

UNIWIN VERSION 9.7.0

ANALYSE FACTORIELLE DE DONNEES MIXTES

Révision : 02/09/2023

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichier GIFL	3
L'option Rapports	7
L'option Graphiques	8
Une rapide interprétation des résultats.....	14
Exemple 2 : Fichier AUTOS.....	15
L'option Rapports	16
L'option Graphiques	18
Les variables internes créées par la procédure	20

Définition

L'Analyse Factorielle de Données Mixtes (AFDM) est une méthode spécialement conçue pour permettre l'étude simultanée de variables quantitatives et qualitatives (données dites mixtes) mesurées sur une population d'individus en tant qu'éléments actifs dans une même analyse.

Cette analyse prend en compte les variables quantitatives comme une analyse en composantes principales normée (ACP) et les variables qualitatives comme une analyse des correspondances multiples (ACM).

Après l'affichage du tableau et de l'histogramme des inerties, vous pouvez choisir le nombre d'axes factoriels à extraire.

Un rapport général de synthèse est proposé ainsi que les graphiques des cercles factoriels, des représentations simultanées des variables quantitatives et qualitatives, des plans factoriels des individus et des modalités. De nombreux outils d'aide à l'interprétation sont fournis, notamment pondérations des indicatrices des variables qualitatives, liaisons entre les variables initiales, contributions des variables à l'inertie et pour les plans factoriels les contributions et cosinus carrés.

Entrée des données

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

Cette boîte de dialogue permet de sélectionner les variables quantitatives et les variables qualitatives. Elle permet également d'indiquer les noms des variables contenant les libellés des individus, des variables quantitatives et qualitatives ainsi que les libellés des modalités de ces variables qualitatives.

Si plusieurs variables qualitatives interviennent dans l'analyse, les libellés des modalités doivent être indiqués dans l'ordre des variables qualitatives sélectionnées.

Données manquantes

Dans cette procédure les données manquantes ne sont pas autorisées.

Exemple 1 : Fichier GIFI

Nous utiliserons le fichier GIFI contenant les données de Russet, popularisées en France par Tenenhaus, qui les emprunta à Gifi.

Ces données comportent plusieurs groupes de variables mais un seul de ces groupes sera pris en compte ici.

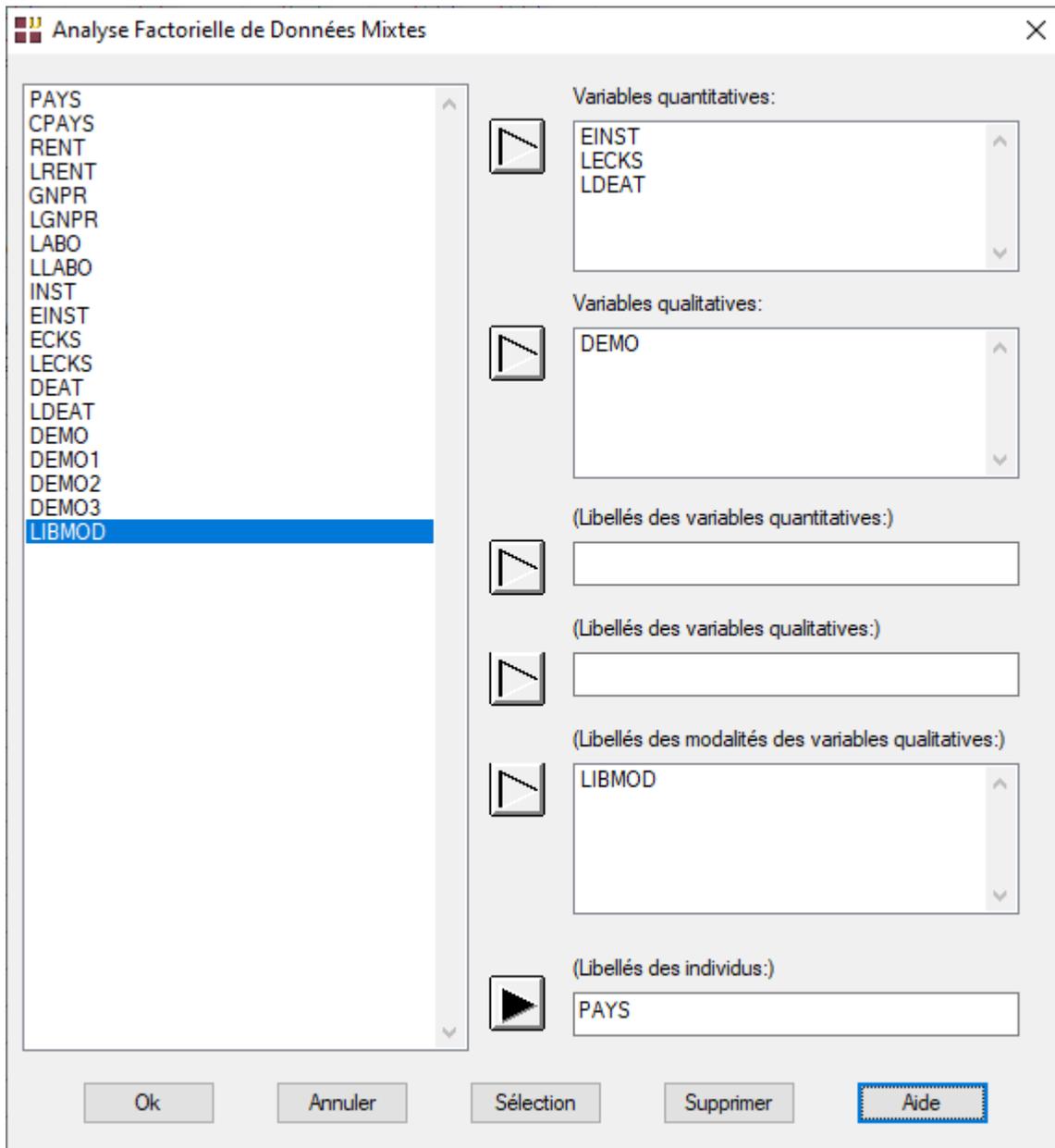
On dispose pour 47 pays de quatre mesures de l'instabilité politique :

- La variable INST est une fonction du nombre de responsables du pouvoir exécutif et du nombre d'années pendant lesquelles le pays a été indépendant entre 1945 et 1961. Cet indice varie entre 0 (très stable) et 17 (très instable).
- La variable ECKS est l'indice d'Eckstein calculé sur la période 1946-1961. Il mesure le nombre de conflits violents entre communautés sur cette période.
- La variable DEAT est le nombre de personnes tuées lors de manifestations violentes sur la période 1950-1962.
- La variable DEMO classe les pays en trois groupes : démocratie stable, démocratie instable et dictature.

Nous utiliserons les transformations des données introduites par Gifi :

Variable d'origine	Transformation	Variable transformée
Inst	Exp (Inst-16,3)	EINST
Ecks	Ln (Ecks + 1)	LECKS
Deat	Ln (Deat + 1)	LDEAT

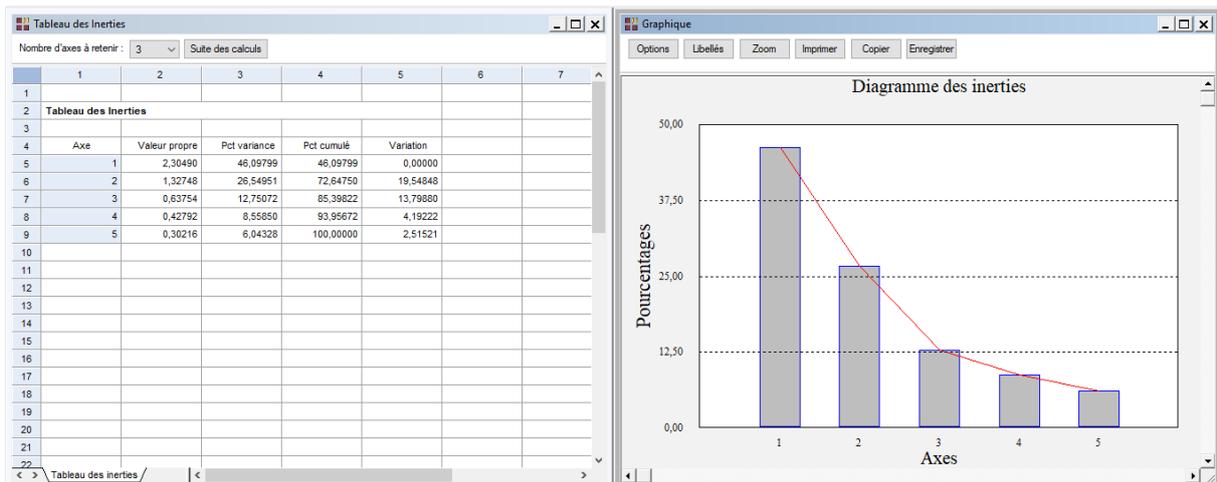
Cliquons sur l'icône AFDM dans le ruban Décrire et renseignons la boîte de dialogue comme montré ci-après :



La variable PAYS contient les libellés des 47 pays et la variable LIBMOD les libellés des trois modalités de la variable DEMO.

Cliquons sur Ok pour exécuter le traitement de l'analyse.

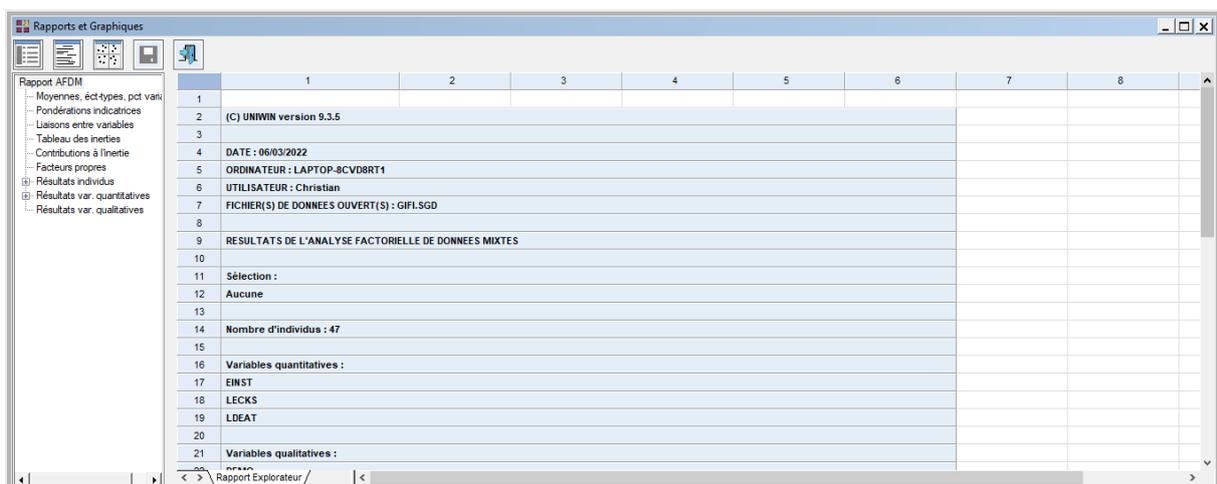
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.



L'option 'Nombre d'axes à retenir' permet de préciser le nombre d'axes à extraire.

Cliquons sur le bouton 'Suite des calculs'.

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.

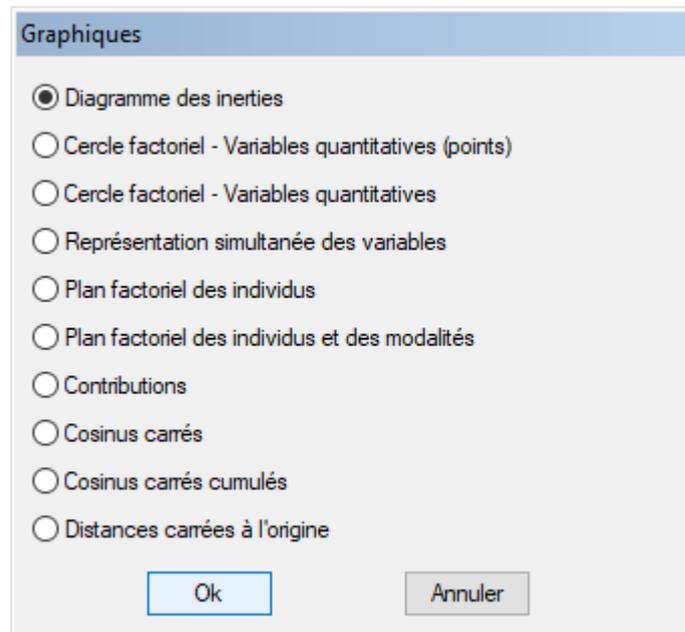
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :

The 'Rapports' dialog box contains the following options:

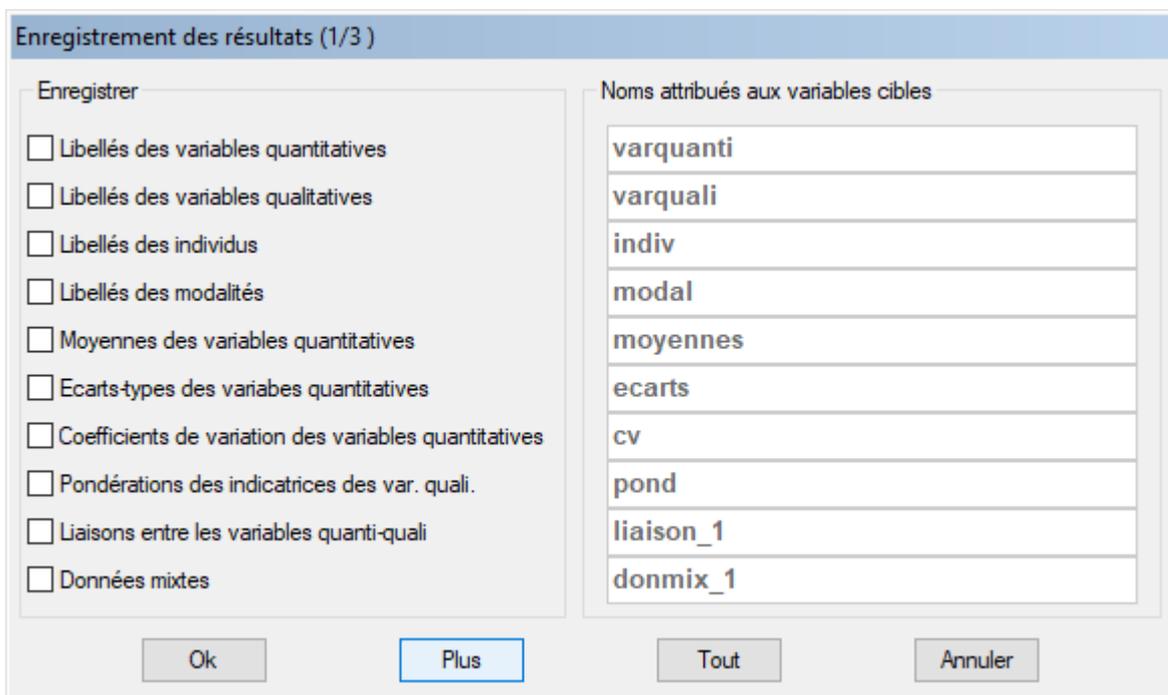
- Rapport Explorateur
- Rapport Général
- Rapport Html

Buttons: Ok, Annuler

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.



Note : le bouton 'Plus' permet d'afficher la suite de la liste des variables.

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 AFDM : Explorateur, Général, HTML.

	1	2	3	4	5	6	7	8
1								
2	RESULTATS INDIVIDUS POUR LE FACTEUR : 1							
3								
4	DISTANCE*2 = CARRES DES DISTANCES A L'ORIGINE							
5	COMP. PRIN. = COMPOSANTES PRICIPALES - COORDONNEES DES INDIVIDUS							
6	CONTRIB. = CONTRIBUTIONS A L'INERTIE							
7	COSINUS*2 = COSINUS CARRES							
8	COS*2 CUM. = SOMMES CUMULEES DES COSINUS CARRES							
9								
10								
11		INDIV/DU	DISTANCE*2	COMP. PRIN.	CONTRIB.	COSINUS*2	COS*2 CUM.	
12	RFA	1	4,79300	-0,85917	0,68142	0,15401	0,15401	
13	Argentine	2	6,65123	1,47707	2,01396	0,32802	0,32802	
14	Australie	3	7,03498	-2,56261	6,06200	0,93348	0,93348	
15	Autriche	4	4,60082	-0,02187	0,62353	0,14681	0,14681	
16	Belgique	5	3,53472	-0,91854	0,77884	0,23870	0,23870	
17	Bolivie	6	6,75392	2,29963	4,88163	0,78299	0,78299	
18	Brsil	7	5,29308	0,83841	0,64888	0,13280	0,13280	
19	Canada	8	3,68589	-1,21758	1,36850	0,40221	0,40221	
20	Chili	9	3,32965	0,17289	0,02759	0,00898	0,00898	
21	Colombie	10	6,54695	1,62213	2,42896	0,40191	0,40191	

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
33														
34	PONDERATIONS DES INDICATRICES DES VARIABLES QUALITATIVES													
35														
36														
37		PONDERATIONS												
38	Stable		0,56493											
39	Instable		0,50529											
40	Dictature		0,65233											
41														
42	LIAISONS ENTRE LES VARIABLES INITIALES													
43														
44	> VARIABLES QUANTITATIVES ET QUANTITATIVES : CARRE DU COEFFICIENT DE CORRELATION													
45	> VARIABLES QUANTITATIVES ET QUALITATIVES : CARRE DU RAPPORT DE CORRELATION													
46														
47														
48		ENST	LECKS	LDEAT	DEMO									
49	ENST	1,00000	0,32608	0,06984	0,16899									
50	LECKS	0,32608	1,00000	0,62274	0,36614									
51	LDEAT	0,06984	0,62274	1,00000	0,34366									
52	DEMO	0,16899	0,36614	0,34366										
53														

(C) UNIWIN version 9.3.5

DATE : 06/03/2022
 ORDINATEUR : LAPTOP-8CVDBRT1
 UTILISATEUR : Christian
 FICHIER(S) DE DONNEES OUVERT(S) : GIF1.SGD

RESULTATS DE L'ANALYSE FACTORIELLE DE DONNEES MIXTES

Sélection :
 Aucune

Nombre d'individus : 47

Variables quantitatives :
 ENST
 LECKS
 LDEAT

Variables qualitatives :
 DEMO

MOYENNES, ECARTS-TYPES ET COEFFICIENTS DE VARIATION DES VARIABLES QUANTITATIVES

	MOYENNES	ECARTS-TYPES	COEF. DE VARIATION
ENST	0,20199	0,24234	119,97601

Ce rapport donne les informations suivantes :

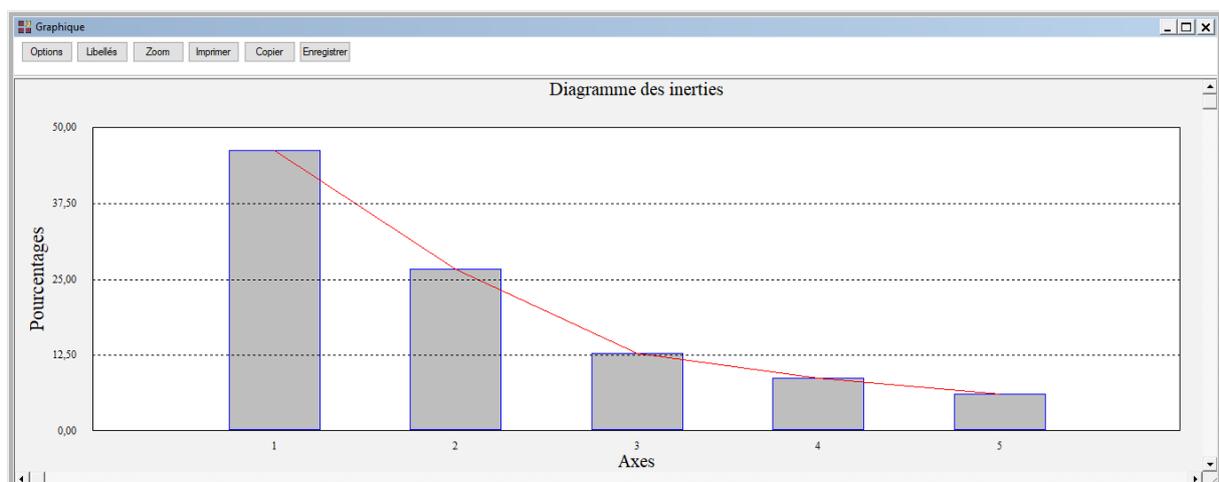
- Moyennes, écarts-types et coefficients de variation pour les variables quantitatives
- Pondérations des indicatrices des variables qualitatives
- Liaisons entre variables quantitatives (carrés des coefficients de corrélation)
- Liaisons entre variables qualitatives et variables quantitatives (carrés des rapports de corrélation)
- Tableau des inerties
- Contributions des variables quantitatives à l'inertie (carrés des corrélations entre les variables et les facteurs)
- Contributions des variables qualitatives à l'inertie (carrés des rapports de corrélation entre les variables et les facteurs)
- Facteurs propres
- Résultats pour les individus (carré de la distance à l'origine, coordonnée sur la composante principale, contribution à l'inertie, cosinus carré, cosinus carré cumulé)
- Affichage des éventuels points multiples
- Résultats pour les variables quantitatives (coordonnée, facteur propre, contribution à l'inertie, cosinus carré, cosinus carré cumulé)
- Résultats pour les modalités des variables qualitatives (coordonnées sur les composantes principales)

L'option Graphiques

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

- Diagramme des inerties

Ce graphique affiche les pourcentages d'inertie pour chacun des axes factoriels.

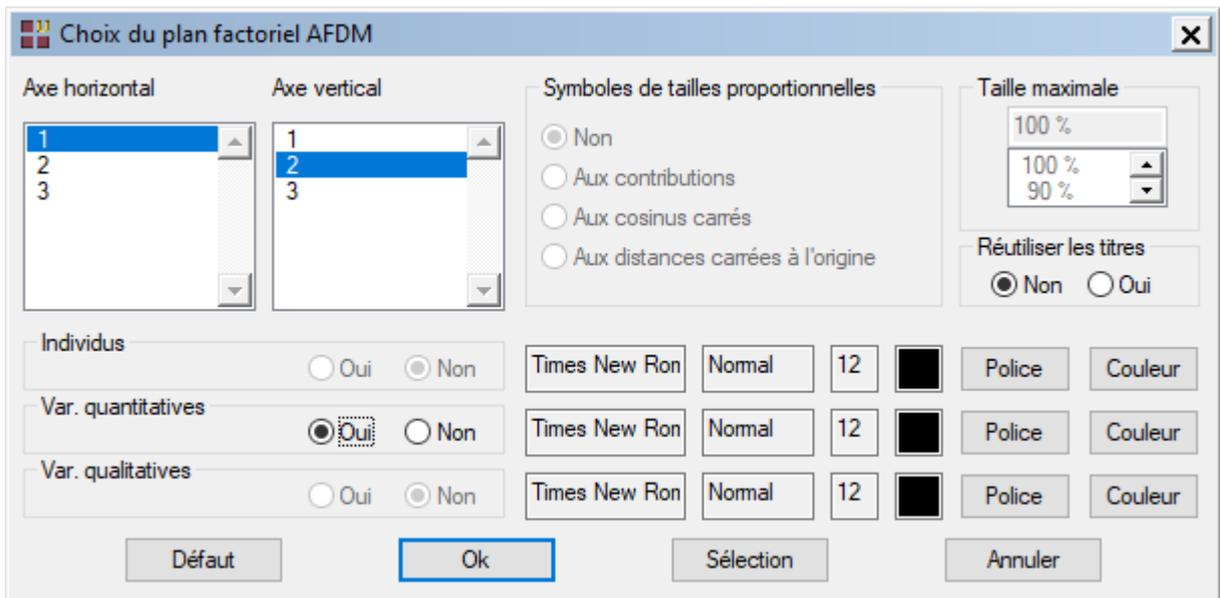


- Les options Cercles factoriels

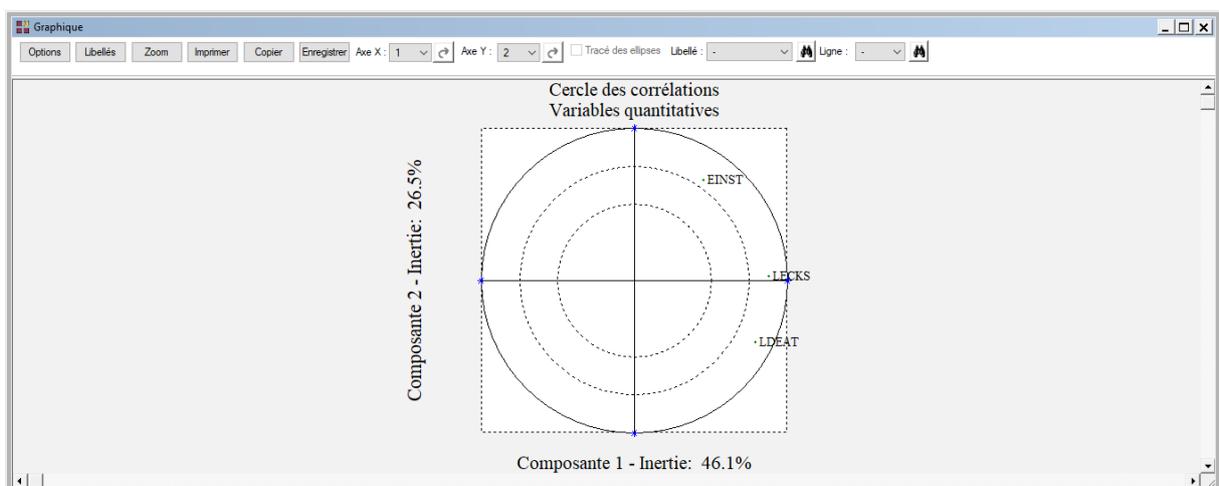
Ces options permettent d'afficher le cercle de corrélations des variables quantitatives avec les facteurs et de choisir si on désire tracer les droites reliant les points à l'origine du cercle. L'option sans ces droites (points) est utile lorsqu'il y a un grand nombre de variables quantitatives représentées.

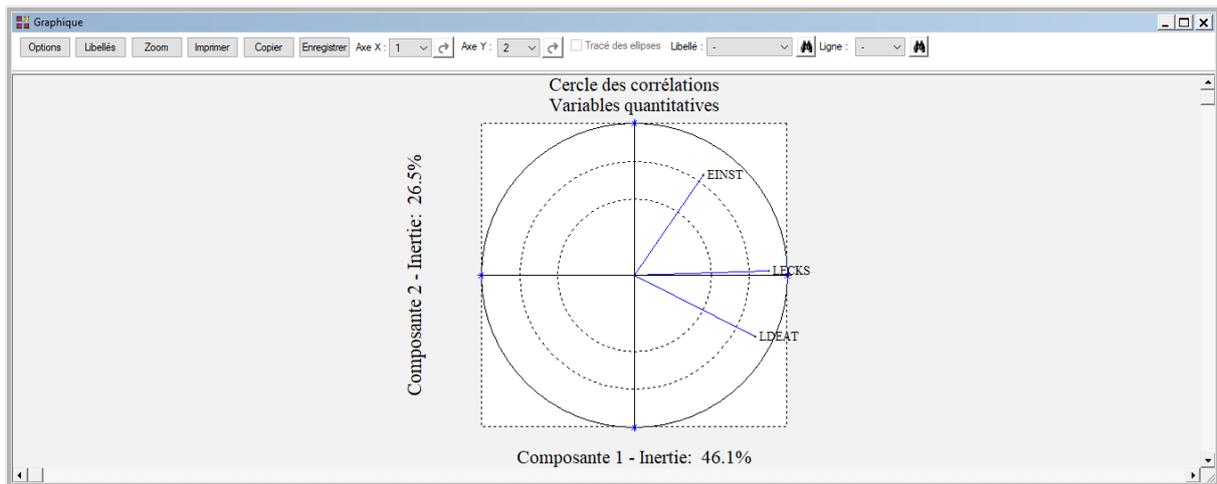
Choisissons les graphiques sans les droites puis avec les droites.

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.





- L'option Représentation simultanée des variables

Cette option permet de représenter de façon simultanée les variables quantitatives et les variables qualitatives par leurs contributions à l'inertie c'est-à-dire par les carrés des coefficients de corrélation entre variables quantitatives et les carrés des rapports de corrélation entre variables qualitatives et variables quantitatives.

Choix du plan factoriel AFDM

Axe horizontal: 1, 2, 3
Axe vertical: 1, 2, 3

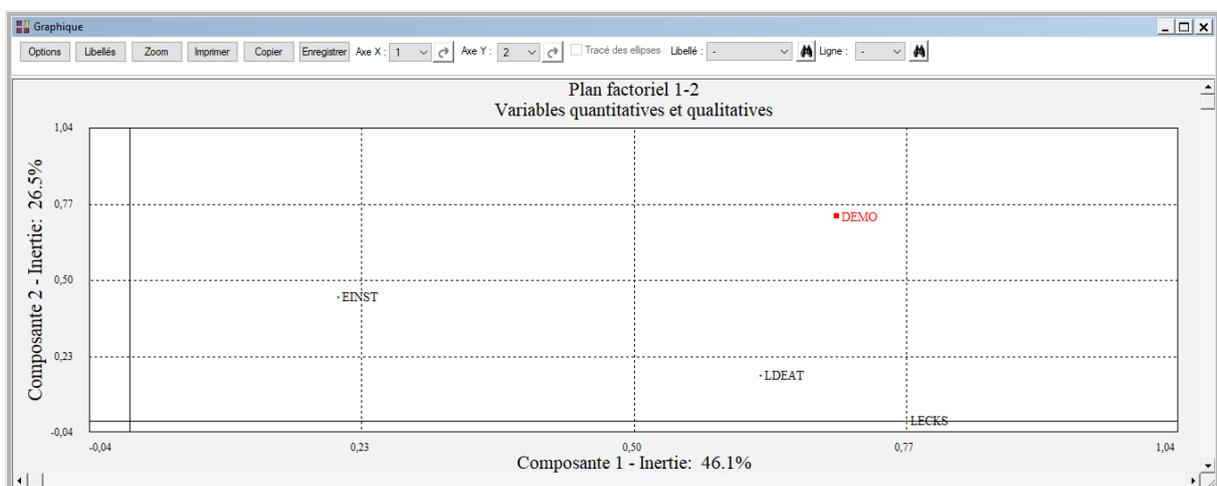
Symboles de tailles proportionnelles:
 Non
 Aux contributions
 Aux cosinus carrés
 Aux distances carrées à l'origine

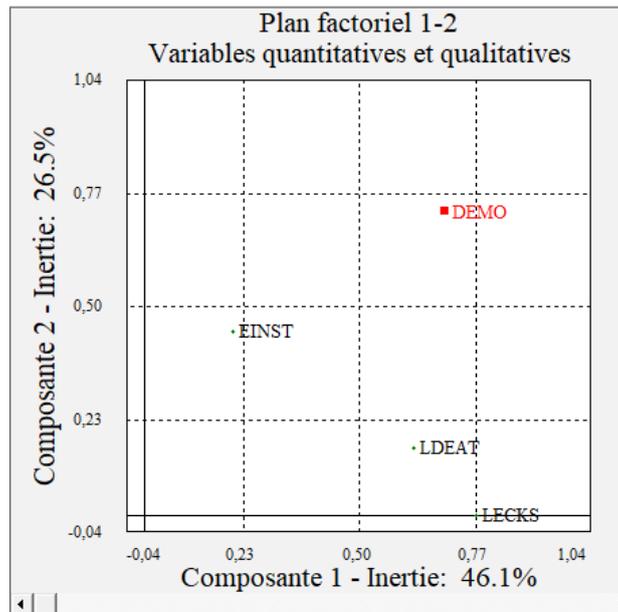
Taille maximale: 100%, 100%, 90%

Réutiliser les titres: Non Oui

Individus: Oui Non
 Var. quantitatives: Oui Non
 Var. qualitatives: Oui Non

Défaut Ok Sélection Annuler

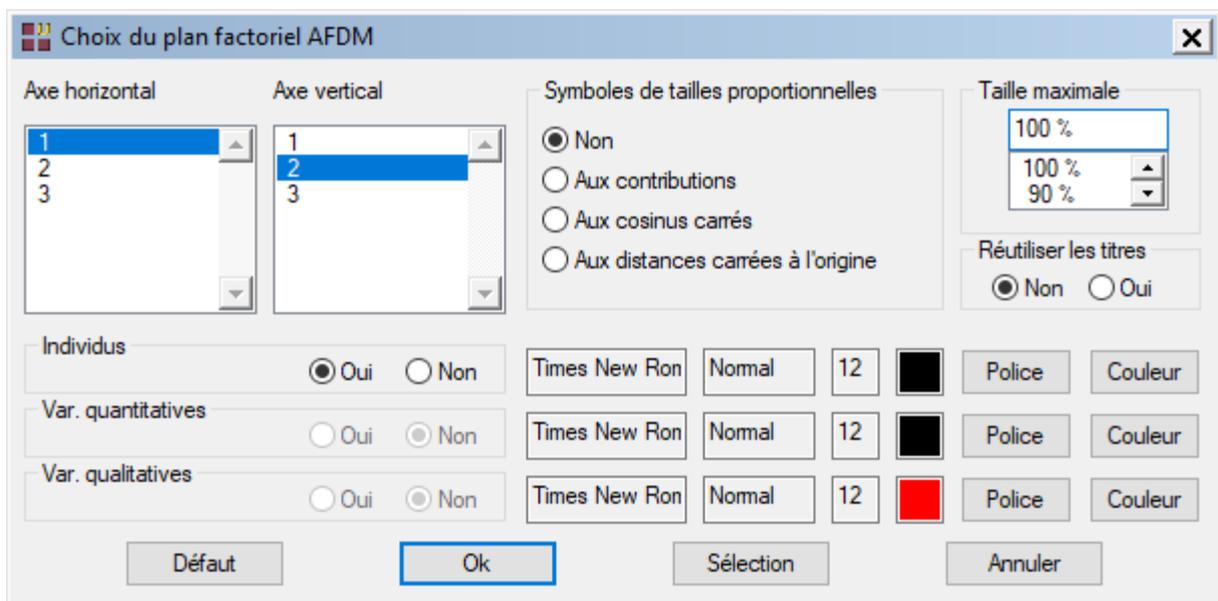




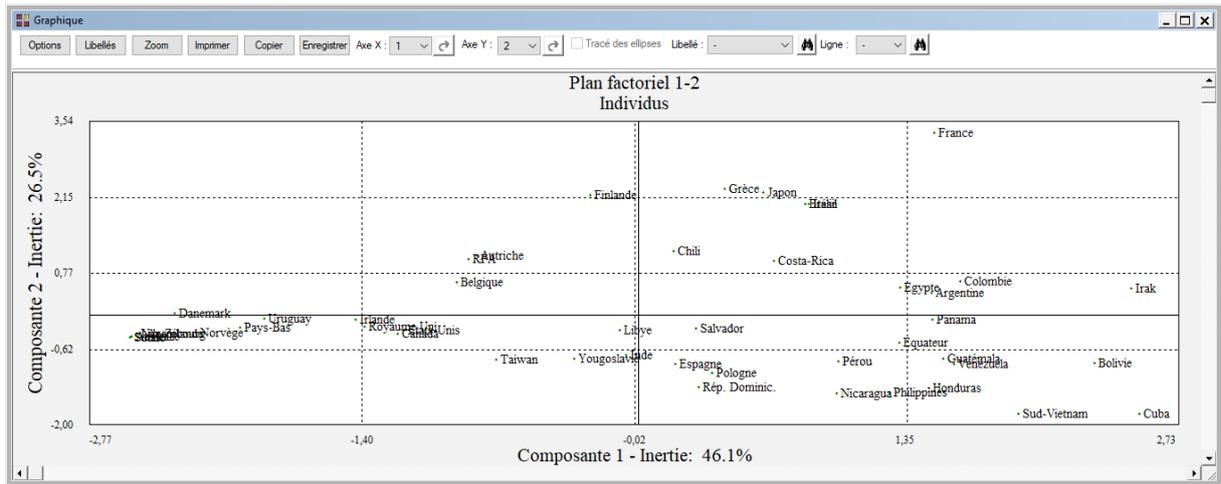
Le graphique est ici représenté avec des échelles orthonormées (ratio 1/1).

- L'option Plan factoriel des individus

Cette option permet d'afficher des plans factoriels des individus. 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 individus, 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 carrées 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



A noter que les options d'exploration graphique décrites dans le 'Manuel de Utilisateur' sont toutes accessibles, notamment l'interrogation d'un point et le zoom.

A noter également 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'individus : 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 GIFL.

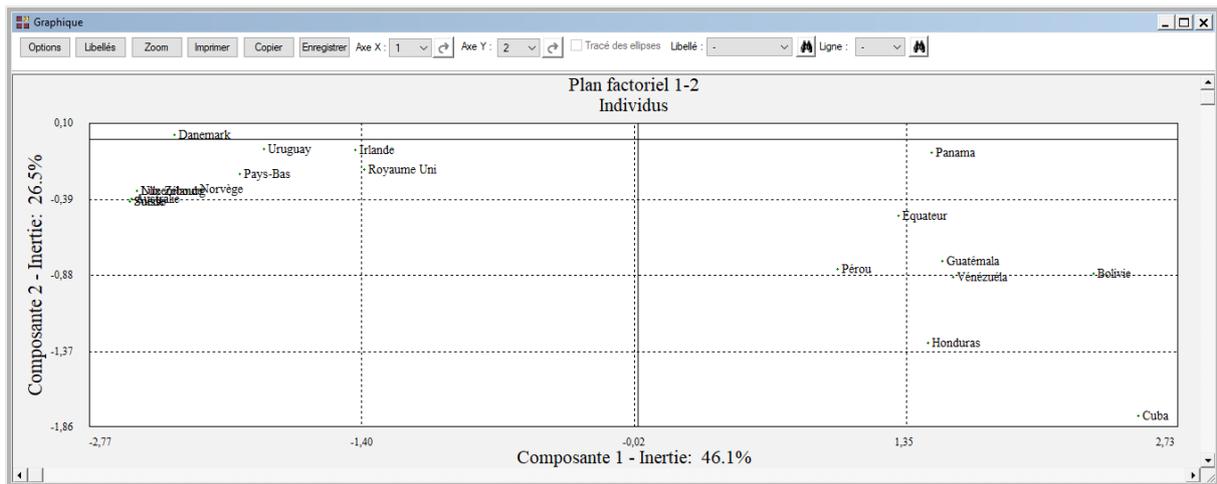
Comme exemple, mettons en œuvre une sélection basée sur la variable 'cosinus_1' qui contient les cosinus carrés sur le premier axe factoriel.

Définition de la sélection

Et cosinus_1 >= 0.5

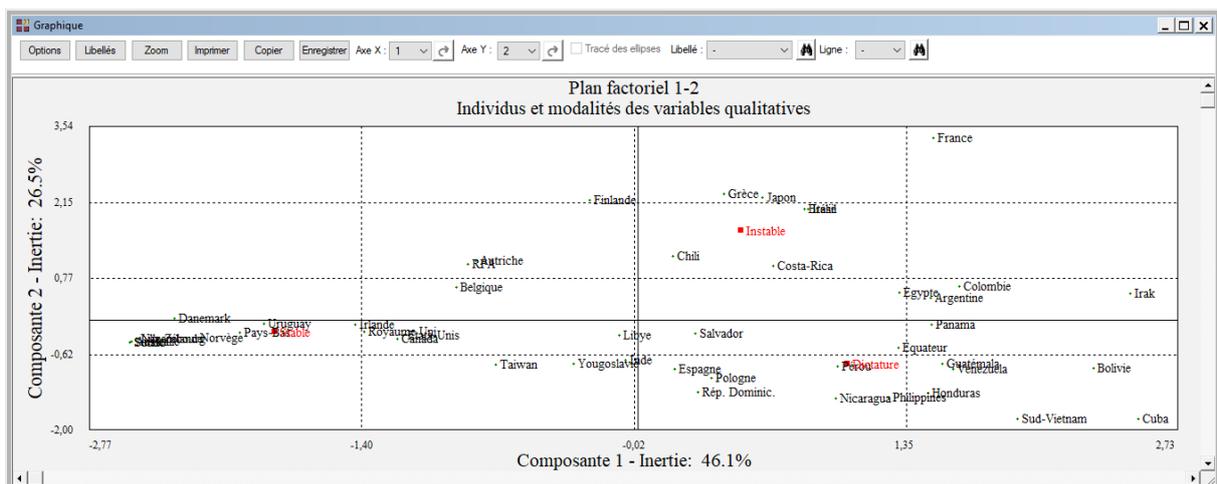
Liaison	Variable	Relation	Valeur ou variable
Et	coiner_3	=	CPAYS
Et non	coiner_4	<>	DEAT
Ou	contrib_1	<	DEMO
Ou non	contrib_2	<=	DEMO1
	contrib_3	>	DEMO2
	cosinus_1	>=	DEMO3
	cosinus_2	début	ECKS

Buttons: Ok, Annuler, Ajouter, Aide



- L'option Plan factoriel des individus et des modalités

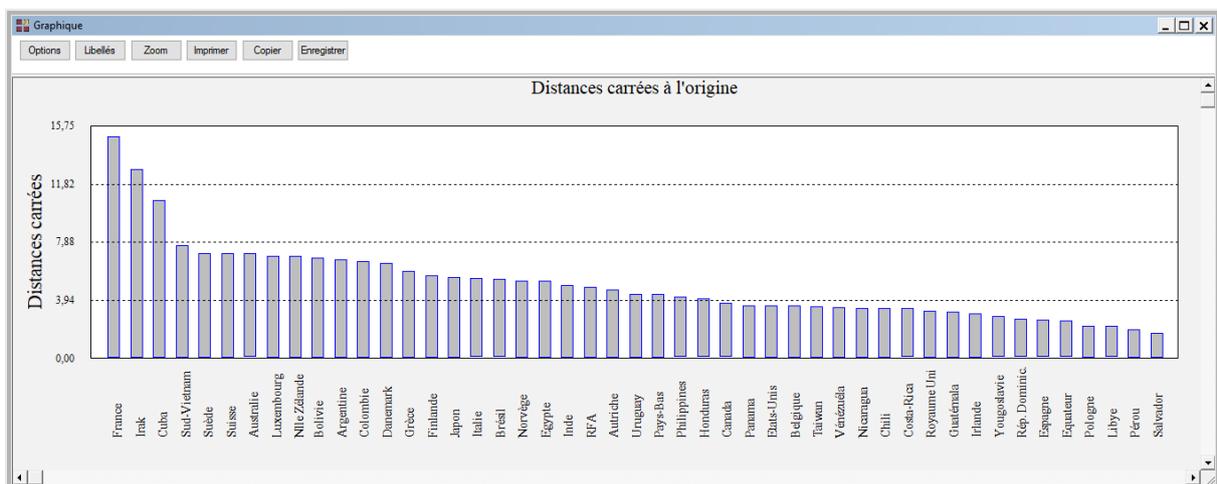
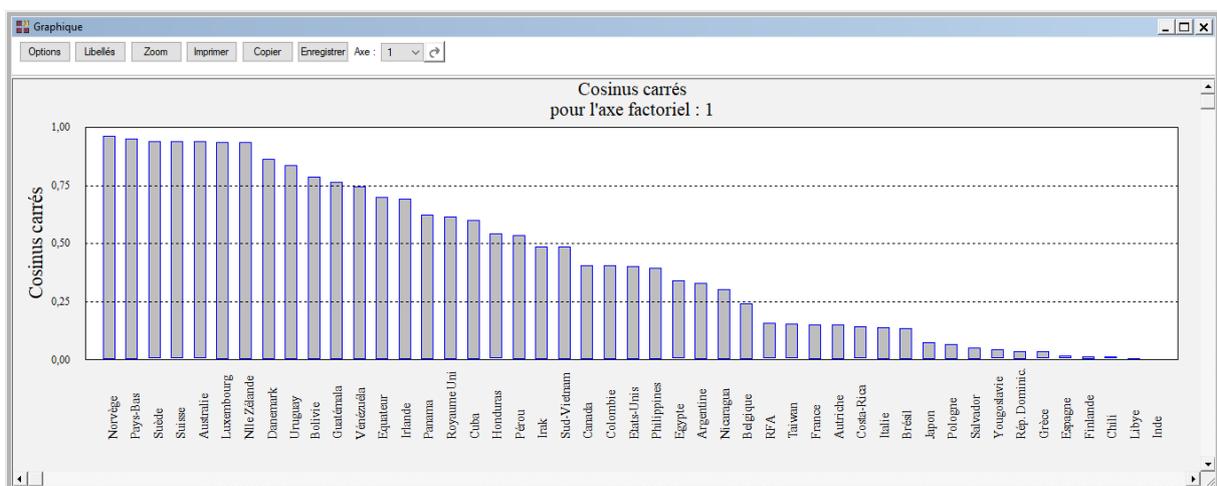
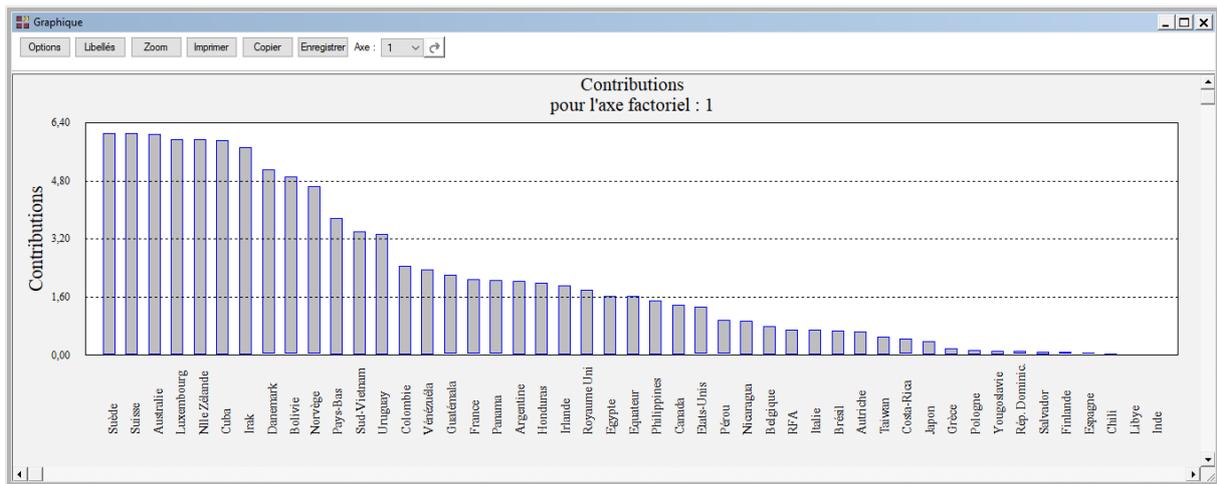
Cette option permet d'afficher des plans factoriels des individus et des barycentres des modalités des variables qualitatives.



- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés et Distances carrées à l'origine

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

Ils mettent en évidence les individus qui contribuent fortement à la définition des axes factoriels, qui sont bien projetés dans les plans factoriels et qui sont éloignés de l'origine.



Une rapide interprétation des résultats

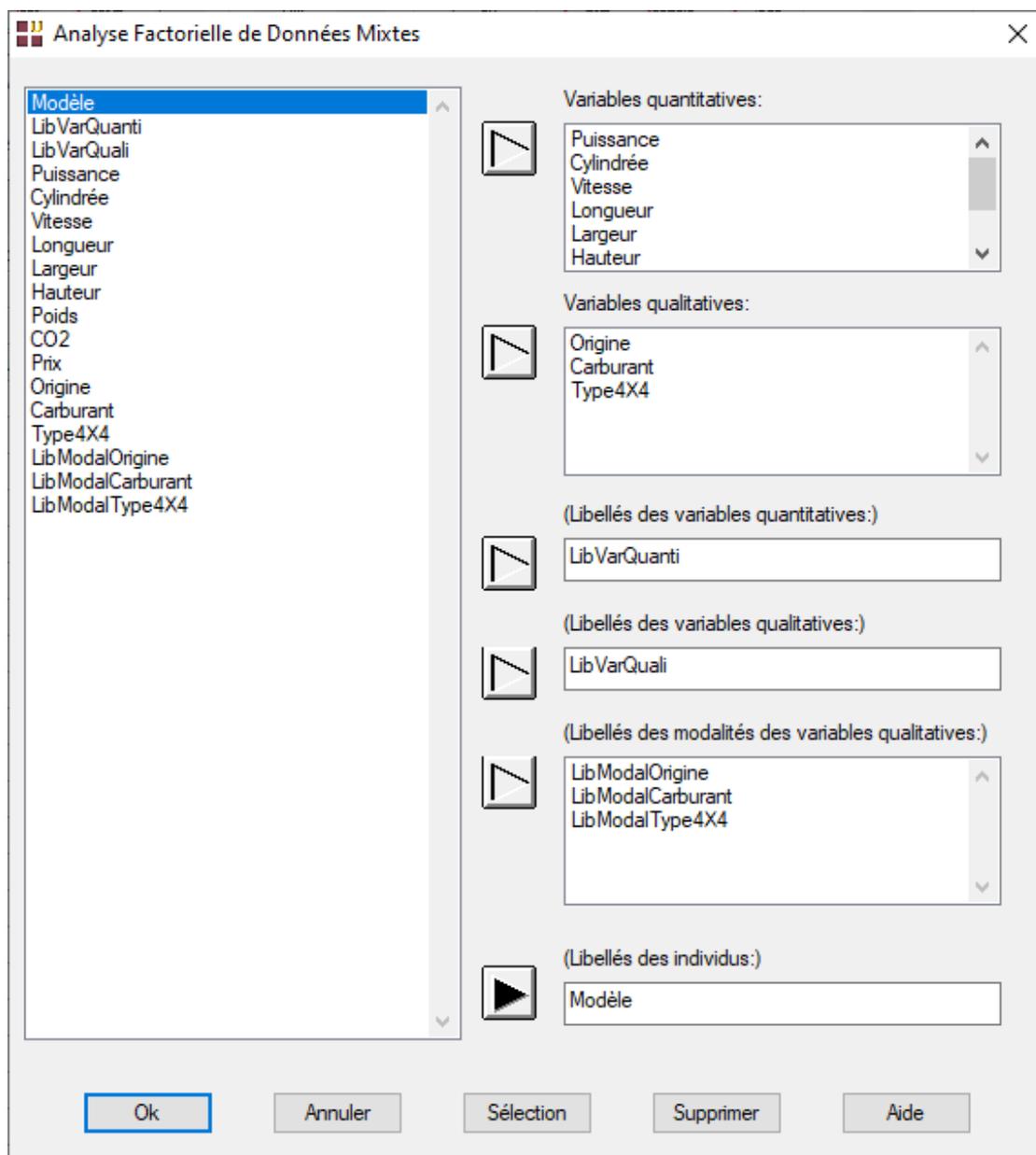
Le plan factoriel des individus (pays) et des modalités de la variable DEMO positionne les divers pays par rapport aux modalités 'démocratie stable', 'démocratie instable' ou 'dictature'. A noter que les données datent de 1961, beaucoup de pays ont évolué depuis cette date.

Le plan factoriel 1-2 oppose les pays pauvres essentiellement agricoles (ayant souvent un régime de dictature) aux pays riches et industriels (à démocratie stable). Parmi les pays à démocratie instable, on note la France qui a compté beaucoup de changements de gouvernements et le Japon dont le PNB en 1961 était inférieur à celui de la Yougoslavie.

Exemple 2 : Fichier AUTOS

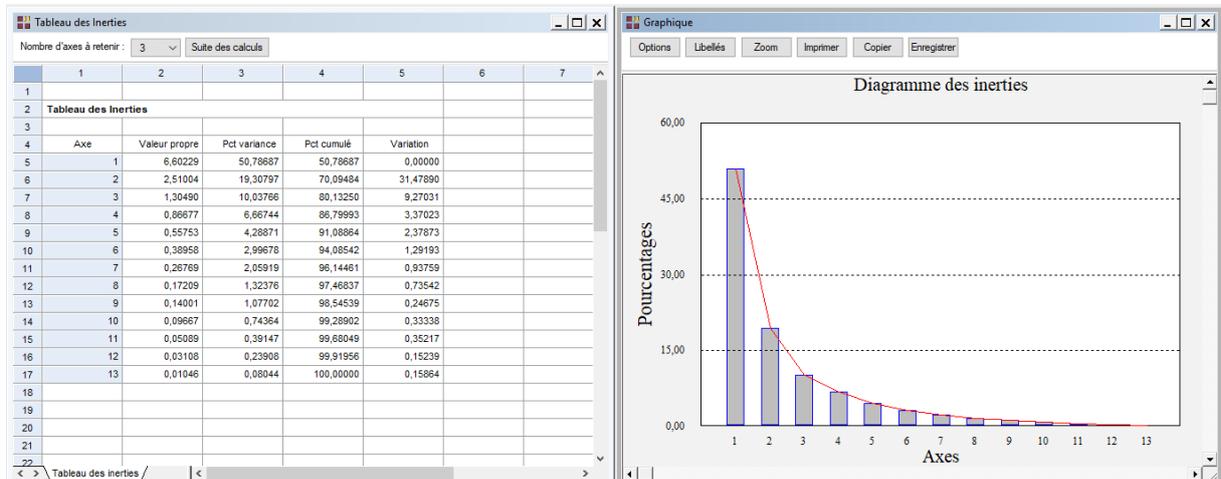
Ce fichier décrit 38 modèles de véhicules via 9 variables quantitatives (puissance, cylindrée, vitesse, longueur, largeur, hauteur, poids, CO2, prix) et 3 variables qualitatives (origine, carburant, type 4x4).

Cliquons sur l'icône AFDM dans le ruban Décrire et renseignons la boîte de dialogue comme montré ci-après :



Cliquons sur Ok pour exécuter le traitement de l'analyse.

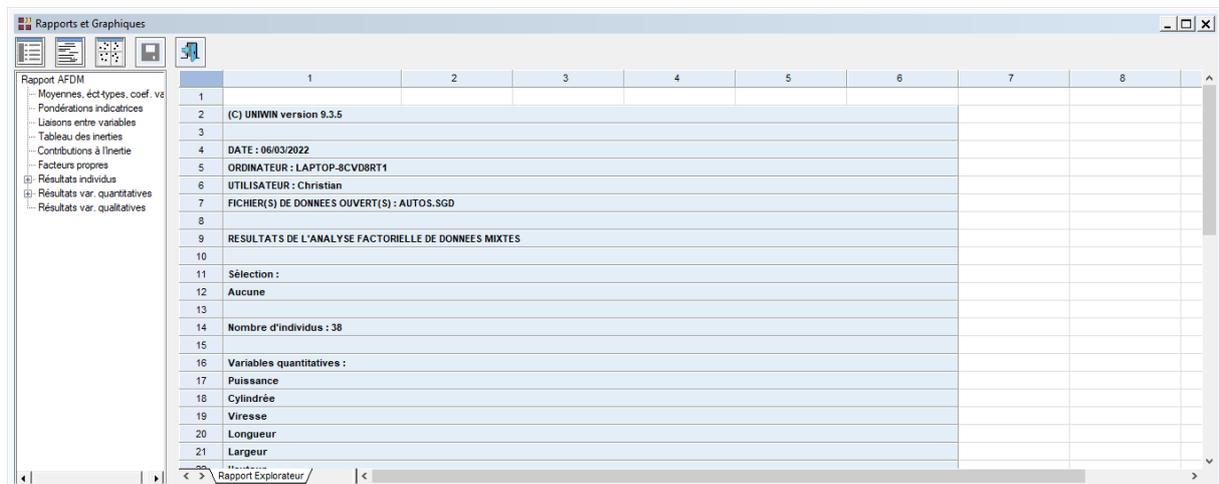
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.



L'option 'Nombre d'axes à retenir' permet de préciser le nombre d'axes à extraire.

Cliquons sur le bouton 'Suite des calculs'.

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



L'option Rapports

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

Rapports et Graphiques

Rapport AFDM

- Moyennes, écarts-types, pct vari
- Ponderations indicatrices
- Tableaux des inerties**
- Contributions à l'inertie
- Facteurs propres
- Résultats individus
- Résultats var. quantitatives
- Résultats var. qualitatives

	1	2	3	4	5	6	7	8
1								
2	LIAISONS ENTRE LES VARIABLES INITIALES							
3								
4	> VARIABLES QUANTITATIVES ET QUALITATIVES : CARRE DU COEFFICIENT DE CORRELATION							
5	> VARIABLES QUANTITATIVES ET QUALITATIVES : CARRE DU RAPPORT DE CORRELATION							
6								
7								
8		Puissance	Cylindrée	Vitesse	Longueur	Largeur	Hauteur	Poids
9	Puissance	1,00000	0,84301	0,84744	0,78638	0,70313	0,03327	0,73866
10	Cylindrée	0,84301	1,00000	0,58215	0,73641	0,72458	0,25571	0,79151
11	Vitesse	0,84744	0,58215	1,00000	0,74621	0,59593	-0,32202	0,47700
12	Longueur	0,78638	0,73641	0,74621	1,00000	0,87398	0,08673	0,81755
13	Largeur	0,70313	0,72458	0,59593	0,87398	1,00000	0,22519	0,82326
14	Hauteur	0,03327	0,25571	-0,32202	0,08673	0,22519	1,00000	0,51930
15	Poids	0,73866	0,79151	0,47700	0,81755	0,82326	0,51930	1,00000
16	CO2	0,90078	0,75346	0,65109	0,72545	0,67731	0,18326	0,74935
17	Prix	0,85200	0,86383	0,64734	0,81552	0,85107	0,23564	0,86140
18	Origine	0,09156	0,10016	0,00828	0,05376	0,05139	0,18409	0,13479
19	Carburant	0,03092	0,00002	0,03965	0,00471	0,00362	0,14335	0,05937
20	Type4x4	0,02160	0,05704	0,02567	0,05464	0,08172	0,51870	0,22877
21								

Rapport Explorateur /

Rapports et Graphiques

	1	2	3	4	5	6	7	8	9	10	11	12
33												
34	MOYENNES, ECARTS-TYPES ET POURCENTAGES DE VARIATION DES VARIABLES QUANTITATIVES											
35												
36												
37		MOYENNES	ECARTS-TYPES	PCT VARIATION								
38	Puissance	154,44737	68,21047	44,16422								
39	Cylindrée	2191,60526	937,49724	42,77674								
40	Vitesse	200,89474	30,35560	15,11020								
41	Longueur	436,39474	44,96873	10,30460								
42	Largeur	177,55283	9,32370	5,25123								
43	Hauteur	151,81579	11,34125	7,47040								
44	Poids	1427,60526	343,14387	24,03633								
45	CO2	193,10526	53,51822	27,71453								
46	Prix	29493,15789	15579,77655	52,82505								
47												
48	PONDERATIONS DES INDICATRICES DES VARIABLES QUALITATIVES											
49												
50		PONDERATIONS										
51												
52	Origine_Autres	0,51299										
53	Origine_Europe	0,62828										

Rapport Général /

Rapports et Graphiques

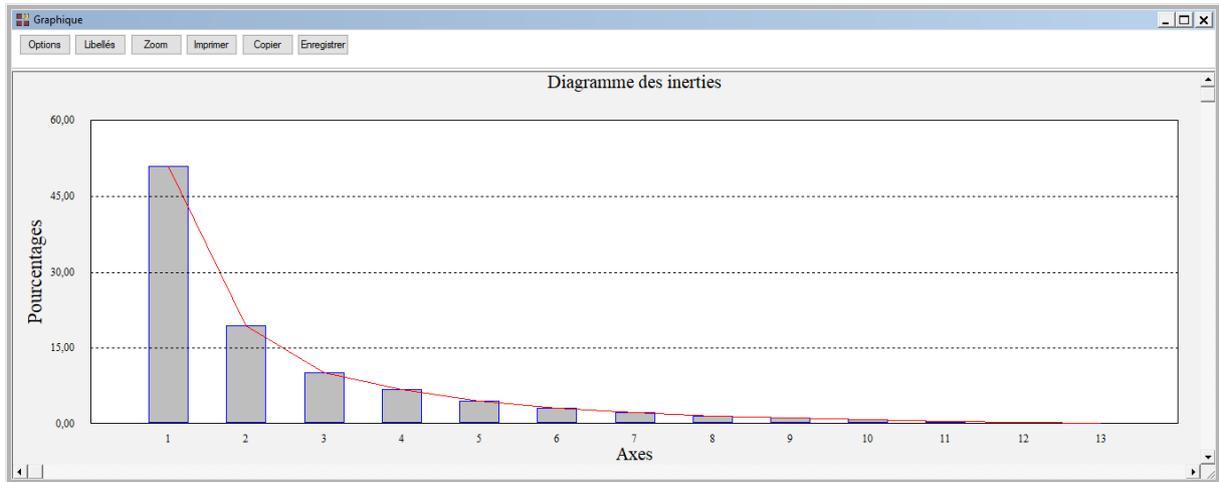
RESULTATS INDIVIDUS POUR LE FACTEUR : 1

DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE
 COMP. PRIN. = COMPOSANTES PRICIPALES - COORDONNEES DES INDIVIDUS
 CONTRIB. = CONTRIBUTIONS A L'INERTIE
 COSINUS*2 = COSINUS CARRÉS
 COS*2 CUM. = SOMMES CUMULEES DES COSINUS CARRÉS

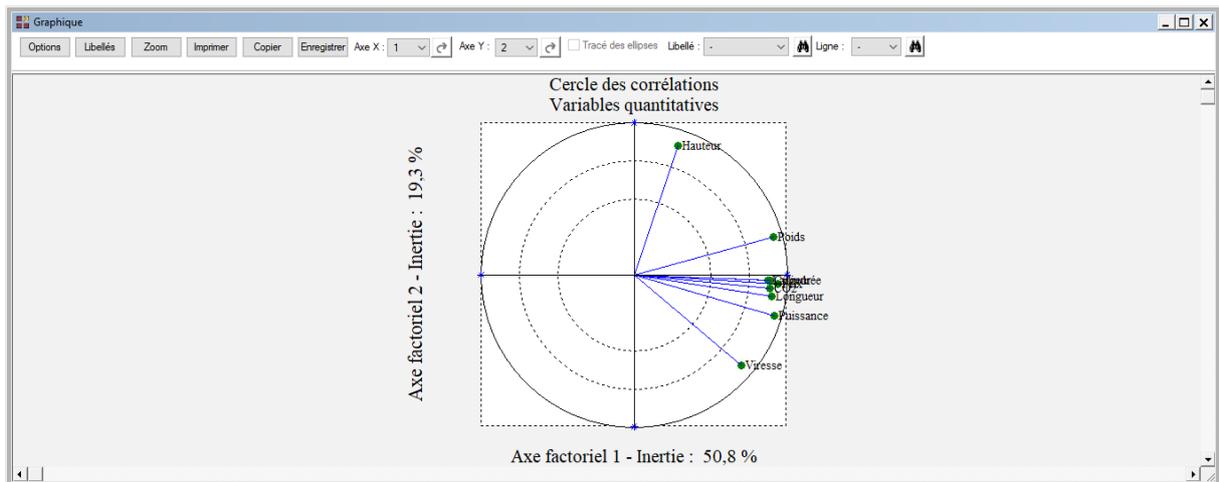
	INDIVIDU	DISTANCE*2	COMP. PRIN.	CONTRIB.	COSINUS*2	COS*2 CUM.
	SANTA_FE	1	16,92816	1,06073	0,44847	0,06647
	MURANO	2	24,59873	3,85505	5,92354	0,60415
	LANDCRUI	3	48,58410	5,30879	11,23345	0,58009
	OUTLAND	4	14,43820	1,75986	1,23447	0,21451
	X-TRAIL	5	13,34388	0,63717	0,16182	0,03042
	CITRONC5	6	6,40025	1,26375	0,63657	0,24953
	P607	7	8,94643	1,86083	1,38017	0,38705
	VELSATIS	8	6,77756	1,00598	0,40336	0,14931
	LAGUNA	9	4,27757	-0,17622	0,01238	0,00726
	P407	10	4,37143	-0,09857	0,00387	0,00222
	CITRONC4	11	4,85353	-0,95650	0,36466	0,18850

L'option Graphiques

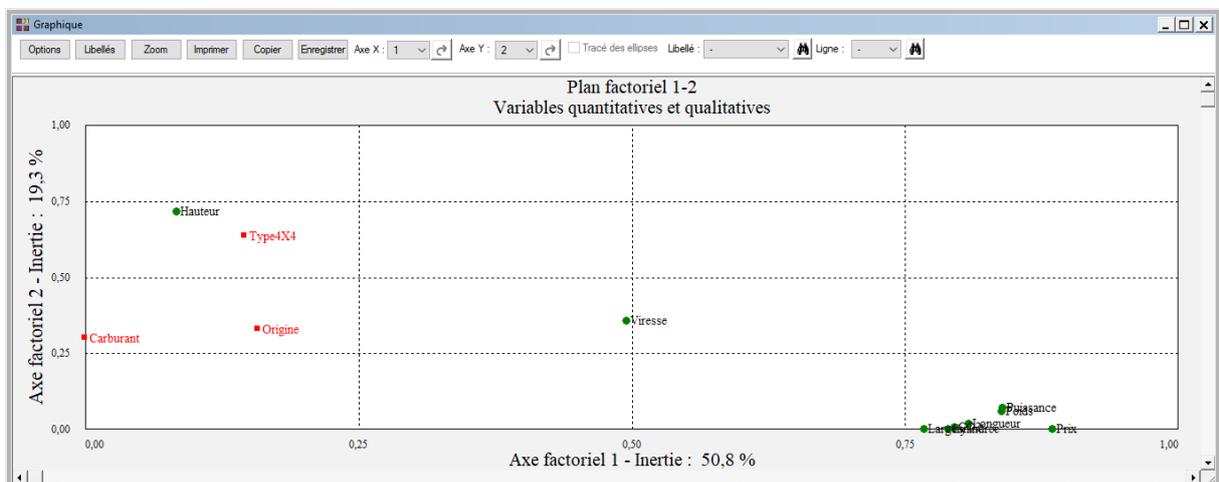
- Diagramme des inerties



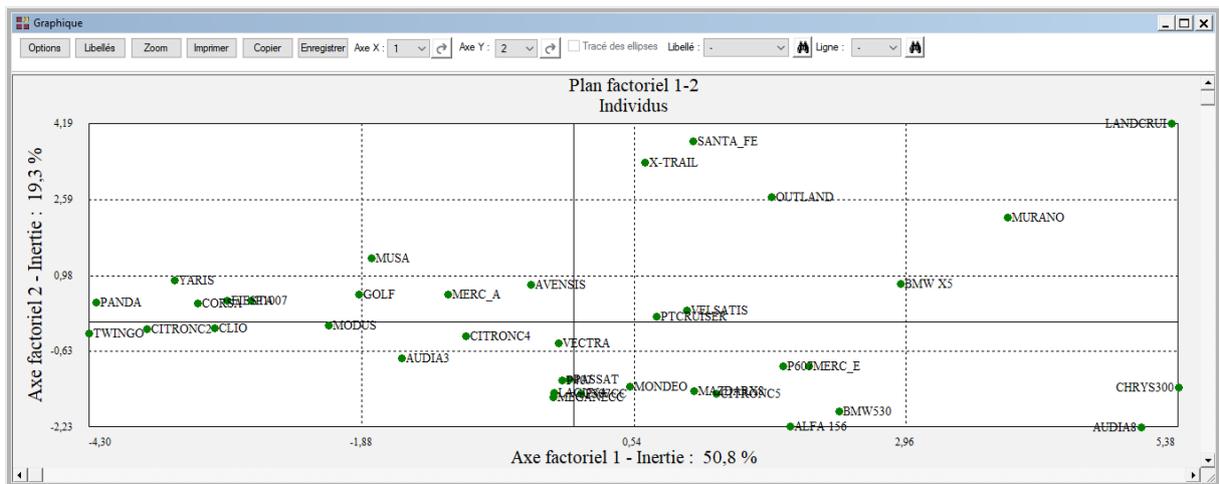
- Les options Cercles factoriels



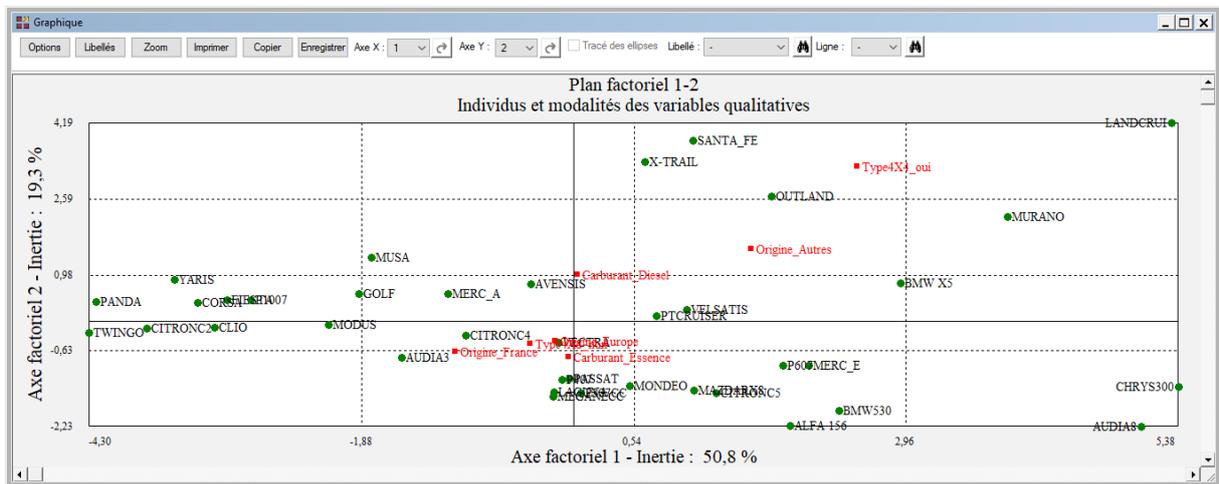
- L'option Représentation simultanée des variables



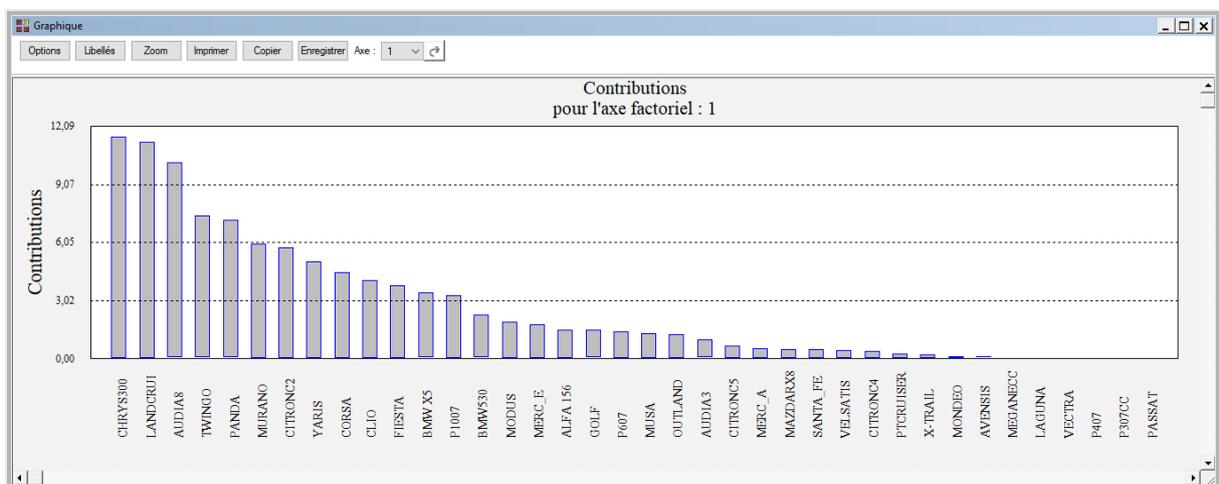
- L'option Plan factoriel des individus

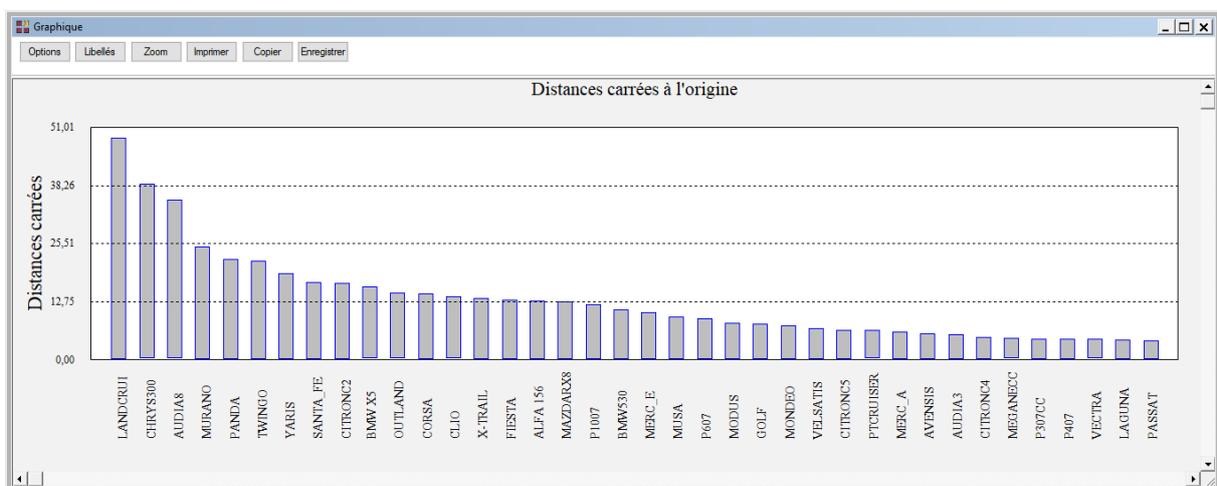
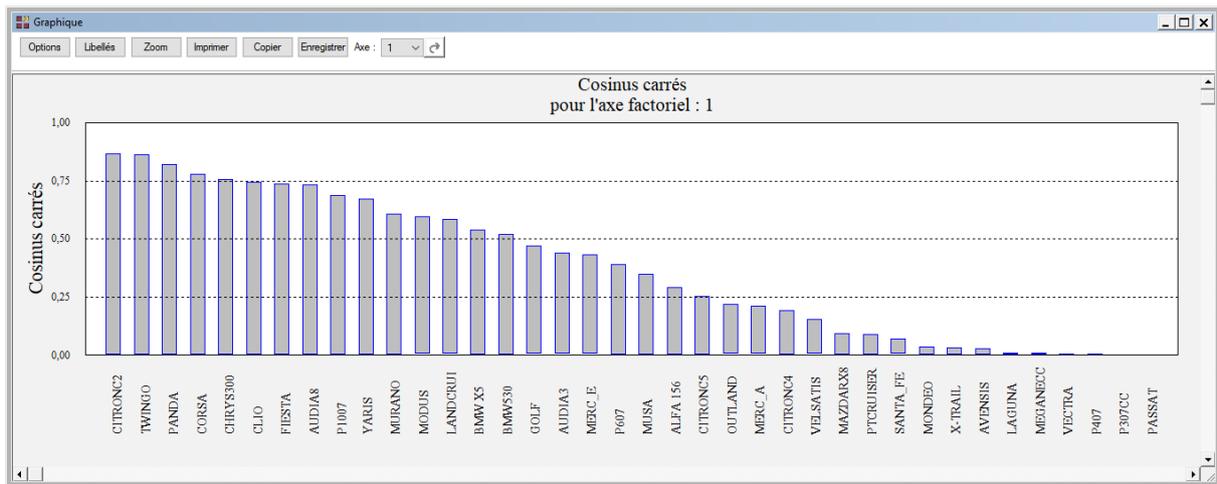


- L'option Plan factoriel des individus et des modalités



- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés et Distances carrées à l'origine





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
varquanti	Libellés des variables quantitatives
varquali	Libellés des variables qualitatives
indiv	Libellés des individus
modal	Libellés des modalités des variables qualitatives
moyennes	Moyennes des variables quantitatives
ecarts	Ecarts-types des variables quantitatives
cv	Coefficients de variation des variables quantitatives
pond	Pondérations des indicatrices des variables qualitatives
liaison	Liaisons entre les variables quantitatives et qualitatives
donmix	Tableau des données mixtes
numaxe	Numéros des axes

valpro	Valeurs propres
pctvar	Pourcentages de variance expliquée
pctvcum	Pourcentages cumulés de variance expliquée
variation	Variations de la variance expliquée
vectpro	Vecteurs propres
coiner	Tableau des contributions des variables à l'inertie
vdistances	Distances carrées à l'origine (variables quantitatives)
vcos	Cosinus carrés (variables)
vcontrib	Contributions à l'inertie (variables)
distances	Distances carrées à l'origine (individus)
cosinus	Cosinus carrés (individus)
contrib	Contributions à l'inertie (individus)
facteur	Facteurs colonnes
pcomp	Composantes principales
mvq	Moyennes des variables qualitatives, par modalité
vcoor	Coordonnées des variables