX
Ford Sierra
Fiat Tipo
Renault 21
Ford Fiesta
Renault Espace
Opel Omega
Nissan Vanette
VW Caravelle
-3,83
-1,89
0,05
1,99
3,94
Composante 1 - Inertie : 76,8 %

Exemple 4 : Fichier VEHICULE2 avec validation croisée

Reprenons cet exemple avec le fichier VEHICULE2 et demandons maintenant la mise en œuvre de la validation croisée pour définir le nombre de composantes à retenir.

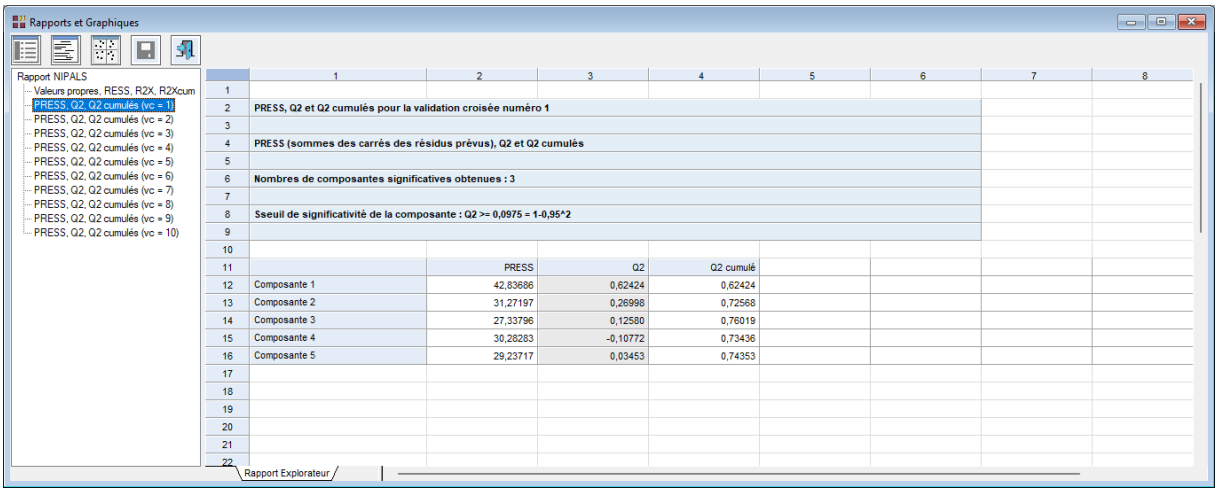
Choisissons la méthode basée sur un nombre de groupes égal à 7 et demandons la réalisation de 10 validations croisées.



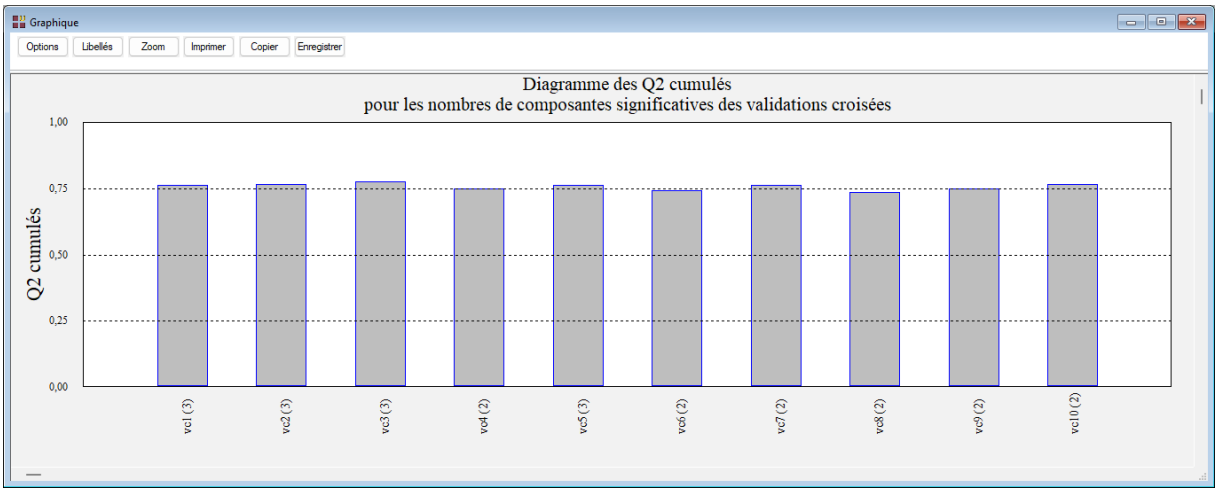
Après quelques instants, le rapport affiche :

- Valeurs propres, RESS, R2X et R2X cumulés
- PRESS, Q2 et Q2 cumulés pour chacune des validations croisées

Les tableaux de Q2 précisent également les nombres de composantes à retenir.



Une synthèse graphique est disponible via l'option 'Graphique des Q2 cumulés'. Les libellés de l'axe horizontal indiquent le numéro de la validation croisée et le nombre de composantes à retenir. Par exemple 'vc6 (3)' indique la validation croisée n°6 pour laquelle 3 composantes sont à retenir.



Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure. Ces variables peuvent notamment être utilisées avec l'option 'Sélection'. A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

Variable	*Contenu*
valpro	Valeurs propres
poidsvar	Poids des variables
scores	Scores des observations
corvar	Corrélations variables-composantes
contrib	Contributions des observations

cosinus	Cosinus carrés des observations
distmod	Distances normalisées au modèle des observations
distori	Distances carrées à l'origine des observations
ecart	Ecarts entre données observées et reconstituées
recons	Données reconstituées
nbccv	Nombres de composantes des validations croisées
presscv	PRESS des validations croisées
Q2cv	Q2 des validations croisées
Q2cvcum	Q2 cumulés des validations croisées

Références

Wold, H. (1966) Estimation of principal components and related models by iterative least squares. In Multivariate Analysis (Ed., P.R. Krishnaiah), Academic Press, NY, 391-420.

Tenenhaus M. (1998) La Régression PLS: théorie et pratique Paris: Editions Technip

H. T. Eastment & W. J. Krzanowski (1982) Cross-Validatory Choice of the Number of Components From a Principal Component Analysis, Technometrics, 24:1, 73-77