

UNIWIN VERSION 9.7.0

ANALYSE NIPALS

Révision : 02/09/2023

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 :

Analyse NIPALS

Variables quantitatives :

(Libellés des variables :)

(Libellés des observations :)

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

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

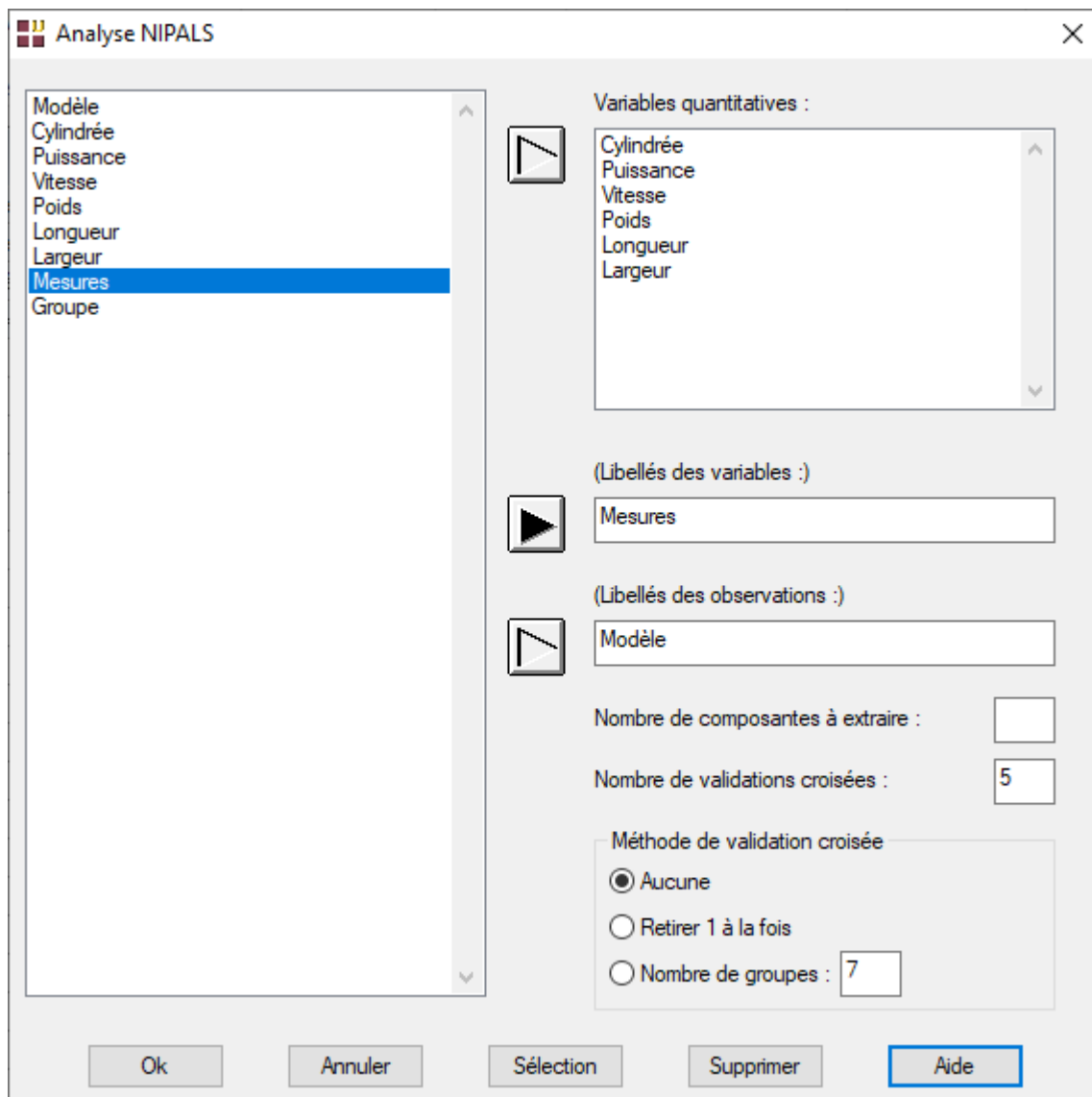
Nous utiliserons le fichier VEHICULE pour illustrer ce premier exemple. Ce fichier sans données manquantes contient les informations suivantes concernant 24 modèles de voitures :

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.

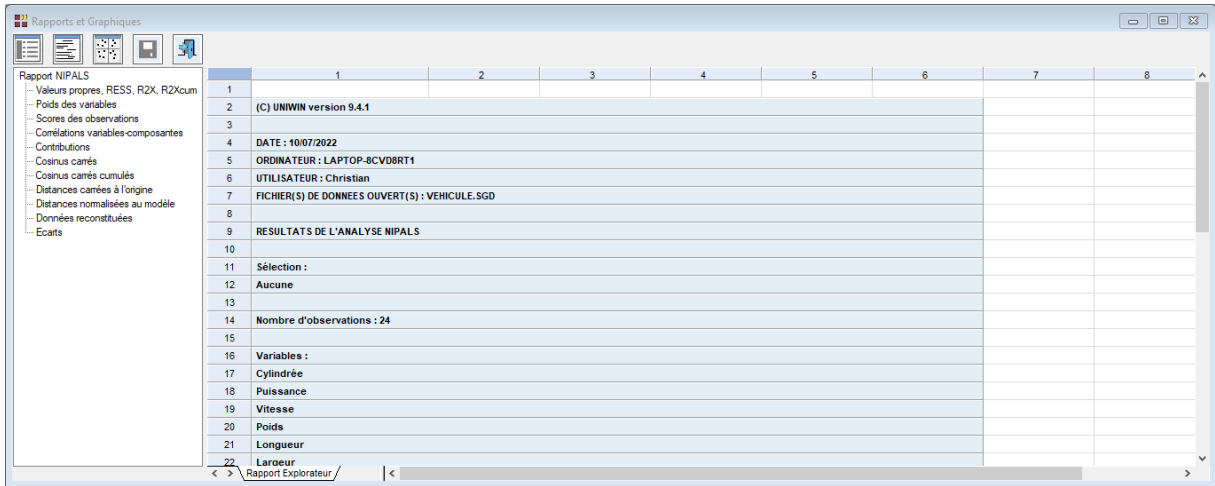
Sélectionnons les variables 'Cylindrée', 'Puissance', 'Vitesse', 'Poids', 'Longueur' et 'Largeur' comme variables quantitatives, la variable 'Mesures' pour les libellés des variables et la variable 'Modèles' pour les libellés des observations.


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.




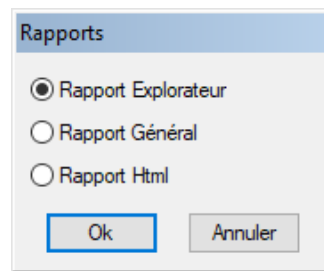
Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul de l'Analyse NIPALS.

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

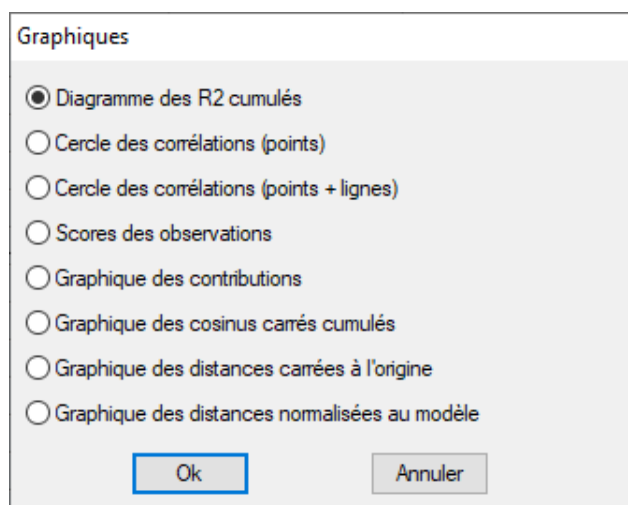



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

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.

Enregistrement des résultats (1/1)

Enregistrer

- Valeurs propres
- Poids des variables
- Corrélations variables-composantes
- Scores des observations
- Contributions
- Cosinus carrés
- Distances carrées à l'origine
- Distances normalisées au modèle
- Données reconstituées
- Ecart

Noms attribués aux variables cibles

valpro
poidsvar_1
corvar_1
scores_1
contrib_1
cosinus_1
distori
distmod_1
recons_1
ecart_1

Ok Plus Tout Annuler

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

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

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

Rapports et Graphiques

Rapport NIPALS

- Valeurs propres, RESS, R2X, R2Xcum
- Poids des variables
- Scores des observations**
- Corrélations variables-composantes
- Contributions
- Cosinus carrés
- Cosinus carrés cumulés
- Distances carrées à l'origine
- Distances normalisées au modèle
- Données reconstituées
- Ecart

	1	2	3	4	5	6	7	8
1								
2	Scores des observations - Composantes principales							
3								
4								
5		Composante 1	Composante 2	Composante 3	Composante 4	Composante 5	Composante 6	
6	Honda-Civic	-1,96032	0,31322	0,51994	-0,39849	0,11416	-0,10802	
7	R19	-0,76212	-0,13006	0,43199	0,20848	0,22145	-0,00205	
8	Fiat-Tipo	-1,26262	-0,42503	0,45946	-0,18595	0,16765	-0,19524	
9	405	-0,26904	-0,45489	0,18172	0,59871	-0,02511	-0,02773	
10	R21	0,17674	-0,62374	-0,06394	0,62020	0,32317	-0,12965	
11	BX	-0,49401	-0,20323	0,14735	0,40663	0,06154	-0,17166	
12	BMW-530i	3,86252	0,61786	-0,50468	-0,13720	0,11721	-0,14145	
13	Rover-827i	3,12662	0,75531	-0,01400	0,00724	0,04527	0,06373	
14	R25	3,36718	0,59666	0,61504	-0,18795	-0,07971	0,04009	
15	Opel-Omega	1,46803	-0,75920	0,50079	0,39406	-0,28945	0,08760	
16	405-Break	0,57418	0,14023	0,34120	0,18810	-0,14516	0,09914	
17	Ford-Sierra	0,72378	-0,42540	0,10686	0,31284	-0,10465	0,06837	
18	BMW-325ix	1,67020	1,33394	-0,95692	-0,15566	-0,08441	0,15087	
19	Audi-90-Quattro	1,38374	1,06648	0,14295	0,03457	-0,53312	0,05399	
20	Ford-Scorpio	2,73621	-0,11912	-0,37730	-0,04705	0,72539	0,06824	
21	Espace	0,90066	-0,86975	0,25510	-0,38865	-0,14957	-0,12390	
22	Nissan-Vanette	-0,02415	-1,78234	-1,22188	-0,09406	-0,41931	-0,26573	

	1	2	3	4	5	6	7	8	9	10	11	12	13
31													
32	Valeurs propres, RESS, R2X et R2X cumulés												
33													
34	Valeurs propres, sommes des carrés des résidus (RESS), inerties expliquées par les X (R2X, R2X cumulés)												
35													
36													
37		Valeurs propres	RESS	R2X	R2X cumulé								
38	Composante 1	4,65602	30,91151	0,77600	0,77600								
39	Composante 2	0,91522	9,86142	0,15254	0,92854								
40	Composante 3	0,24043	4,33152	0,04007	0,96861								
41	Composante 4	0,10271	1,96921	0,01712	0,98573								
42	Composante 5	0,06466	0,48210	0,01078	0,99651								
43	Composante 6	0,02096	0,00000	0,00349	1,00000								
44													
45	Poids des variables												
46													
47		Composante 1	Composante 2	Composante 3	Composante 4	Composante 5	Composante 6						
49	Cylindrée	0,44420	0,03391	-0,40135	-0,05347	0,79843	0,00972						
50	Puissance	0,41447	0,42123	-0,03896	-0,48866	-0,30763	0,56199						
51	Vitesse	0,34351	0,66352	0,36937	0,32045	-0,00668	-0,45009						
52	Poids	0,43033	-0,25524	-0,48425	-0,12212	-0,47378	-0,52516						
	Rapport Général /												

Poids des variables						
	Composante 1	Composante 2	Composante 3	Composante 4	Composante 5	Composante 6
Cylindrée	0,44420	0,03391	-0,40135	-0,05347	0,79843	0,00972
Puissance	0,41447	0,42123	-0,03896	-0,48866	-0,30763	0,56199
Vitesse	0,34351	0,66352	0,36937	0,32045	-0,00668	-0,45009
Poids	0,43033	-0,25524	-0,48425	-0,12212	-0,47378	-0,52516
Longueur	0,43022	-0,29556	0,04304	0,71251	-0,16277	0,43766
Largeur	0,37765	-0,47818	0,68163	-0,36482	0,12986	-0,11898

Scores des observations - Composantes principales						
	Composante 1	Composante 2	Composante 3	Composante 4	Composante 5	Composante 6
Honda-Civic	-1,98032	0,31322	0,51994	-0,39849	0,11416	-0,10802
R19	-0,76212	-0,13006	0,43199	0,20848	0,22145	-0,00205
Fiat-Tipo	-1,26262	-0,42503	0,45946	-0,18595	0,16765	-0,19524
405	-0,26804	-0,45489	0,18172	0,59871	-0,02511	-0,02773
R21	0,17674	-0,62374	-0,06394	0,62020	0,32317	-0,12865

Ce rapport affiche les informations suivantes :

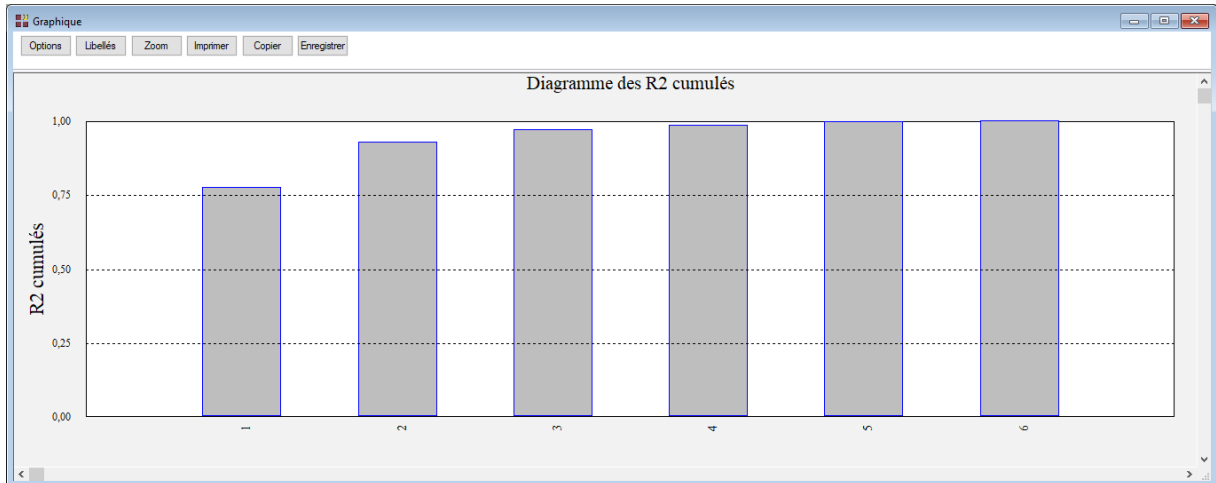
- Valeurs propres, RESS, R2X et R2X cumulés
- Poids des variables
- Scores des observations
- Corrélations entre les variables et les composantes principales
- Contributions des observations
- Cosinus carrés des observations
- Cosinus carrés cumulés des observations
- Distances carrées à l'origine des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

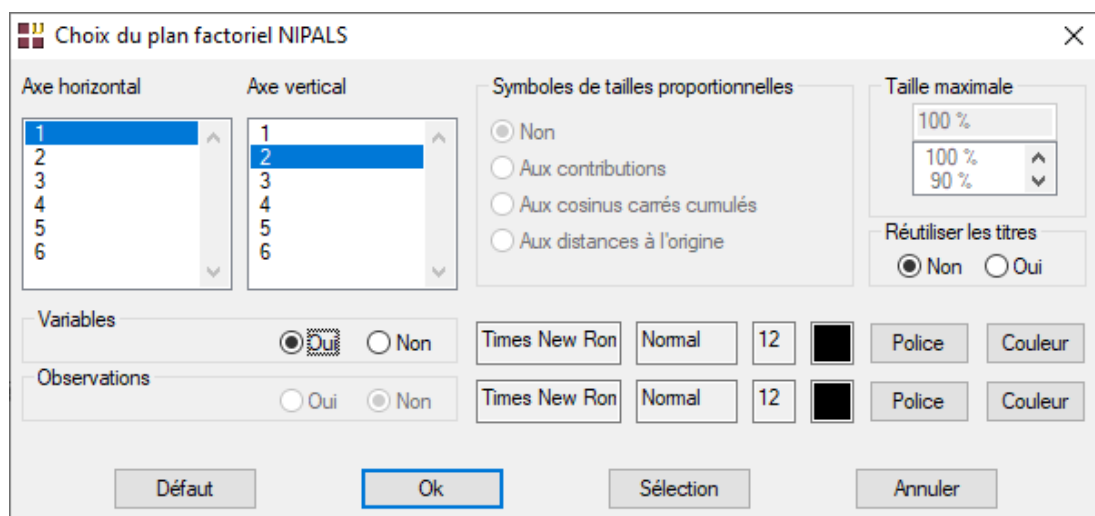
Cette option affiche les valeurs des R2 cumulés pour les axes extraits.



- 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.



Choix du plan factoriel NIPALS

Axe horizontal: 1, 2, 3, 4, 5, 6

Axe vertical: 1, 2, 3, 4, 5, 6

Symboles de tailles proportionnelles:

- Non
- Aux contributions
- Aux cosinus carrés cumulés
- Aux distances à l'origine

Taille maximale: 100 % (100 %, 90 %)

Réutiliser les titres: Non Oui

Variables: Oui Non

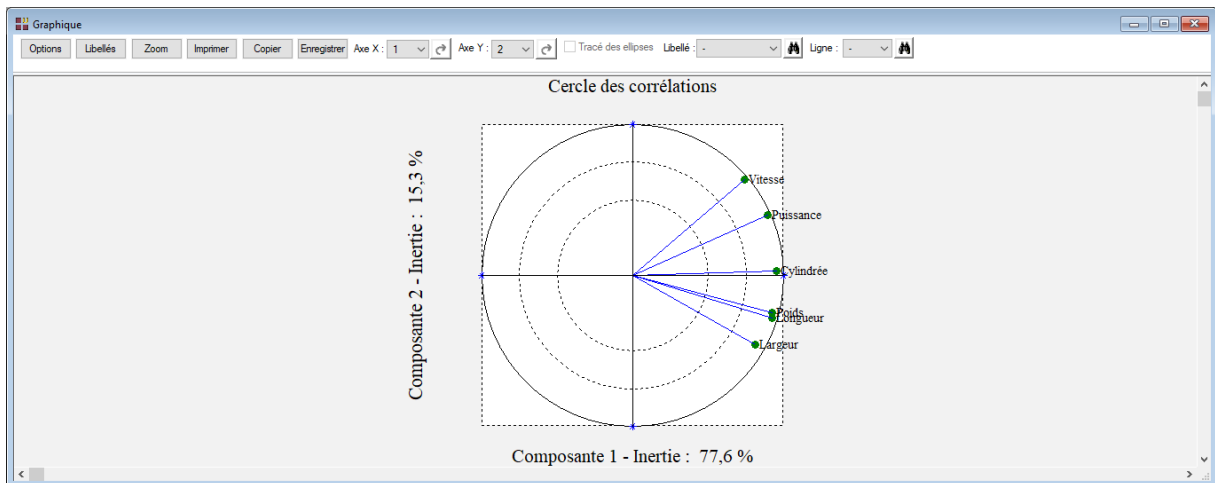
Observations: Oui Non

Times New Ron Normal 12 [Police] [Couleur]

Times New Ron Normal 12 [Police] [Couleur]

Défaut Ok Sélection Annuler

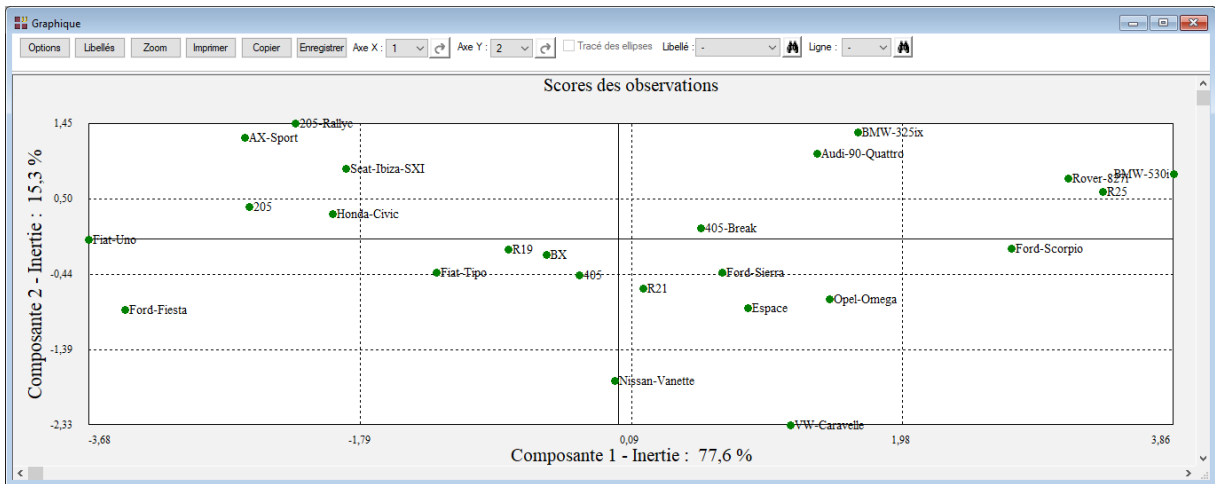
Un exemple du cercle factoriel obtenu est montré ci-après.



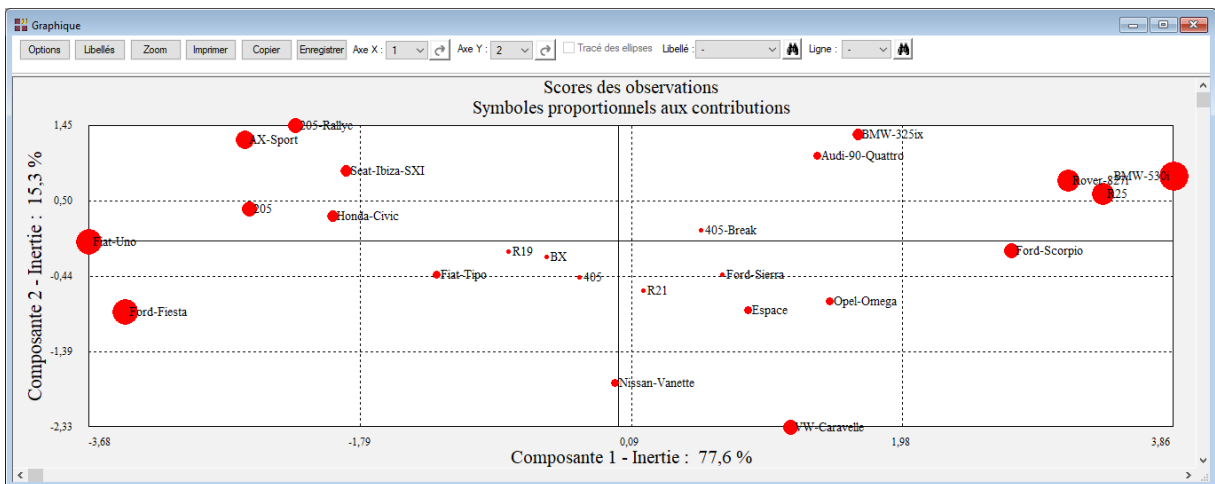
- L'option Scores des observations

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.



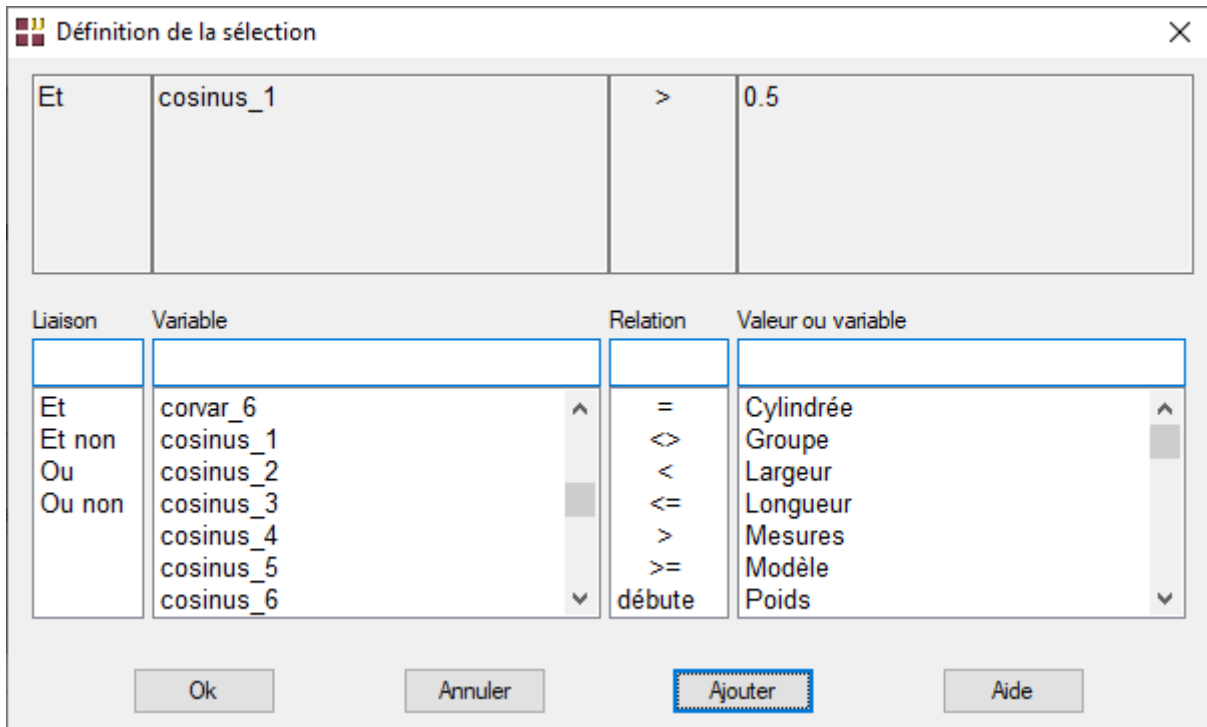
Affichons ce même graphique avec des symboles de points proportionnels aux contributions.



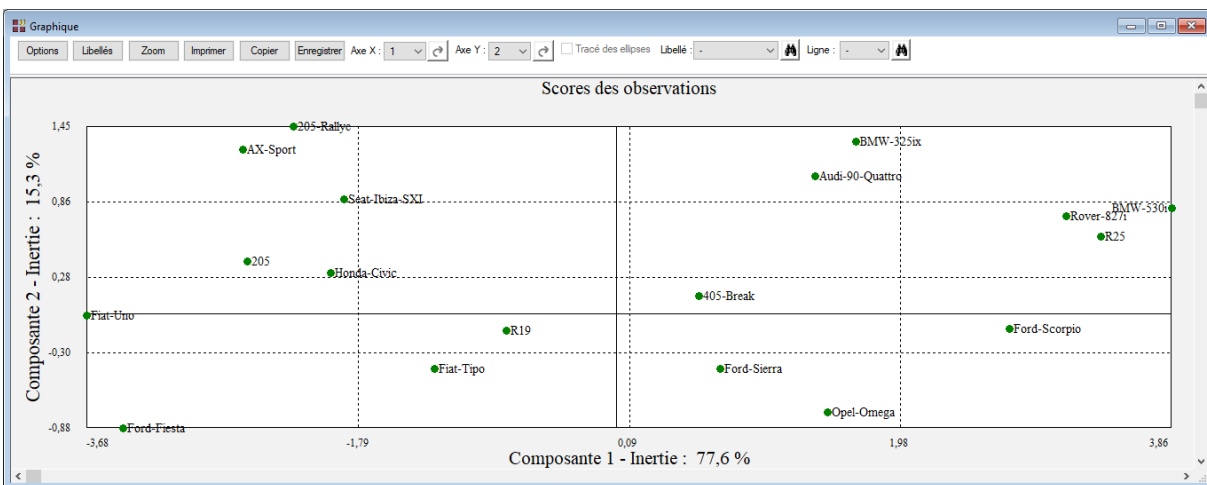
A noter le bouton 'Sélection' qui permet de n'afficher que les points qui satisfont une condition logique, par exemple une condition basée sur les contributions ou les cosinus carrés. Ceci est très pratique dans le cas d'un nuage comportant beaucoup d'observations : seuls les points significatifs d'un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



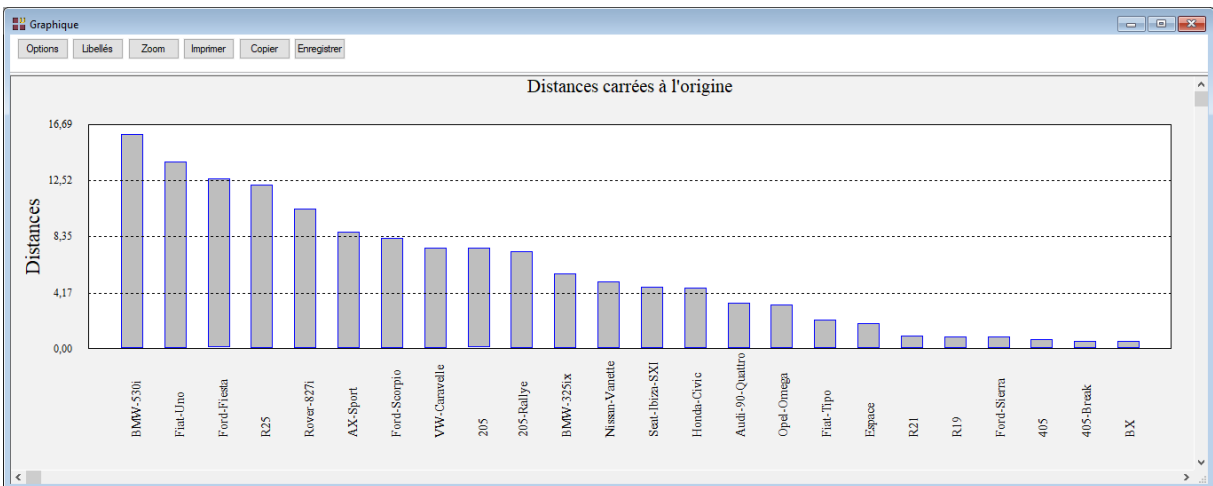
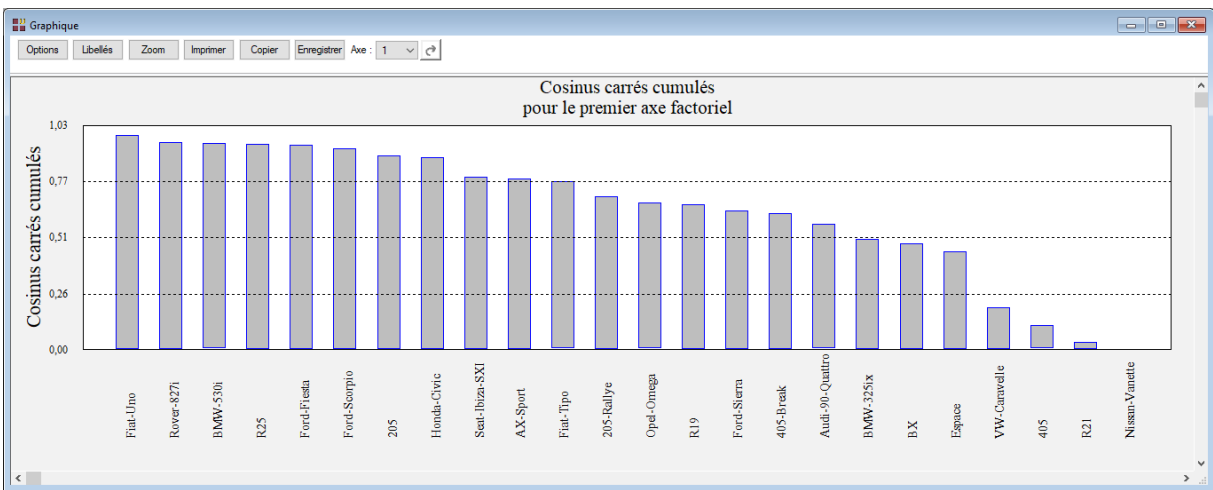
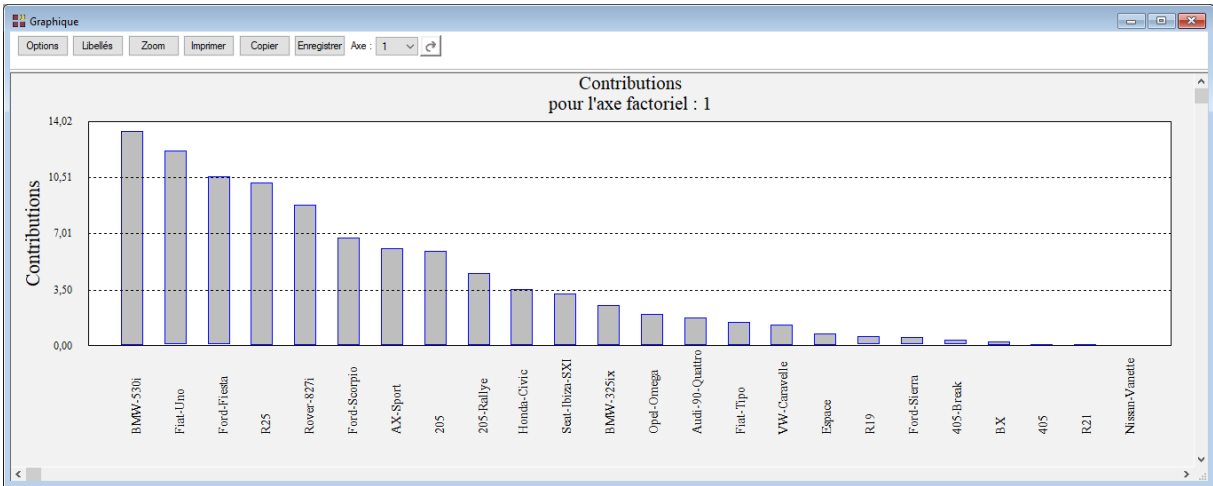
La sélection définie comporte 18 éléments.

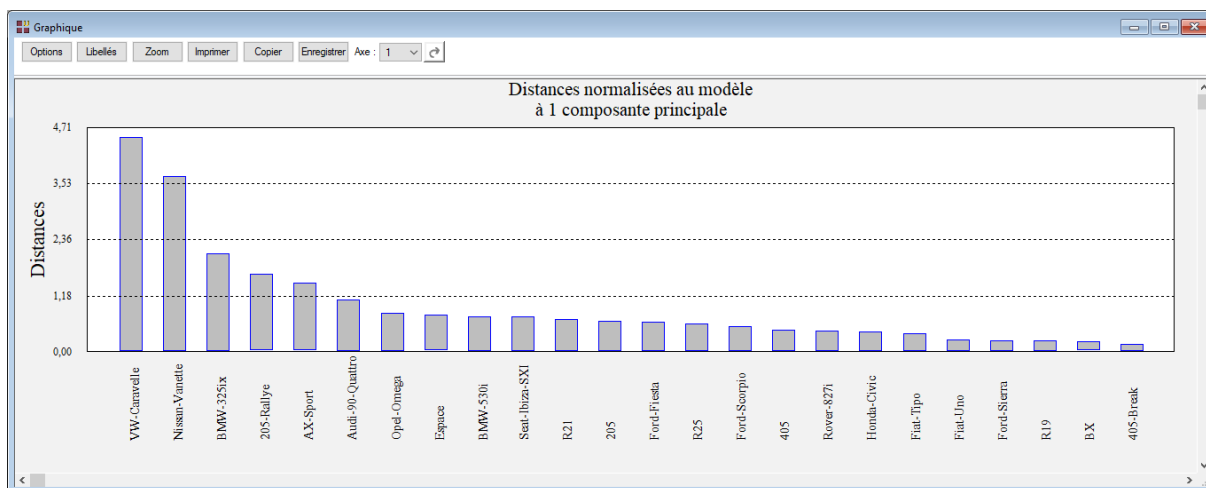


- Les options Contributions, Cosinus carrés cumulés, Distances carrées à l'origine, Distances normalisées au modèle

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.





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

	1	2	3	4	5	6	7	8
1								
2	PRESS, Q2 et Q2 cumulés pour la validation croisée numéro 1							
3								
4	PRESS (sommes des carrés des résidus prévus), Q2 et Q2 cumulés							
5								
6	Nombre de composantes significatives obtenues : 3							
7								
8	Seuil de significativité de la composante : Q2 >= 0,0975 = 1-0,95^2							
9								
10								
11								
12	Composante 1	PRESS	Q2	Q2 cumulé				
13	Composante 2	44,88501	0,67475	0,67475				
14	Composante 3	27,21345	0,39371	0,80280				
15	Composante 4	22,56617	0,17077	0,83648				
16	Composante 5	21,86689	0,03099	0,84154				
17		20,25019	0,07383	0,85326				
18								
19								
20								
21								
22								

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

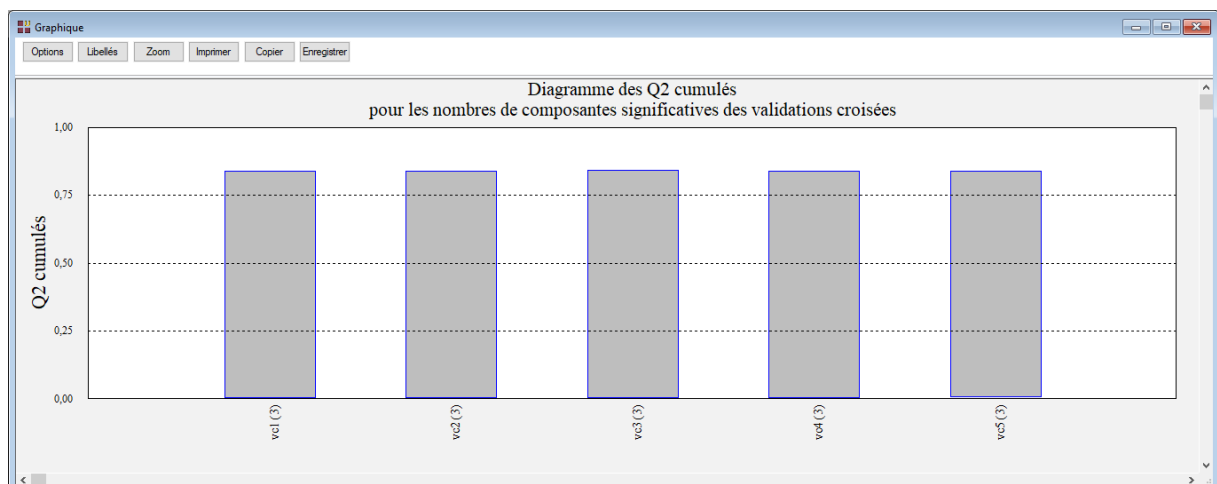
(Libellés des observations :)
Modèle

Nombre de composantes à extraire :

Nombre de validations croisées :

Méthode de validation croisée
 Aucune
 Retirer 1 à la fois
 Nombre de groupes :

Ok Annuler Sélection Supprimer Aide



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.

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 : 3

Nombre de validations croisées : 5

Méthode de validation croisée
 Aucune
 Retirer 1 à la fois
 Nombre de groupes : 7

Ok Annuler Sélection Supprimer Aide

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.

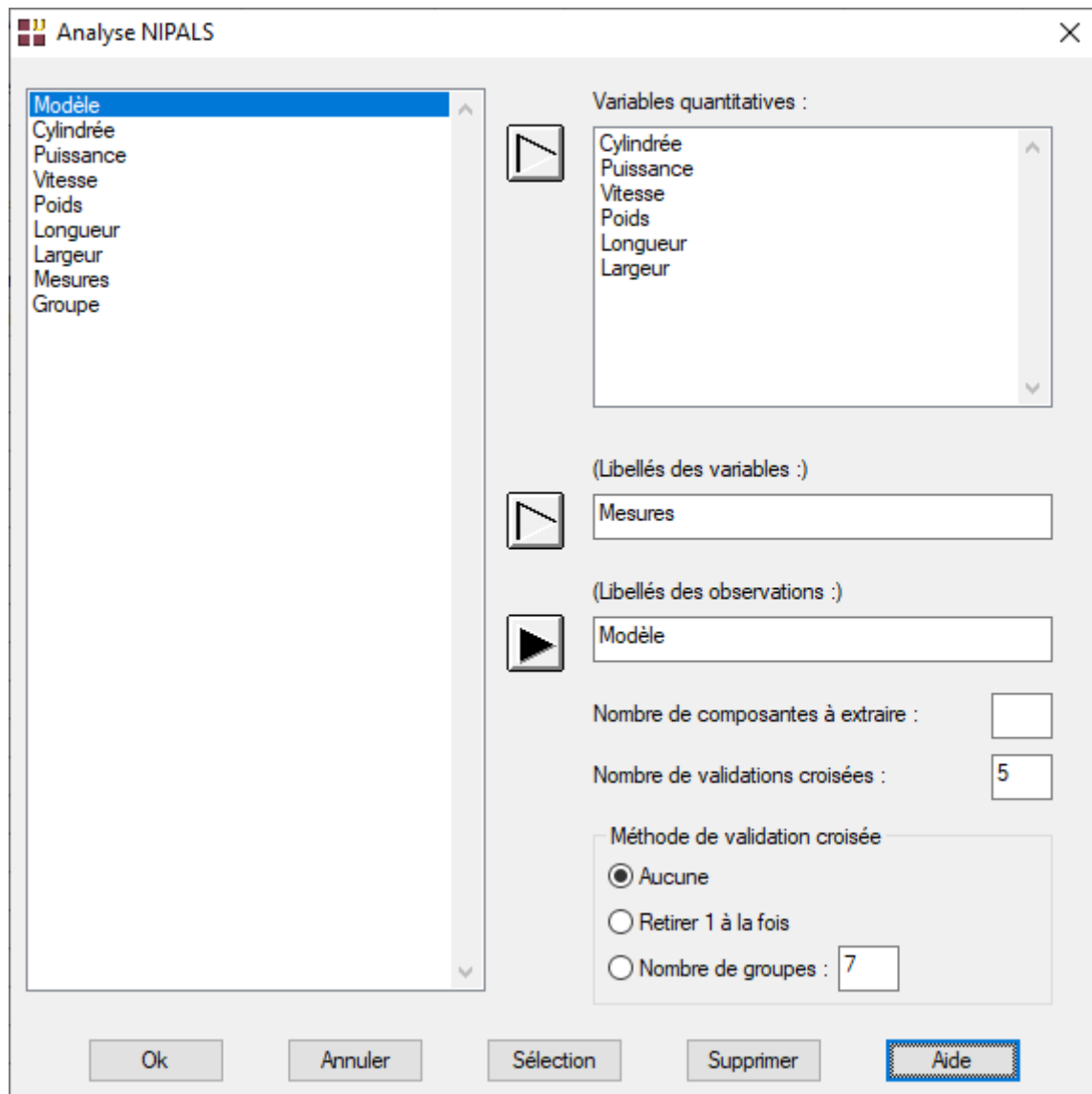
Cliquons sur l'icône NIPALS dans le ruban Décrire.

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

Rapports et Graphiques

Rapport NIPALS

- Valeurs propres, RESS, R2X, R2Xcum
- Poids des variables
- Scores des observations
- Données reconstituées
- Ecarts

1								
2	(C) UNIWIN version 9.4.1							
3								
4	DATE : 10/07/2022							
5	ORDINATEUR : LAPTOP-8CVD8RT1							
6	UTILISATEUR : Christian							
7	FICHIER(S) DE DONNÉES OUVERT(S) : VEHICULE2.SGD							
8								
9	RESULTATS DE L'ANALYSE NIPALS							
10								
11	Sélection :							
12	Aucune							
13								
14	Nombre d'observations : 24							
15								
16	Variables :							
17	Cylindrée							
18	Puissance							
19	Vitesse							
20	Poids							
21	Longueur							
22	Largeur							

Rapport Explorateur

Rapports et Graphiques

Rapport NIPALS

- Valeurs propres, RESS, R2X, R2Xcum
- Poids des variables
- Scores des observations
- Données reconstituées
- Ecarts

1								
2	Données reconstituées							
3								
4								
5		Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur	
6	Honda Civic	1292,16450	90,57145	174,30829	850,22938	368,64079	165,99927	
7	Renault 19	1725,39308	98,45000	181,39741	959,25310	414,01576	168,53631	
8	Fiat Tipo	1583,52050	82,19925	165,01010	964,10246	396,42252	170,48219	
9	Peugeot 405	1782,66880	89,33772	180,33220	1119,88247	442,02139	169,26051	
10	Renault 21	2050,69945	86,47287	180,14108	1127,01625	437,68737	169,77107	
11	Citroen BX	1753,84179	89,21982	182,49038	1050,51171	423,83302	168,12802	
12	BMW 530i	2799,87503	185,80772	224,86377	1507,25393	472,24305	174,74149	
13	Rover-827i	2663,50840	162,74865	222,07097	1361,51983	467,35648	174,65036	
14	Renault 25	2488,85594	179,31560	216,90397	1325,50756	467,74584	179,25835	
15	Opel Omega	1994,57496	121,34792	190,34094	1276,19829	474,96901	177,21851	
16	Peugeot 405 Break	1896,33072	124,33650	193,96129	1118,28070	428,14689	170,82771	
17	Ford Sierra	1972,34529	115,23470	185,28395	1181,40912	449,44692	172,01783	
18	BMW 325ix	2363,01917	168,97022	205,84211	1305,36917	430,88685	163,19259	
19	Audi 90 Quattro	1970,79111	133,85075	212,87353	1219,67905	436,21528	168,51604	
20	Ford Scorpio	2930,84047	148,48235	210,90811	1337,18589	468,93292	176,13510	
21	Renault Espace	1964,55788	119,30345	175,61801	1191,83803	435,72483	177,11120	
22	Nissan Vanette	1945,29620	89,02886	144,06043	1442,45141	455,51736	169,05352	

Rapport Explorateur

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

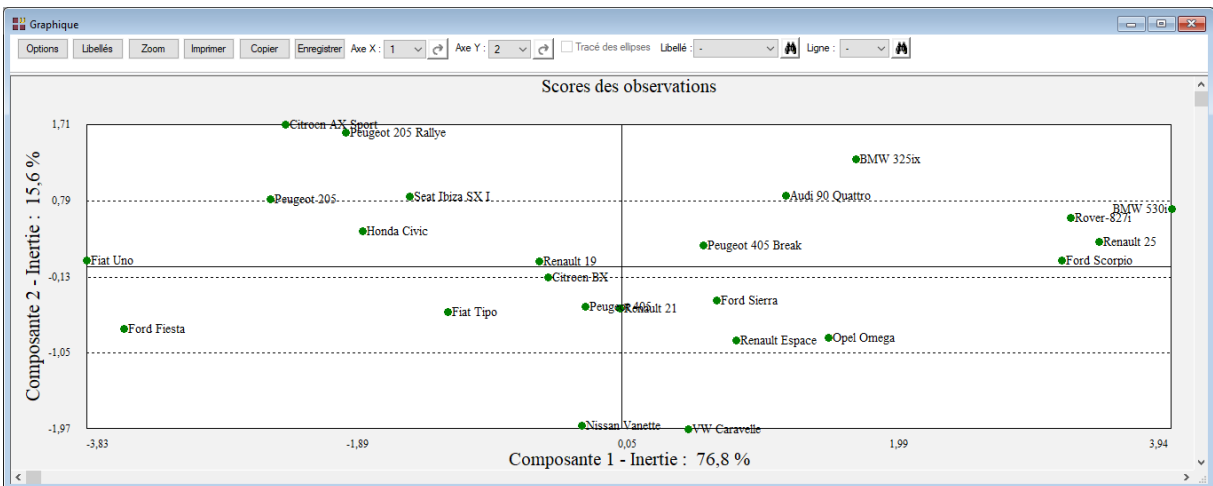
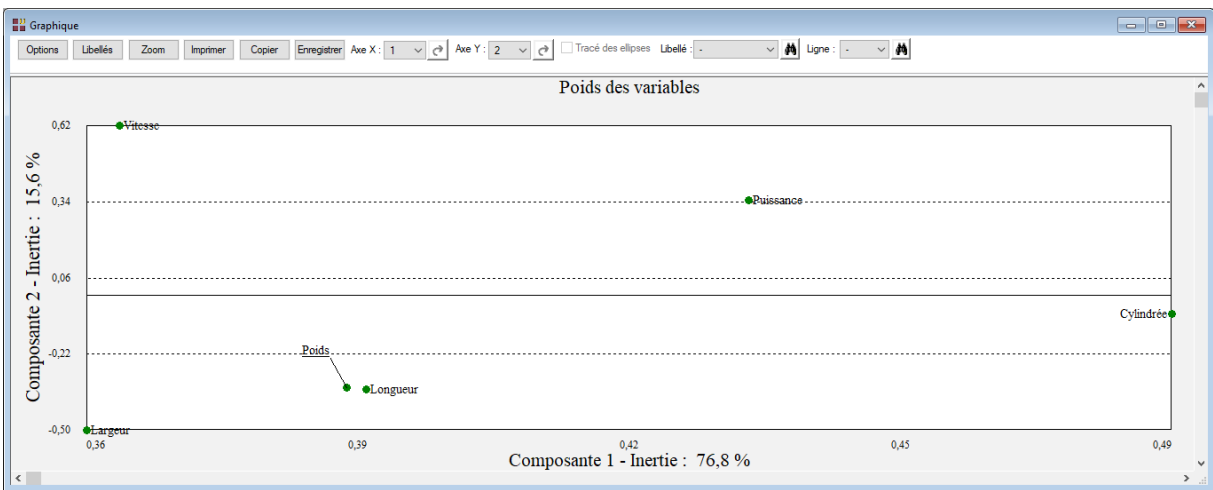
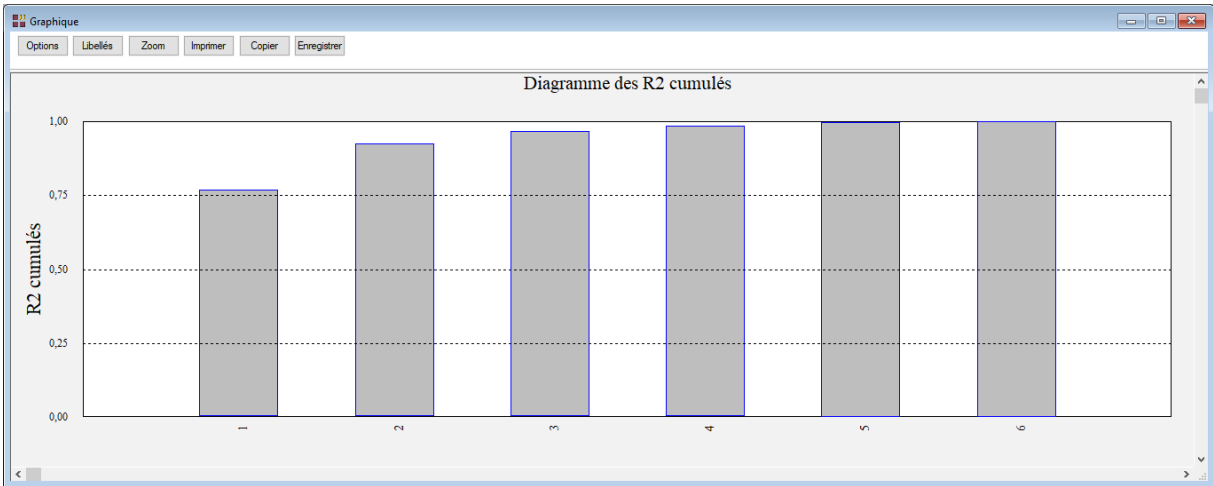
Graphiques

Diagramme des R2 cumulés

Poids des variables

Scores des observations

Ok Annuler



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.

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 : 10

Méthode de validation croisée
 Aucune
 Retirer 1 à la fois
 Nombre de groupes : 7

Ok Annuler Sélection Supprimer Aide

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.

Rapports et Graphiques

Rapport NIPALS

Valeurs propres, PRESS, R2X, R2Ycum

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, Q2 cumulés (vc=6)

PRESS, Q2, Q2 cumulés (vc=7)

PRESS, Q2, Q2 cumulés (vc=8)

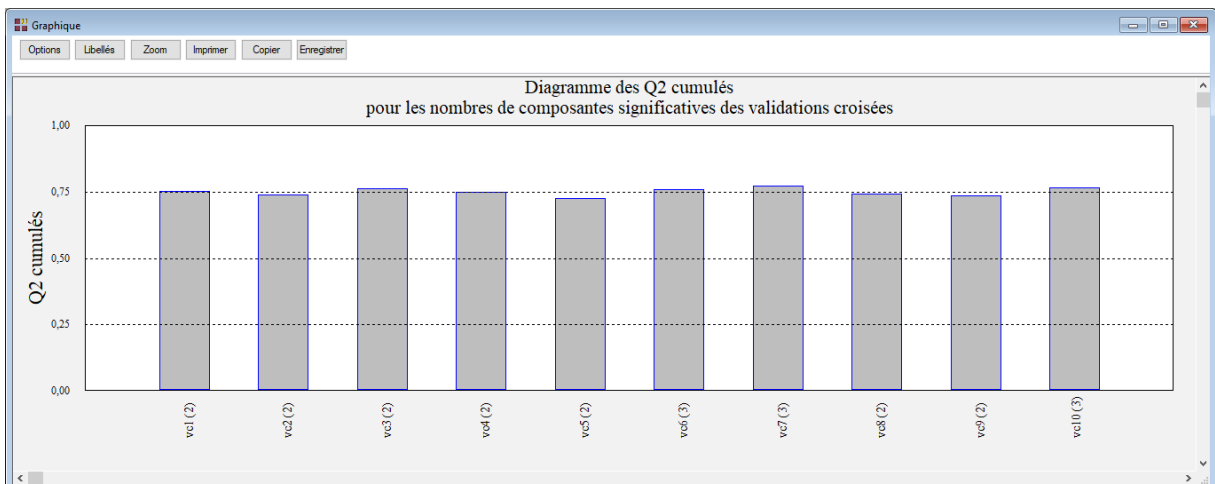
PRESS, Q2, Q2 cumulés (vc=9)

PRESS, Q2, Q2 cumulés (vc=10)

	1	2	3	4	5	6	7	8
1								
2	PRESS, Q2 et Q2 cumulés pour la validation croisée numéro 1							
3								
4	PRESS (sommes des carrés des résidus prévus), Q2 et Q2 cumulés							
5								
6	Nombre de composantes significatives obtenues : 2							
7								
8	Seuil de significativité de la composante : $Q2 \geq 0,0975 = 1-0,95^2$							
9								
10								
11		PRESS	Q2	Q2 cumulé				
12	Composante 1	41,88969	0,63255	0,63255				
13	Composante 2	26,49271	0,31982	0,75006				
14	Composante 3	26,69713	0,06302	0,76581				
15	Composante 4	30,09165	-0,12715	0,73604				
16	Composante 5	28,58071	0,05021	0,74929				
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

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	Ecart 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