

UNIWIN VERSION 9.7.0

ANALYSE EN COMPOSANTES PRINCIPALES

Révision : 02/09/2023

Définition.....	1
Entrée des données	2
Données manquantes	4
Entrée des données des individus supplémentaires.....	4
Exemple 1 : Fichier OCDE (variables et individus de base)	5
L'option Rapports	9
L'option Graphiques	10
Tracé des ellipses de tolérance.....	17
Exemple 2 : Fichier OCDE (variables et individus de base et supplémentaires)	19
L'option Rapports	25
L'option Graphiques	26
Une rapide interprétation des résultats.....	35
Exemple 3 : Fichier LEVURE.....	35
Les variables internes créées par la procédure	42

Définition

La méthode d'Analyse en Composantes Principales (ACP) permet d'étudier un tableau individus x variables dans le cas où toutes les variables sont quantitatives.

La méthode permet d'obtenir une carte des individus en fonction de leurs proximités et une carte des variables en fonction de leurs corrélations. Il est également possible d'obtenir des représentations simultanées (Biplots).

La possibilité d'analyser des individus et de variables supplémentaires (quantitatives ou qualitatives) est offerte.

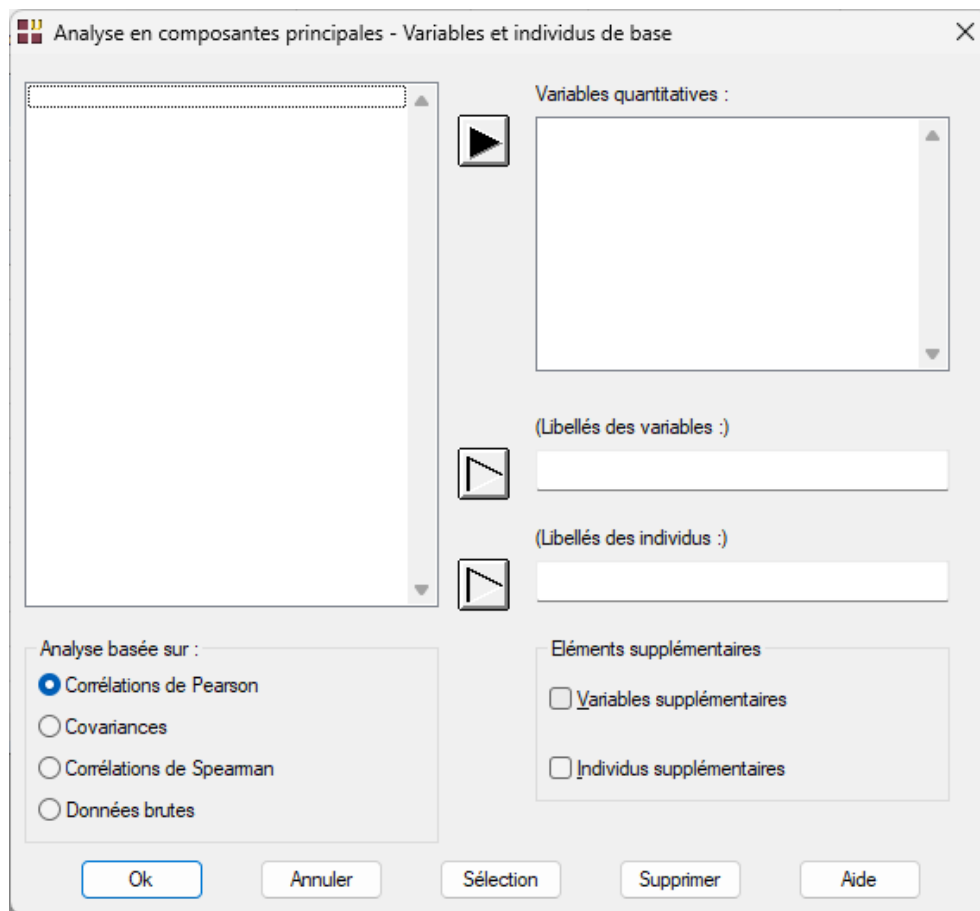
L'analyse peut être réalisée en utilisant la matrice des corrélations de Pearson ou de Spearman, la matrice des covariances ou les données brutes.

Après 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 plans factoriels des individus, des cercles des corrélations, du Biplot, de ceux relatifs aux individus et variables supplémentaires ainsi que des graphiques des contributions, des cosinus carrés et des distances carrées à l'origine.

Entrée des données

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



Cette boîte de dialogue permet de définir les variables quantitatives de base (actives) et les individus de base (actifs) de l'analyse.

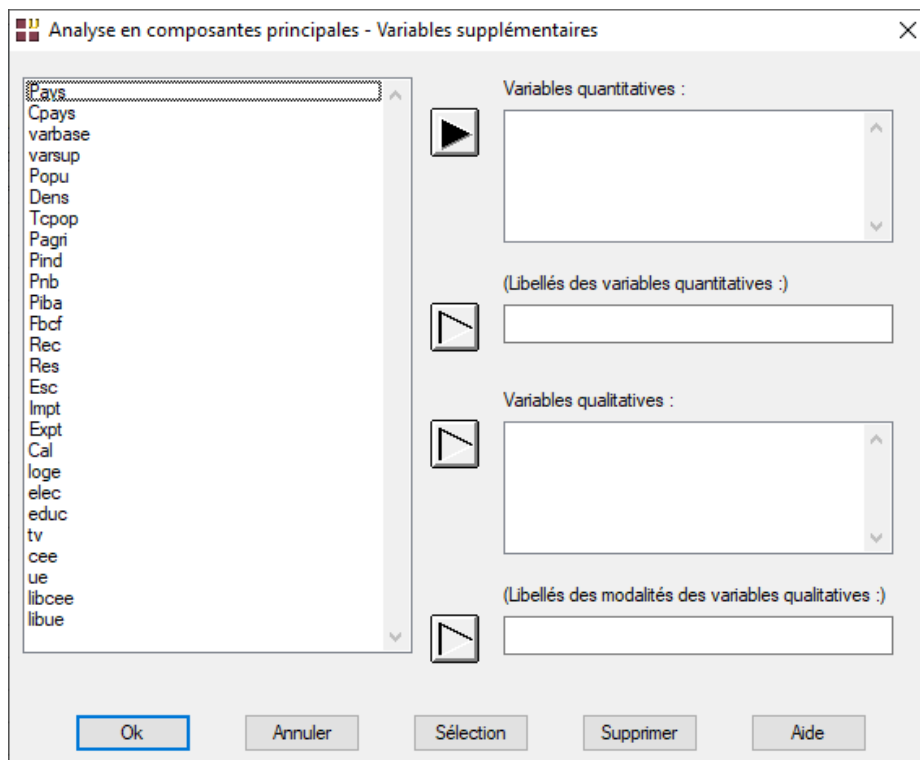
Elle permet également, en option, d'indiquer les noms des variables contenant les libellés des variables quantitatives et les libellés des individus.

Une option de transformation des variables (centrage et réduction, centrage, pas de transformation) des données est proposée.

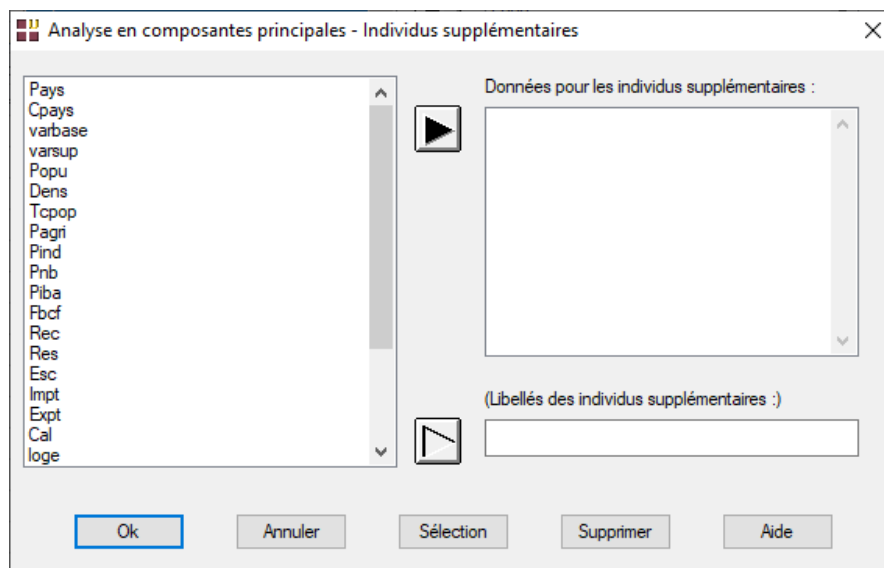
Enfin, vous pouvez cocher l'une ou les deux cases indiquant si vous avez des variables supplémentaires (passives) et/ou des individus supplémentaires (passifs) à étudier.

Dans ces cas des boîtes de dialogue complémentaires s'affichent. Elles sont montrées ci-après sur le fichier exemple OCDE.

La boîte de dialogue 'Variables supplémentaires' permet de définir des variables quantitatives supplémentaires (passives), des variables qualitatives supplémentaires (passives), la variable optionnelle contenant les libellés des variables quantitatives supplémentaires ainsi que la variable optionnelle contenant les libellés associés aux différentes modalités des variables qualitatives supplémentaires.



La boîte de dialogue 'Individus supplémentaires' permet de sélectionner les variables contenant les données pour les individus supplémentaires (passifs) et de définir la variable optionnelle contenant les libellés des individus supplémentaires (passifs).



Dans toutes ces boîtes de dialogue, les zones de libellés sont optionnelles : UNIWIN génère automatiquement des libellés en cas de besoin.

Données manquantes

Les données manquantes sont permises pour les variables de base et supplémentaires. Elles sont remplacées par les valeurs moyennes des variables respectives. Les données manquantes ne sont pas admises pour les données supplémentaires.

Entrée des données des individus supplémentaires

L'entrée des données pour les individus supplémentaires peut se faire de trois façons différentes :

- Les données supplémentaires sont dans d'autres variables du même fichier de données. Dans la boîte de dialogue 'Individus supplémentaires', il faut sélectionner les variables définissant les données supplémentaires en respectant l'ordre défini lors de la sélection des variables de base : la première variable sélectionnée définit les données supplémentaires pour la première variable de base, la deuxième variable sélectionnée définit les données supplémentaires pour la deuxième variable de base...
- Les données supplémentaires sont dans les mêmes variables que les variables contenant les données de base. Il faut alors utiliser le bouton 'Sélection' dans les boîtes de dialogue. Dans la boîte de dialogue 'Variables de base', sélectionner les variables puis cliquer sur le bouton 'Sélection' et définir la sélection logique désirée. Dans la boîte de dialogue 'Individus supplémentaires', sélectionner ces mêmes variables dans le même ordre puis cliquer sur le bouton 'Sélection' et définir la sélection logique.
- Les données supplémentaires sont dans un deuxième fichier. Il faut alors avant de débiter l'ACP, ouvrir les deux fichiers simultanément (le fichier des données de base et le fichier des données supplémentaires). Dans la boîte de dialogue 'Individus supplémentaires', il faut sélectionner les variables définissant les données supplémentaires dans ce deuxième fichier en respectant l'ordre défini lors de la sélection des variables de base : la première variable sélectionnée dans le deuxième fichier définit les données supplémentaires pour la première variable sélectionnée dans le premier fichier, la deuxième variable sélectionnée dans le deuxième fichier définit les données supplémentaires pour la deuxième variable sélectionnée dans le premier fichier...

Exemple 1 : Fichier OCDE (variables et individus de base)

Nous utiliserons le fichier OCDE pour illustrer ce premier exemple. Ce fichier contient les informations suivantes concernant 18 pays membres ou non de la CEE et de l'Union Européenne :

<i>Pays</i>	<i>Code Pays</i>	<i>Membre CEE</i>	<i>Membre UE</i>
Allemagne	D	Oui	Oui
Autriche	A	Non	Oui
Belgique	B	Oui	Oui
Canada	CDN	Non	Non
Danemark	DK	Oui	Oui
Espagne	E	Oui	Oui
USA	USA	Non	Non
Finlande	SF	Non	Oui
France	F	Oui	Oui
Grèce	GR	Oui	Oui
Irlande	SE	Oui	Oui
Italie	I	Oui	Oui
Japon	JAP	Non	Non
Norvège	N	Non	Non
Pays-Bas	NL	Oui	Oui
Portugal	P	Oui	Oui
Royaume-Uni	RUN	Oui	Non
Suède	S	Non	Oui

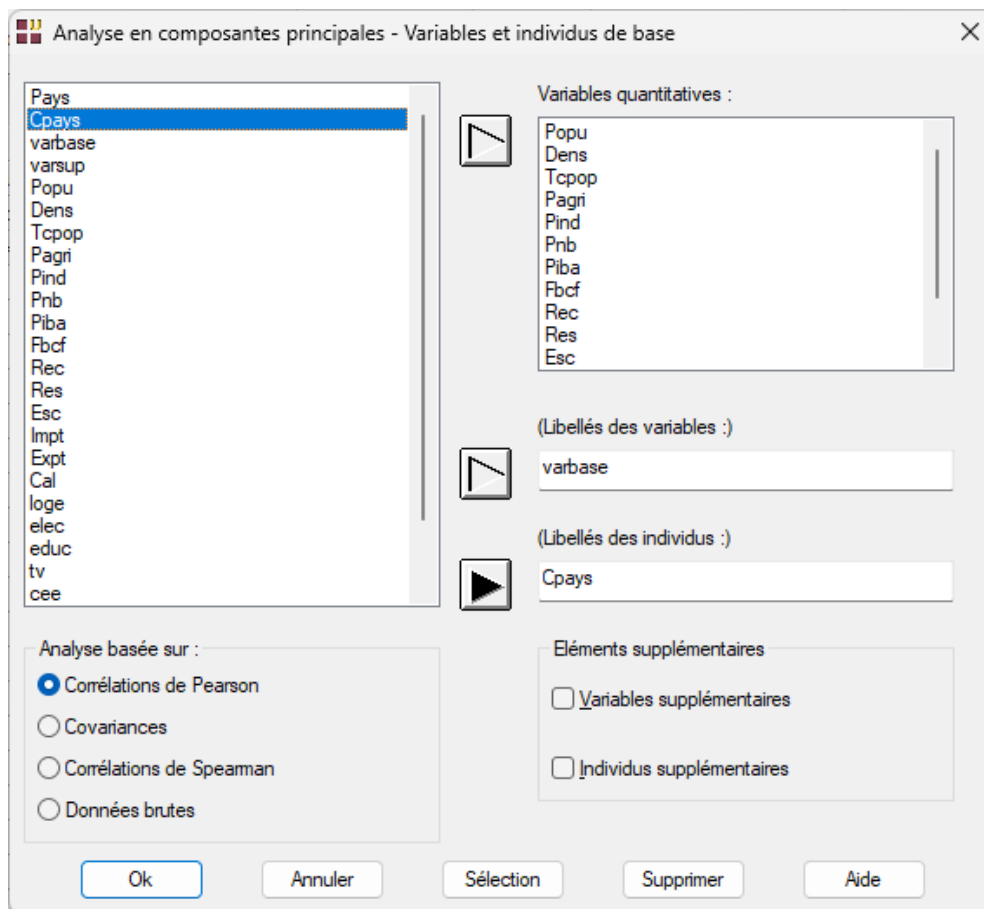
Les informations suivantes ont été collectées :

<i>Information</i>	<i>Variable</i>
Nom du pays (libellés longs)	Pays
Code du pays (libellés courts)	Cpays
Population (en milliers)	Popu
Densité de population (au km ²)	Dens
Taux de croissance de la population	Tcpop
Pct de la population dans l'agriculture	Pagri
Pct de la population dans l'industrie	Pind
Produit national brut en dollars par habitant	Pnb
Pct produit intérieur brut en agriculture	Piba
Formation brute de capital fixe (en pct du pnb)	Fbcf
Recettes (en pct du pnb)	Rec
Réserves (en millions de dollars)	Res
Taux d'escompte	Esc
Importations (en millions de dollars)	Impt
Exportations (en millions de dollars)	Expt
Calories par jour et par habitant	Cal
Logements (nb achevés par 1000 habitants)	loge

Electricité (consommation par habitant, par an, en kWh)	elec
Dépenses en éducation (en pct du pnb)	educ
Nombre de téléviseurs (par 1000 habitants)	tv
Membre de la CEE	cee
Membre de l'UE	ue
Libellés longs des variables de base Popu à Cal	varbase
Libellés longs des variables supplémentaires loge à tv	varsup
Libellés des modalités CEE	libcee
Libellés des modalités UE	libue

Cliquons sur l'icône ACP dans le ruban Décrire.

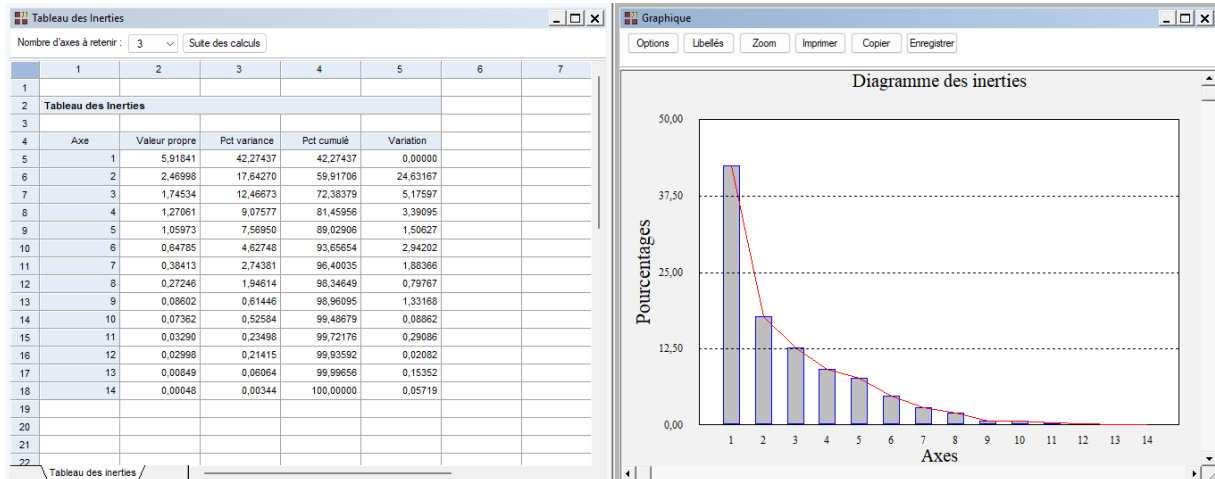
La première boîte de dialogue montrée ci-après apparaît.



Nous choisissons les variables de Popu à Cal comme variables quantitatives de base (actives), la variable varbase comme variable contenant les libellés longs des variables quantitatives de base, la variable Cpays pour les libellés courts des individus (les pays), une analyse avec centrage et réduction (corrélations de Pearson) et ne cochons pas les cases Variables supplémentaires et Individus supplémentaires.

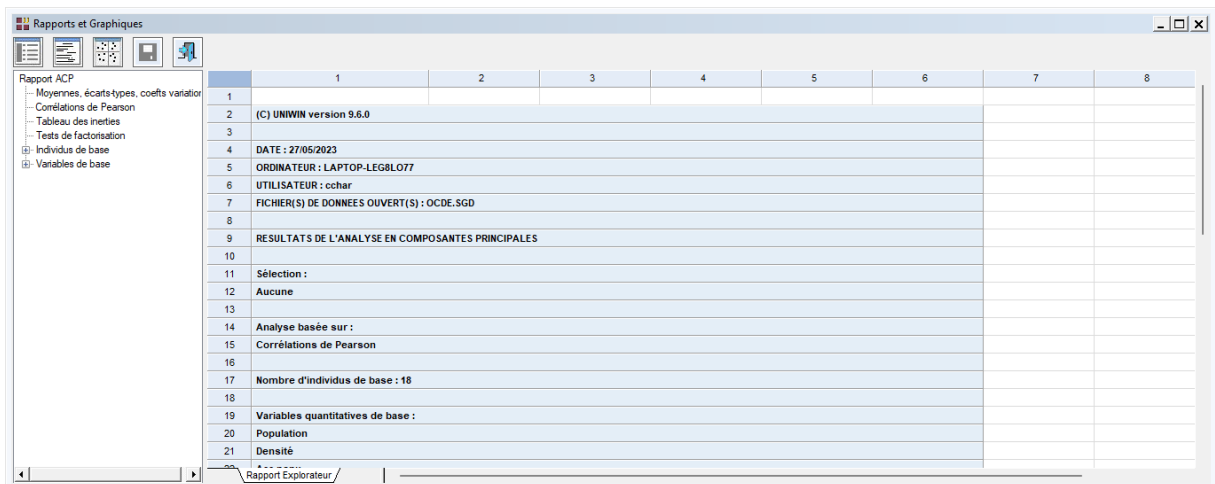
Cliquons sur Ok pour exécuter le traitement de l'analyse.


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.




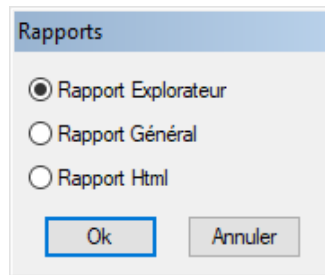
L'option 'Nombre d'axes à retenir' permet de préciser le nombre de composantes principales à extraire.


Cliquons sur le bouton 'Suite des calculs'. Après quelques instants, l'écran suivant s'affiche :

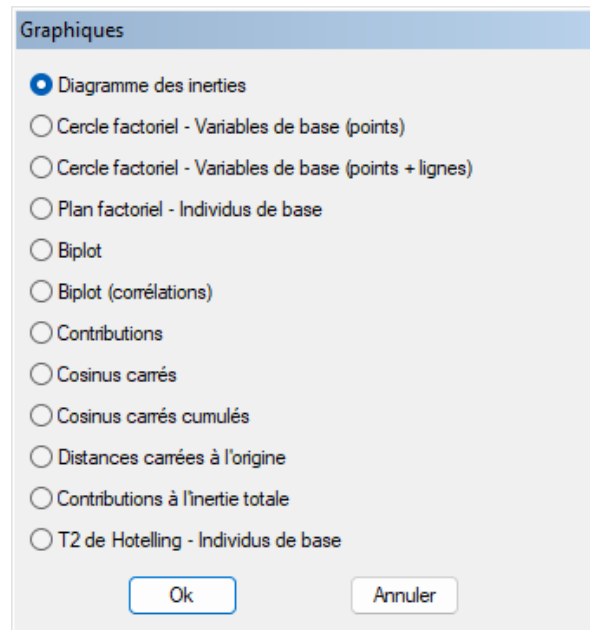



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

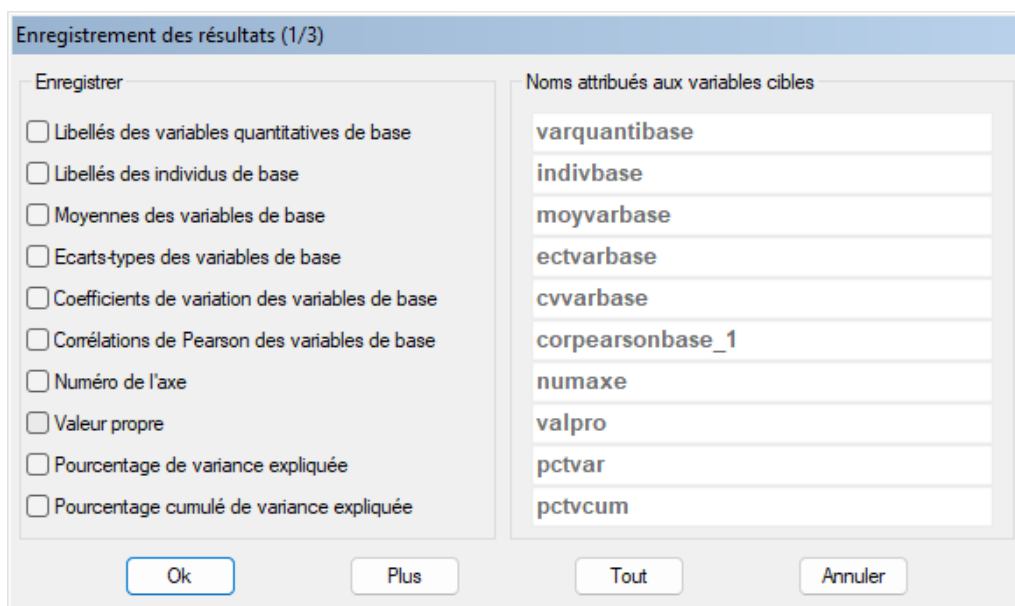
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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'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 ACP : Explorateur, Général, HTML.

Rapports et Graphiques

Rapport ACP

- Moyennes, écarts-types, coeffs variation
- Corrélation de Pearson
- Tableau des inerties
- Tests de factorisation
- Individus de base
- Variables de base
 - Axe Factoriel 1
 - Axe Factoriel 2
 - Axe Factoriel 3

	1	2	3	4	5	6	7	8
1	RESULTATS VARIABLES DE BASE POUR LE FACTEUR : 1							
2								
3								
4	DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE							
5	COORDONNEE = COORDONNÉES DE VARIABLES							
6	CORRELATION = CORRELATIONS SIMPLES VARIABLES-FACTEURS							
7	FACTEUR = FACTEURS PRINCIPAUX							
8	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE							
9	COSINUS*2 = COSINUS CARRÉS							
10	COS*2 CUMULE = COSINUS CARRÉS CUMULÉS POUR 1 AXE							
11								
12								
13		Variable	Distance*2	Coordonnée	Corrélation	Facteur	Contribution	Cosinus*
14	Population	1	1	0,70564	0,70564	0,29006	8,41321	0,4976
15	Densité	2	1	0,26076	0,26076	0,10719	1,14887	0,0676
16	Acc popu	3	1	0,53864	0,53864	0,22141	4,90217	0,2901
17	Pct popu agri	4	1	-0,83393	-0,83393	-0,34279	11,75057	0,6954
18	Pct popu ind	5	1	0,53576	0,53576	0,22023	4,84993	0,2870
19	Pnb	6	1	0,80363	0,80363	0,33033	10,91197	0,6458
20	Pct pnb agri	7	1	-0,86819	-0,86819	-0,35687	12,73572	0,7537
21	Fbctf	8	1	-0,25678	-0,25678	-0,10555	1,11405	0,0659

Rapport Explorateur

Rapports et Graphiques

	1	2	3	4	5	6	7	8	9	10	11	12
131	RESULTATS INDIVIDUS POUR LE FACTEUR : 1											
132												
133	DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE											
134	CONTRIBUTION TOTALE = CONTRIBUTIONS A L'INERTIE TOTALE											
135	COMPOSANTE = COMPOSANTES PRINCIPALES - COORDONNÉES DES INDIVIDUS											
136	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE											
137	COSINUS*2 = COSINUS CARRÉS											
138	COS*2 CUMULE = COSINUS CARRÉS CUMULÉS POUR 1 AXE											
139	T2 = T2 DE HOTELLING POUR 1 AXE											
140												
141												
142		Individu	Distance*2	Contribution totale	Composante	Contribution	Cosinus*2	Cos*2 cumulé	T2			
143	D	1	17,51839	8,45270	3,54682	12,50332	0,71810	0,71810	2,12556			
144	A	2	4,67459	1,25076	-1,14584	1,30495	0,28087	0,28087	0,22184			
145	B	3	7,64381	3,25365	0,80195	0,63921	0,08414	0,08414	0,10866			
146	CDN	4	12,66400	3,14936	1,53887	2,35369	0,18700	0,18700	0,40013			
147	DK	5	8,53781	2,69209	-0,33676	0,11272	0,01328	0,01328	0,01916			
148	E	6	9,49664	3,96332	-2,21666	4,88364	0,51740	0,51740	0,83022			
149	USA	7	47,32904	26,85170	5,72414	32,56611	0,69230	0,69230	5,53624			
150	SF	8	5,42973	2,63319	-2,05769	4,20829	0,77980	0,77980	0,71541			
151	F	9	4,25526	1,38218	1,48525	2,19254	0,51841	0,51841	0,37273			

Rapport Général

Rapports et Graphiques

RESULTATS INDIVIDUS POUR LE FACTEUR : 1

DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE
 CONTRIBUTION TOTALE = CONTRIBUTIONS A L'INERTIE TOTALE
 COMPOSANTE = COMPOSANTES PRINCIPALES - COORDONNÉES DES INDIVIDUS
 CONTRIBUTION = CONTRIBUTIONS A L'INERTIE
 COSINUS*2 = COSINUS CARRÉS
 COS*2 CUMULE = COSINUS CARRÉS CUMULÉS POUR 1 AXE
 T2 = T2 DE HOTELLING POUR 1 AXE

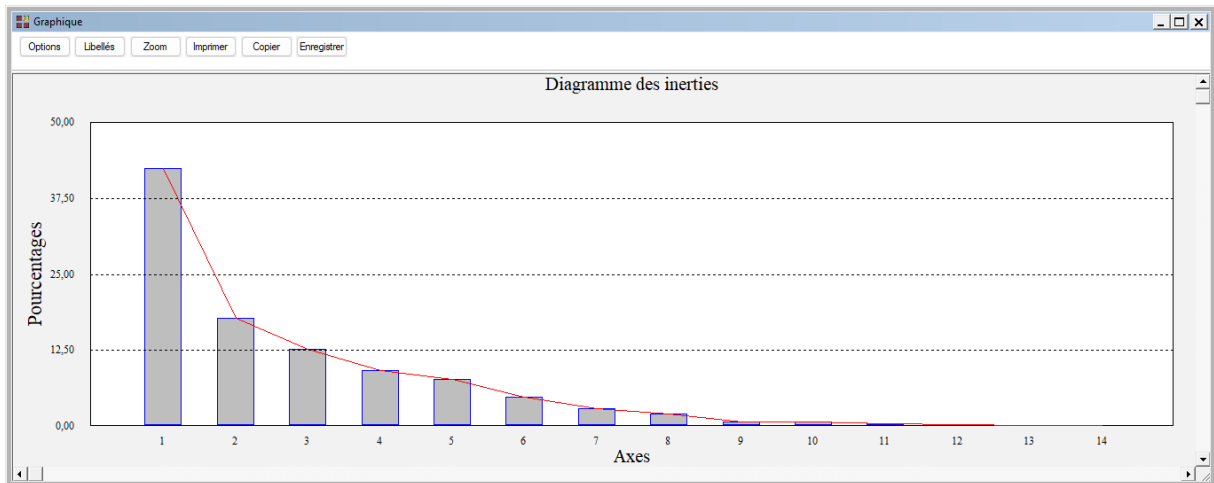
	Individu	Distance*2	Contribution totale	Composante	Contribution	Cosinus*2	Cos*2 cumulé	T2
D	1	17,51839	8,45270	3,54682	12,50332	0,71810	0,71810	2,12556
A	2	4,67459	1,25076	-1,14584	1,30495	0,28087	0,28087	0,22184
B	3	7,64381	3,25365	0,80195	0,63921	0,08414	0,08414	0,10866
CDN	4	12,66400	3,14936	1,53887	2,35369	0,18700	0,18700	0,40013
DK	5	8,53781	2,69209	-0,33676	0,11272	0,01328	0,01328	0,01916
E	6	9,49664	3,96332	-2,21666	4,88364	0,51740	0,51740	0,83022
USA	7	47,32904	26,85170	5,72414	32,56611	0,69230	0,69230	5,53624
SF	8	5,42973	2,63319	-2,05769	4,20829	0,77980	0,77980	0,71541
F	9	4,25526	1,38218	1,48525	2,19254	0,51841	0,51841	0,37273

L'option Graphiques

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

- Diagramme des inerties

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



- Les options Cercle factoriel

Ces options permettent d'afficher le cercle de corrélations des variables de base 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 de base avec lignes.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.

Choix du plan factoriel ACP

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 %

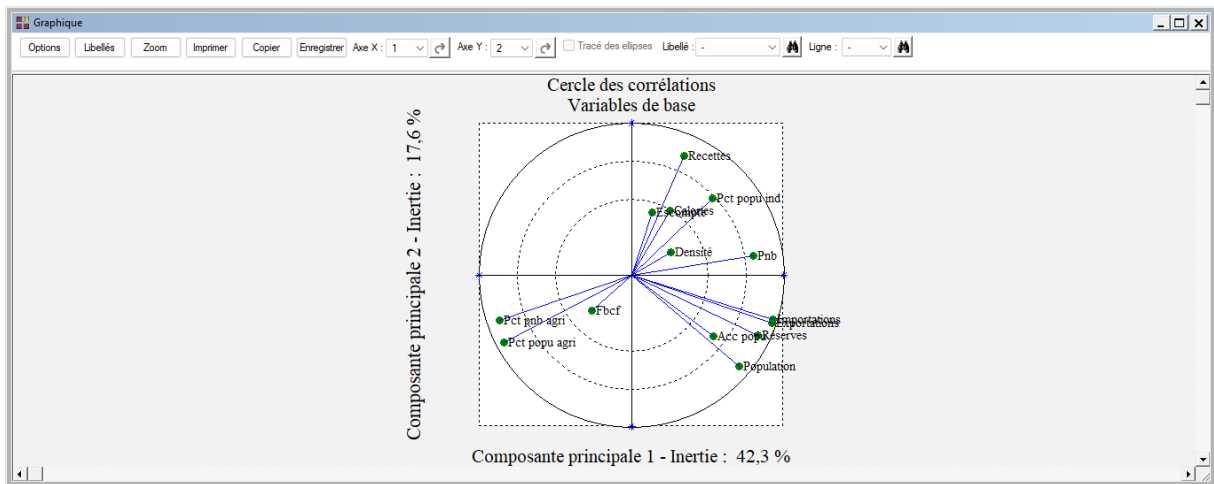
Réutiliser les titres : Non, Oui

Individus de base : Oui, Non
Individus supp. : Oui, Non
Var. quanti. de base : Oui, Non
Var. quanti. supp. : Oui, Non
Var. quali. supp. : Oui, Non

Times New Ron, Normal, 12, Police, Couleur

Défaut, Ok, Sélection, Annuler

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 Plan factoriel

Cette option permet d'afficher le plan factoriel des individus de base.

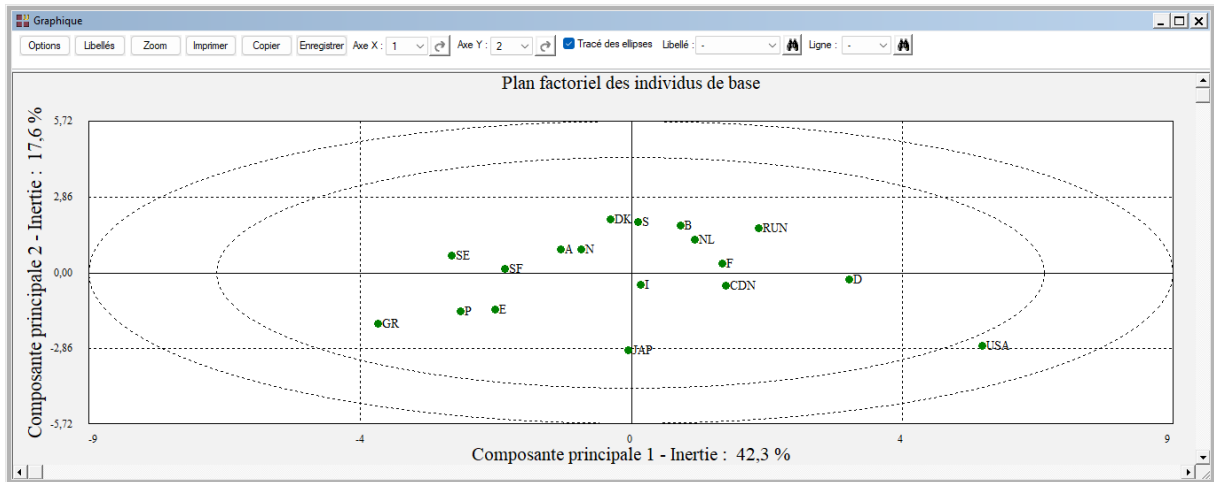
Une boîte de dialogue permettant de choisir le plan factoriel s'affiche alors.

Times New Ron	Normal	12	■	Police	Couleur
Times New Ron	Normal	12	■	Police	Couleur
Times New Ron	Normal	12	■	Police	Couleur
Times New Ron	Normal	12	■	Police	Couleur
Times New Ron	Normal	12	■	Police	Couleur

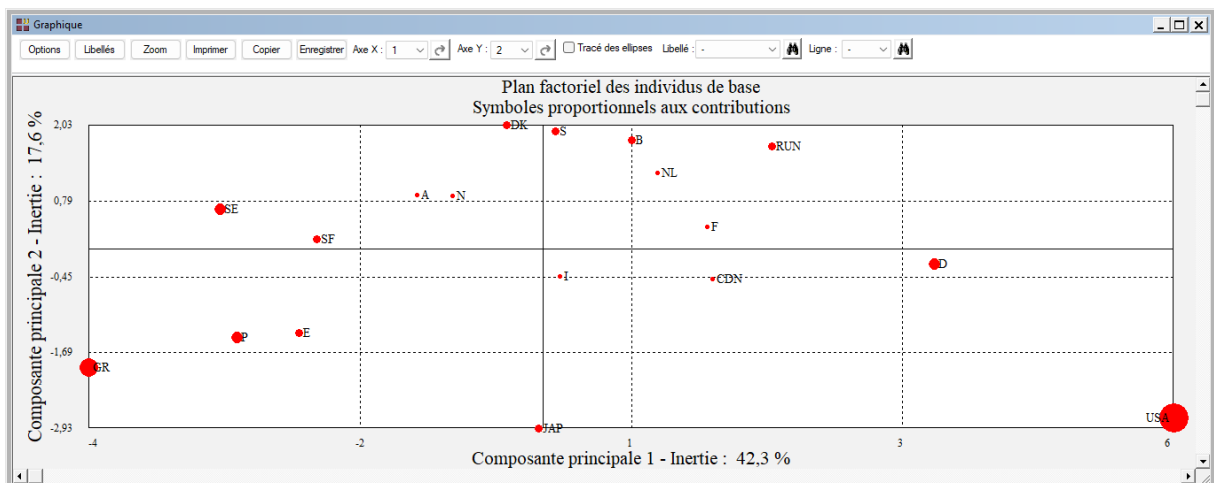
Elle permet également de préciser si l'on désire afficher les libellés des individus, de choisir la couleur et la police, d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail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

Des ellipses de confiance à 95% et 99% ont été ajoutées dans le graphique.



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'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 OCDE.

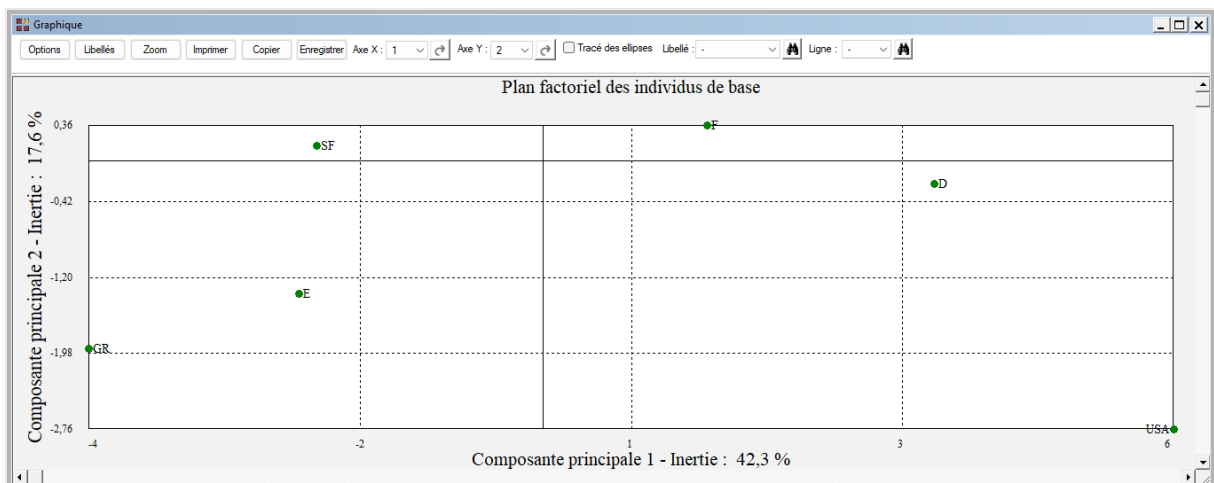
Comme exemple, mettons en œuvre une sélection basée sur la variable 'cosindivbase_1' qui contient les cosinus carrés sur le premier axe factoriel.

Définition de la sélection

Et cosindivbase_1 > 0.5

Liaison	Variable	Relation	Valeur ou variable
Et	correlsimplevarbase_2	=	Cal
Et non	correlsimplevarbase_3	<>	cee
Ou	cosindivbase_1	<	composante_1
Ou non	cosindivbase_2	<=	composante_2
	cosindivbase_3	>	composante_3
	cosvarbase_1	>=	contribindivbase_1
	cosvarbase_2	début	contribindivbase_2

Ok Annuler Ajouter Aide



- Les options Biplot et Biplot (corrélations)

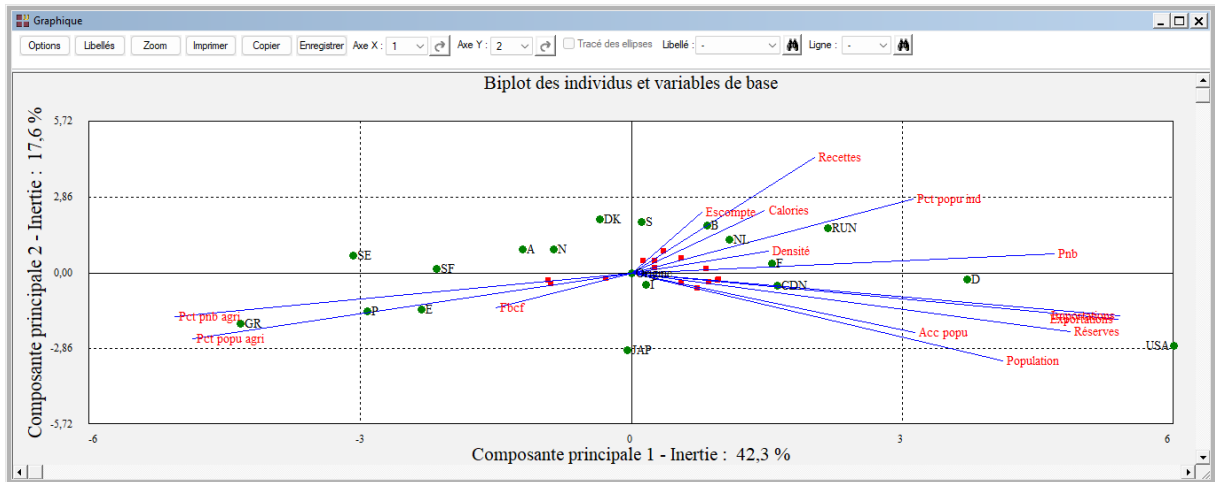
L'option Biplot permet d'obtenir un graphique donnant une représentation simultanée des individus et des directions des variables d'origine dans le plan factoriel choisi.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Un exemple du Biplot est montré ci-après.

A noter que les droites indiquant les directions des variables d'origine dans le plan factoriel ont des longueurs proportionnelles aux corrélations visualisées dans le cercle factoriel.

L'option d'interrogation des points individus et des points variables est possible. Par contre le zoom ne l'est pas.

La boîte de dialogue 'Echelles des axes' permet de définir un graphique orthonormé.



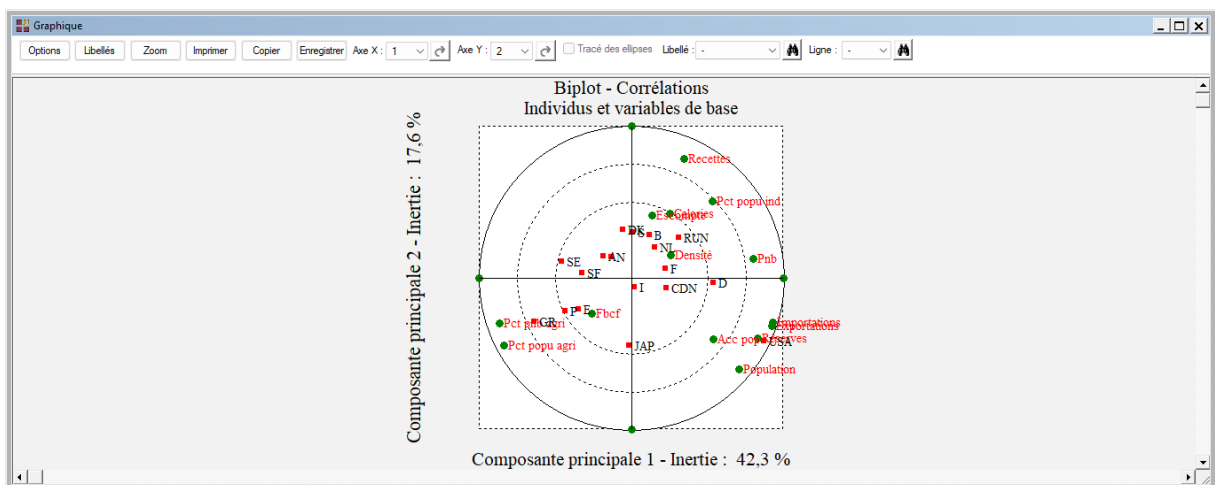
L'option Biplot (corrélations) permet d'obtenir un graphique donnant une représentation simultanée des variables indicatrices des individus et des variables d'origine dans un cercle des corrélations.

Les variables indicatrices des individus sont représentées comme si elles étaient des variables supplémentaires dans l'analyse.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Un exemple du Biplot (corrélations) est montré ci-après.

L'option d'interrogation des points individus et des points variables est possible.

L'option zoom est également disponible.

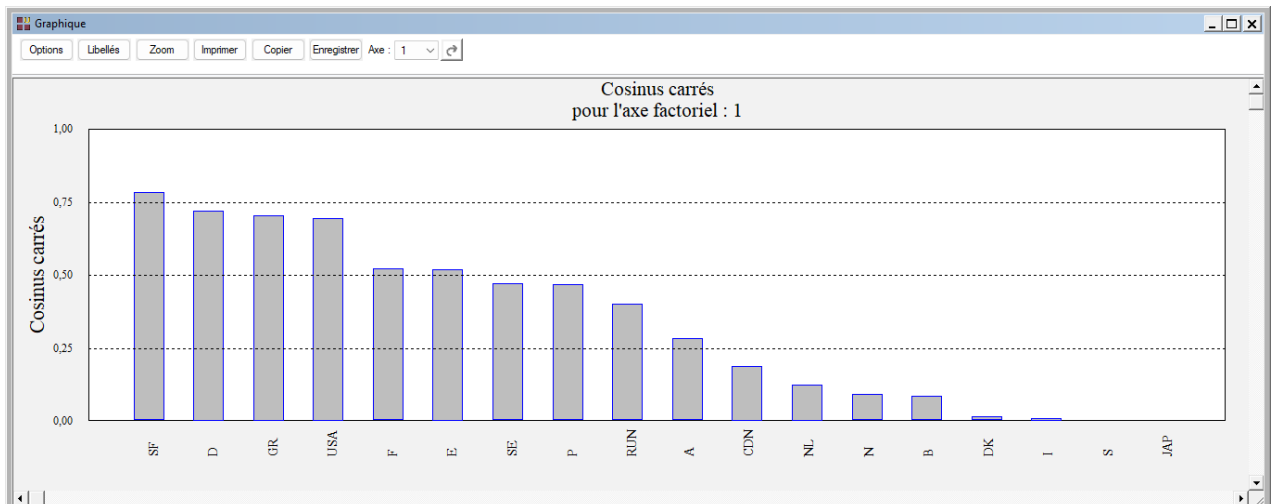
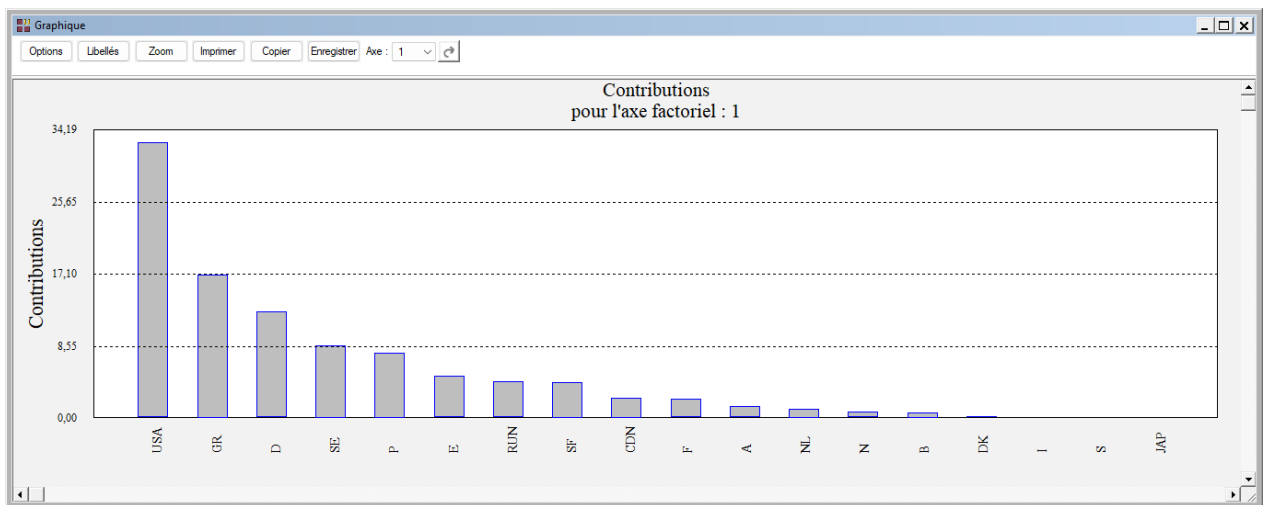


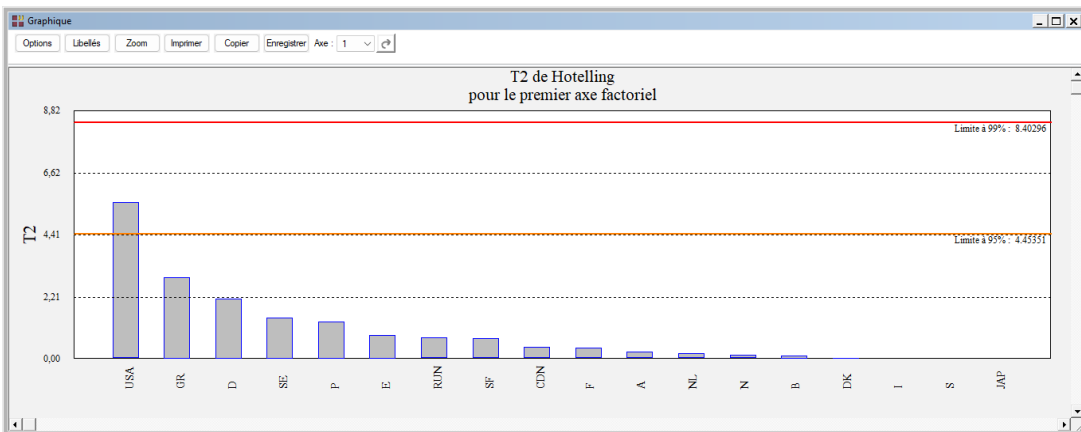
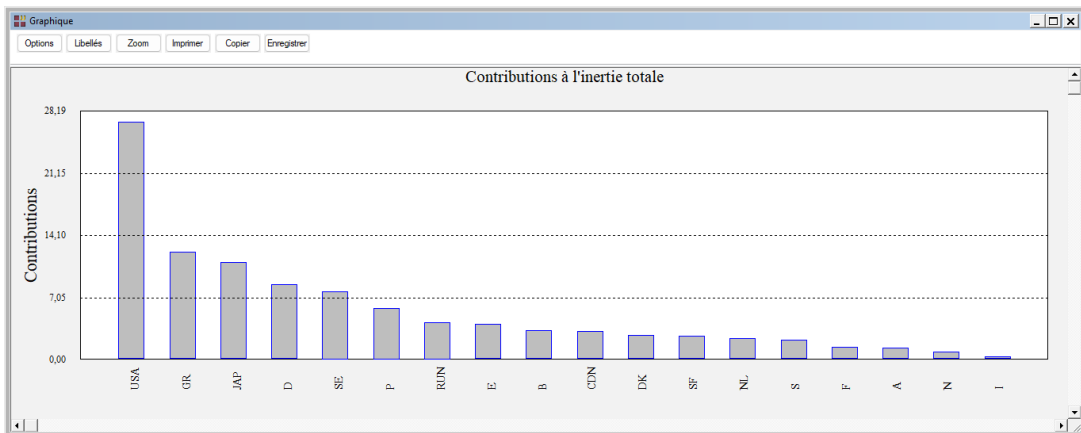
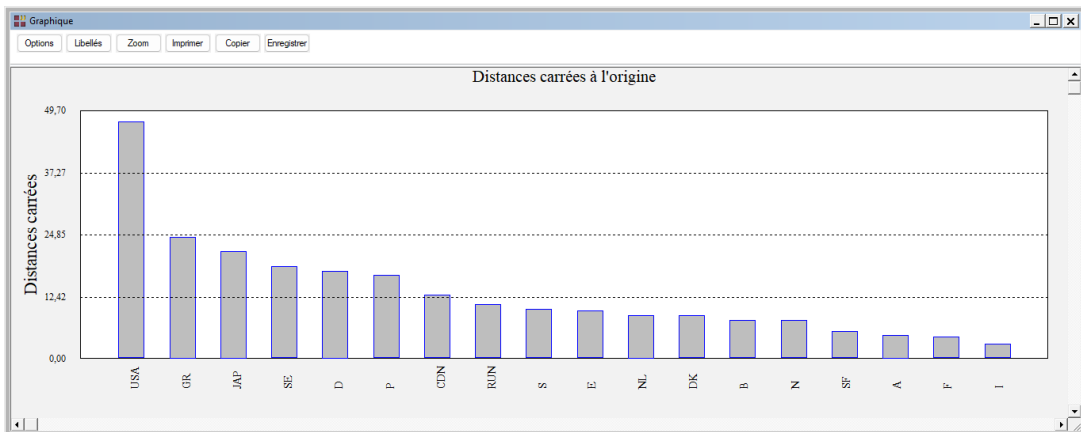
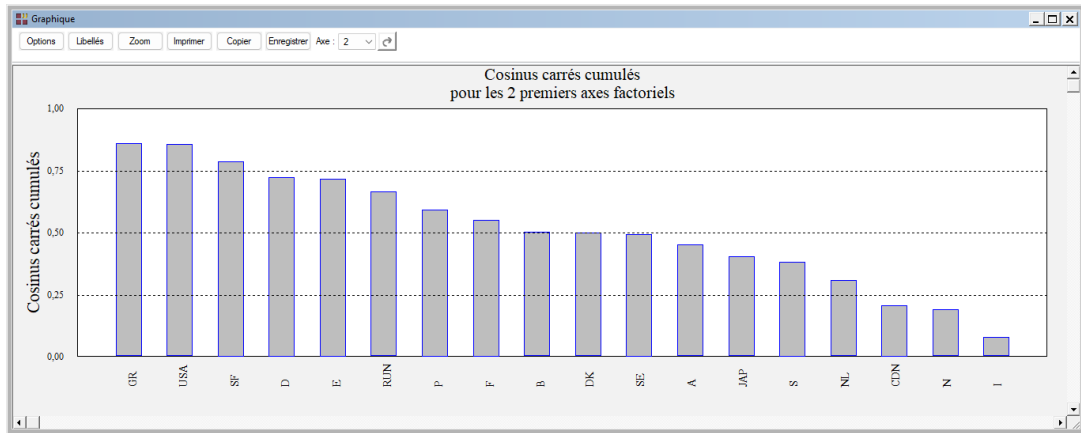
- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés, Distances carrées à l'origine, Contributions à l'inertie totale, T2 de Hotelling

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

Elles 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.

La sixième option affiche les T2 de Hotelling ainsi que les limites à 95% (individu suspect) et 99% (individu extrême).




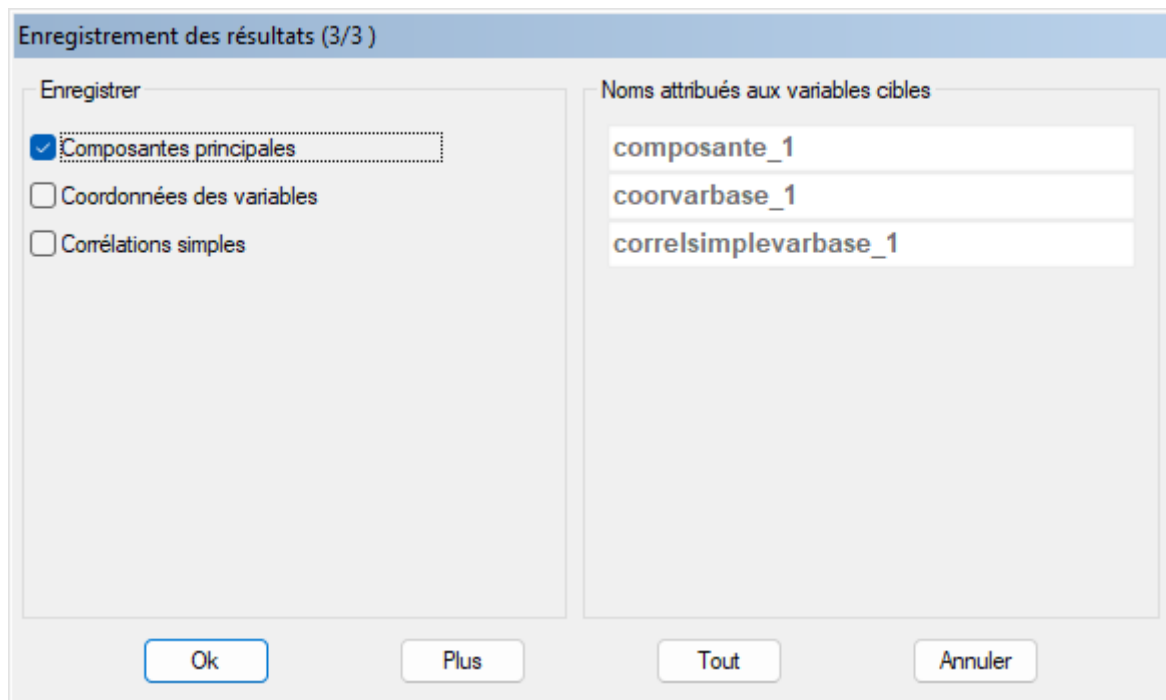


Tracé des ellipses de tolérance

Il est possible d'afficher les ellipses de tolérance contenant un pourcentage donné des observations autour de groupes définis par une variable qualitative.

Après avoir exécuté l'ACP, enregistrer les composantes principales en cliquant sur le

bouton Enregistrer  :



Préciser le nom du fichier des résultats ou accepter son nom par défaut ACP_Ocde.SGD pour notre exemple.

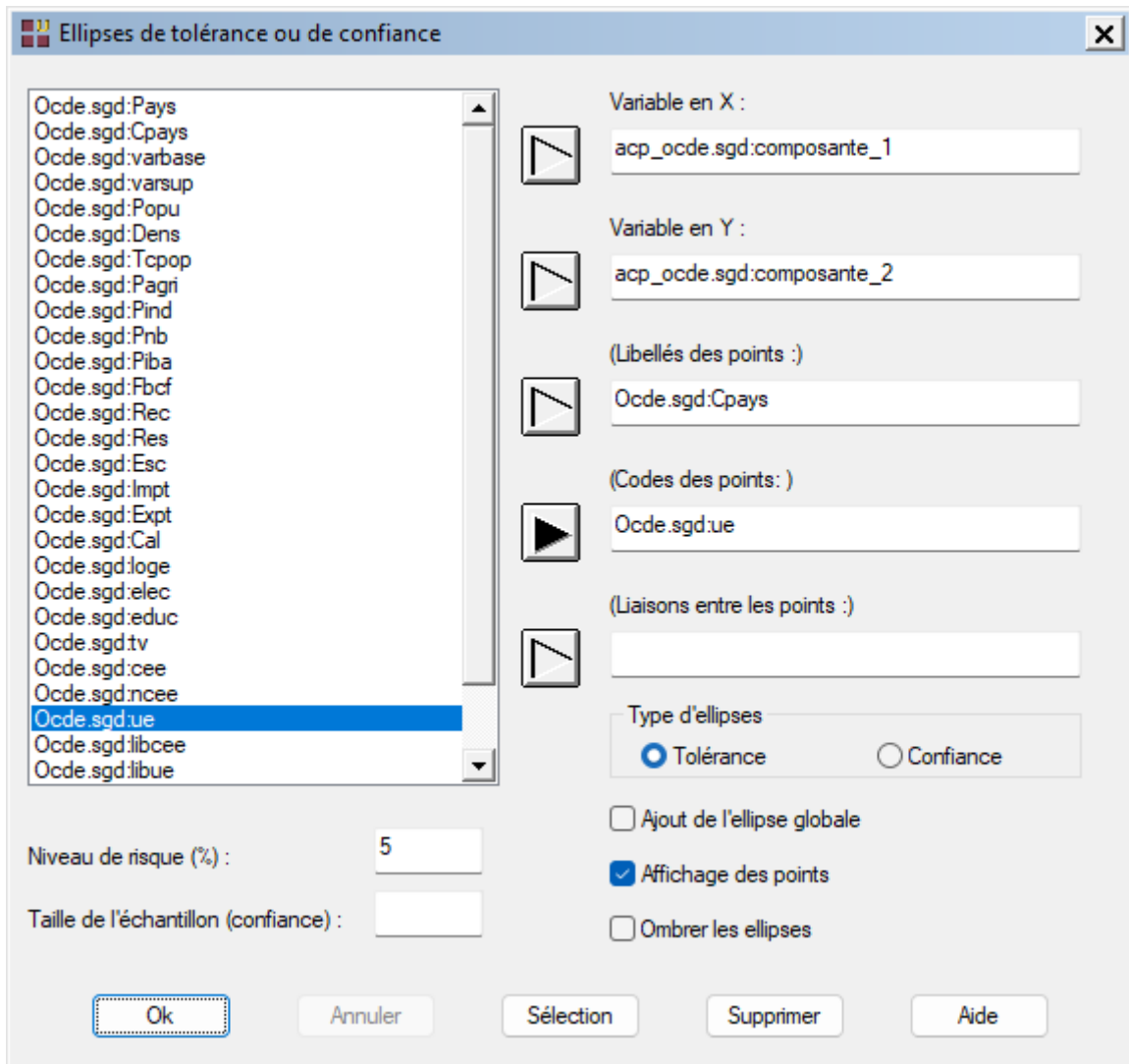
Cliquer sur l'icône ELL dans le ruban Décrire, quitter la procédure en cours et renseigner la boîte de dialogue Ellipses comme montré ci-après.

Choisir pour l'axe X la première composante principale et pour l'axe Y la deuxième composante principale.

Choisir la variable qualitative définissant les groupes, dans notre exemple 'ue' qui indique l'appartenance ou non à l'Union Européenne.

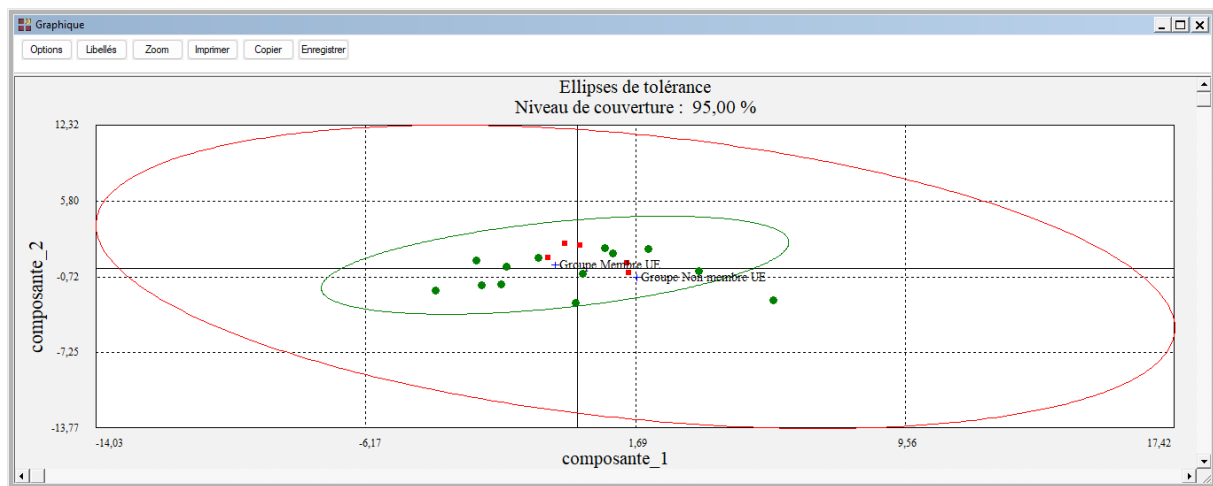
Choisir des ellipses de tolérance.

Définir le niveau de risque, par exemple 5%, indiquant que les ellipses contiendront 95% des observations pour chacun des groupes.

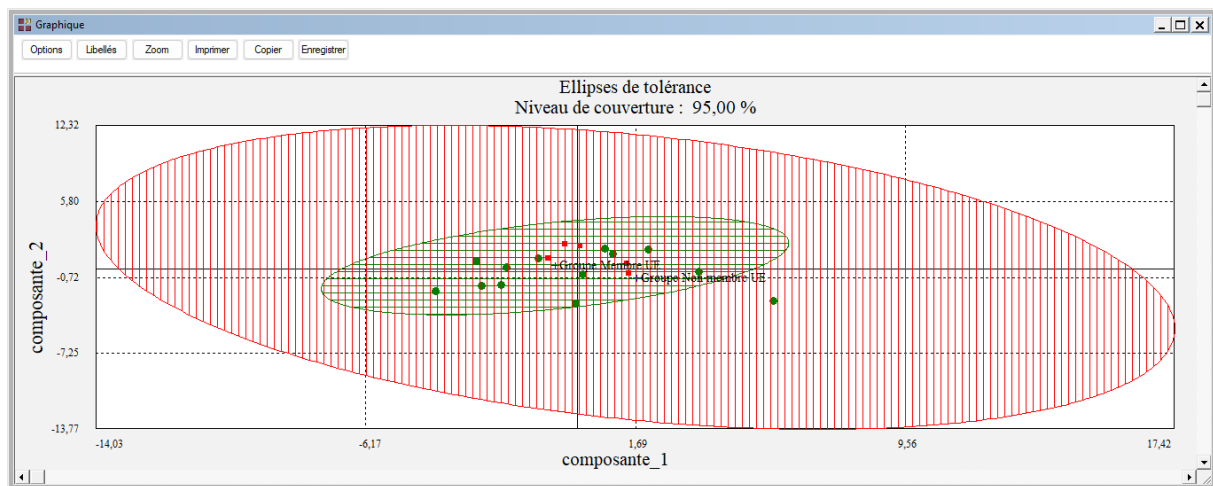


Cliquer sur Ok pour afficher le rapport et le graphique.

	Effectif	Moyenne X	Moyenne Y	Ecart-type X	Ecart-type Y	Corrélation X-Y
Groupe Membre UE	13	-0,65156	0,27678	2,14044	1,32017	0,47229
Groupe Non membre UE	5	1,89406	-0,72482	2,53806	2,08795	-0,35449
Groupe Global	18	0,00000	0,00000	2,43278	1,57162	0,00000



Diverses options d'affichage sont disponibles : ajout de l'ellipse globale, affichage des points, ombrage des ellipses et pour les libellés : libellés des points, codes des points ou numéros des points.

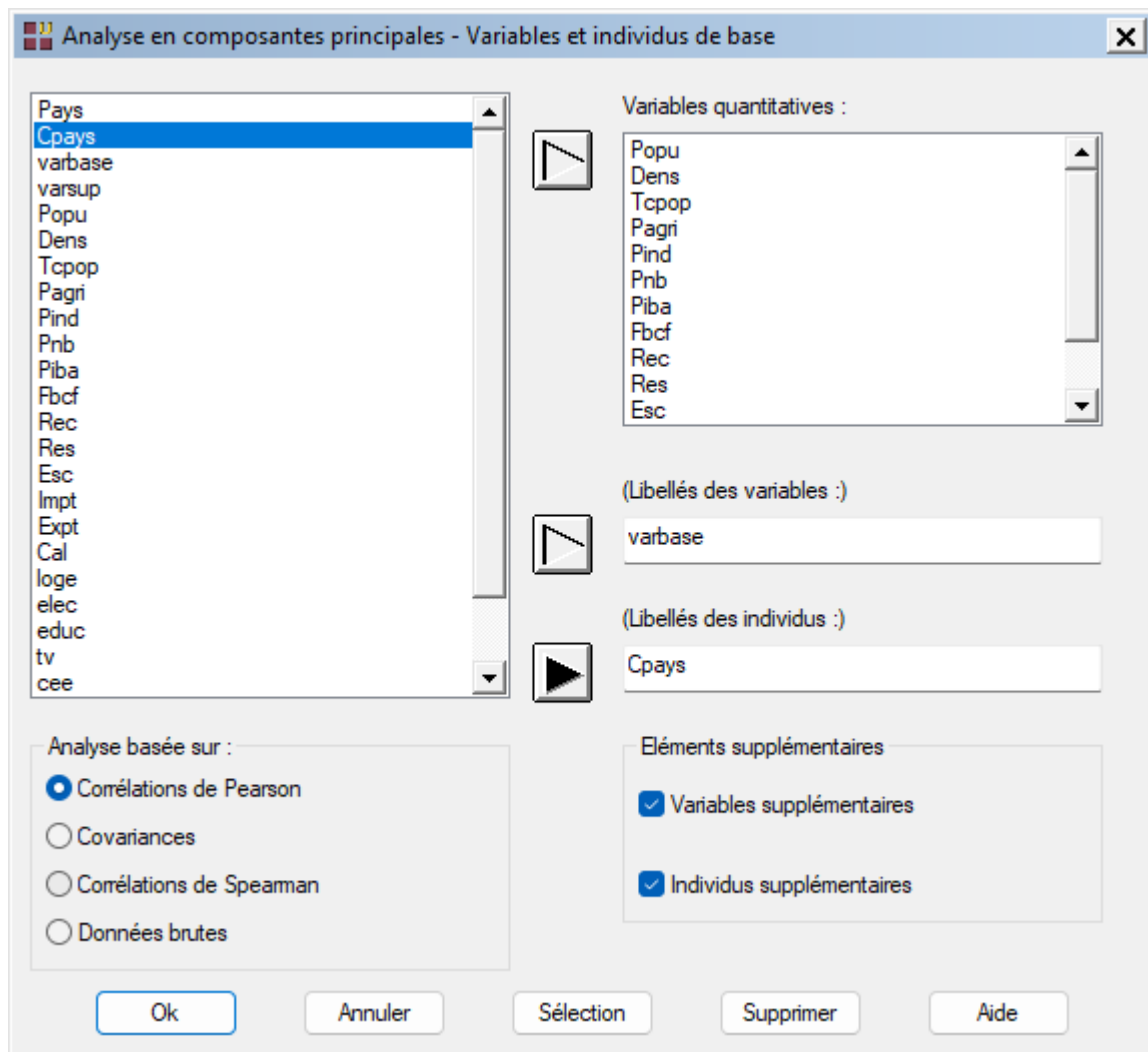


Exemple 2 : Fichier OCDE (variables et individus de base et supplémentaires)

Nous utiliserons à nouveau le fichier OCDE pour ce deuxième exemple.

Nous allons mettre en œuvre une ACP contenant à la fois des individus de base (les pays membres de la CEE), des individus supplémentaires (les pays qui ne sont pas membres de la CEE), des variables quantitatives de base (les variables Popu à Cal), une variable qualitative supplémentaire (la variable ue), des variables quantitatives supplémentaires (les variables loge à tv).

La première boîte de dialogue montrée ci-après apparaît.



Nous choisissons les variables de Popu à Cal comme variables quantitatives de base (actives), la variable varbase comme variable contenant les libellés longs des variables quantitatives de base, la variable Cpays pour les libellés courts des individus (les pays), une analyse avec centrage et réduction et cochons les cases Variables supplémentaires et Individus supplémentaires.

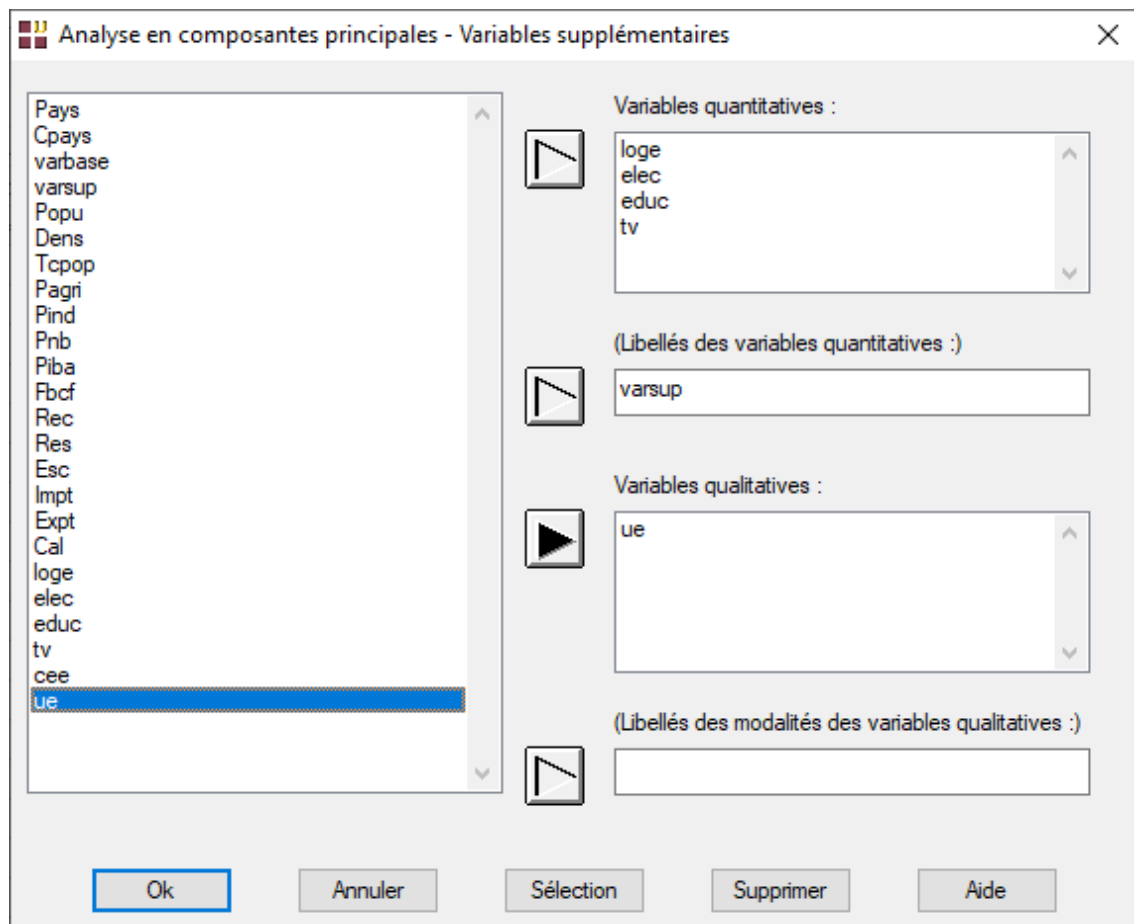
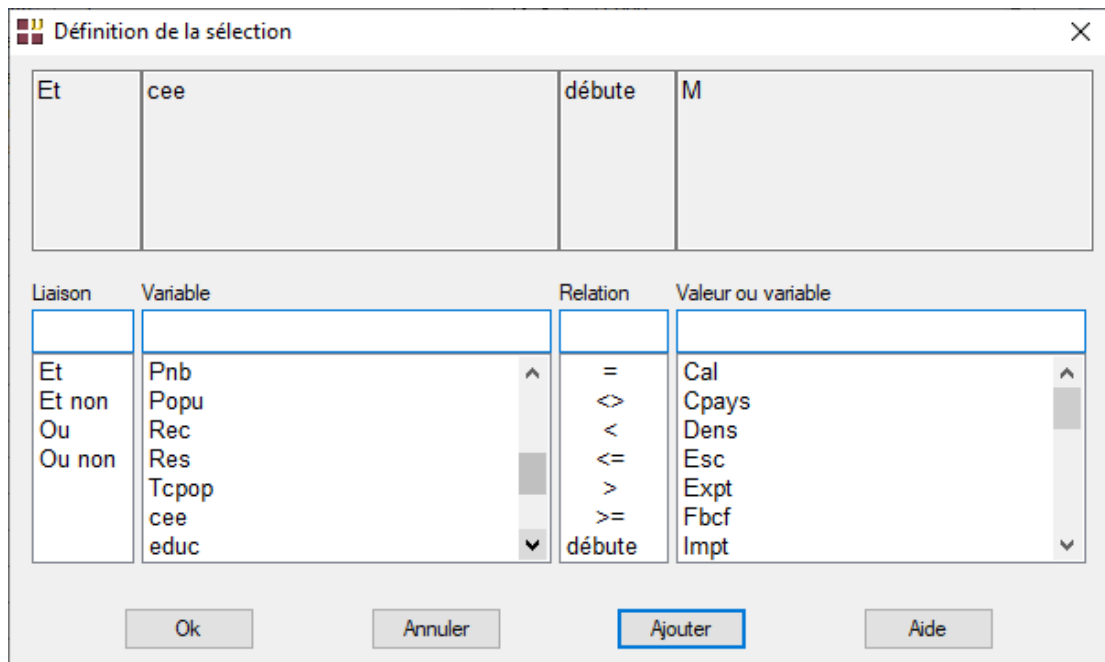
Cliquons sur le bouton Sélection pour sélectionner les individus de base, c'est-à-dire les pays membres de la CEE.

La boîte de dialogue montrée ci-après indique la sélection effectuée.

Cliquons sur Ok. Un message s'affiche ensuite et indique que 11 des 18 pays ont été sélectionnés.

Cliquons sur Ok et la deuxième boîte de dialogue apparaît.

Choisissons les variables loge à tv comme variables quantitatives supplémentaires, la variable varsup comme variable contenant les libellés longs des variables quantitatives supplémentaires, la variable ue comme variable passive qualitative.



Cliques sur le bouton 'Sélection' et définissons à nouveau la sélection comme pour les données de base. Nous sélectionnons à nouveau les 11 pays membres de la CEE.

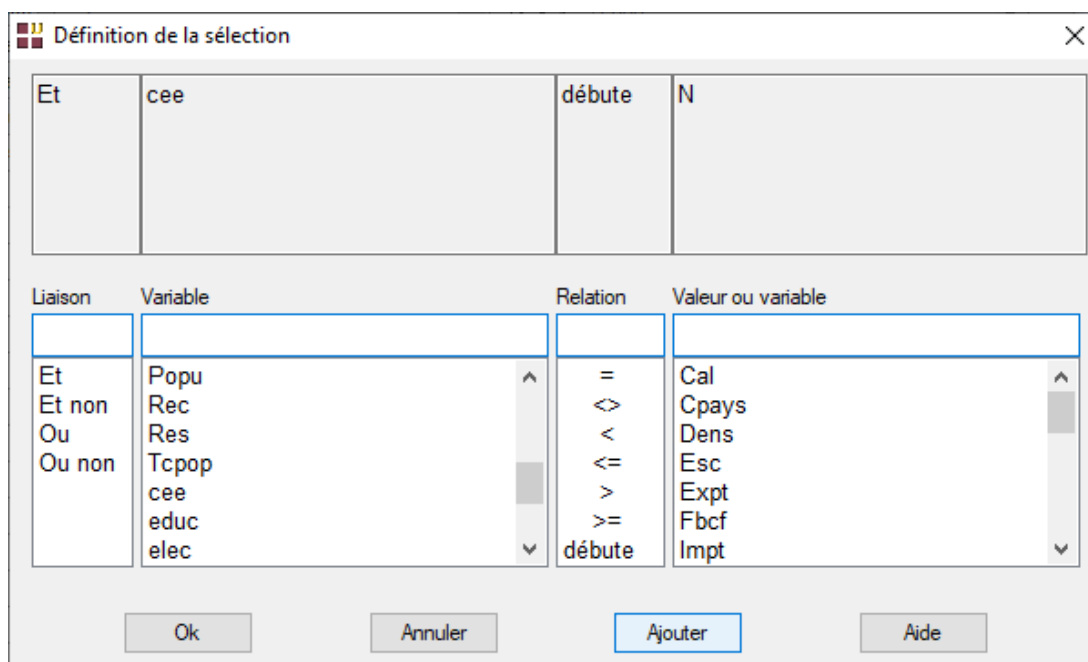
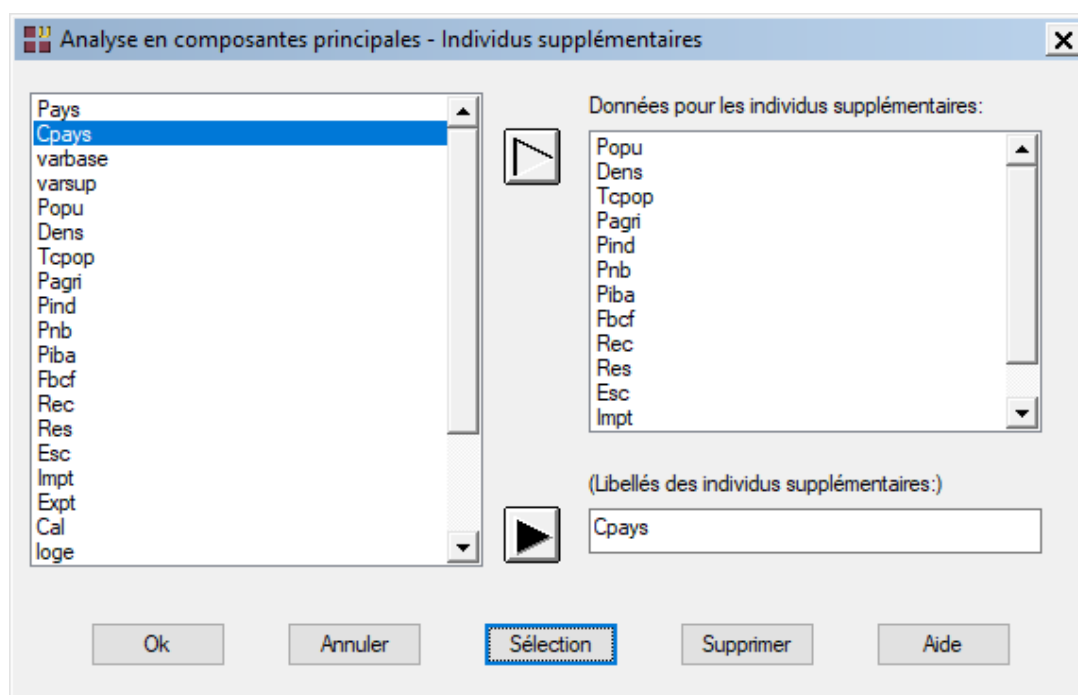
Cliques sur Ok et la troisième boîte de dialogue apparaît.

Choisissons à nouveau les variables Popu à Cal comme variables contenant les données des individus supplémentaires, la variable Cpays comme variable contenant les libellés des individus (les pays).

Cliquons sur le bouton 'Sélection' et définissons la sélection des individus supplémentaires (passifs), ici les pays qui ne sont pas membres de la CEE. La boîte de dialogue de cette sélection est montrée ci-après.

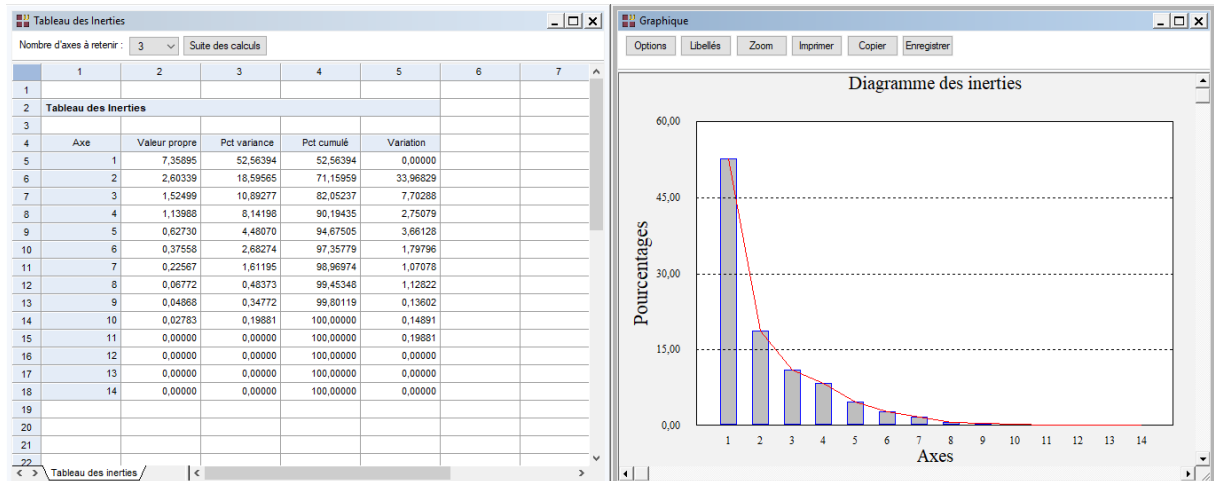
Un message indique que la sélection est constituée de 7 pays.

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



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

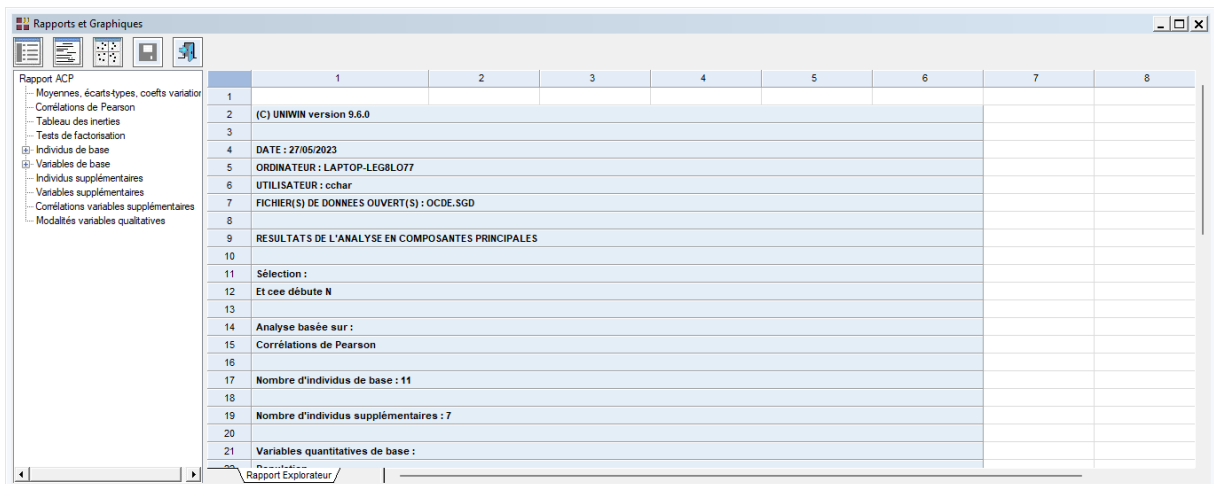
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.





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

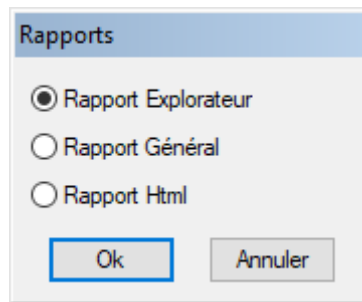
Cliquons sur le bouton 'Suite des calculs'.


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

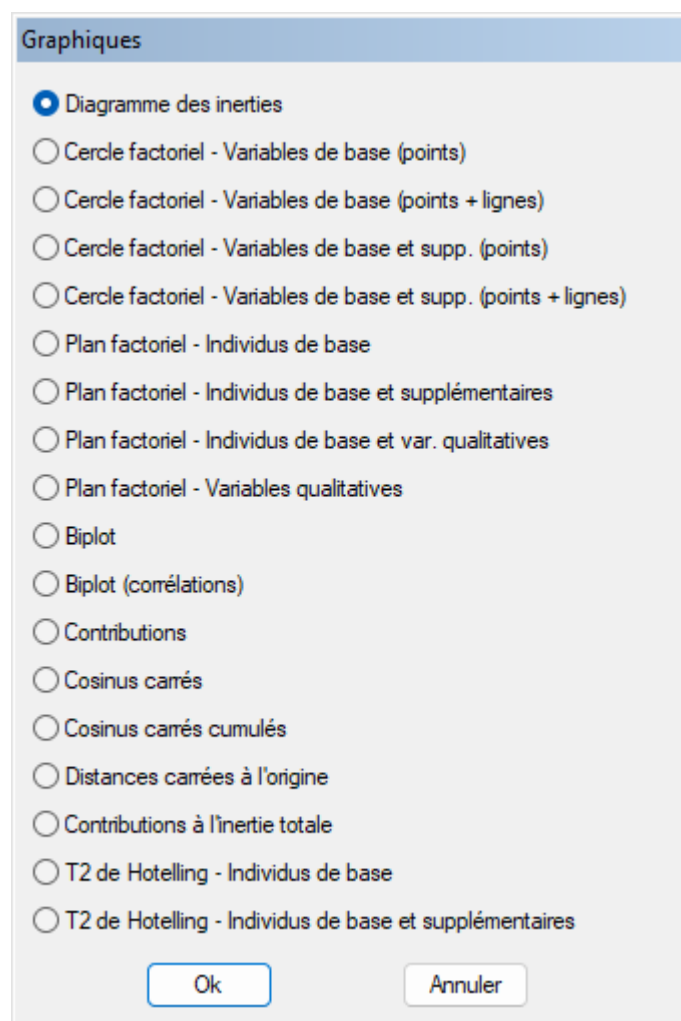



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

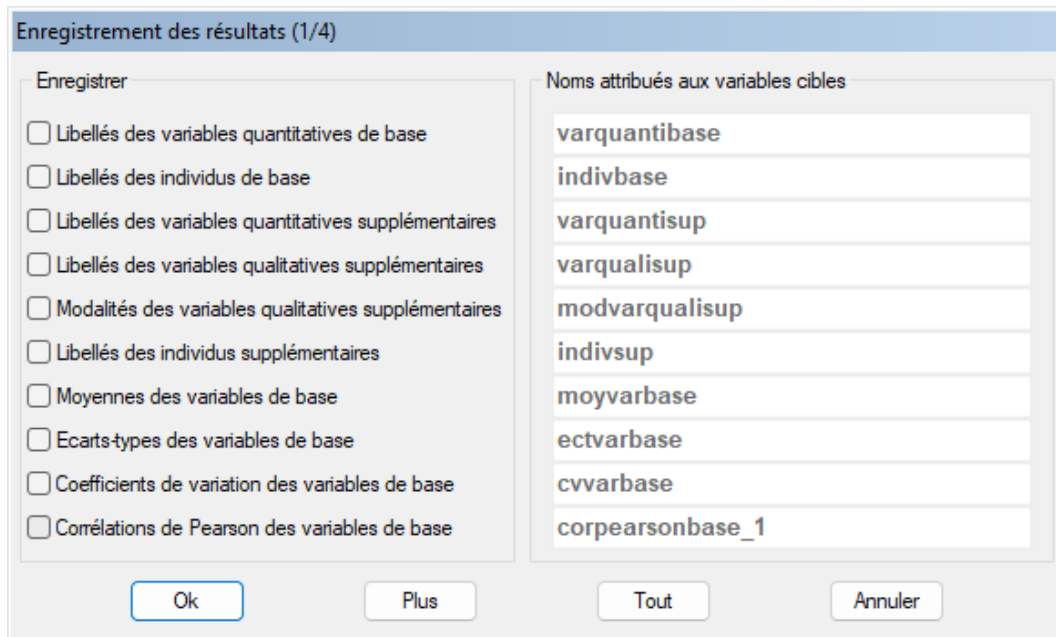
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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.



Le bouton 'Plus' permet d'afficher la suite de la liste des variable et le bouton 'Tout' permet de sélectionner toutes les 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.

L'impression des rapports fait appel à la procédure 'Aperçu avant impression'. Pour des informations sur cette procédure, voir le 'Manuel de l'Utilisateur'.

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

Rapports et Graphiques

Rapport ACP

- Moyennes, écarts-types, coeffs variat
- Corrélations de Pearson
- Tableau des inerties
- Tests de factorisation
- Individus de base
- Variables de base
 - Plus facteurs**
 - Axe factoriel 2
 - Axe factoriel 3
- Individus supplémentaires
- Variables supplémentaires
- Corrélations variables supplémentaires
- Modalités variables qualitatives

	1	2	3	4	5	6	7	8
1								
2	RESULTATS VARIABLES DE BASE POUR LE FACTEUR : 1							
3								
4	DISTANCE*2 = CARRES DES DISTANCES A L'ORIGINE OU AU BARYCENTRE							
5	COORDONNEE = COORDONNEES DES VARIABLES							
6	CORRELATION = CORRELATIONS SIMPLES VARIABLES-FACTEURS							
7	FACTEUR = FACTEURS PRINCIPAUX							
8	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE							
9	COSINUS*2 = COSINUS CARRES							
10	COS*2 CUMULE = COSINUS CARRES CUMULES POUR 1 AXE							
11								
12								
13		Variable	Distance*2	Coordonnée	Corrélation	Facteur	Contribution	Cosinus*
14	Population	1	1	-0,66180	-0,66180	-0,24396	5,95172	0,4379
15	Densité	2	1	-0,70402	-0,70402	-0,25953	6,73534	0,4956
16	Acc popu	3	1	-0,42921	-0,42921	-0,15822	2,50338	0,1842
17	Pct popu agri	4	1	0,88082	0,88082	0,32470	10,54295	0,7758
18	Pct popu ind	5	1	-0,90559	-0,90559	-0,33383	11,14408	0,8200
19	Pnb	6	1	-0,80996	-0,80996	-0,29858	8,91473	0,6560
20	Pct pnb agri	7	1	0,96519	0,96519	0,35580	12,85935	0,9316
21	Fbct	8	1	0,02167	0,02167	0,00799	0,00638	0,0004

	1	2	3	4	5	6	7	8	9	10	11	12
141												
142	RESULTATS INDIVIDUS POUR LE FACTEUR : 1											
143												
144	DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE											
145	CONTRIBUTION TOTALE = CONTRIBUTIONS A L'INERTIE TOTALE											
146	COMPOSANTE = COMPOSANTES PRINCIPALES - COORDONNÉES DES INDIVIDUS											
147	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE											
148	COSINUS*2 = COSINUS CARRÉS											
149	COS*2 CUMULE = COSINUS CARRÉS CUMULES POUR 1 AXE											
150	T2 = T2 DE HOTELLING POUR 1 AXE											
151												
152												
153		Individu	Distance*2	Contribution totale	Composante	Contribution	Cosinus*2	Cos*2 cumulé	T2			
154	D	1	23,61147	18,97253	-4,25941	24,65377	0,76838	0,76838	2,46538			
155	B	2	6,58864	3,17888	-1,29141	2,26628	0,25312	0,25312	0,22663			
156	DK	3	8,78393	5,71295	0,08901	0,01077	0,00090	0,00090	0,00108			
157	E	4	10,55007	7,24619	2,29689	7,16909	0,50006	0,50006	0,71691			
158	F	5	7,64386	4,76971	-1,98397	5,34880	0,51494	0,51494	0,53488			
159	GR	6	22,01811	18,21634	4,00061	21,74889	0,72690	0,72690	2,17489			
160	SE	7	18,39350	13,56819	3,04500	12,59962	0,50409	0,50409	1,25996			
161	I	8	3,72588	2,86102	-0,57297	0,44611	0,08811	0,08811	0,04461			

	12	1	0,01564	0,01564	0,01267	0,01604	0,00024	0,91445
Importations	12	1	0,01564	0,01564	0,01267	0,01604	0,00024	0,91445
Exportations	13	1	-0,00218	-0,00218	-0,00176	0,00031	0,00000	0,90398
Calories	14	1	0,14581	0,14581	0,11807	1,39406	0,02126	0,76968

COMPOSANTES PRINCIPALES ET T2 DE HOTELLING DES INDIVIDUS SUPPLEMENTAIRES

	Individu	Coord. axe 1	Coord. axe 2	Coord. axe 3	T2 axe 1	T2 axe 2	T2 axe 3
A	12	0,96858	0,42631	0,40255	0,11685	0,15129	0,22700
CDN	13	-1,46966	-0,57515	-1,62434	0,26903	0,33171	1,56445
USA	14	-6,33686	-3,72756	0,65324	5,00163	7,63453	7,83390
SF	15	1,94520	0,29977	-0,54195	0,47129	0,48832	0,62554
JAP	16	-0,68849	-3,95051	-2,79013	0,05853	3,01579	6,65300
N	17	0,58559	0,36879	-0,79757	0,04271	0,06848	0,36569
S	18	-0,56636	1,19353	-0,94872	0,03995	0,30988	0,73041

COORDONNÉES DES VARIABLES SUPPLEMENTAIRES

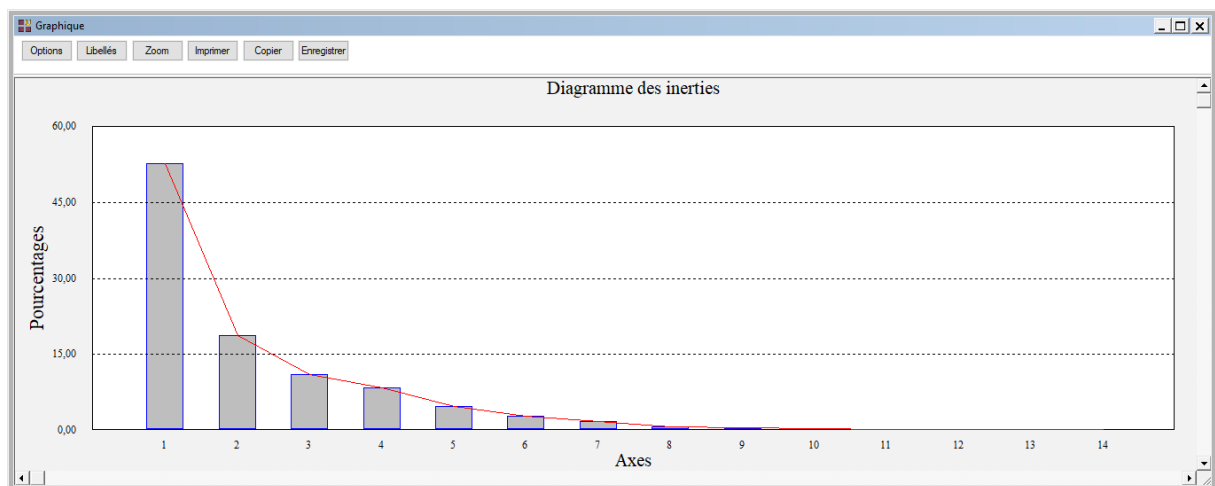
	Composante 1	Composante 2	Composante 3

L'option Graphiques

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

- Diagramme des inerties

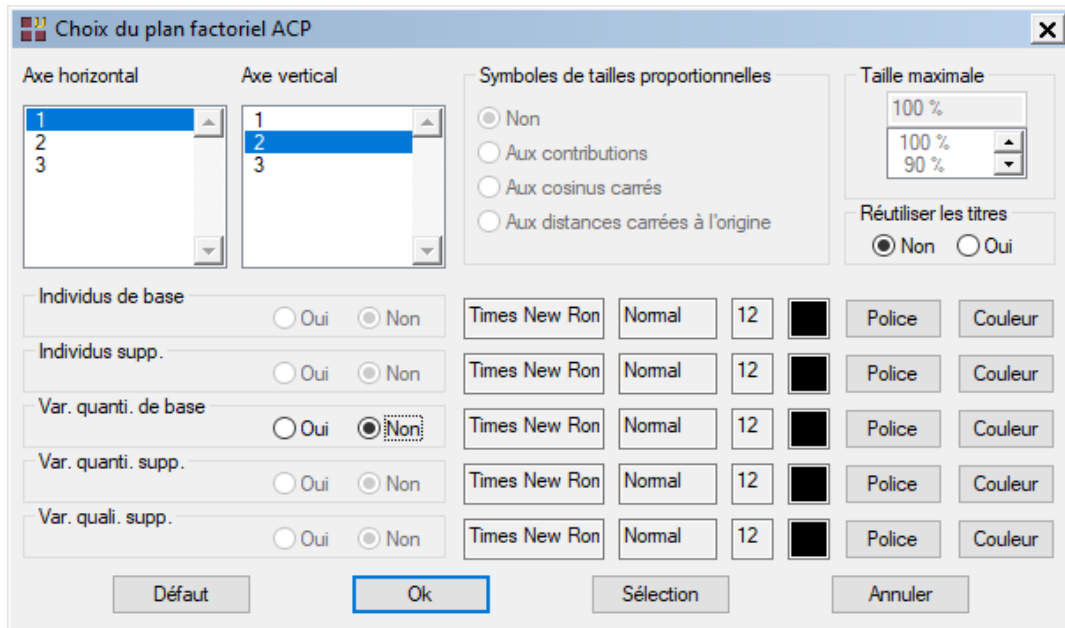
Ce graphique affiche les pourcentages d'inertie pour chacun des axes factoriels.



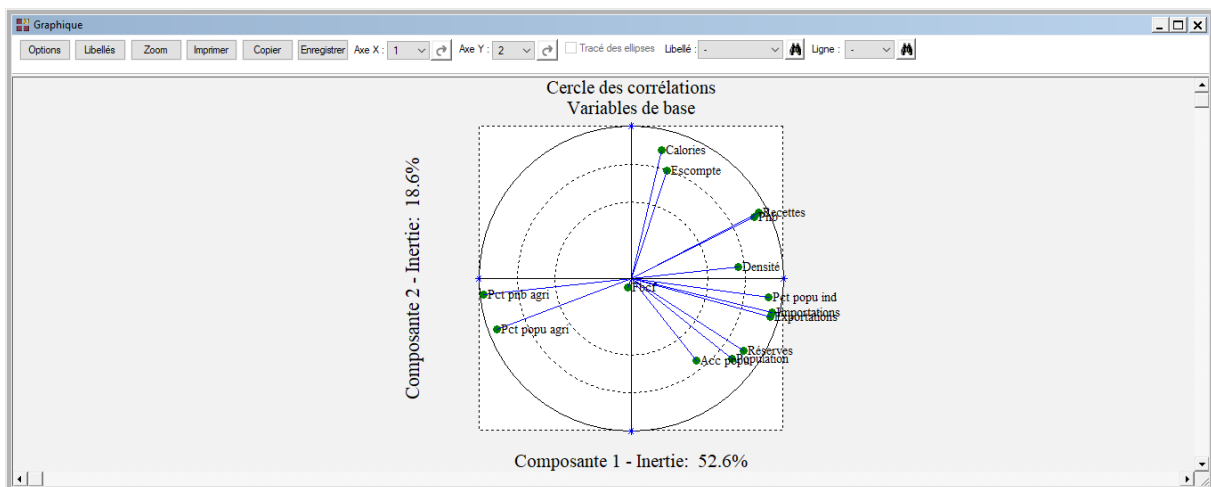
- Les options Cercle factoriel

Ces options permettent d'afficher le cercle de corrélations des variables de base seules ou des variables de base et des variables supplémentaires (s'il y en a) 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 de base avec lignes.

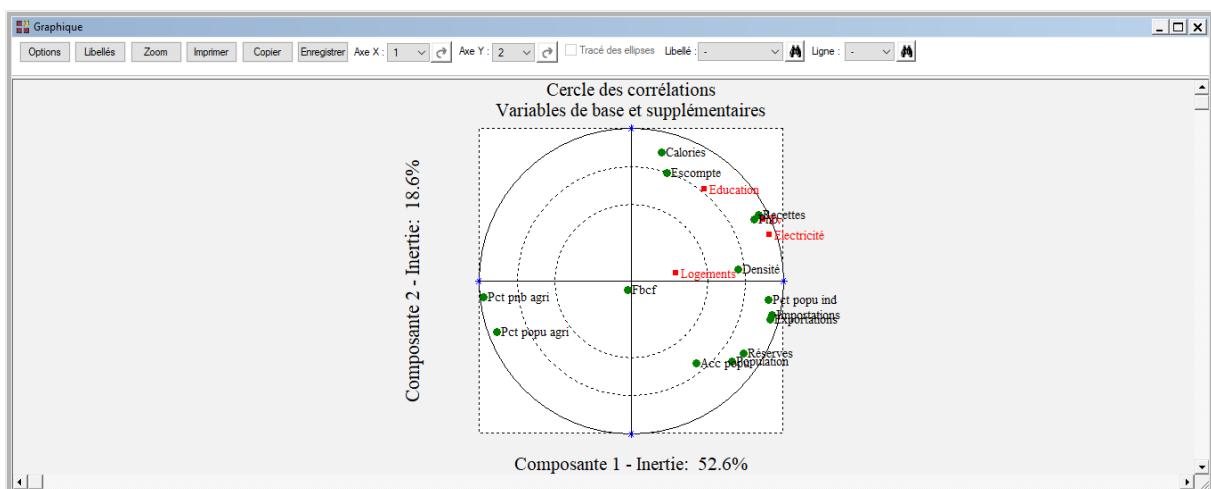
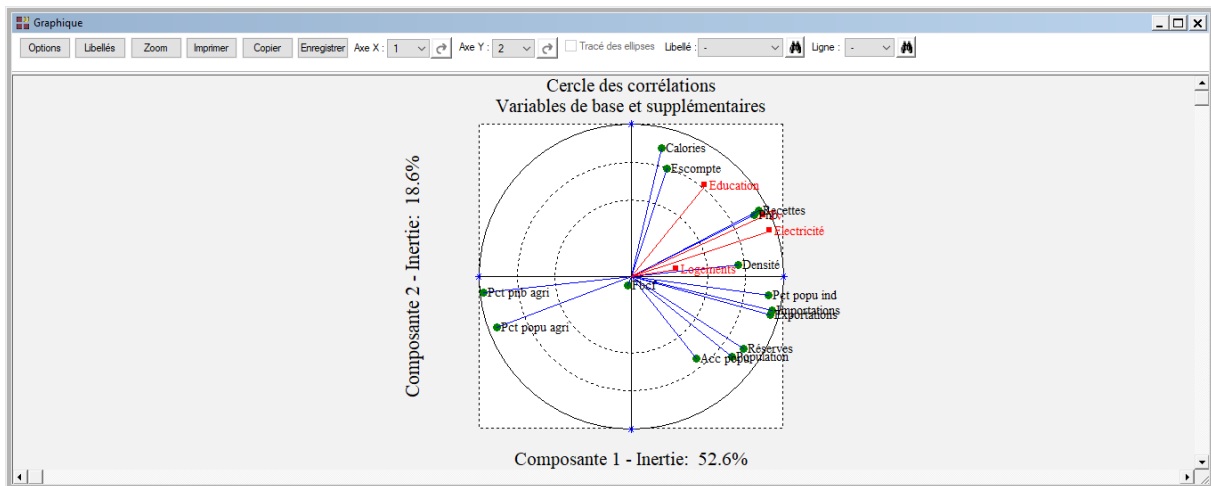
Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.



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.



Visualisons également les corrélations des variables de base et des variables supplémentaires avec ou sans lignes.



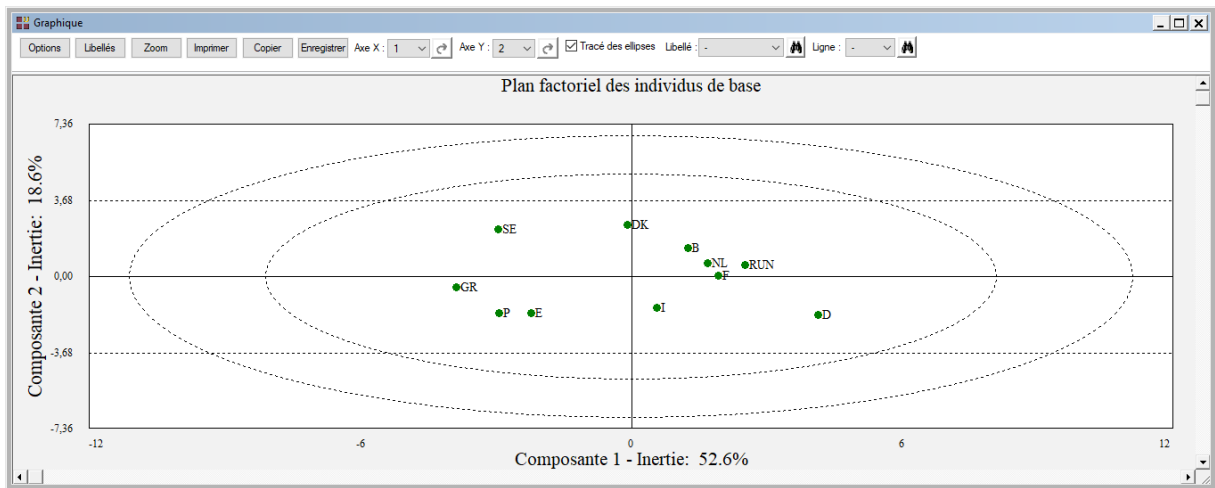
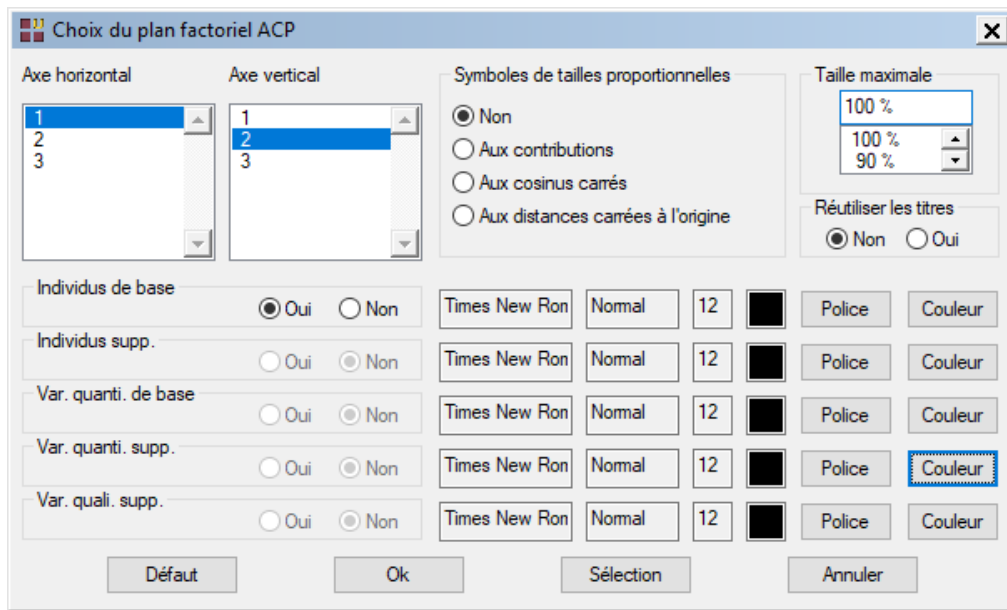
- Les options Plan factoriel et Variables qualitatives

Ces options permettent d'afficher des plans factoriels des individus de base, des individus de base et des individus supplémentaires (s'il y en a), des individus de base et des variables qualitatives (s'il y en a) ainsi que des variables qualitatives seules.

Choisissons la première option. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche alors.

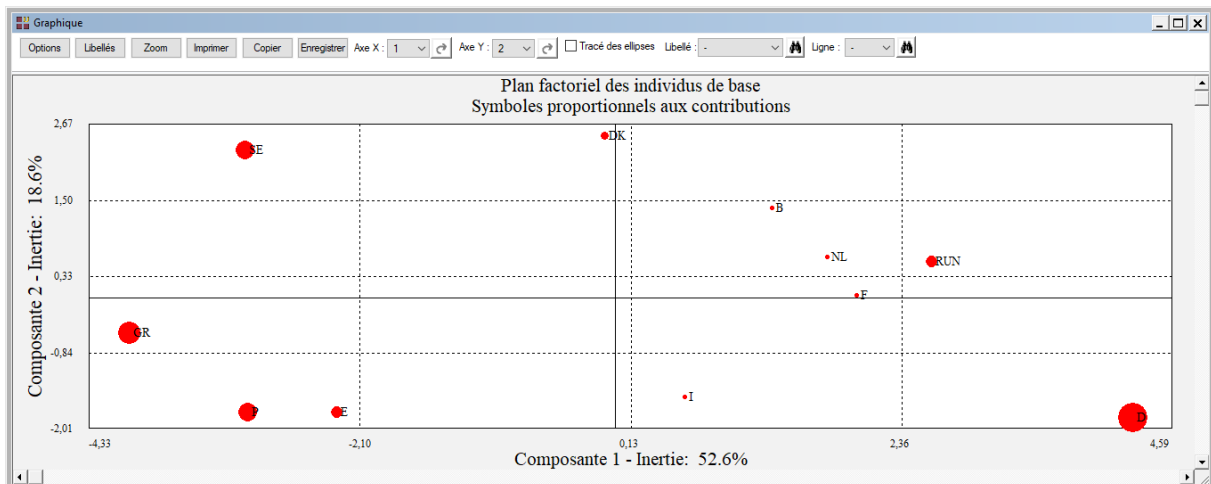
Elle permet également de préciser si l'on désire afficher les libellés des individus, de choisir la couleur et la police, d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail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



Des ellipses de confiance à 95% et 99% ont été ajoutées dans le graphique.

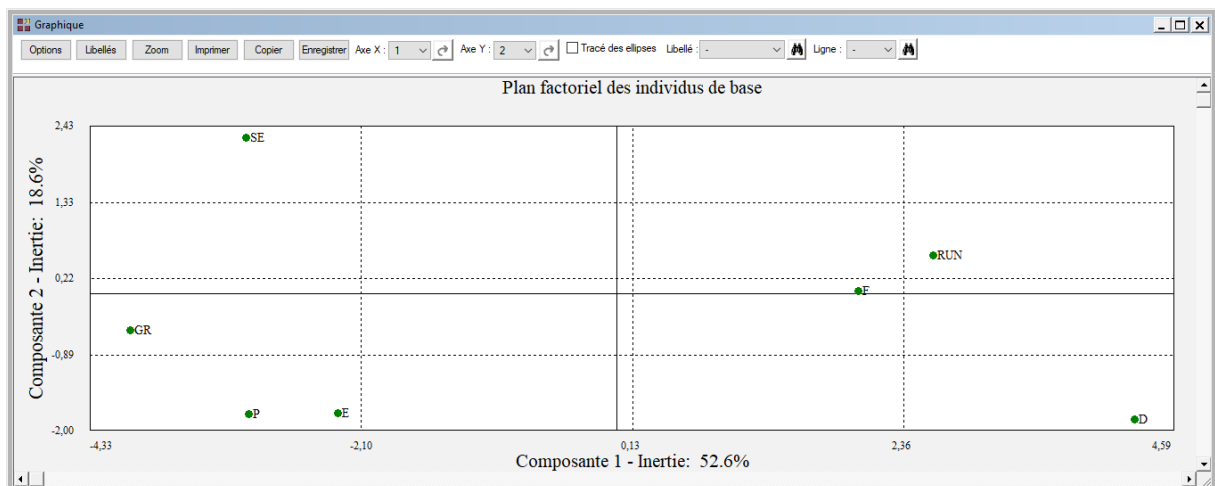
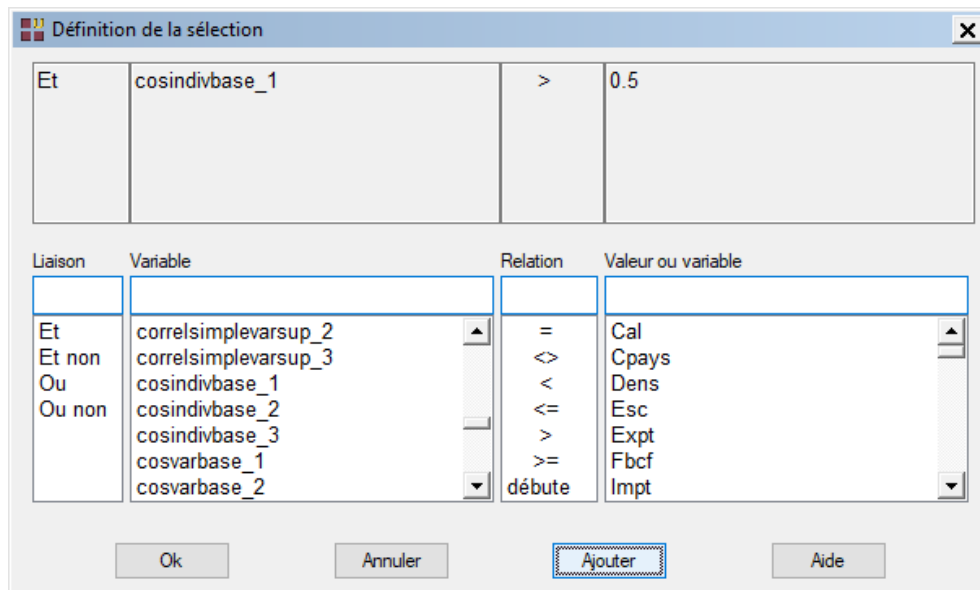
Affichons ce même graphique avec des symboles de points proportionnels aux contributions.



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'individus : seuls les points significatifs d'un point de vue statistique sont affichés.

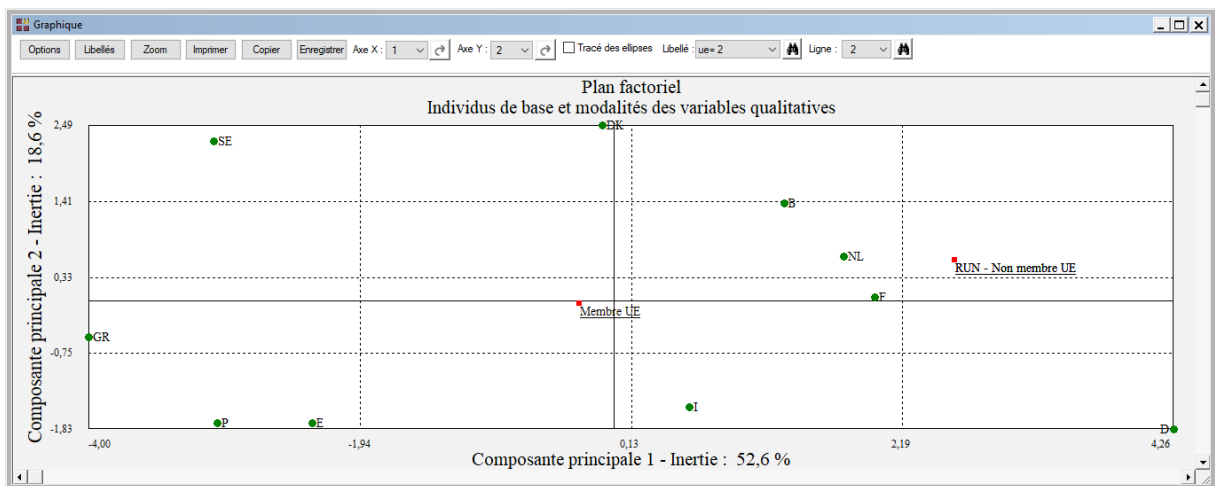
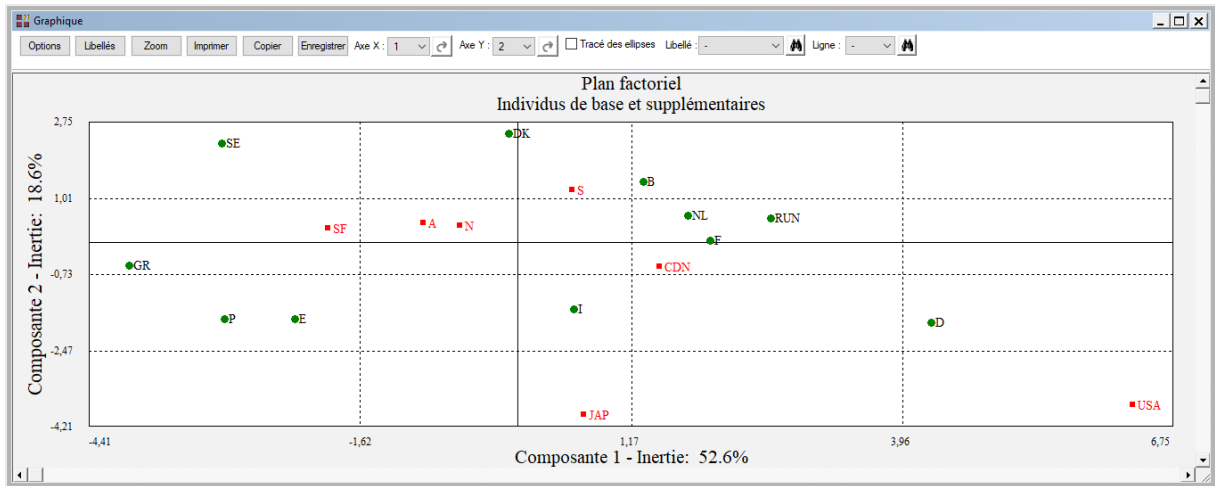
Les principaux résultats de l'analyse sont accessibles dans la boîte de dialogue 'Sélection' en plus des données du fichier OCDE.

Comme exemple, mettons en œuvre une sélection basée sur la variable 'cosindivbase_1' qui contient les cosinus carrés sur le premier axe factoriel.



Des exemples des plans factoriels Individus de base et supplémentaires, Individus de base et Variables qualitatives sont également montrés.

A noter que la modalité 'Non membre UE' est superposée au libellé 'RUN' car seul le Royaume-Uni membre de la CEE n'est pas membre de l'Union Européenne.



- Les options Biplot et Biplot (corrélations)

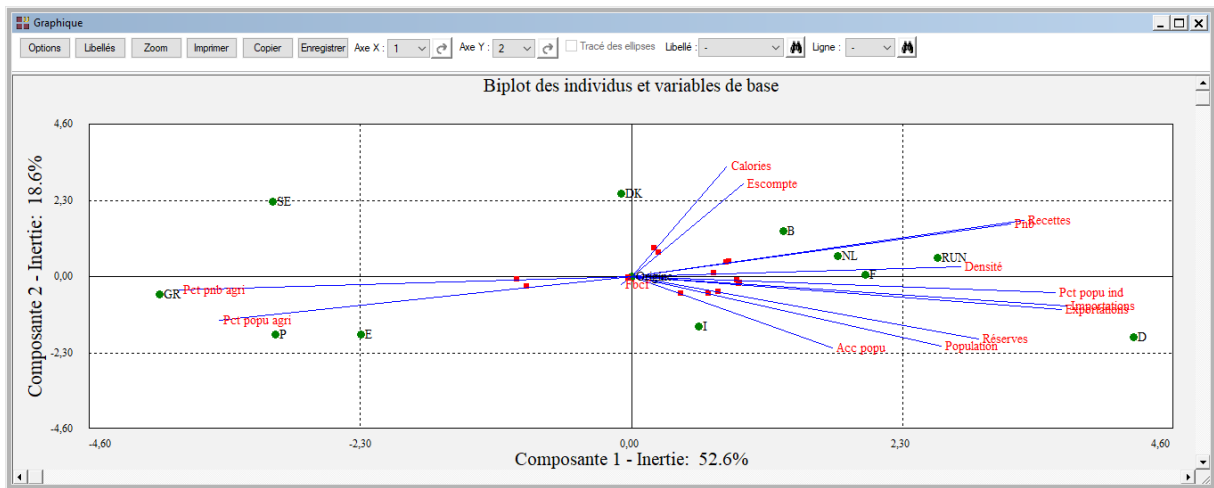
L'option Biplot permet d'obtenir un graphique donnant une représentation simultanée des individus et des directions des variables d'origine dans le plan factoriel choisi.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Un exemple du Biplot est montré ci-après.

A noter que les droites indiquant les directions des variables d'origine dans le plan factoriel ont des longueurs proportionnelles aux corrélations visualisées dans le cercle factoriel.

L'option d'interrogation des points individus et des points variables est possible. Par contre le zoom ne l'est pas.

La boîte de dialogue 'Echelles des axes' permet de définir un graphique orthonormé.



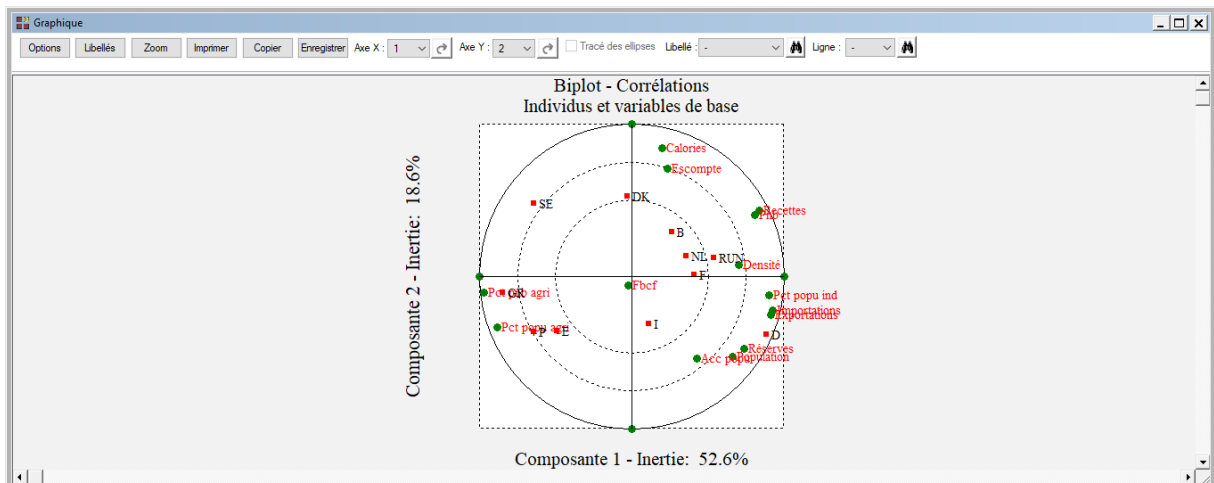
L'option Biplot (corrélations) permet d'obtenir un graphique donnant une représentation simultanée des variables indicatrices des individus et des variables d'origine dans un cercle des corrélations.

Les variables indicatrices des individus sont représentées comme si elles étaient des variables supplémentaires dans l'analyse.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Un exemple du Biplot (corrélations) est montré ci-après.

L'option d'interrogation des points individus et des points variables est possible.

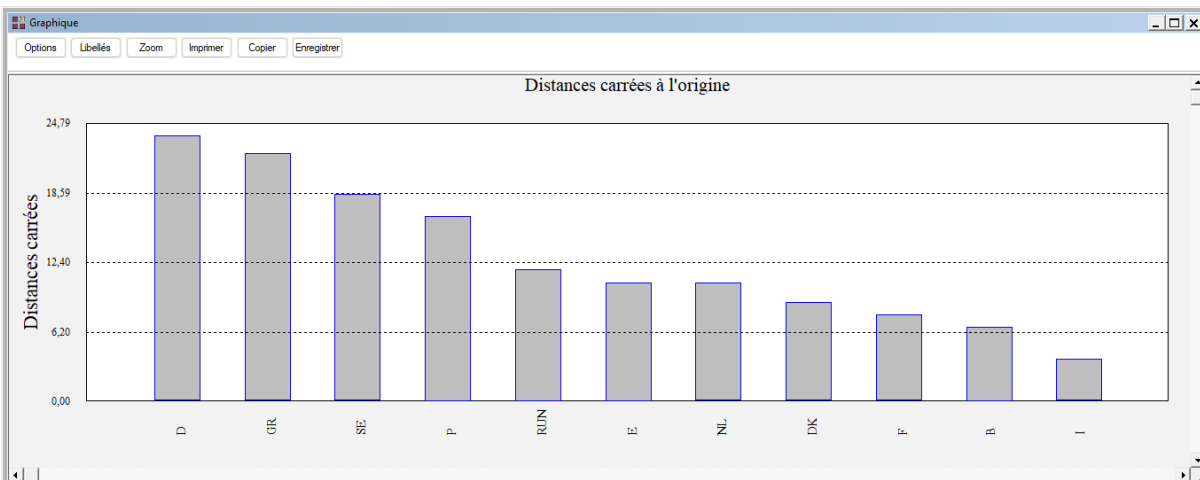
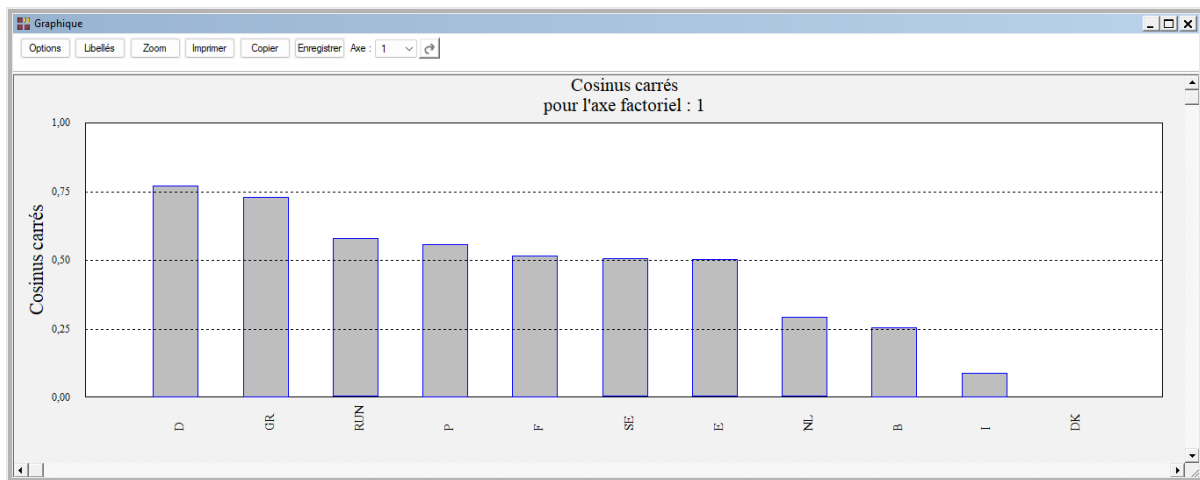
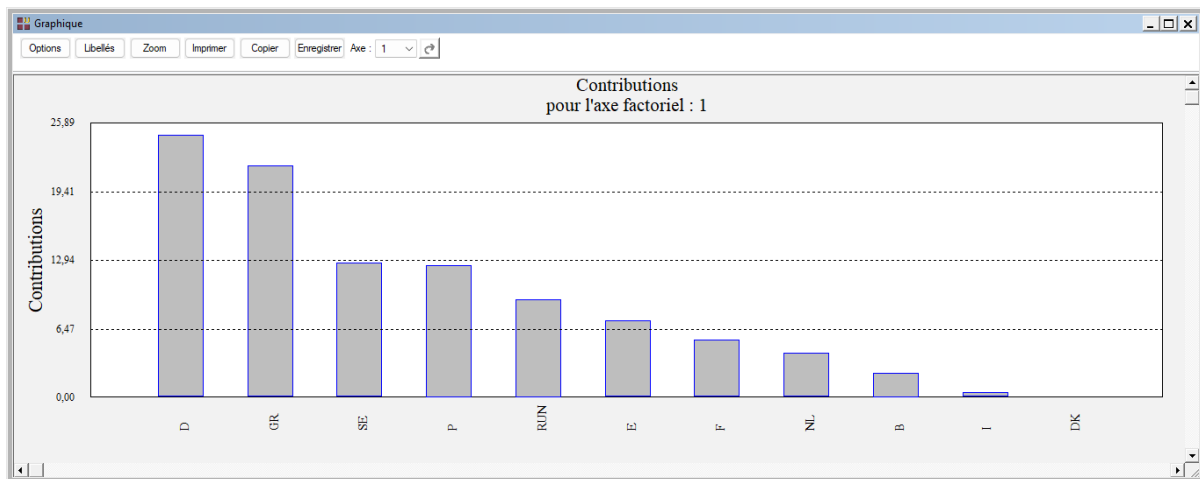
L'option zoom est également disponible.

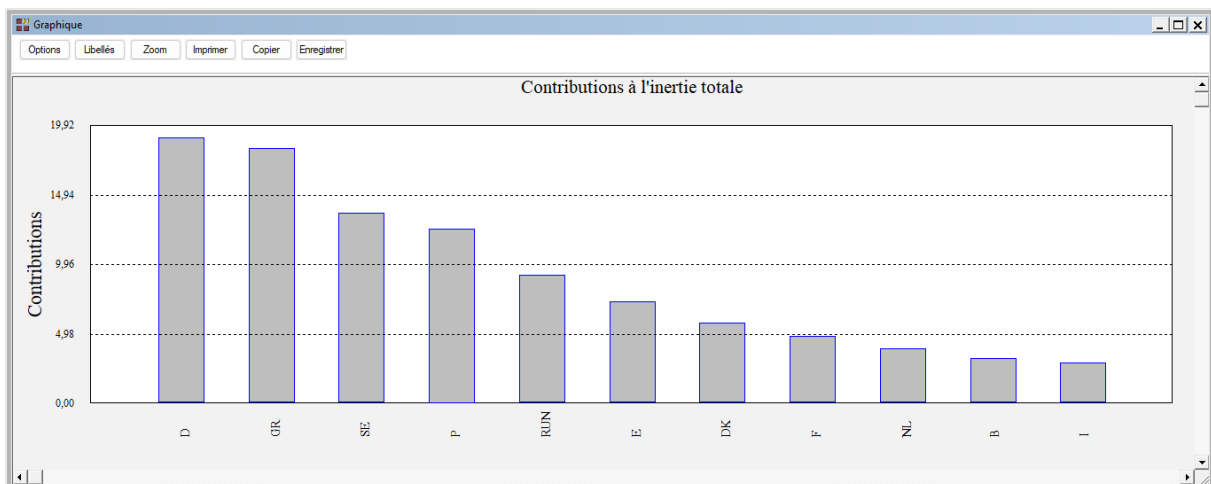
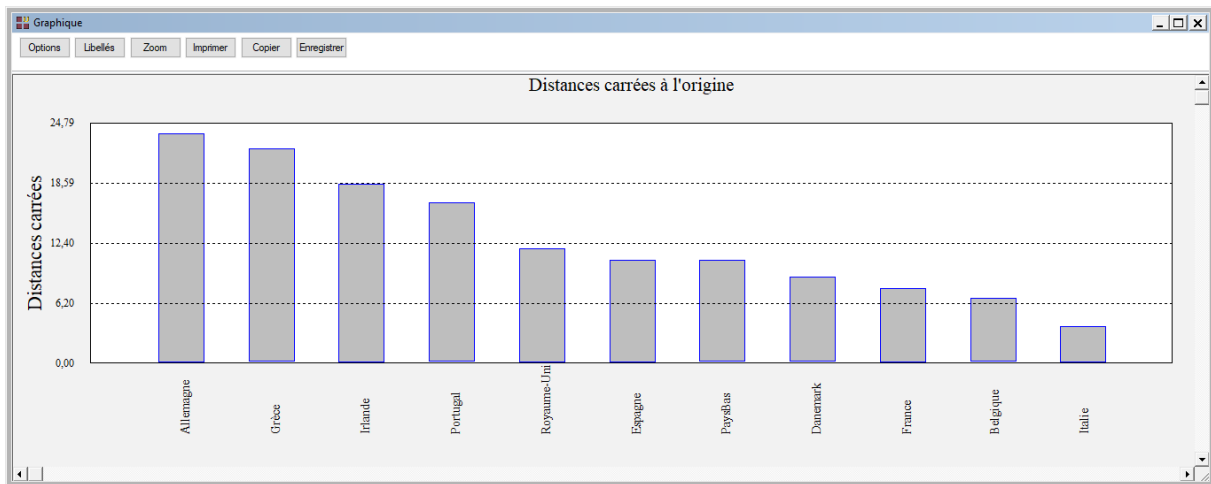


- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés, Distances carrées à l'origine, Contributions à l'inertie totale, T2 de Hotelling

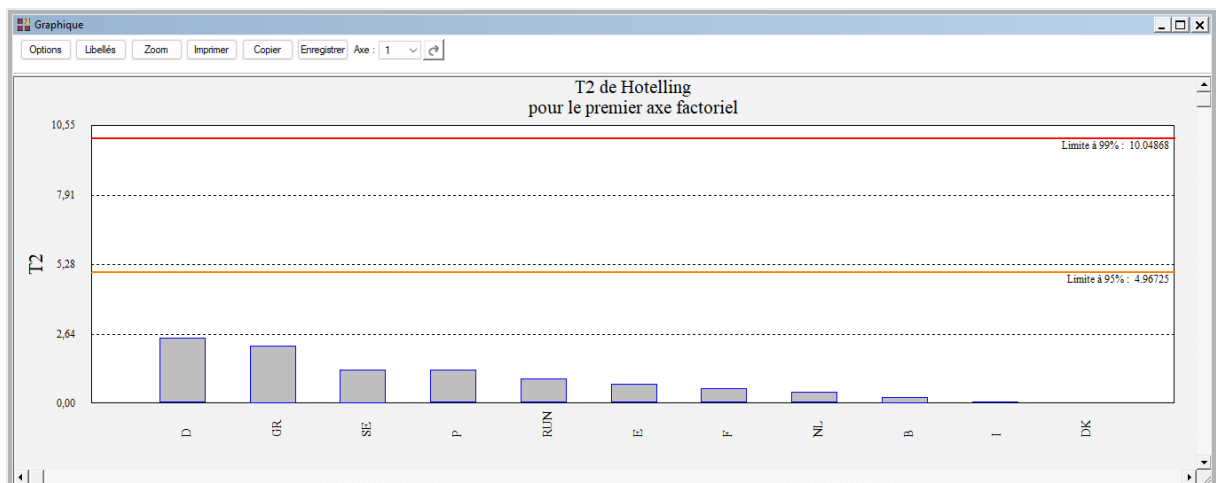
Les cinq premières options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse.

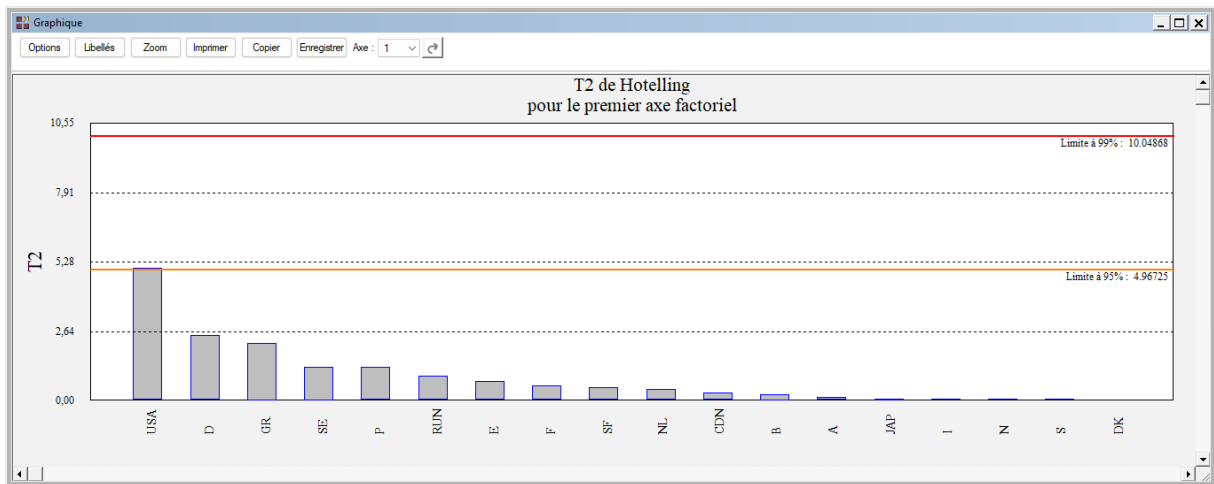
Elles mettent en évidence les individus qui contribuent fortement à la définition des axes factoriels, qui sont bien projetés dans les plans factoriels et qui sont éloignés de l'origine.





Les sixième et septième options affichent les T2 de Hotelling ainsi que les limites à 95% (individu suspect) et 99% (individu extrême).





Une rapide interprétation des résultats

Cet exemple montre que le premier axe factoriel oppose les variables produit national brut, importations, exportations et nombre de téléviseurs aux variables produit intérieur brut agricole et pourcentage de la population dans l'agriculture. Cela peut être analysé comme un facteur de revenu par tête. Du point de vue des pays, il oppose des pays comme la Grèce et les USA. Le deuxième facteur oppose les recettes et les dépenses d'éducation aux réserves et à la population active. Il oppose des pays comme Suède, Norvège et USA, Allemagne. Ce facteur peut s'interpréter comme un facteur niveau des dépenses publiques.

Exemple 3 : Fichier LEVURE

Le pain que nous consommons aujourd'hui est le résultat d'une évolution vieille de 5000 ans.

Cet héritage ancestral est le fruit de la découverte d'un processus pouvant faire lever la pâte.

Plusieurs civilisations, les Egyptiens, les Hébreux, les Grecs, et plus tard les Gaulois et les Ibères, fabriquaient des produits alimentaires fermentés. Le pain était obtenu par des processus empiriques.

Ce n'est qu'au 19^e siècle que les progrès de la science révélèrent les secrets du pouvoir de la levure. C'est le chimiste français Louis Pasteur qui, entre 1857 et 1863, prouva que la fermentation était provoquée par des micro-organismes vivants. Ces contaminants naturels des grains et des fruits ont été identifiés comme étant des champignons microscopiques.



L'industrie de la levure avait cependant démarré en Autriche et en Angleterre. Les progrès décisifs se firent au Danemark et en Allemagne entre 1910 et 1920. Dans la dernière décennie du 20^e siècle, la levure de boulangerie est produite dans le monde sur un rythme de 2,5 millions de tonnes par an. C'est la production de micro-organismes la plus importante qui soit, en raison des énormes progrès techniques et scientifiques que cette industrie a su exploiter ou développer. Elle aura aussi nourri, par ses procédés innovants, toutes les industries de fermentation qui caractérisent nos biotechnologies. Production d'enzymes, d'acides aminés, de vitamines, ou encore de molécules d'intérêt thérapeutiques : hormones, antibiotiques, vaccins...

Les données de ce fichier Levure proviennent de la société Jästbolaget (Suède).

Pendant les quatorze heures nécessaires à la fabrication d'un lot de levure, il est mesuré chaque dix minutes diverses caractéristiques liées au procédé (nos variables). Il y a donc quatre-vingt-trois observations temporelles de ces caractéristiques pour chacun des lots. Trente-trois lots sont collectés.

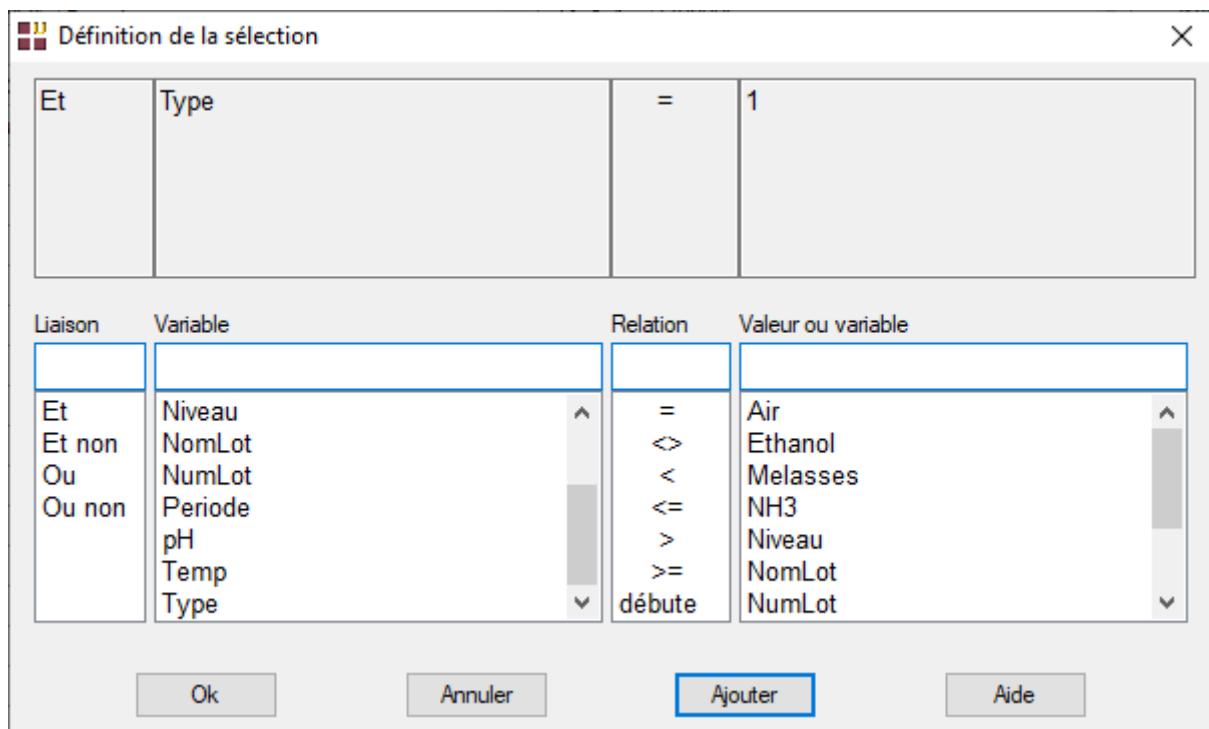
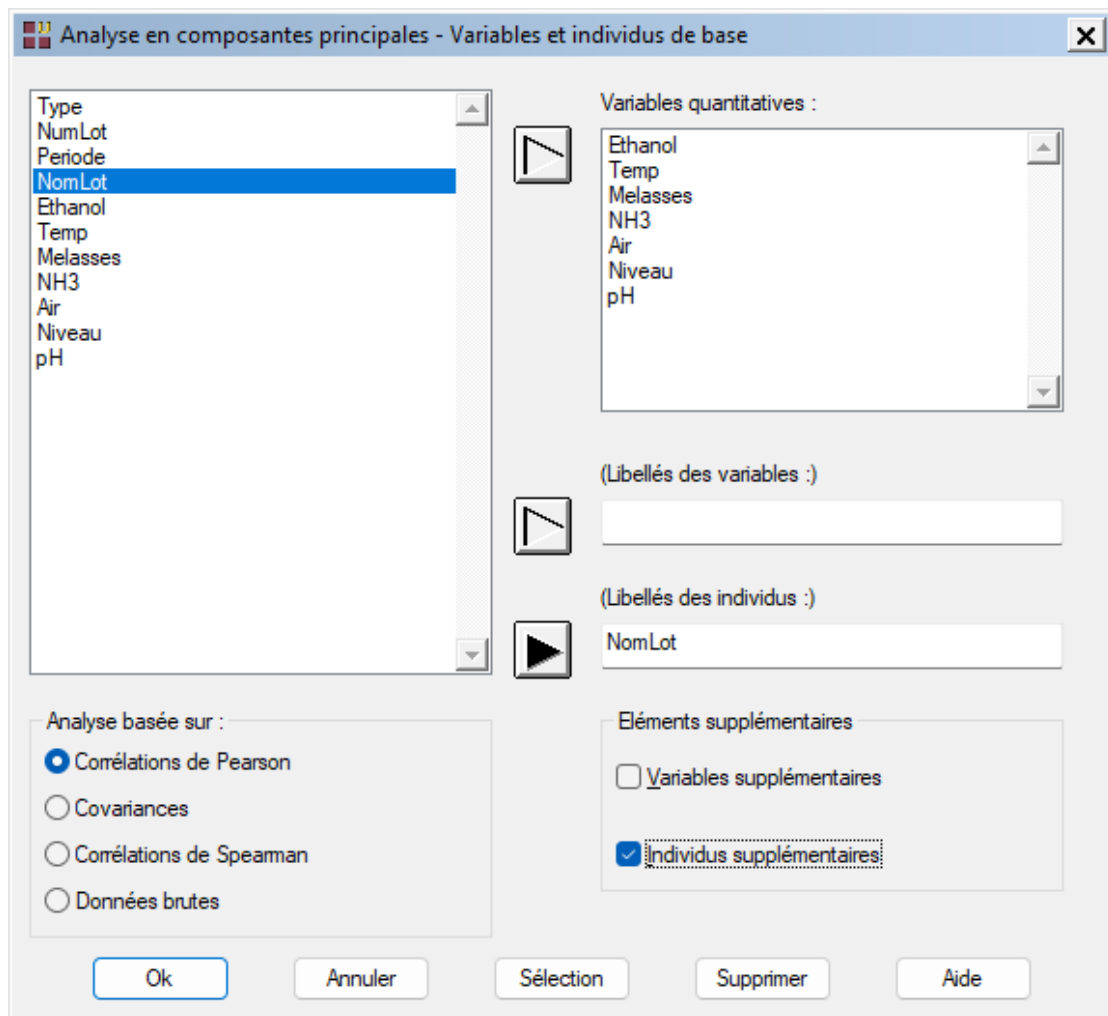
L'objectif est d'aider au pilotage du procédé (MSP) en analysant les trajectoires temporelles des différents lots en mettant en œuvre une analyse multivariée de l'ensemble des caractéristiques (MSP multivariée).

Chaque lot possédant 83 observations, il y a un total de 2.739 individus dans notre analyse.

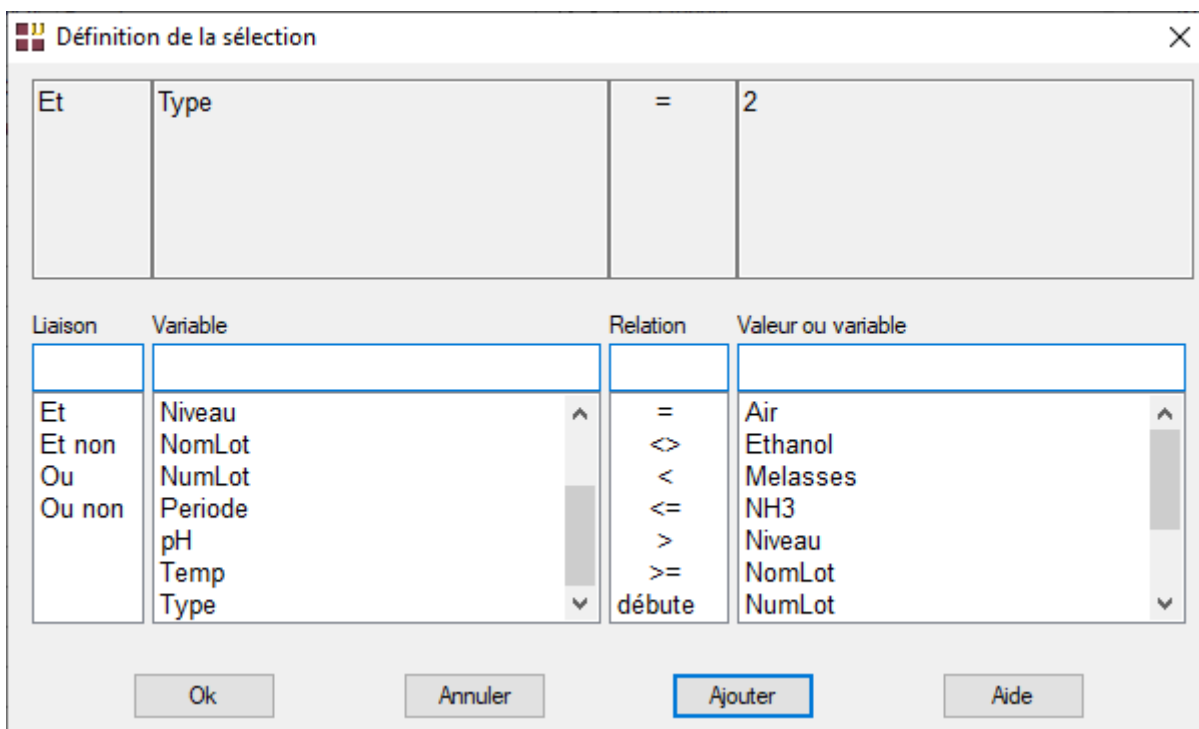
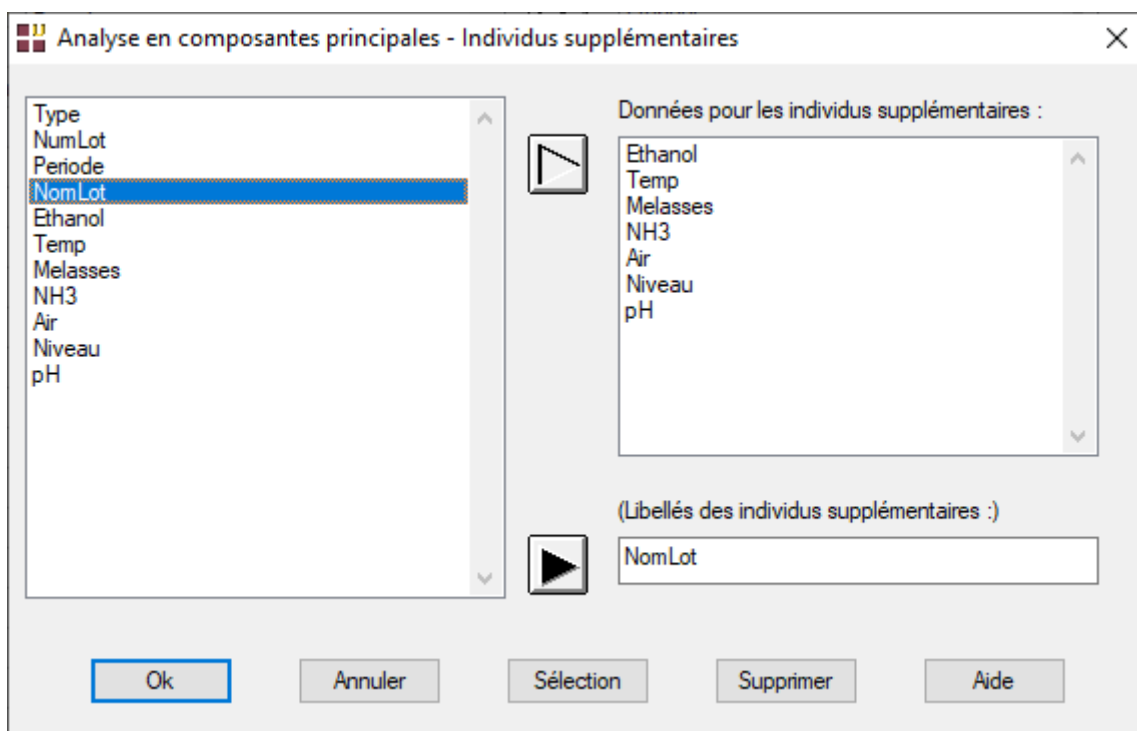
Pour chaque individu, on mesure 7 caractéristiques : Ethanol, Température (Temp), Ammoniac (NH₃), Air, pH, Alimentation en Mélasses (sous-produit du raffinage du sucre de la canne ou de la betterave) et Niveau du réservoir (Niveau).

L'analyse initiale porte sur les 20 lots de référence ou « bons lots » (Type =1) parmi les 33 lots.

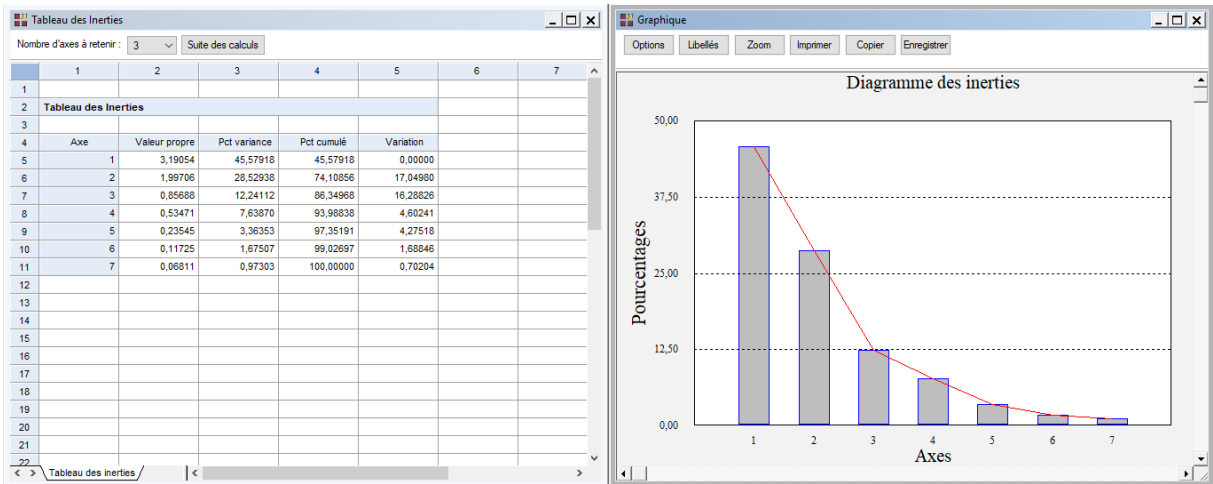
Cliquons sur l'icône ACP ou choisissons Analyse en Composantes Principales dans le ruban Décrire et renseignons la première boîte de dialogue comme montré ci-après.



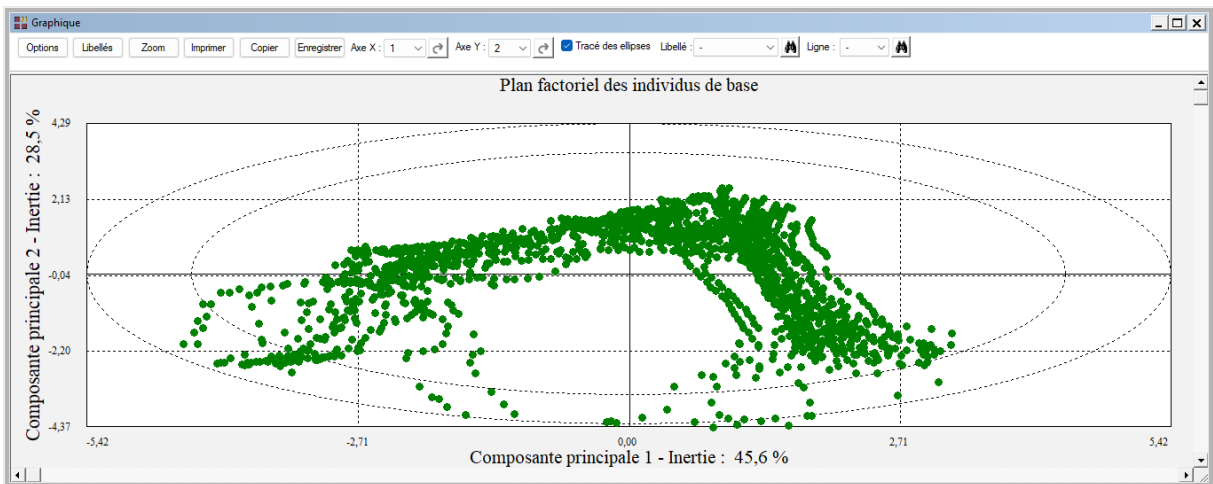
Renseignons la boîte de dialogue des individus supplémentaires (les 13 nouveaux lots).



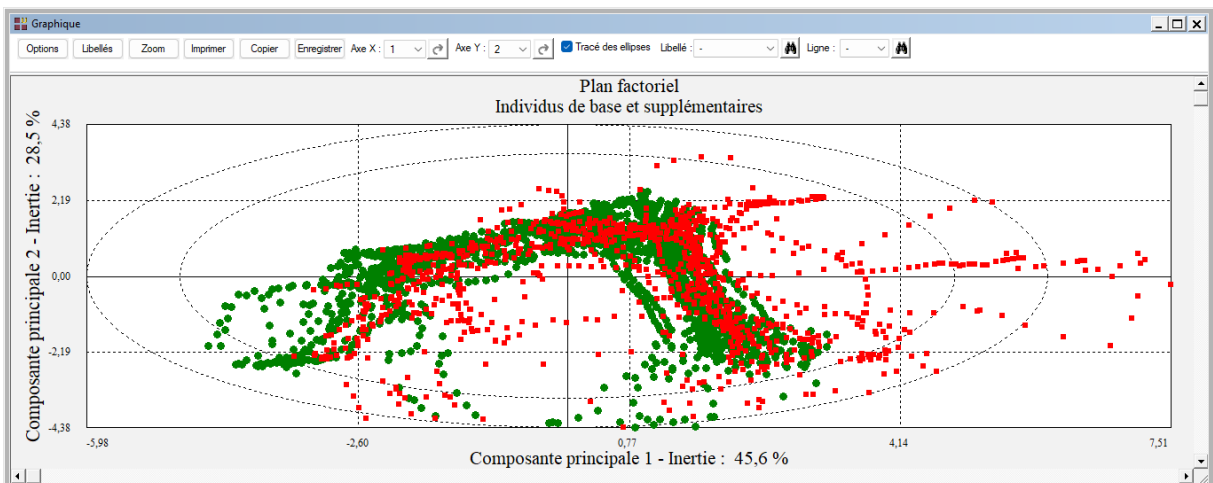
Visualisons le diagramme des inerties. Celui-ci nous indique que 2 composantes principales semblent suffisantes.



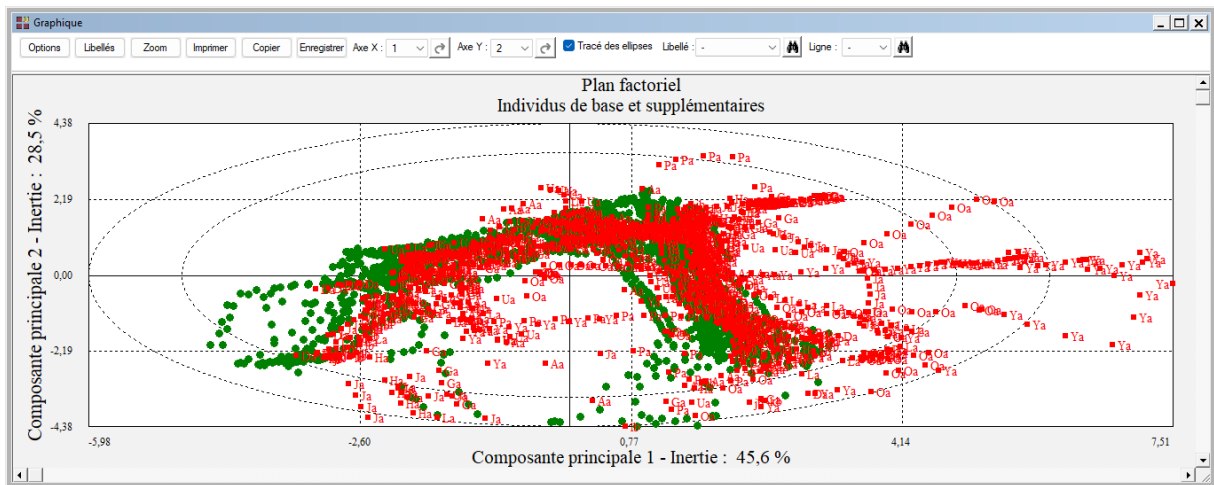
Demandons le plan factoriel 1-2 des individus de base avec les ellipses de confiance. Presque toutes les données de base sont à l'intérieur les ellipses.



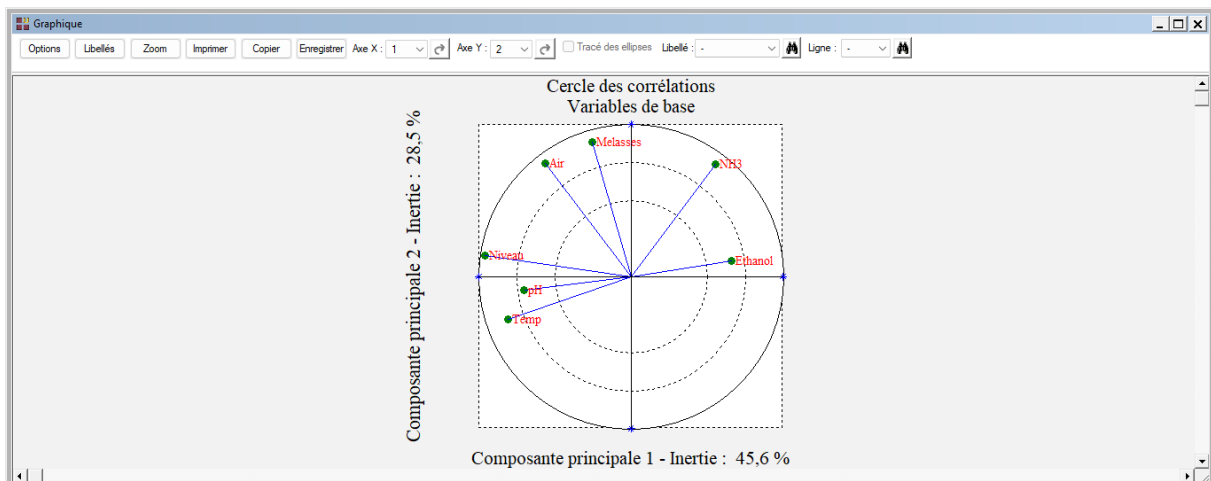
Demandons le plan factoriel 1-2 des individus de base (vert) et supplémentaires (rouge) avec les ellipses de confiance.



Ce graphique montre que plusieurs des nouveaux lots ne suivent pas les trajectoires des lots de référence. Ce sont principalement les lots Oa, Pa et Ya.

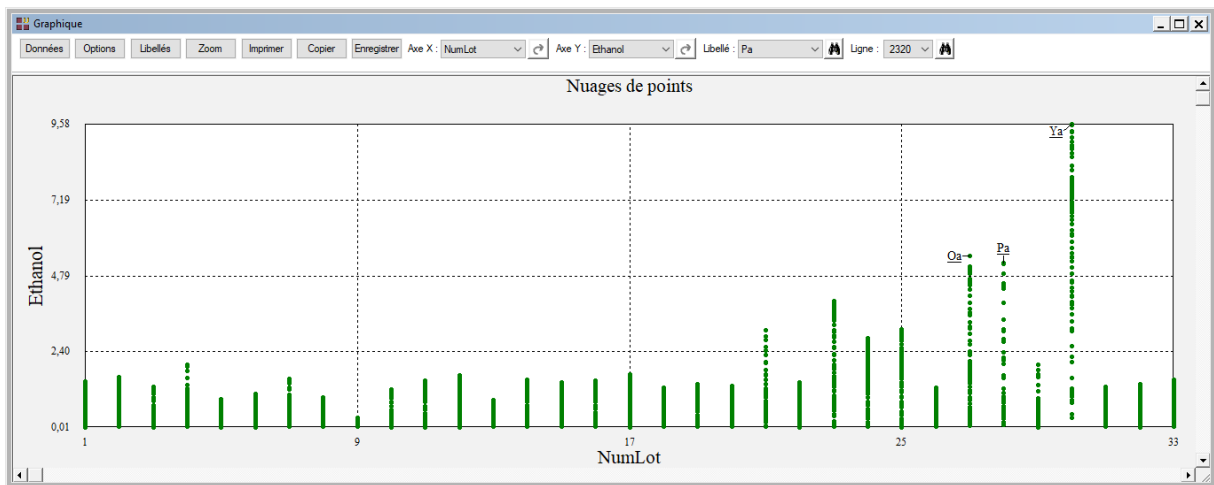
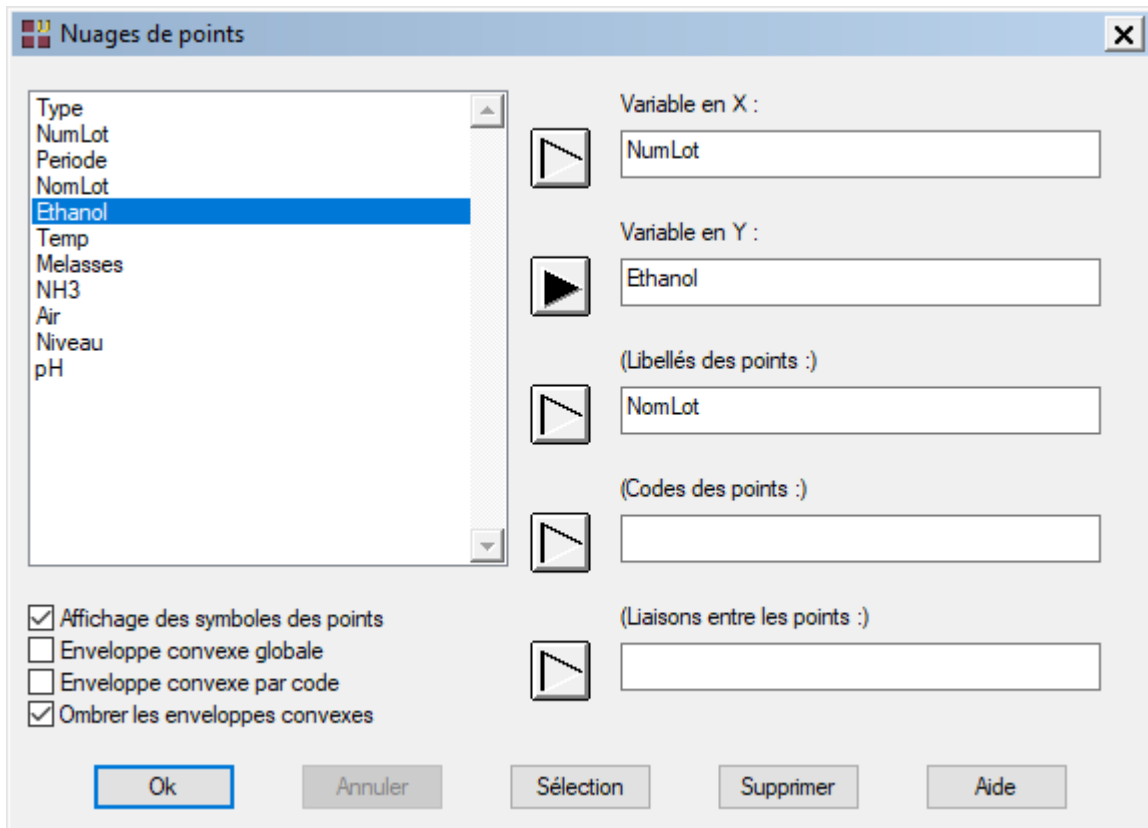


Affichons le cercle des corrélations.



Celui-ci nous indique que les possibles variables influentes sont NH3 et Ethanol.

Un rapide affichage de ces deux variables par l'option 'Graphique – Nuages de points' indique que la variable Ethanol est beaucoup plus importante pour ces trois lots.



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.

<i>Variable</i>	<i>Contenu</i>
varquantibase	Libellés des variables quantitatives de base
varquantisup	Libellés des variables quantitatives supplémentaires
varqualisup	Libellés des variables qualitatives supplémentaires
modvarqualisup	Libellés des modalités des variables qualitatives supplémentaires
indivbase	Libellés des individus de base
indivsup	Libellés des individus supplémentaires
movarbase	Moyennes des variables quantitatives de base
ectvarbase	Ecart-types des variables quantitatives de base
cvvarbase	Coefficients de variation des variables quantitatives de base
corpearsonbase	Matrice des corrélations de Pearson des variables de base
corspearmanbase	Matrice des corrélations des rangs de Spearman des variables de base
covarbase	Matrice des covariances des variables de base
numaxe	Numéros des axes
valpro	Valeurs propres
pctvar	Pourcentages de variance expliquée
pctvcum	Pourcentages cumulés de variance expliquée
variation	Variations de la variance expliquée
distvarbase	Distances carrées à l'origine (variables de base)
cosvarbase	Cosinus carrés (variables de base)
contribvarbase	Contributions à l'inertie (variables de base)
distindivbase	Distances carrées à l'origine (individus de base)
t2indivbase	T2 de Hotelling (individus de base)
cosindivbase	Cosinus carrés (individus de base)
contribindivbase	Contributions à l'inertie (individus de base)
contribtotindivbase	Contributions à l'inertie totale (individus de base)
facteur	Facteurs colonnes
composante	Composantes principales
moyvarqualisup	Moyennes des variables qualitatives supplémentaires
coorvarbase	Coordonnées des variables de base
correlsimplevarbase	Corrélations simples des variables de base
composantesup	Coordonnées des individus supplémentaires
coorvarsup	Coordonnées des variables supplémentaires
correlsimplevarsup	Corrélations simples des variables supplémentaires