UNIWIN VERSION 10.7.0

ANALYSE EN COMPOSANTES PRINCIPALES

Révision : 26/08/2026

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 18
Combiner ACP et Classification Ascendante Hiérarchique 20
Exemple 2 : Fichier OCDE (variables et individus de base et supplémentaires) 22
L'option Rapports 28
L'option Graphiques 29
Une rapide interprétation des résultats 38
Exemple 3 : Fichier LEVURE 38
Les variables internes créées par la procédure 44

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.

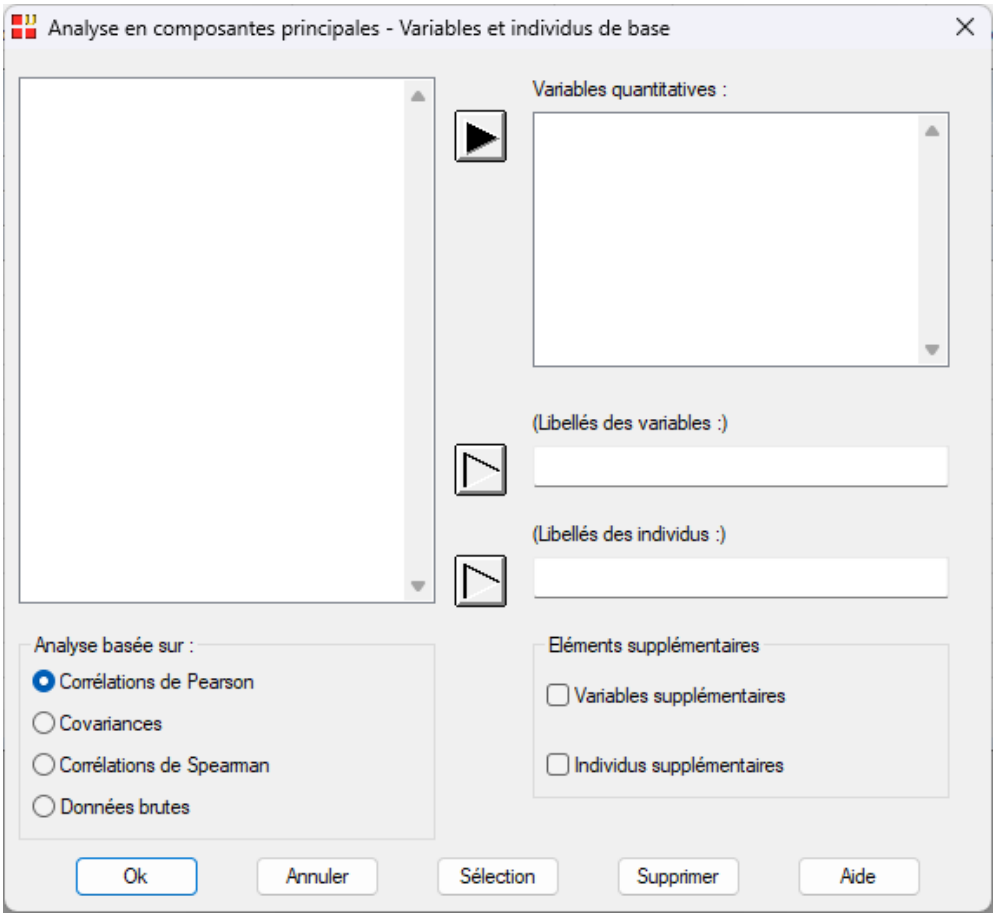
L'analyse peut être réalisée en utilisant la matrice des corrélations de Pearson ou de Spearman, la matrice des covariances ou les données brutes.

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.

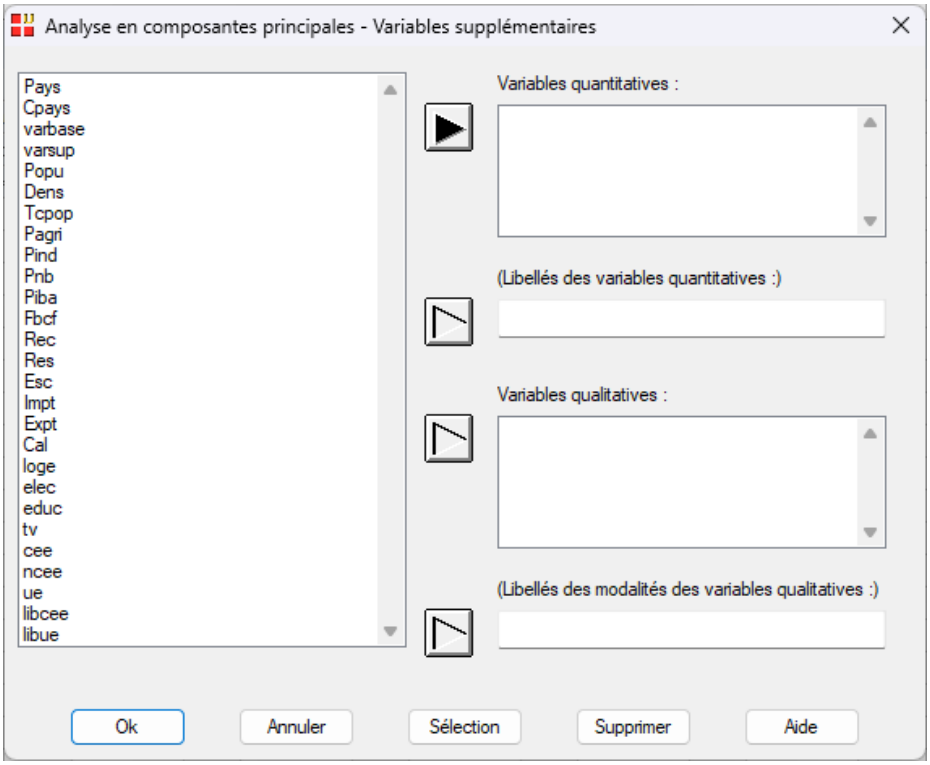
Elle permet également, en option, d'indiquer les noms des variables contenant les libellés des variables quantitatives et les libellés des individus.

L'analyse peut être basée sur la matrice des corrélations de Pearson (option par défaut), des covariances, des corrélations de Spearman ou sur les données brutes.

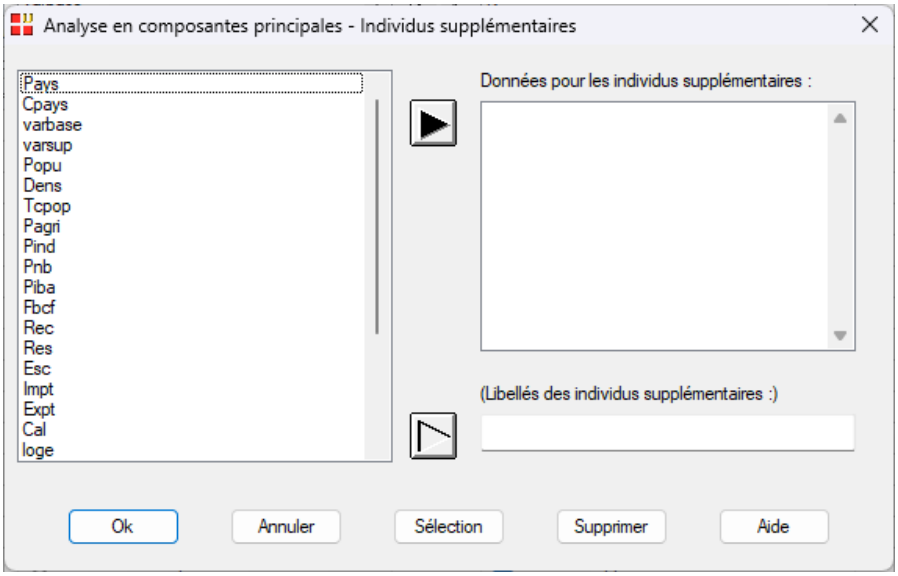
Enfin, vous pouvez cocher l'une ou les deux cases indiquant si vous avez des variables supplémentaires (passives) et/ou des individus supplémentaires (passifs) à étudier.

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 débuter 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 :

Pays	*Code Pays*	*Membre CEE*	*Membre UE*
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

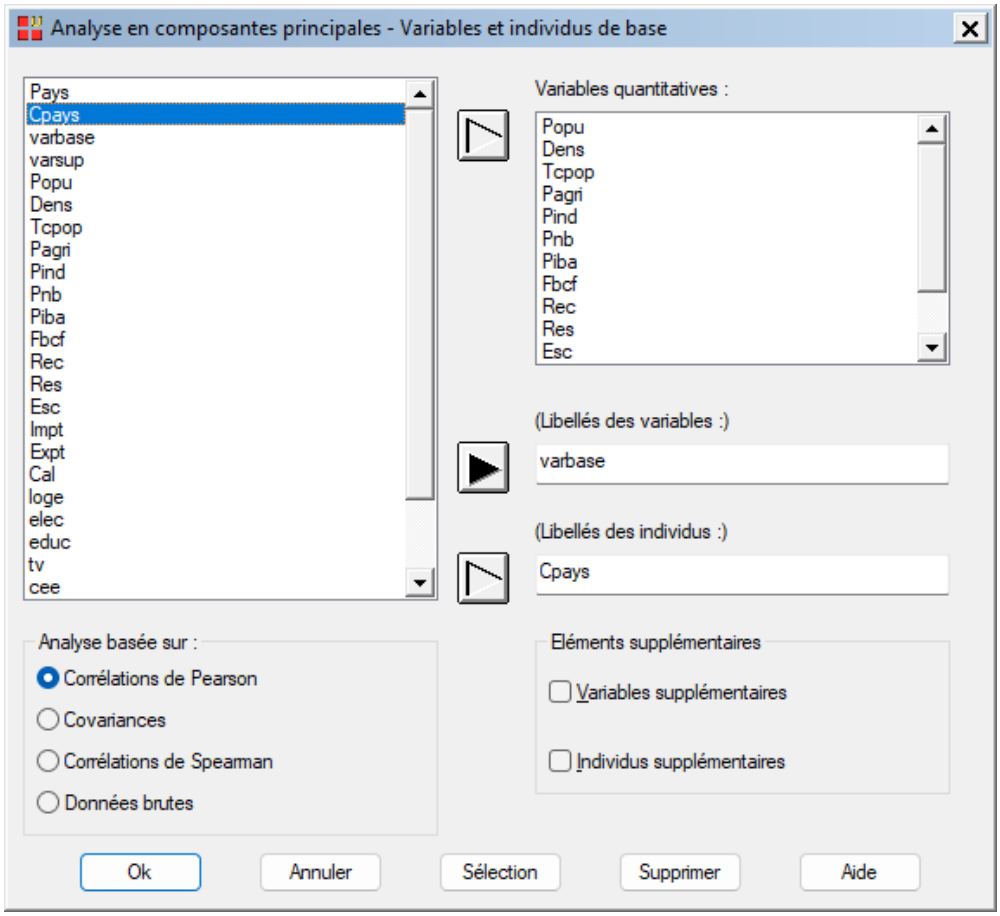
Les informations suivantes ont été collectées :

Information	*Variable*
Nom du pays (libellés longs)	Pays
Code du pays (libellés courts)	Cpays
Population (en milliers)	Popu
Densité de population (au km2)	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.

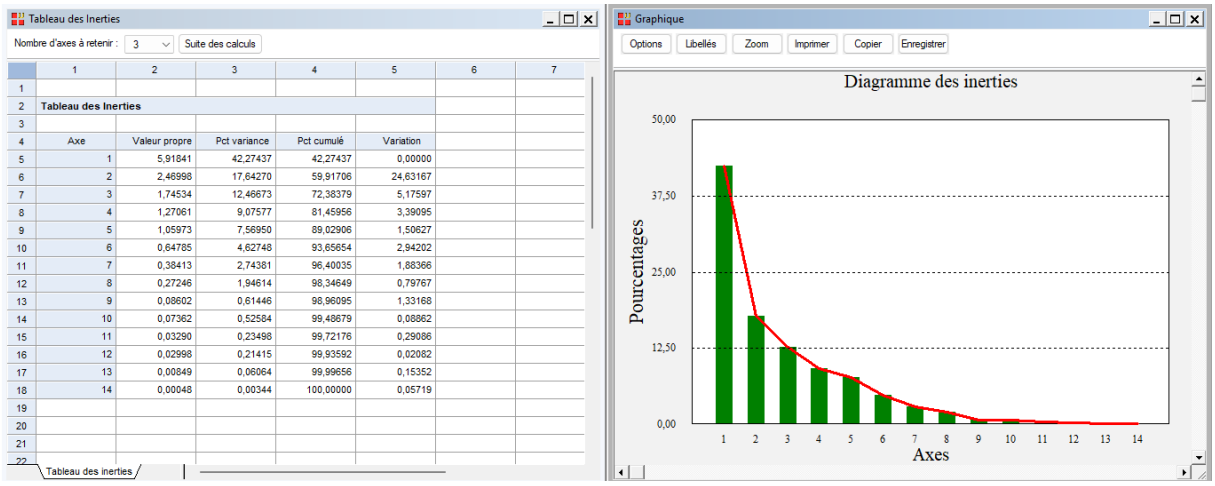
La première boîte de dialogue montrée ci-après apparaît.



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.

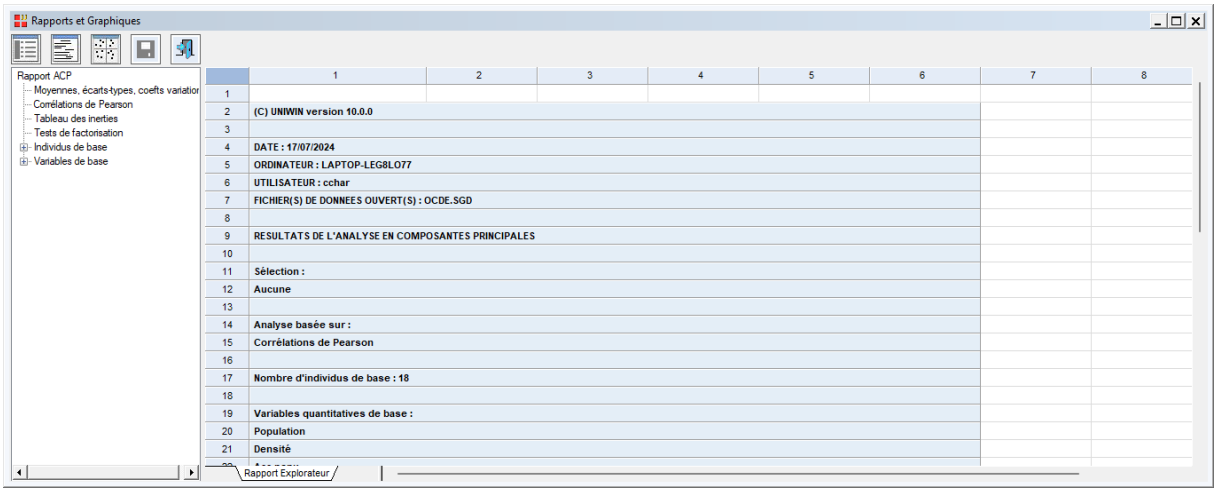
Cliquons sur Ok pour exécuter le traitement de l'analyse.

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.



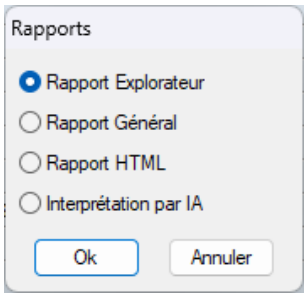
L'option 'Nombre d'axes à retenir' permet de préciser le nombre de composantes principales à extraire.

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

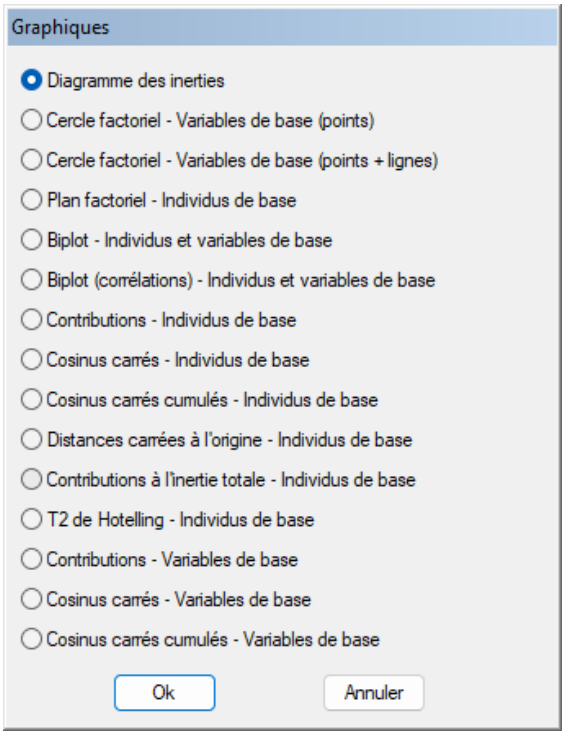


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données' de rappeler la boîte de dialogue d'entrée des données.

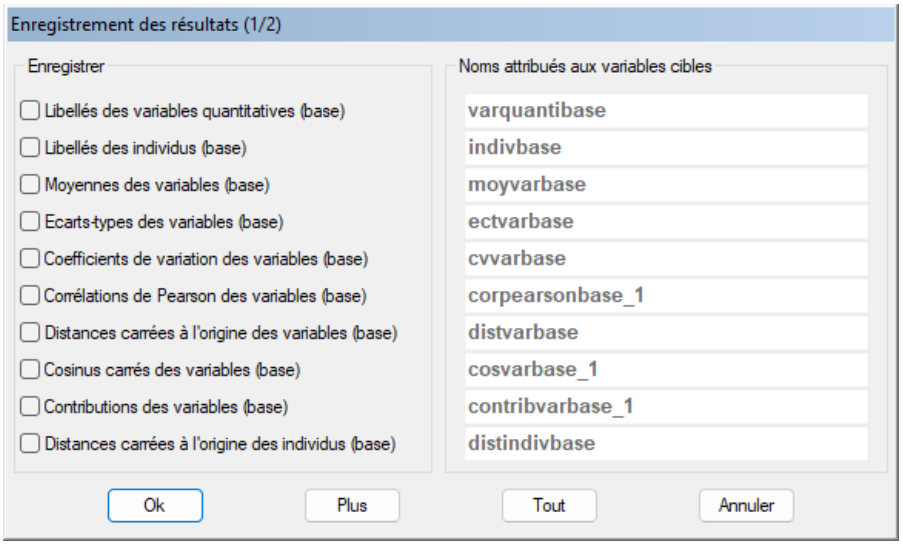
L'icône 'Rapports' affiche la boîte de dialogue des options pour les rapports :



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



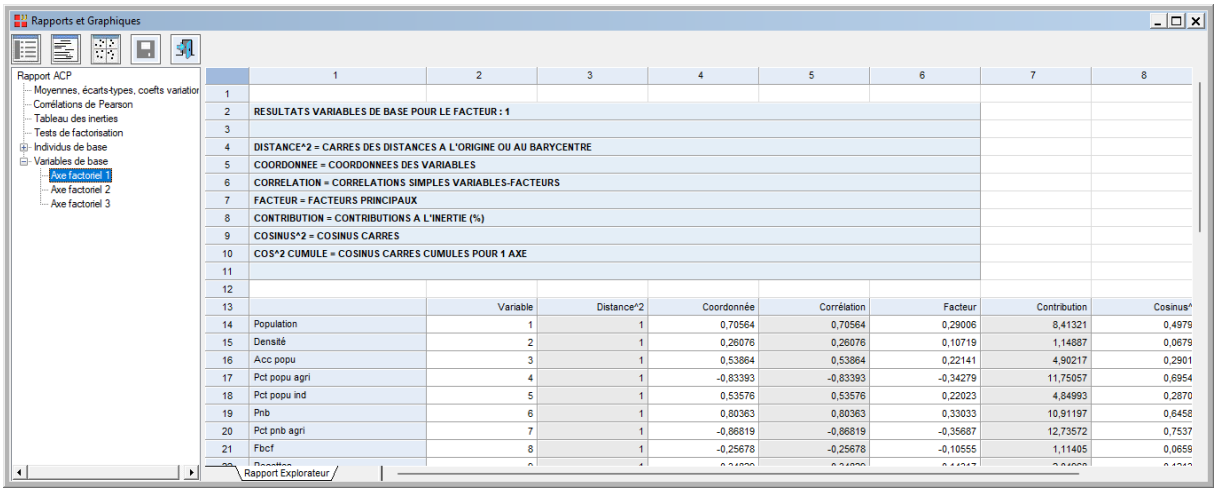
L'icône 'Enregistrer' permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

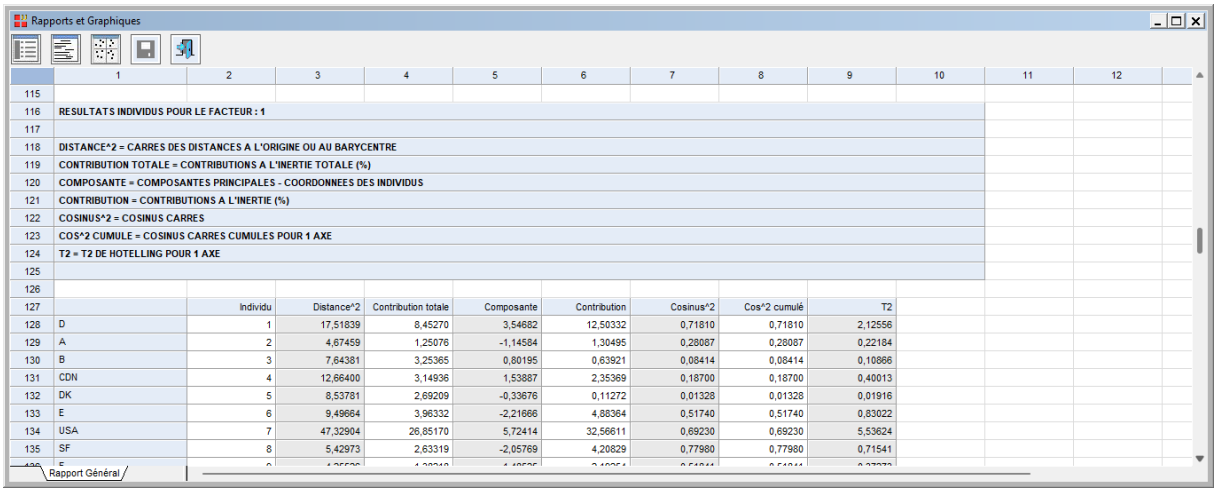


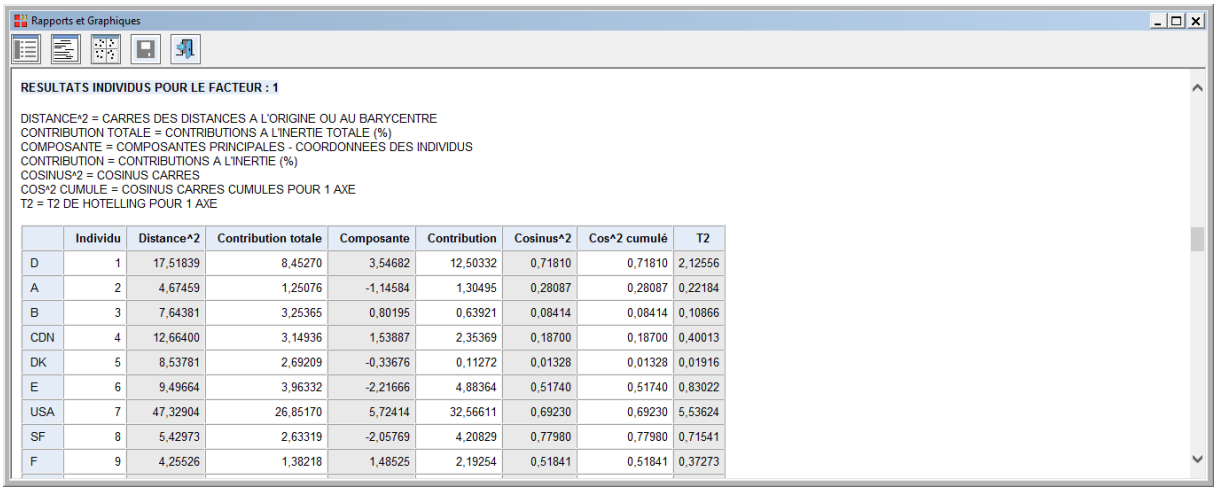
L'option Rapports

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

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





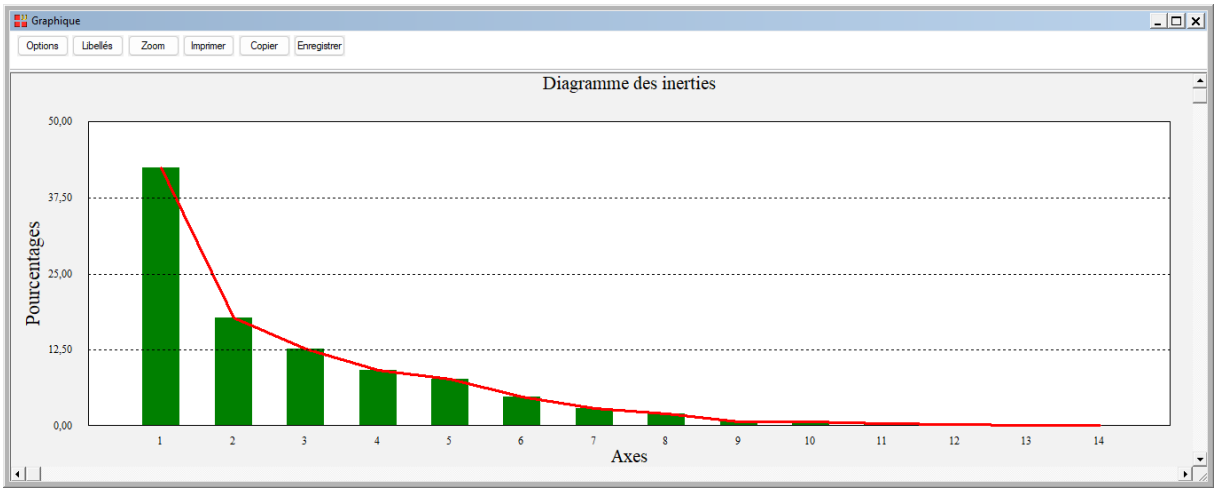


L’option Graphiques

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

- Diagramme des inerties

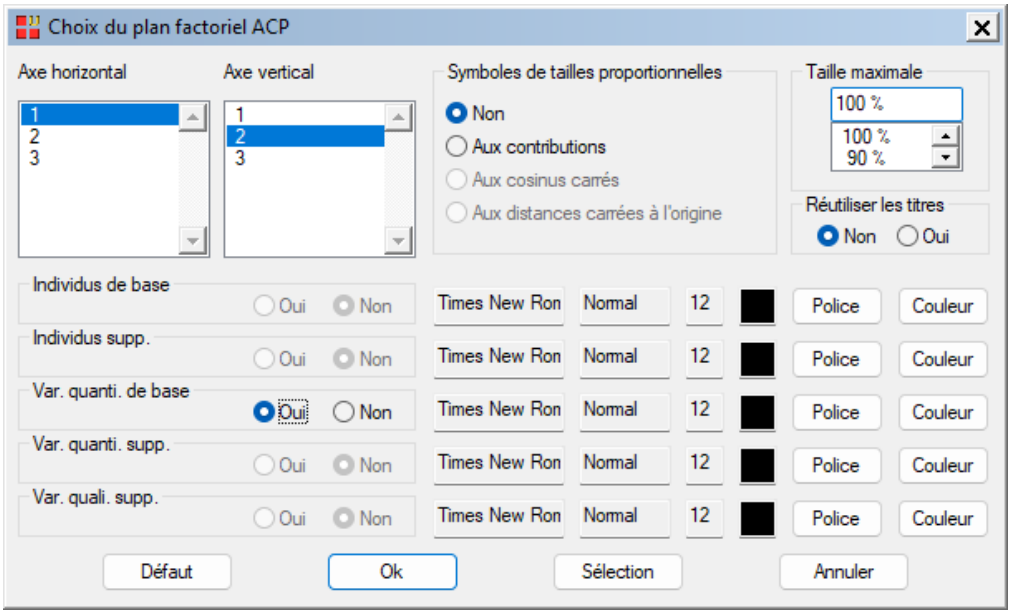
Ce graphique affiche les pourcentages d’inertie pour chacun des axes factoriels.



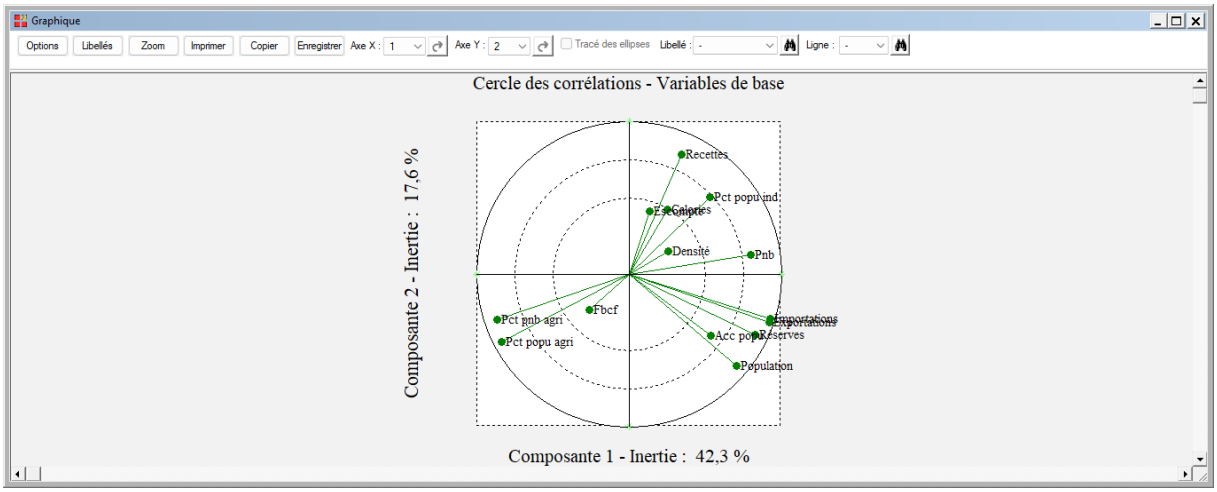
- Les options Cercle factoriel

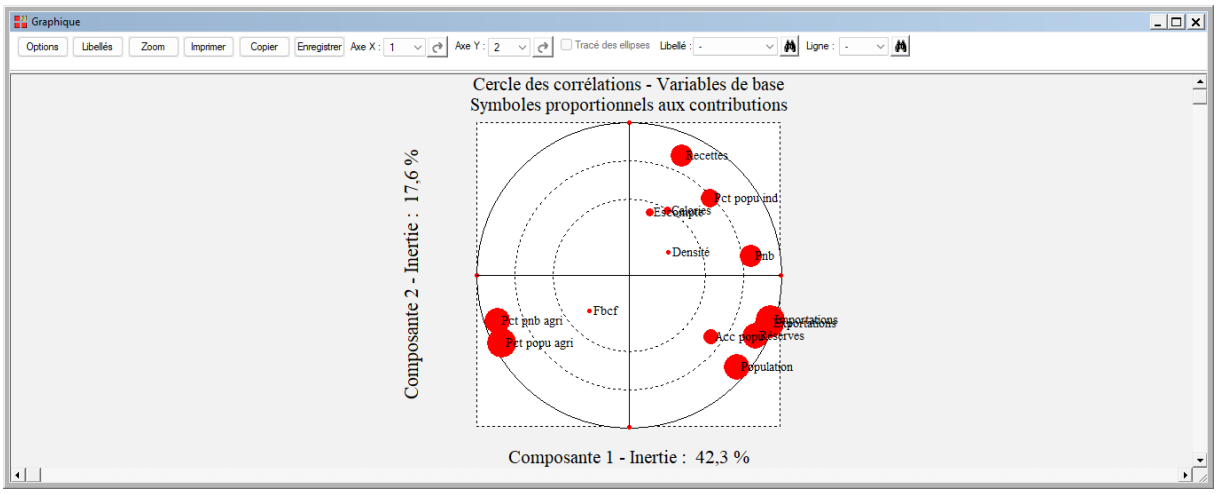
Ces options permettent d’afficher le cercle de corrélations des variables de base et de choisir si on désire tracer les lignes reliant les points à l'origine du cercle. L'option sans ces lignes (points) est utile lorsqu'il y a un grand nombre de variables représentées. Choisissons les variables de base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

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



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. Elle permet également de préciser si les symboles des points doivent être proportionnels aux contributions des variables 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.



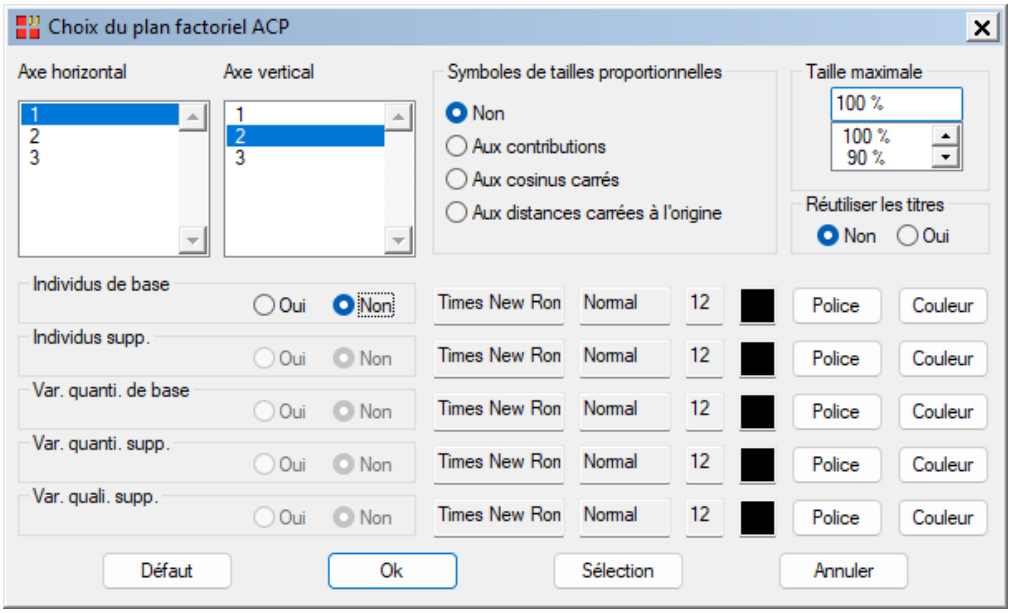


- L'option Plan factoriel

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

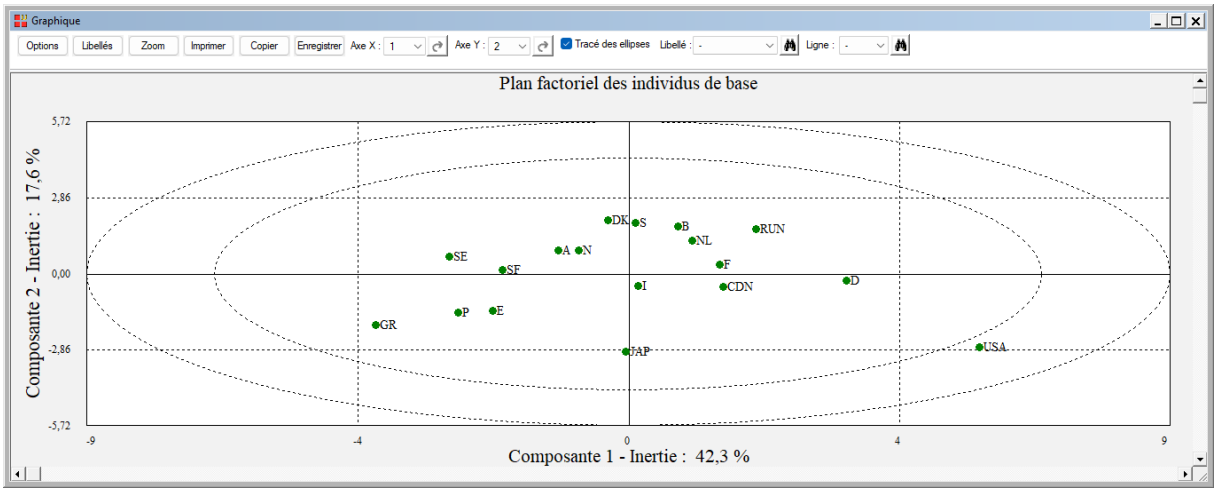
Une boîte de dialogue permettant de choisir le plan factoriel s'affiche alors.

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.

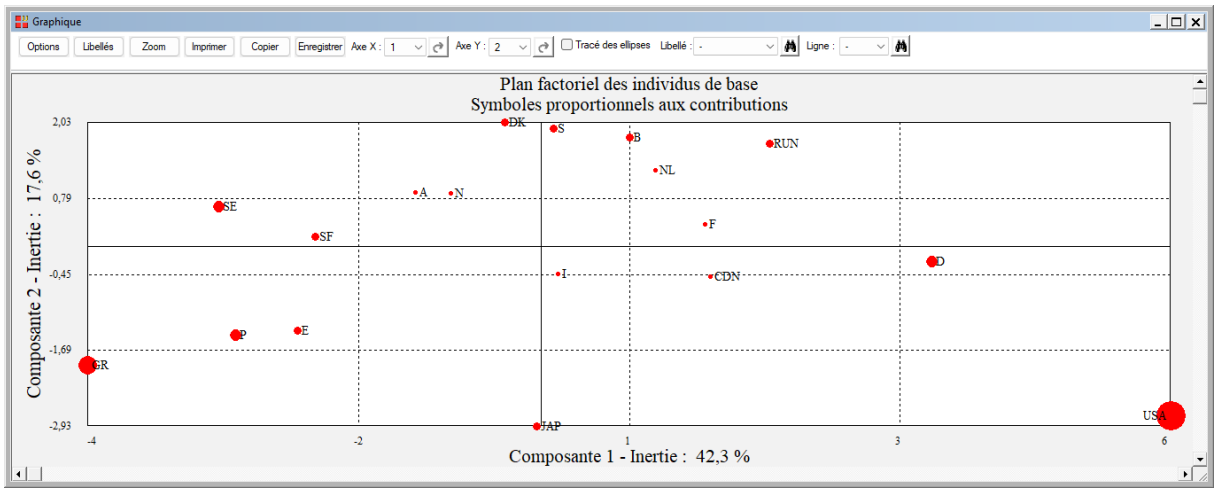


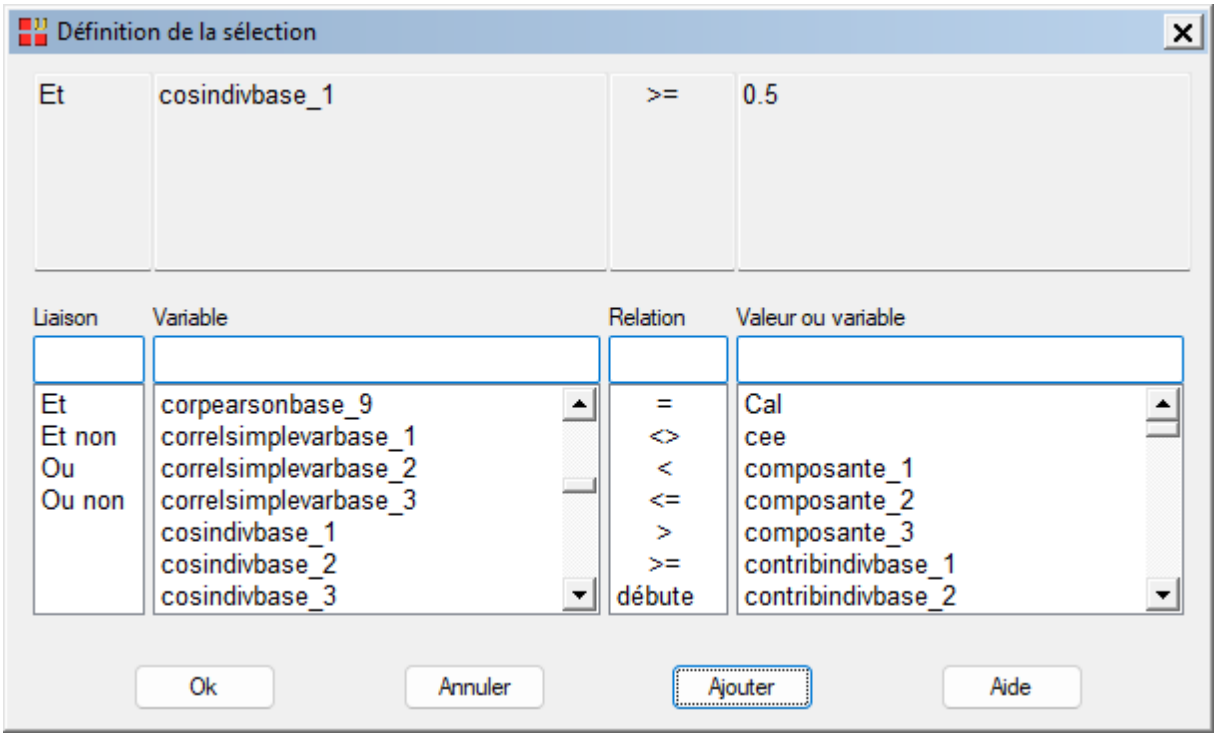
Affichons ce même graphique avec des symboles de points proportionnels aux contributions.

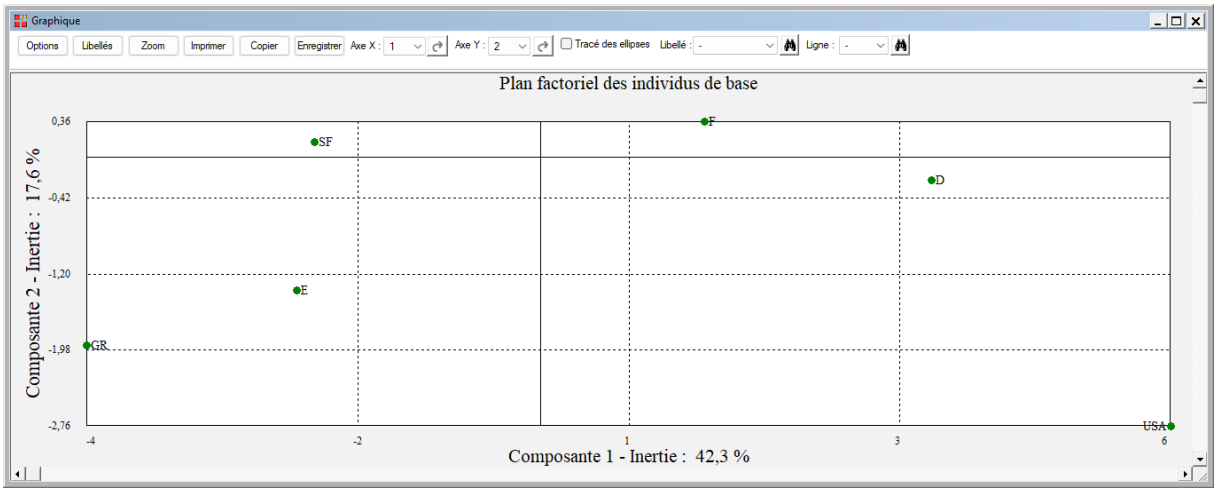
A noter le bouton 'Sélection' qui permet de n'afficher que les points qui satisfont une condition logique, par exemple une condition basée sur les contributions ou les cosinus carrés. Ceci est très pratique dans le cas d'un nuage comportant beaucoup d'individus : seuls les points significatifs d'un point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


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe X : 1
Axe Y : 2
Tracé des ellipses
Libellé :
Ligne :
Plan factoriel des individus de base
Symboles proportionnels aux contributions
Composante 2 - Inertie : 17,6 %
Composante 1 - Inertie : 42,3 %
2,03
0,79
-0,45
-1,69
-2,93
-4
-2
1
3
6
DK
S
B
RUN
NL
A
N
SE
F
SF
D
I
CDN
P
E
GR
JAP
USA


Définition de la sélection
Et
cosindivbase_1
>=
0.5
Liaison
Variable
Relation
Valeur ou variable
Et
Et non
Ou
Ou non
corpearsonbase_9
correlsimplevarbase_1
correlsimplevarbase_2
correlsimplevarbase_3
cosindivbase_1
cosindivbase_2
cosindivbase_3
=
<>
<
<=
>
>=
débute
Cal
cee
composante_1
composante_2
composante_3
contribindivbase_1
contribindivbase_2
Ok
Annuler
Ajouter
Aide


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe X : 1
Axe Y : 2
Tracé des ellipses
Libellé :
Ligne :
Plan factoriel des individus de base
Composante 2 - Inertie : 17,6 %
Composante 1 - Inertie : 42,3 %
0,36
-0,42
-1,20
-1,98
-2,76
-4
-2
1
3
6
F
SF
D
E
GR
USA

- 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.

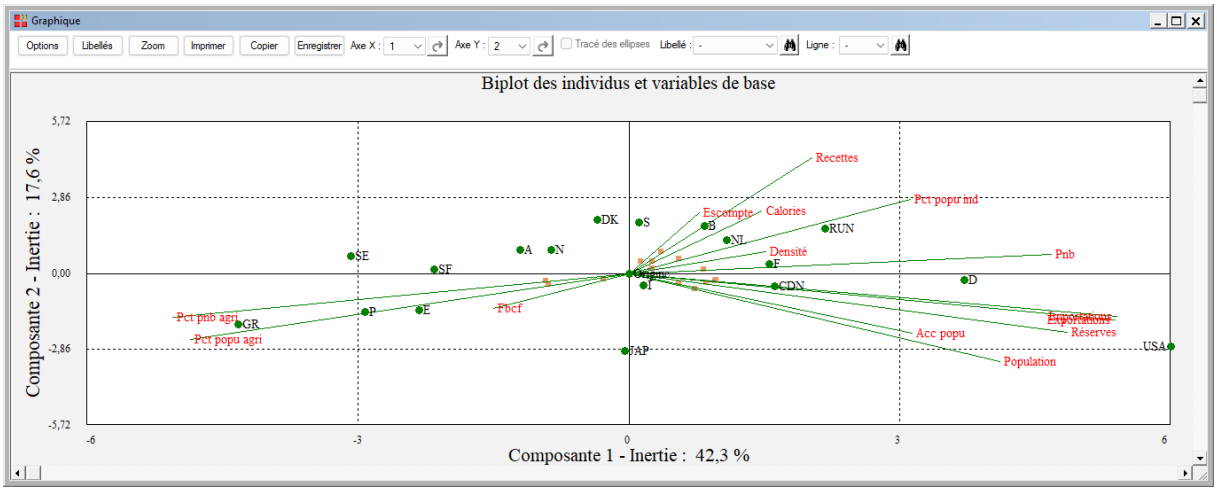
Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Un exemple du Biplot est montré ci-après.

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.

L'option zoom n'est pas active.

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.

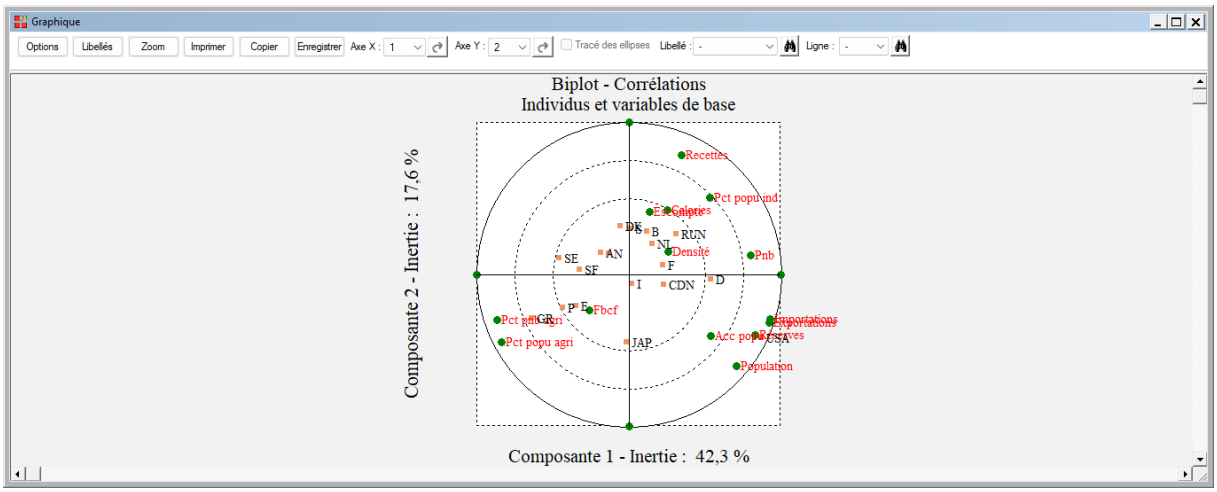
Les variables indicatrices des individus sont représentées comme si elles étaient des variables supplémentaires dans l'analyse.

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.

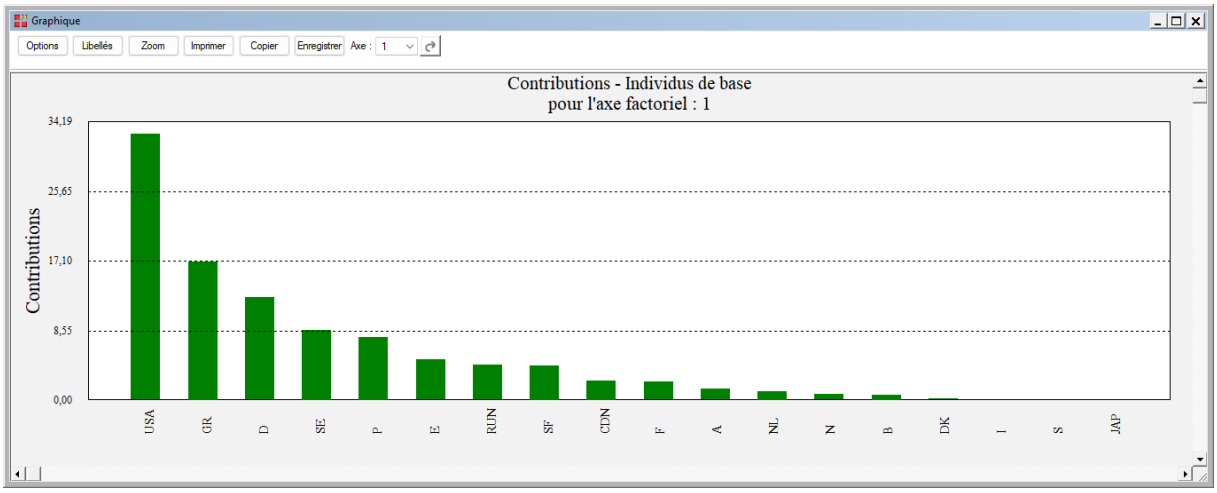


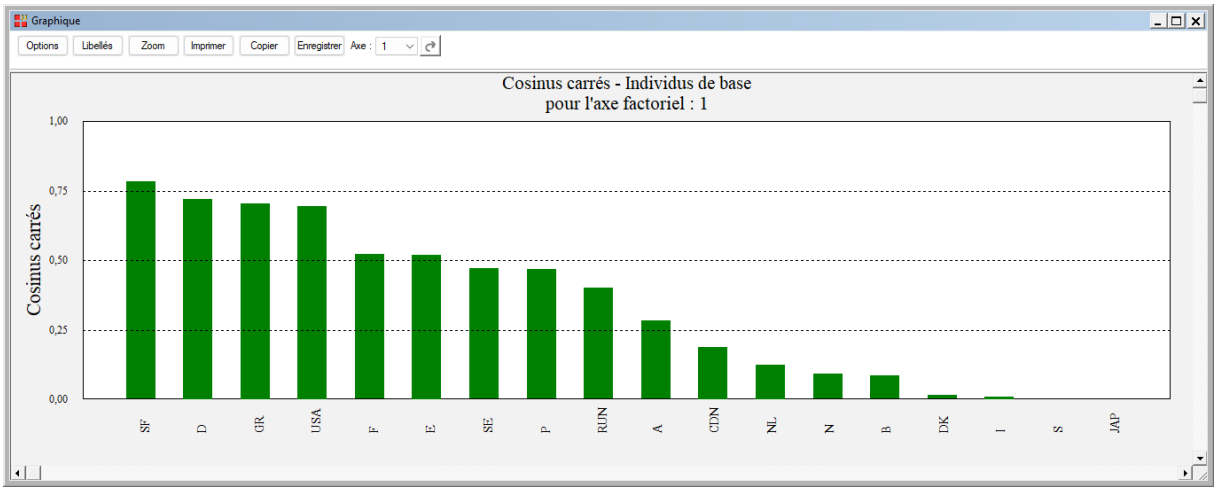
- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés, Distances carrées à l'origine, Contributions à l'inertie totale et T2 de Hotelling pour les individus de base

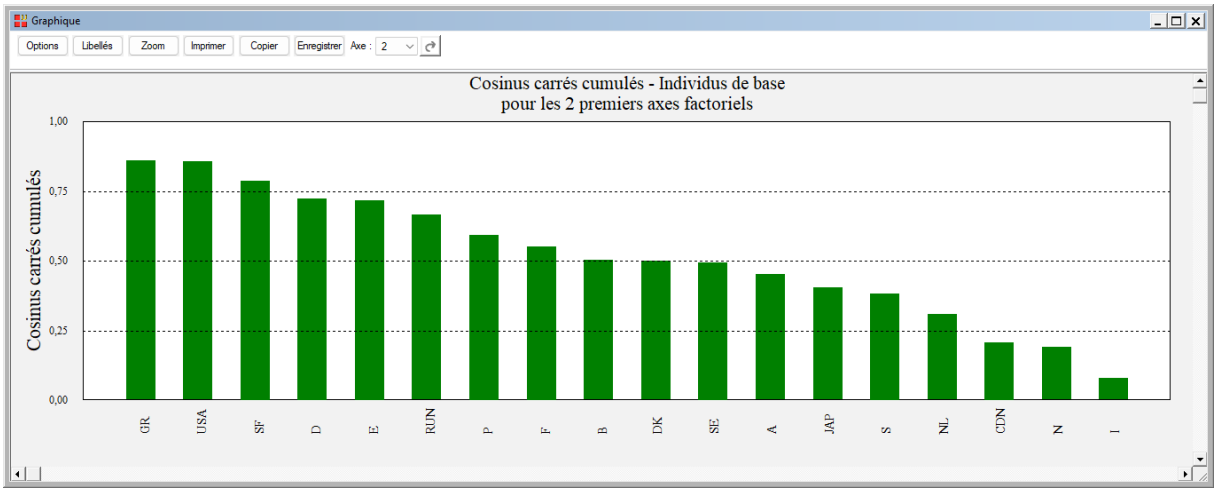
Les cinq premières options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse.

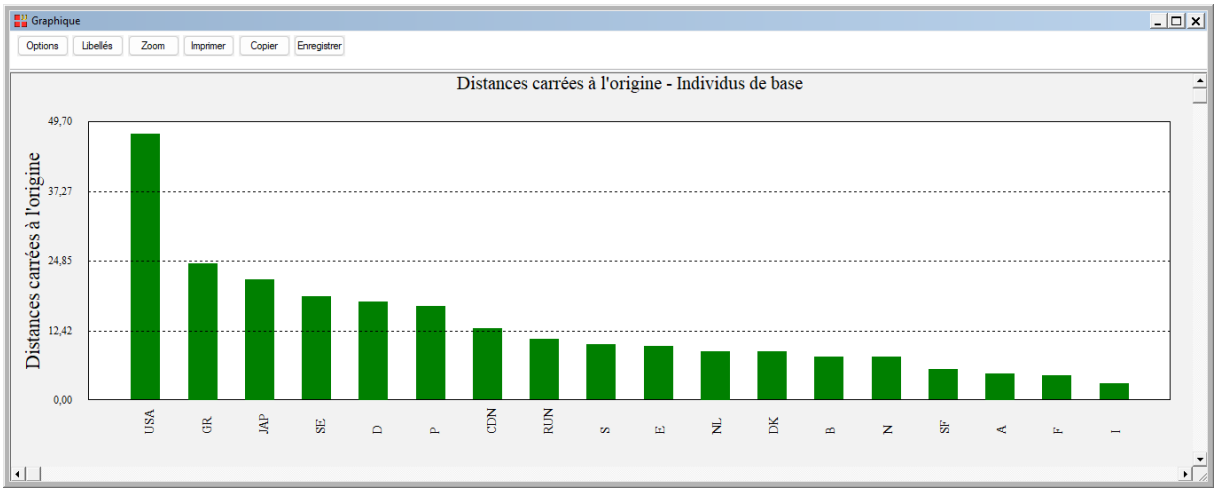
Elles mettent en évidence les individus qui contribuent fortement à la définition des axes factoriels, qui sont bien projetés dans les plans factoriels et qui sont éloignés de l'origine.

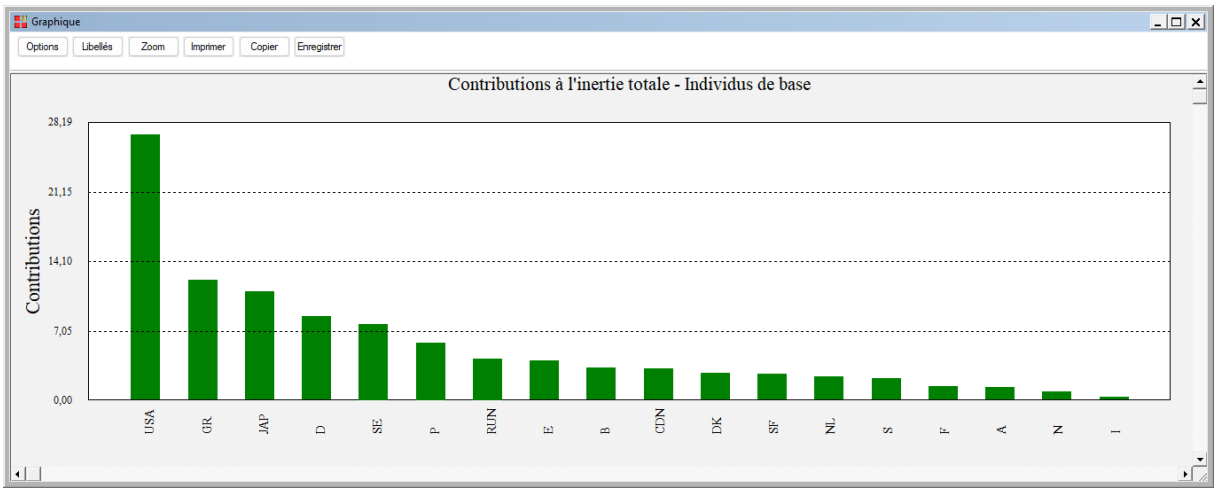
La sixième option affiche les T2 de Hotelling ainsi que les limites à 95% (individu suspect) et 99% (individu extrême).

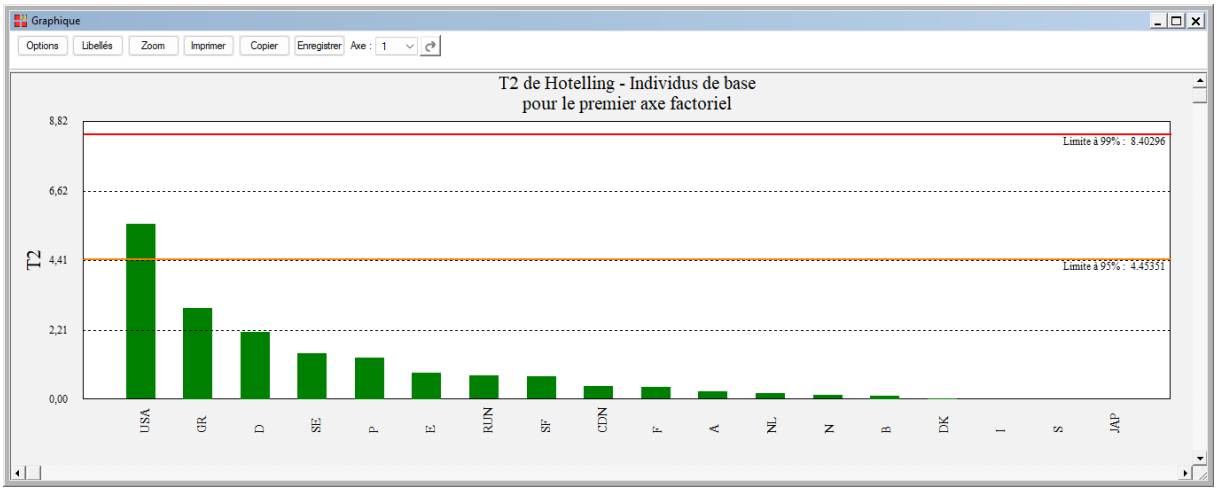



Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe : 1
Cosinus carrés - Individus de base
pour l'axe factoriel : 1
Cosinus carrés
1,00
0,75
0,50
0,25
0,00
SF
D
GR
USA
F
E
SE
P
RUN
A
CDN
NL
N
B
DK
I
S
JAP


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe : 2
Cosinus carrés cumulés - Individus de base
pour les 2 premiers axes factoriels
Cosinus carrés cumulés
1,00
0,75
0,50
0,25
0,00
GR
USA
SF
D
E
RUN
P
F
B
DK
SE
A
JAP
S
NL
CDN
N
I

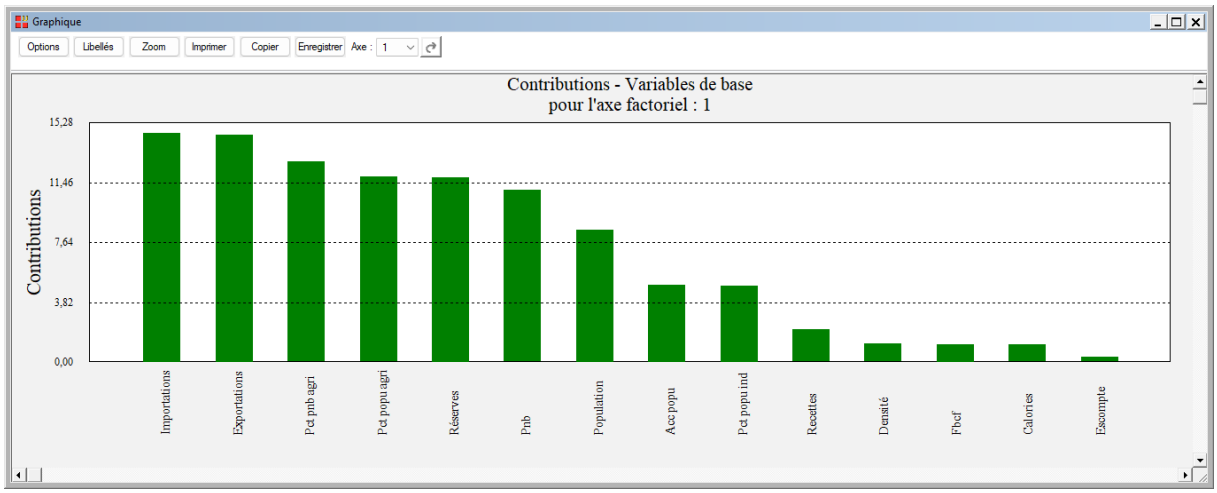

Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Distances carrées à l'origine - Individus de base
Distances carrées à l'origine
49,70
37,27
24,85
12,42
0,00
USA
GR
JAP
SE
D
P
CDN
RUN
S
E
NL
DK
B
N
SF
A
F
I





- Les options Contributions, Cosinus carrés et Cosinus carrés cumulés pour les variables de base

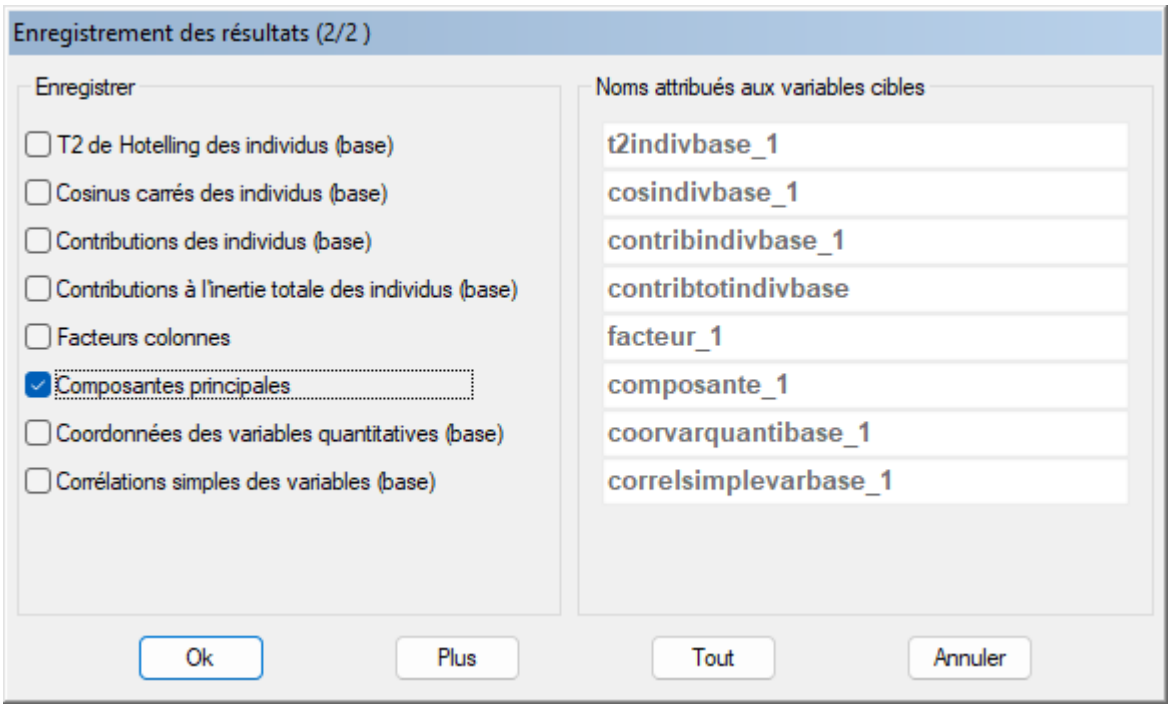
Ces options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse. Elles mettent en évidence les variables qui contribuent fortement à la définition des axes factoriels et qui sont bien projetés dans les plans factoriels.



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.

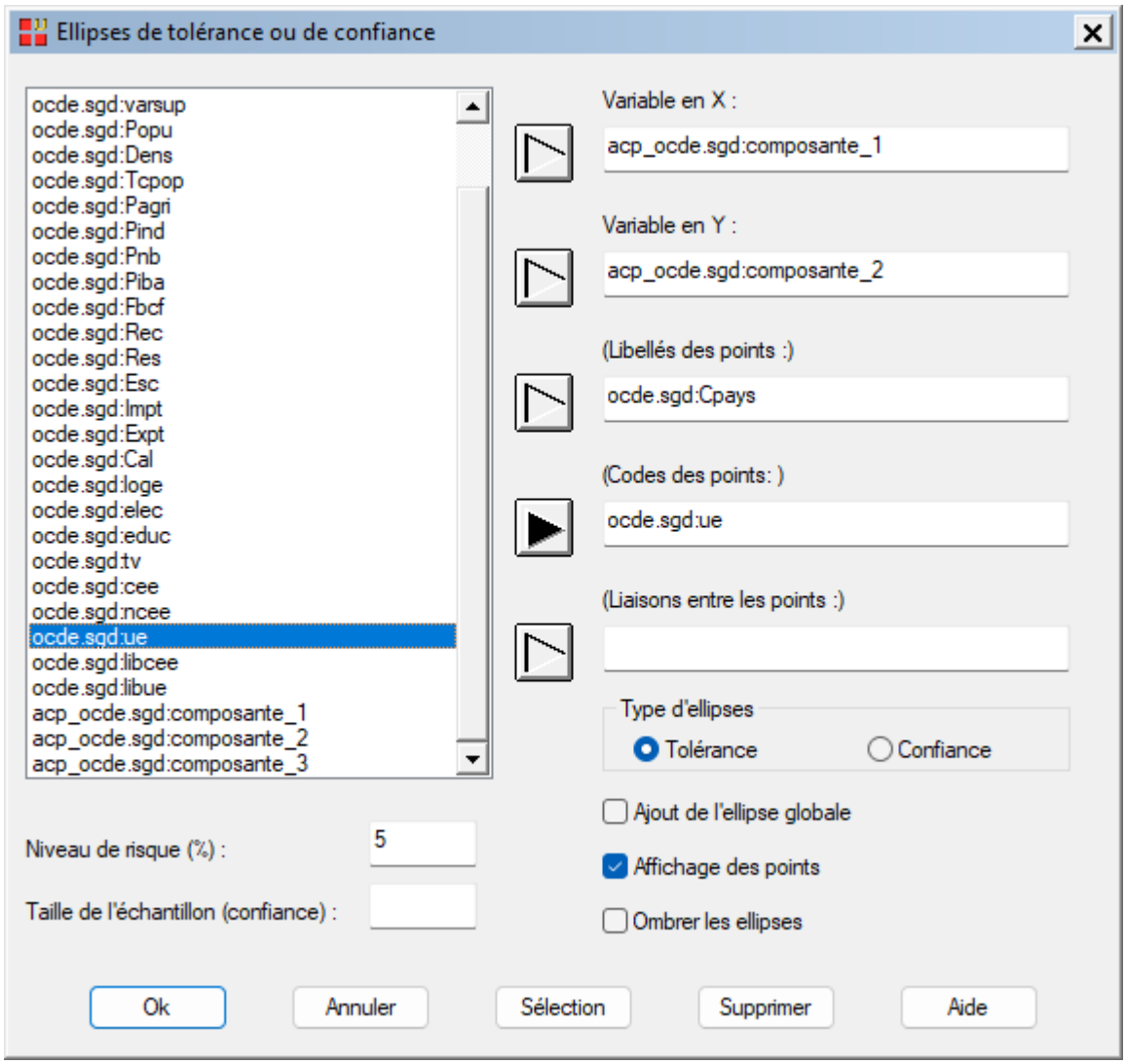
Cliquer sur l'icône ELL dans le ruban Décrire, quitter la procédure en cours et renseigner la boîte de dialogue Ellipses comme montré ci-après.

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

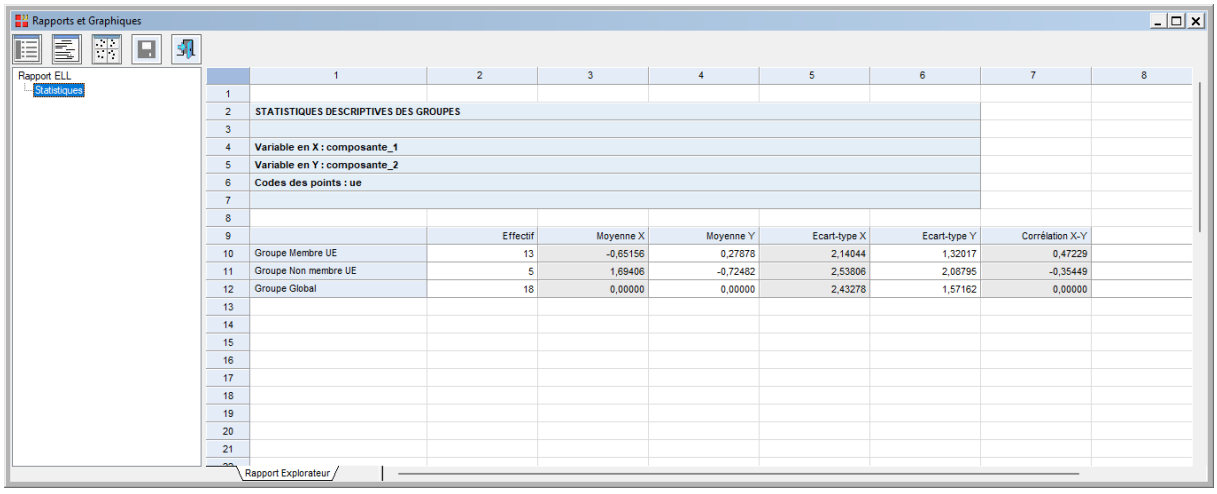
Choisir la variable qualitative définissant les groupes, dans notre exemple 'ue' qui indique l'appartenance ou non à l'Union Européenne.

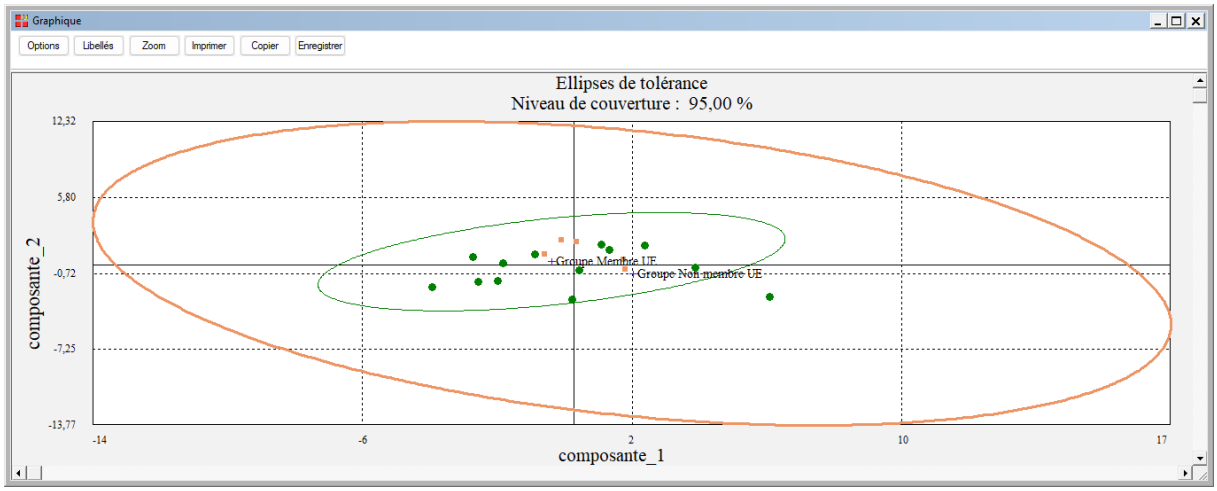
Choisir des ellipses de tolérance.

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

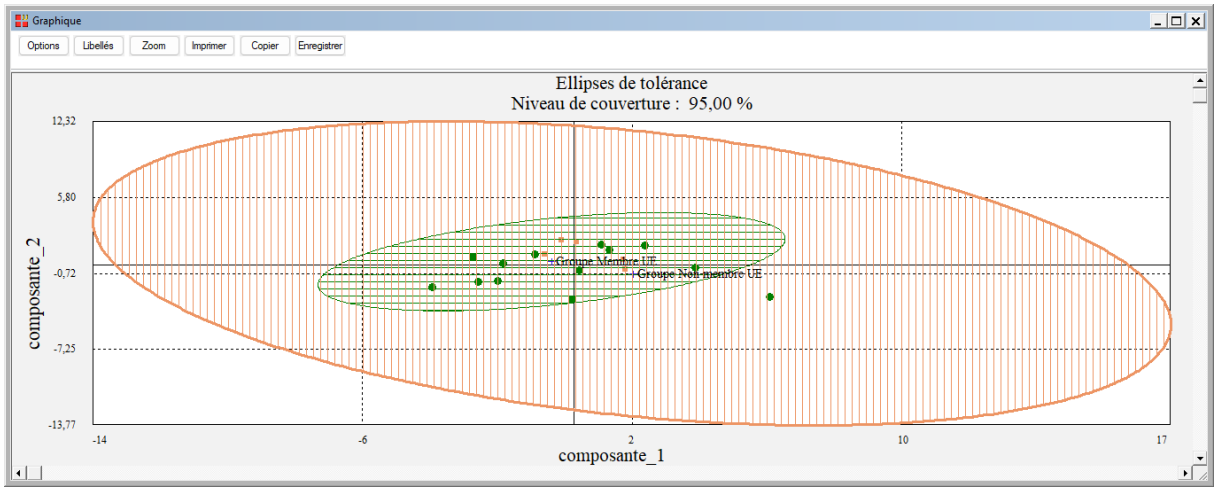


Cliquer sur Ok pour afficher le rapport et le graphique.





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.



Combiner ACP et Classification Ascendante Hiérarchique

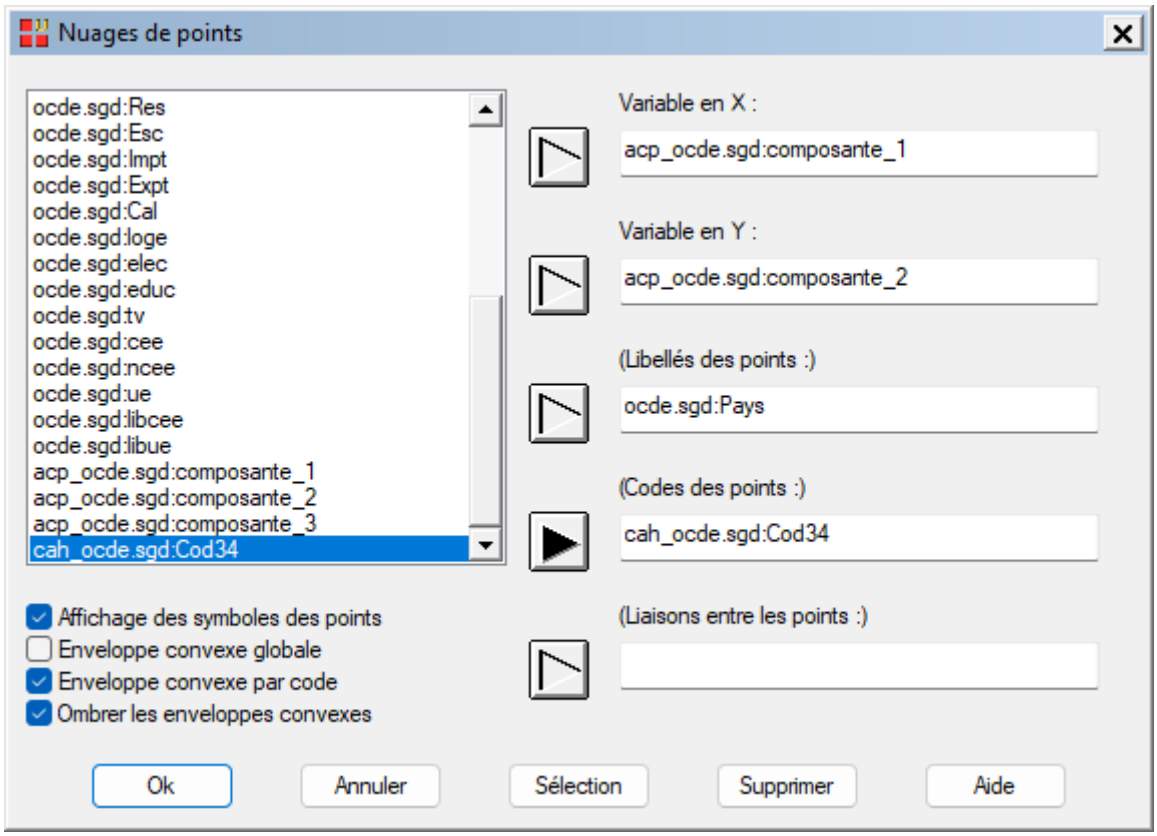
Il est généralement utile pour aider à l'interprétation des plans factoriels issus de l'ACP de combiner cette analyse avec un classification ascendante hiérarchique (CAH).

Reprenons l'analyse effectuée dans l'exemple 1.

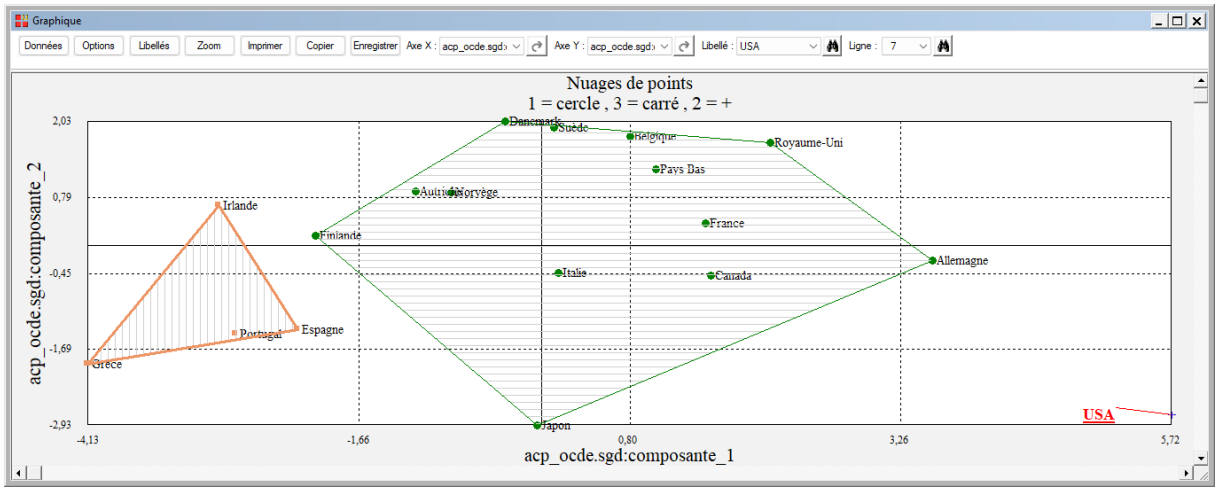
Voici les étapes à réaliser :

- Enregistrer les composantes principales de l'ACP
- Faire la CAH sur les mêmes variables et individus que pour l'ACP
- Choisir le nœud de troncature et générer le rapport associé
- Enregistrer les codes des classes

- Dans le ruban ‘Graphique’, choisir ‘Nuages’ et renseigner la boîte de dialogue comme montré ci-dessous.



- acp_ocde.sgd:composante_1 = première composante principale
- acp_ocde.sgd:composante_2 = deuxième composante principale
- ocde.sgd:Pays = noms des pays
- cah_ocde.sgd:Cod34 = classes des pays pour la troncature choisie au nœud 34



Pour plus de détails concernant la CAH, voir la documentation de cette procédure.

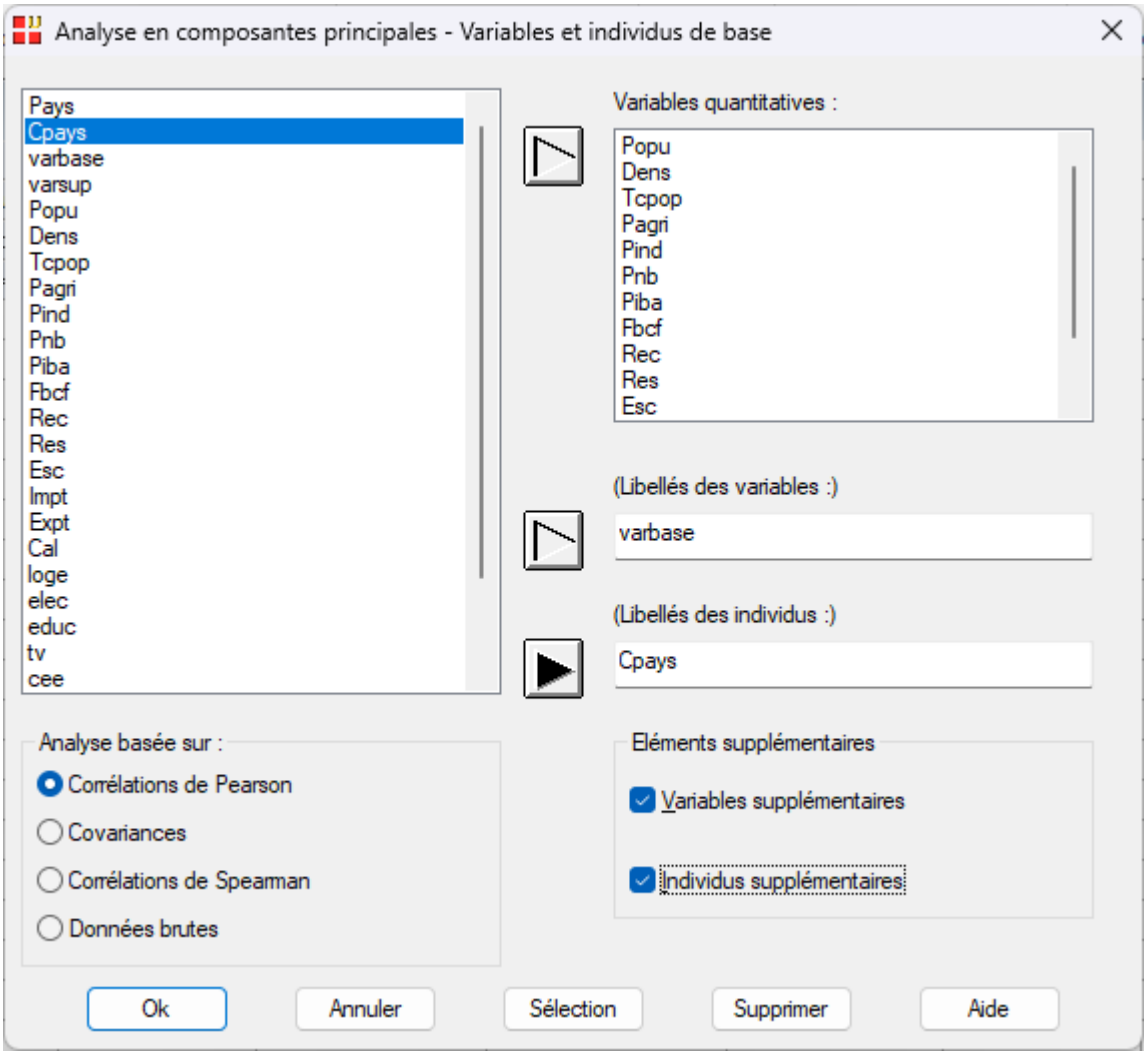
Exemple 2 : Fichier OCDE (variables et individus de base et supplémentaires)

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

Nous allons mettre en œuvre une ACP contenant à la fois :

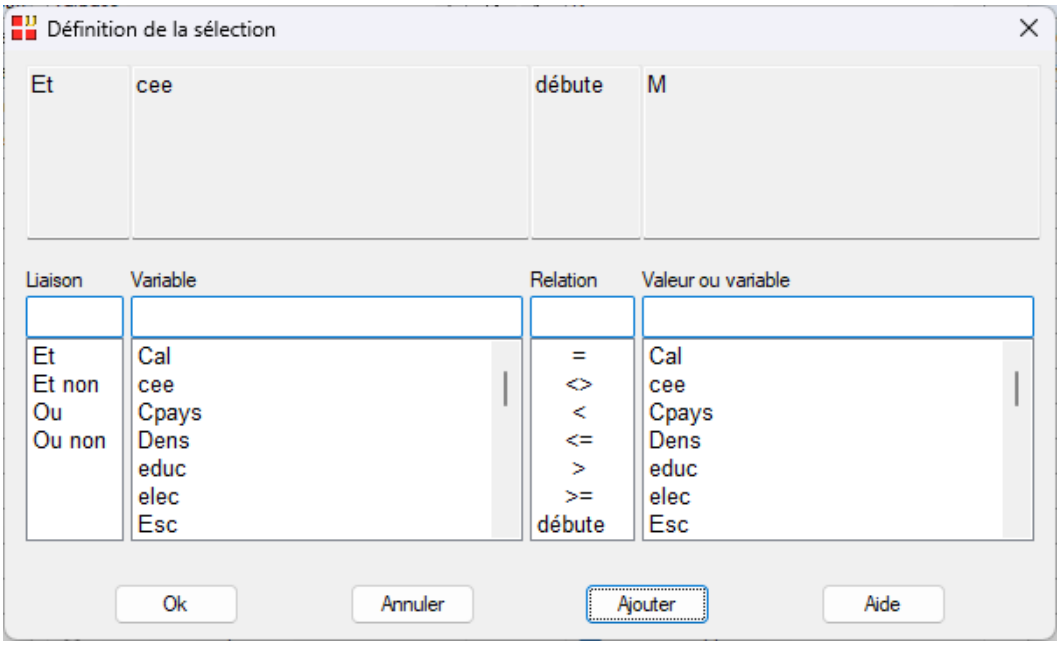
- des individus de base (les pays membres de la CEE)
- des individus supplémentaires (les pays qui ne sont pas membres de la CEE)
- des variables quantitatives de base (les variables Popu à Cal)
- une variable qualitative supplémentaire (la variable ue)
- des variables quantitatives supplémentaires (les 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.



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.

Cliquons 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.

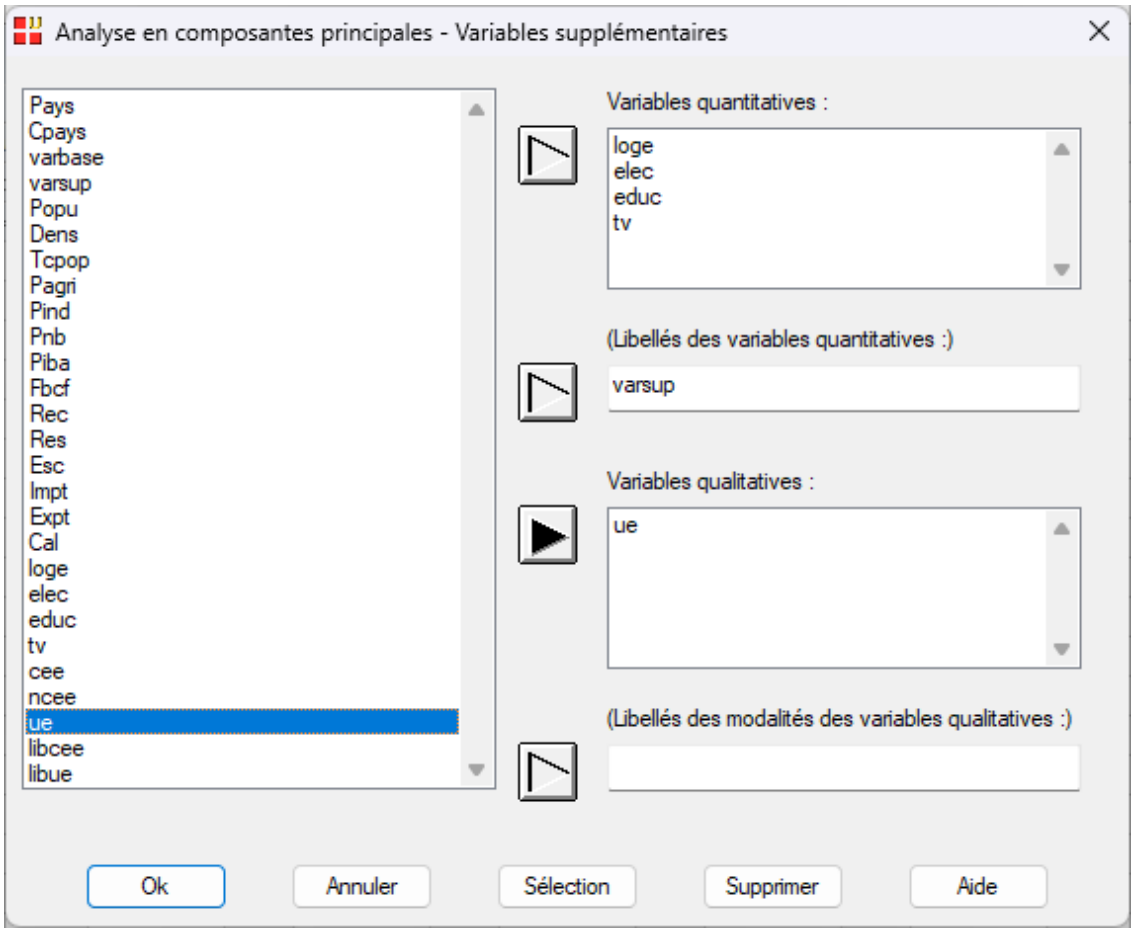
Cliquons sur Ok et la troisième boîte de dialogue apparaît.

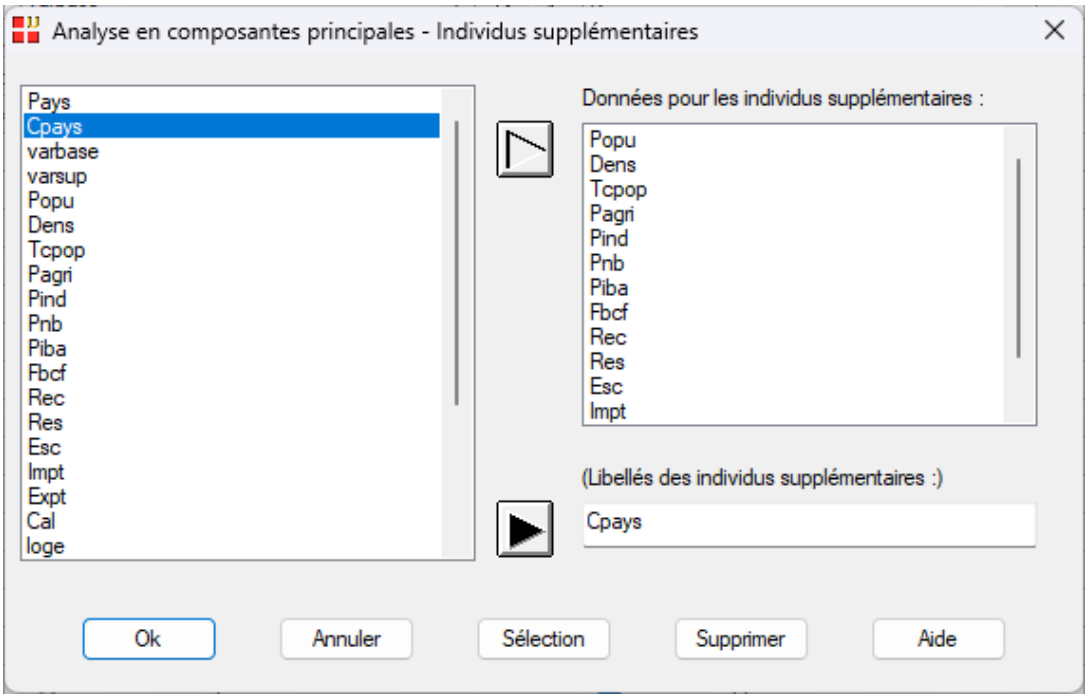
Choisissons à nouveau les variables Popu à Cal comme variables contenant les données des individus supplémentaires, la variable Cpays comme variable contenant les libellés des individus (les pays).

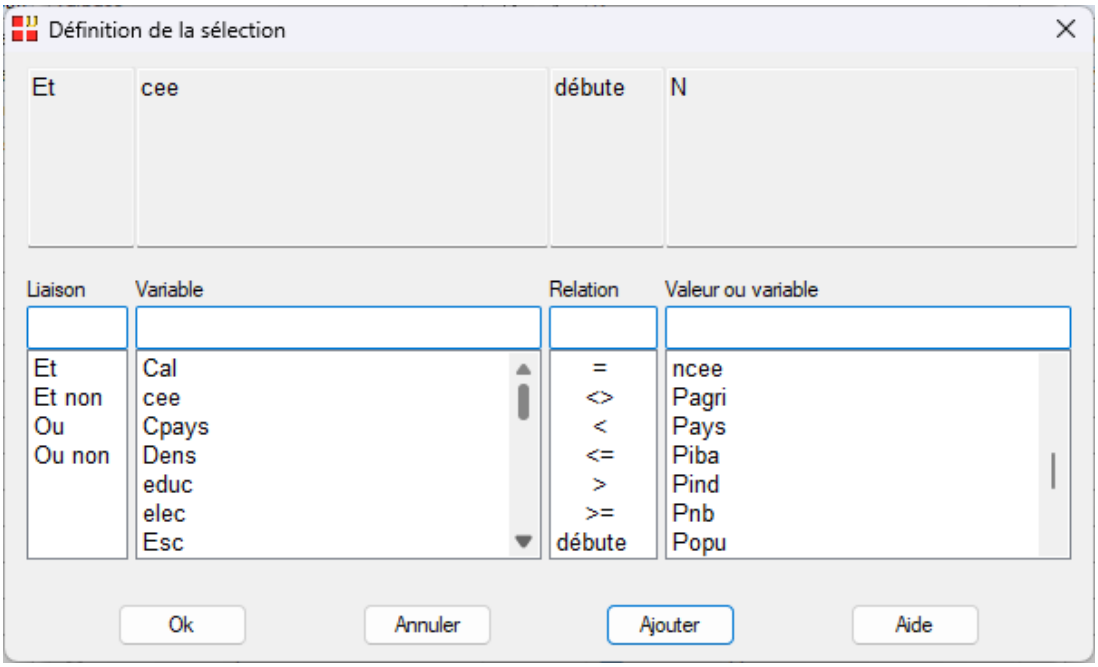
Cliquons sur le bouton 'Sélection' et définissons la sélection des individus supplémentaires (passifs), ici les pays qui ne sont pas membres de la CEE. La boîte de dialogue de cette sélection est montrée ci-après.

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

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

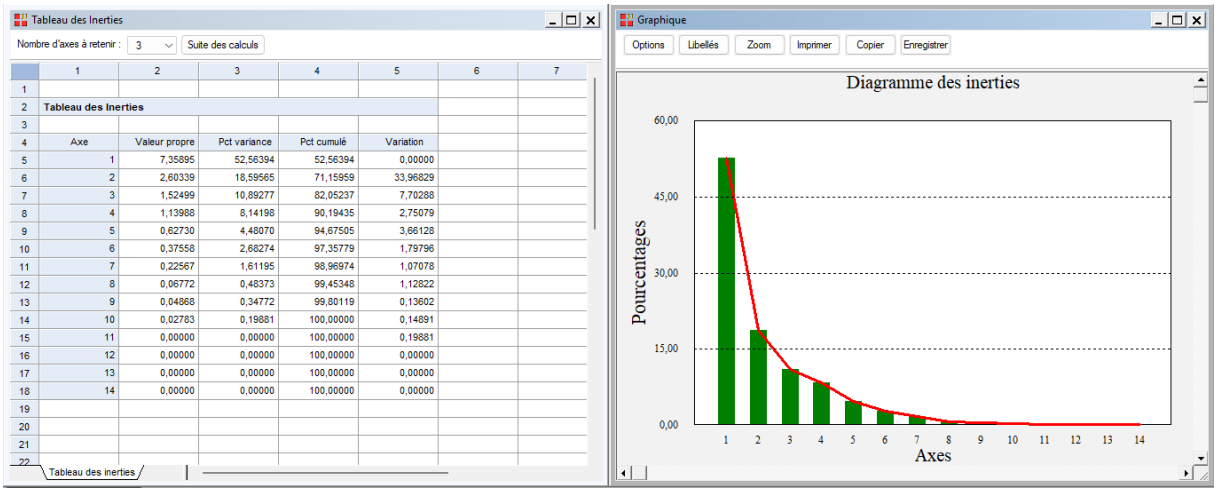

Analyse en composantes principales - Variables supplémentaires
Pays
Cpays
varbase
varsup
Popu
Dens
Tcpop
Pagri
Pind
Pnb
Piba
Fbcf
Rec
Res
Esc
Impt
Expt
Cal
loge
elec
educ
tv
cee
ncee
ue
libcee
libue
Variables quantitatives :
loge
elec
educ
tv
(Libellés des variables quantitatives :)
varsup
Variables qualitatives :
ue
(Libellés des modalités des variables qualitatives :)
Ok
Annuler
Sélection
Supprimer
Aide


Analyse en composantes principales - Individus supplémentaires
Pays
Cpays
varbase
varsup
Popu
Dens
Tcpop
Pagri
Pind
Pnb
Piba
Fbcf
Rec
Res
Esc
Impt
Expt
Cal
loge
Données pour les individus supplémentaires :
Popu
Dens
Tcpop
Pagri
Pind
Pnb
Piba
Fbcf
Rec
Res
Esc
Impt
(Libellés des individus supplémentaires :)
Cpays
Ok
Annuler
Sélection
Supprimer
Aide



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

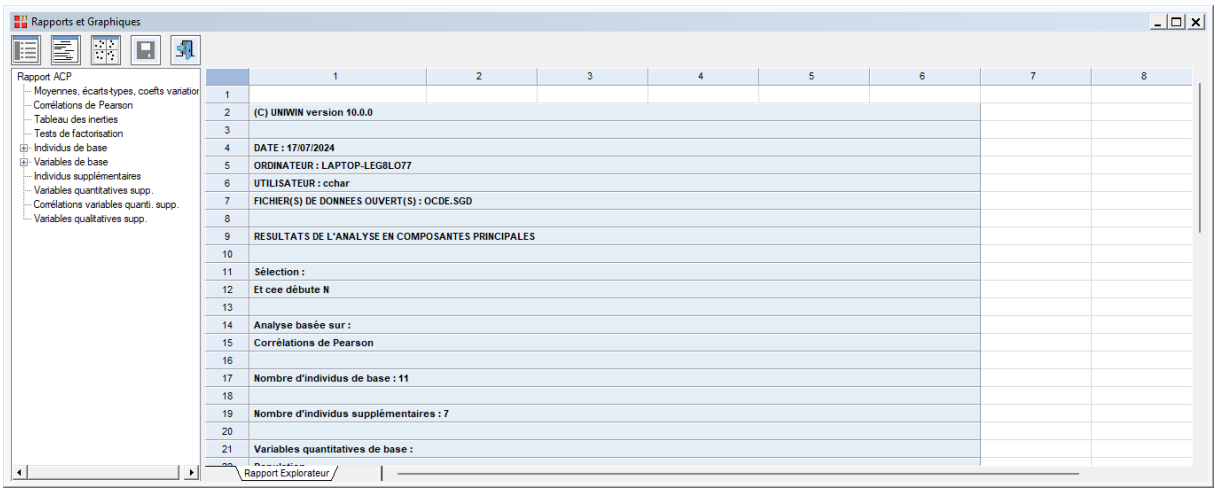
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.



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

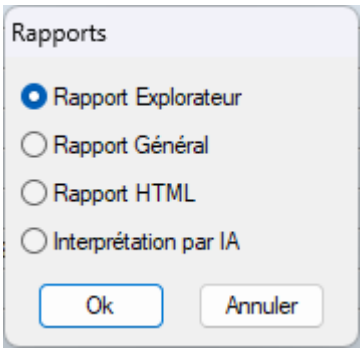
Cliquons sur le bouton 'Suite des calculs'.

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 :

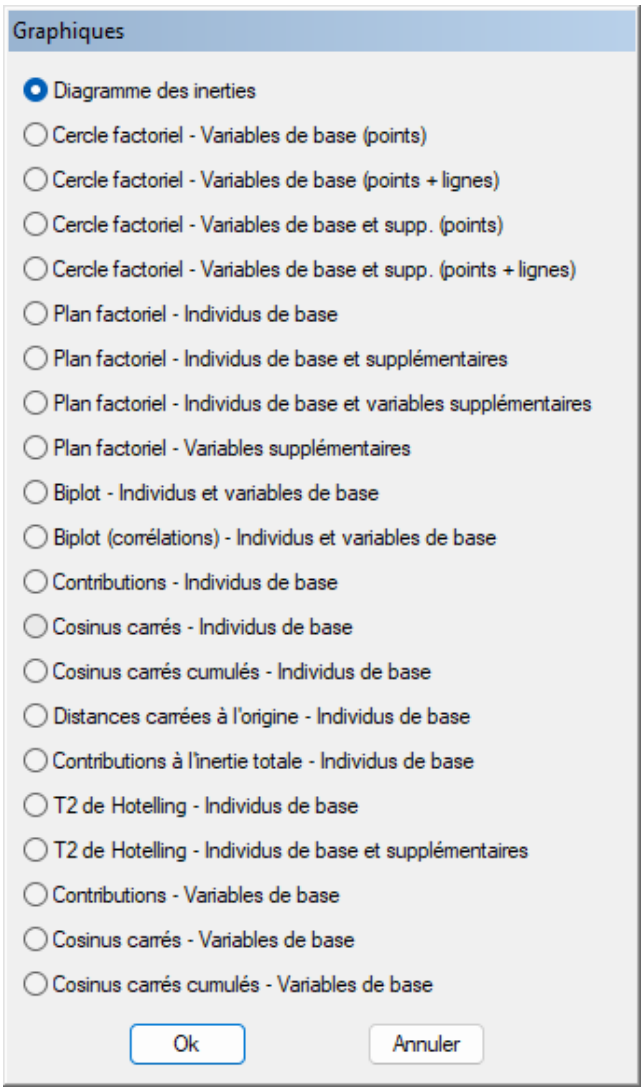


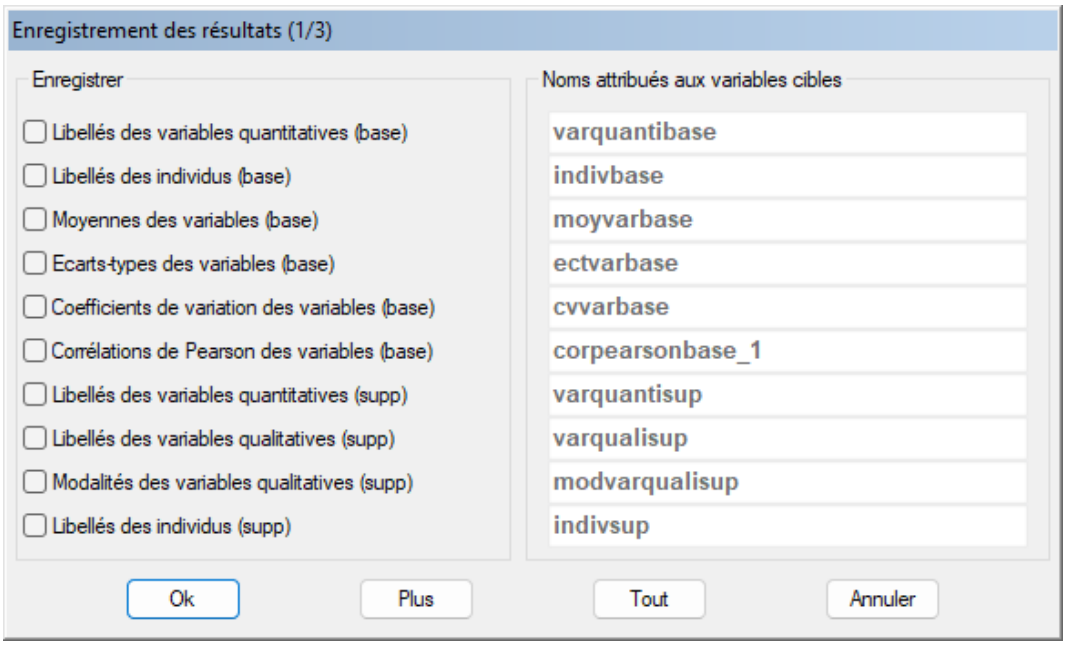
et l'icône 'Graphiques' affiche la boîte de dialogue, montrée ci-après, des options pour les graphiques.

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.

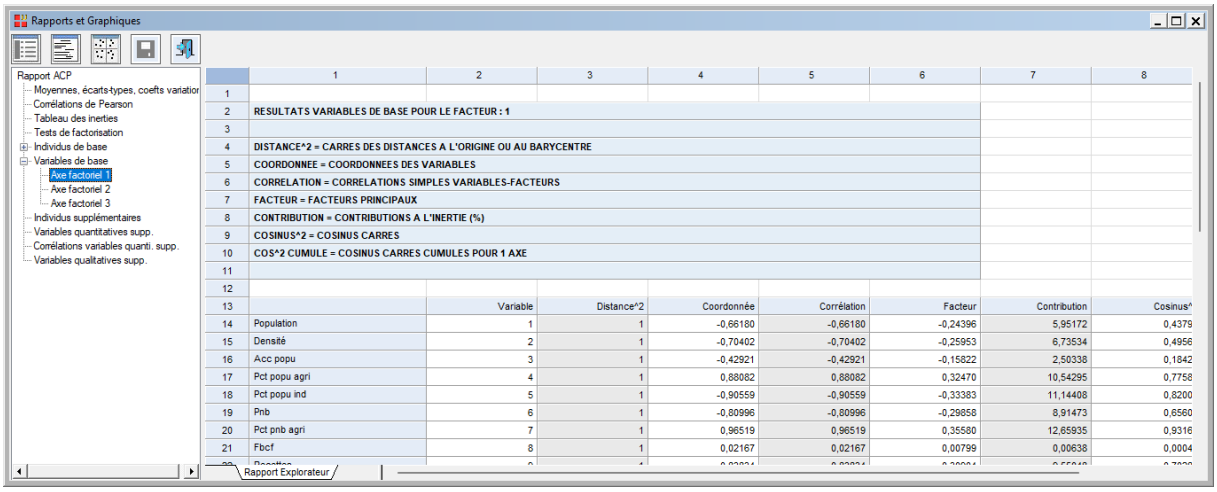

Graphiques
Diagramme des inerties
Cercle factoriel - Variables de base (points)
Cercle factoriel - Variables de base (points + lignes)
Cercle factoriel - Variables de base et supp. (points)
Cercle factoriel - Variables de base et supp. (points + lignes)
Plan factoriel - Individus de base
Plan factoriel - Individus de base et supplémentaires
Plan factoriel - Individus de base et variables supplémentaires
Plan factoriel - Variables supplémentaires
Biplot - Individus et variables de base
Biplot (corrélations) - Individus et variables de base
Contributions - Individus de base
Cosinus carrés - Individus de base
Cosinus carrés cumulés - Individus de base
Distances carrées à l'origine - Individus de base
Contributions à l'inertie totale - Individus de base
T2 de Hotelling - Individus de base
T2 de Hotelling - Individus de base et supplémentaires
Contributions - Variables de base
Cosinus carrés - Variables de base
Cosinus carrés cumulés - Variables de base
Ok
Annuler

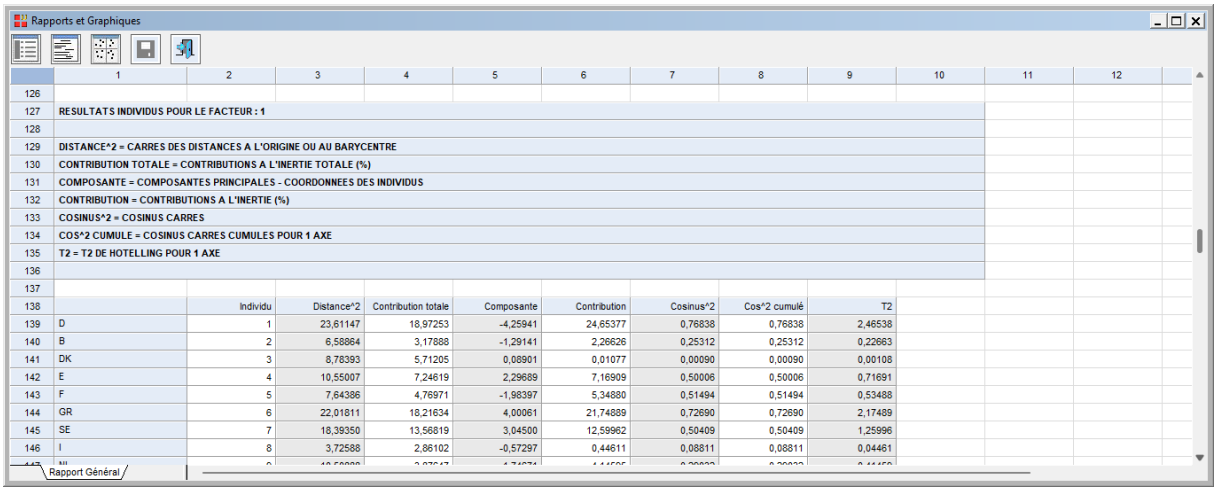

Enregistrement des résultats (1/3)
Enregistrer
Libellés des variables quantitatives (base)
Libellés des individus (base)
Moyennes des variables (base)
Ecarts-types des variables (base)
Coefficients de variation des variables (base)
Corrélations de Pearson des variables (base)
Libellés des variables quantitatives (supp)
Libellés des variables qualitatives (supp)
Modalités des variables qualitatives (supp)
Libellés des individus (supp)
Noms attribués aux variables cibles
varquantibase
indivbase
moyvarbase
ectvarbase
cvvarbase
corpearsonbase_1
varquantisup
varqualisup
modvarqualisup
indivsup
Ok
Plus
Tout
Annuler

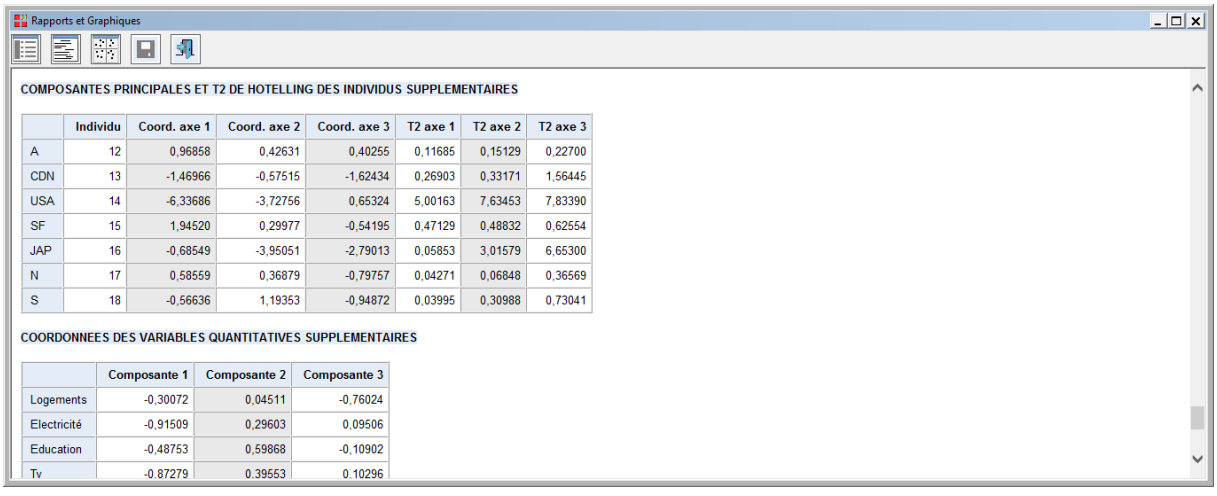
L'option Rapports

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

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





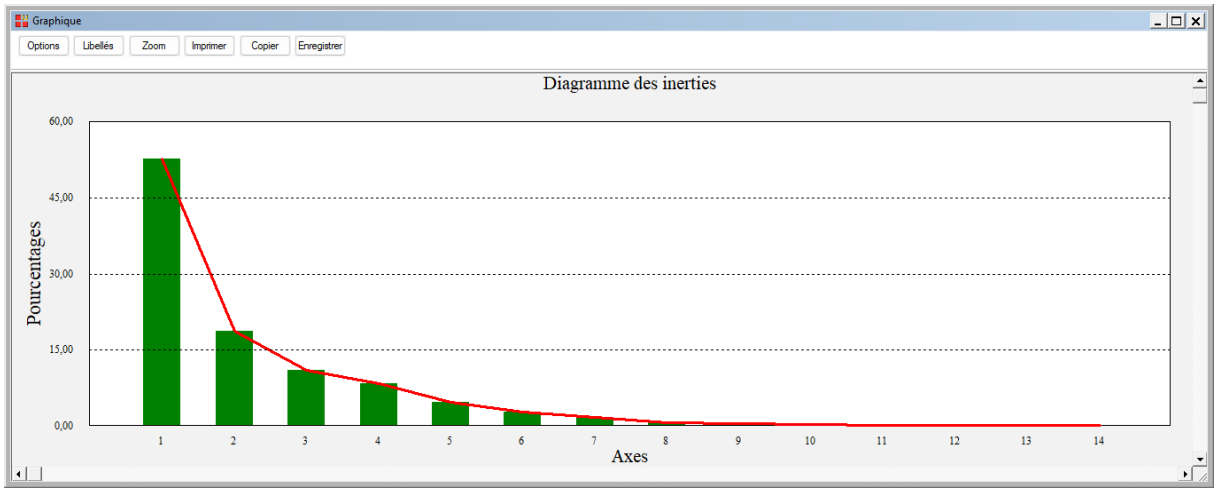


L’option Graphiques

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

- Diagramme des inerties

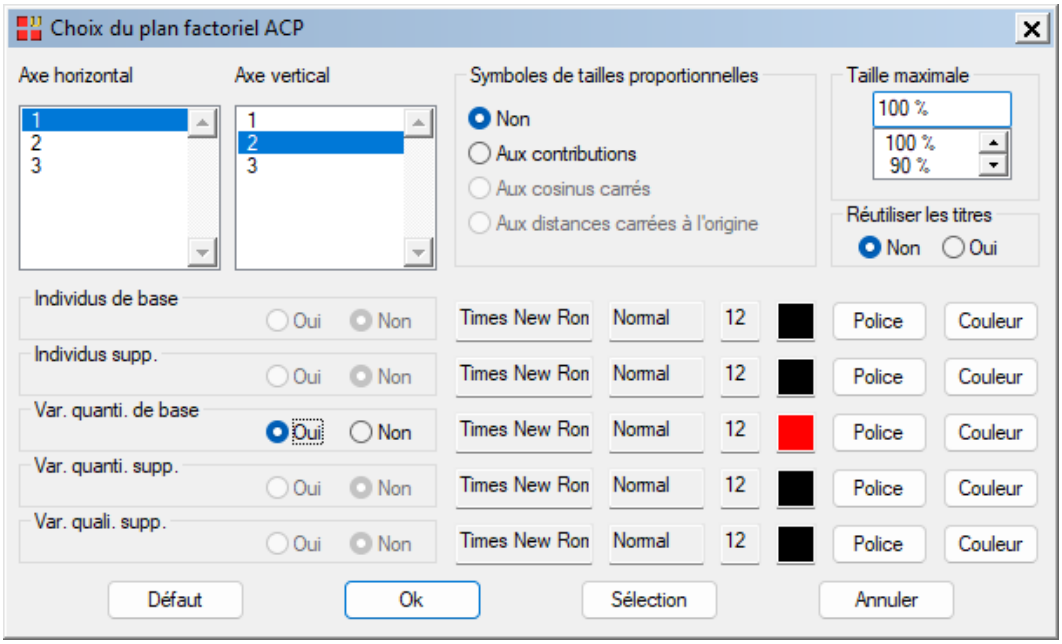
Ce graphique affiche les pourcentages d’inertie pour chacun des axes factoriels.



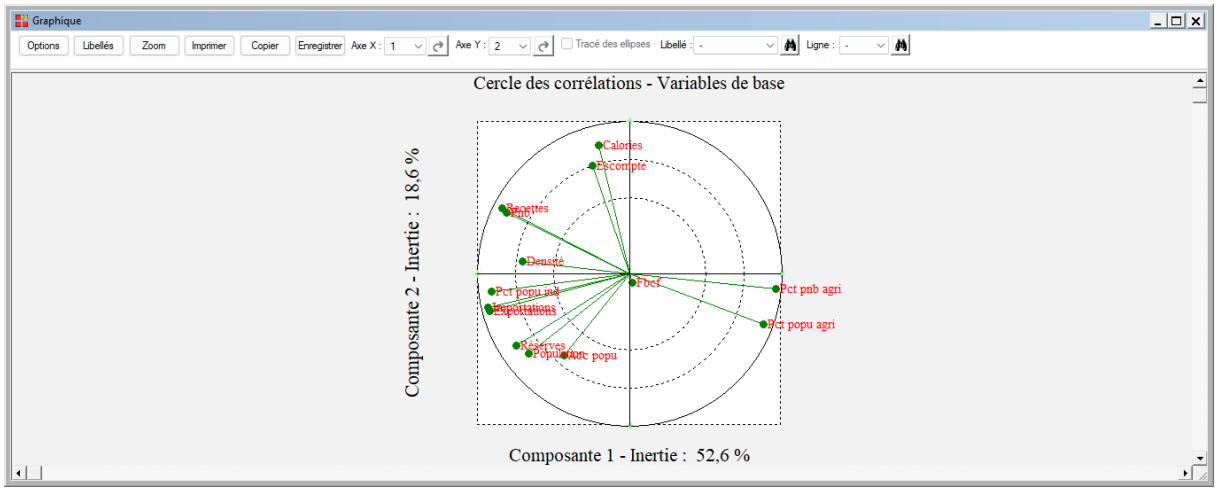
- Les options Cercle factoriel

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.

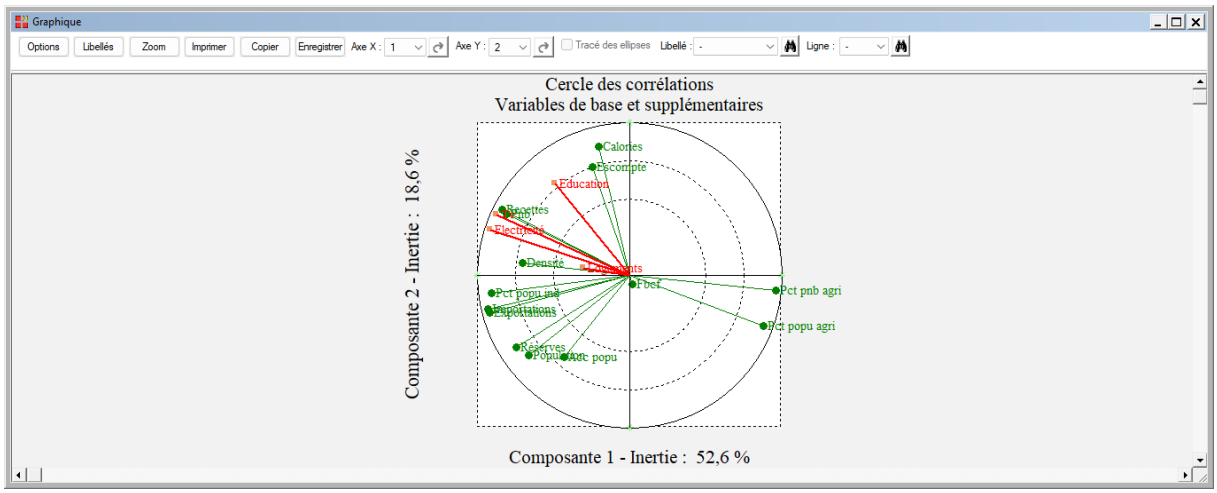
Une boîte de dialogue permettant de choisir le plan factoriel s’affiche.

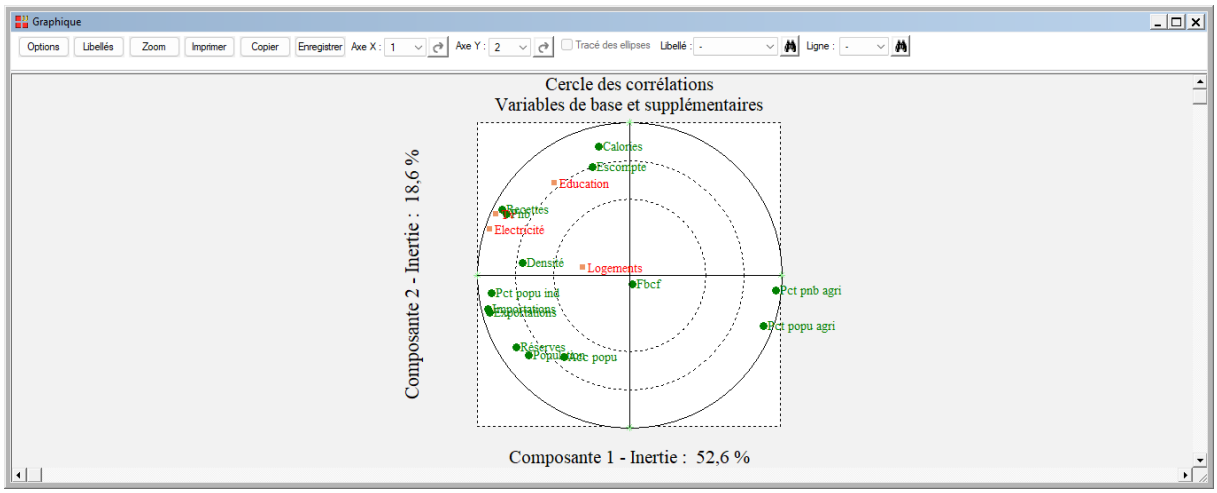


Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de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



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





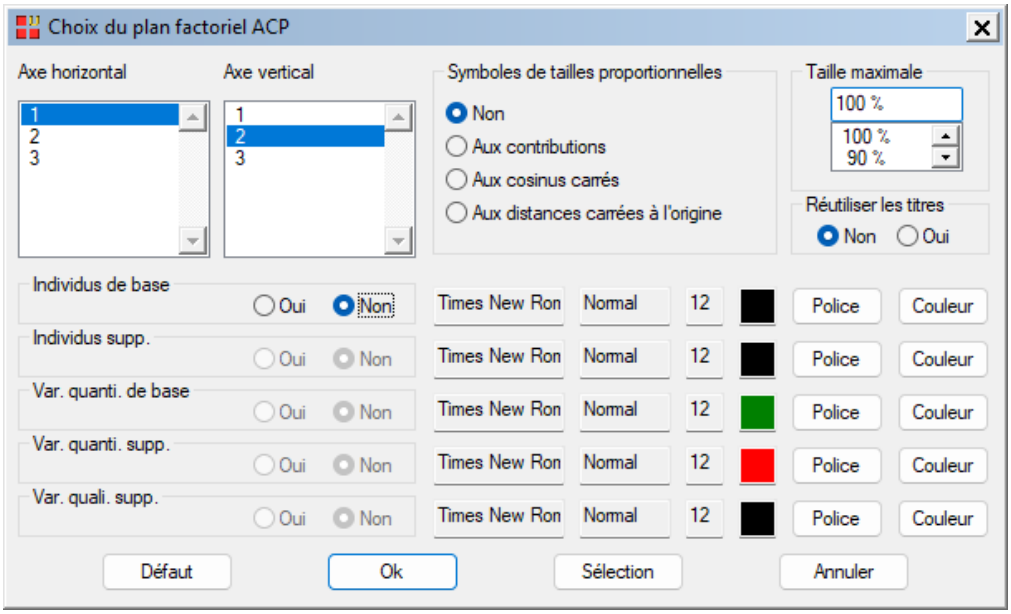
- Les options Plan factoriel et variables supplémentaires

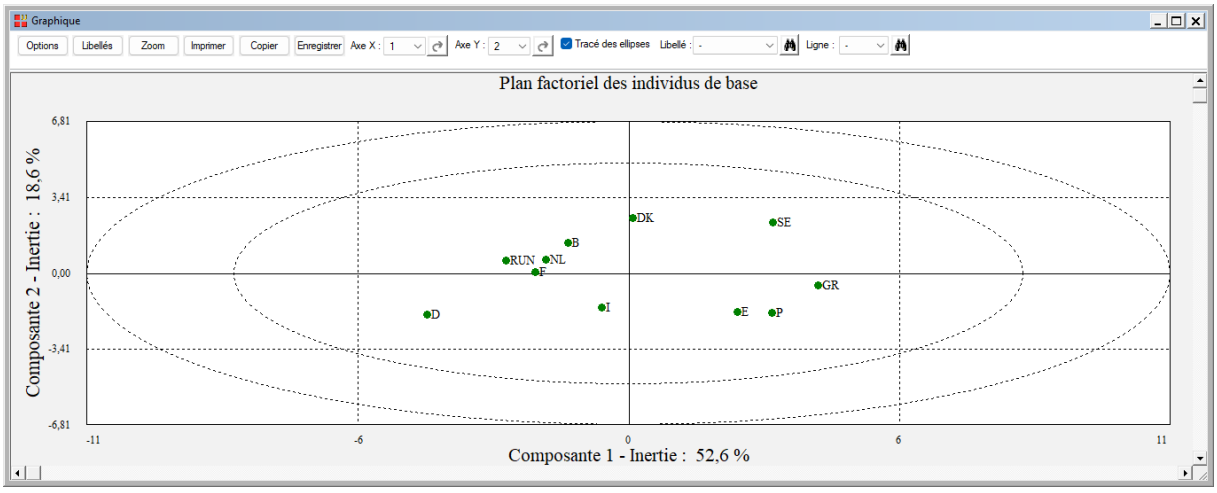
Ces options permettent d'afficher des plans factoriels des individus de base, des individus de base et des individus supplémentaires (s'il y en a), des individus de base et des variables supplémentaires (s'il y en a) ainsi que des variables supplémentaires seules.

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.

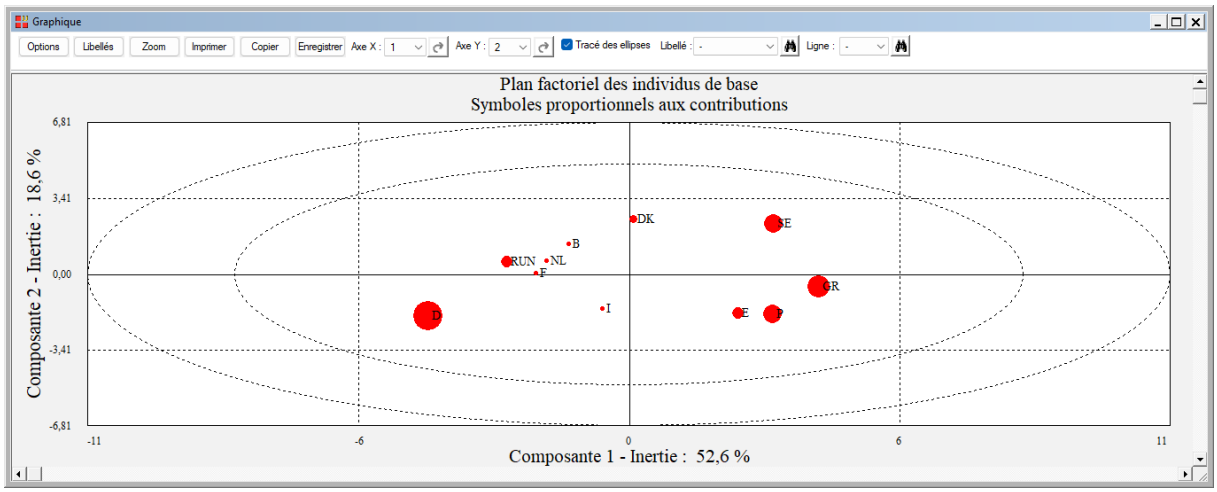
L'option 'Taille maximale' permet de contrôler les tailles des bulles en fonction d'un pourcentage de la taille de la bulle la plus grande.





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

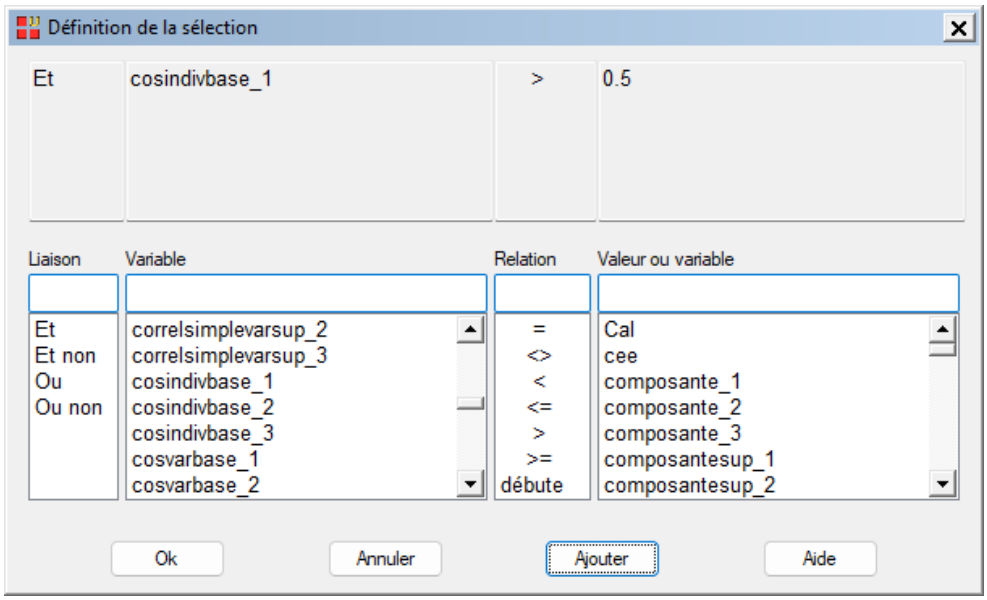
Affichons ce même graphique avec des symboles de points proportionnels aux contributions.

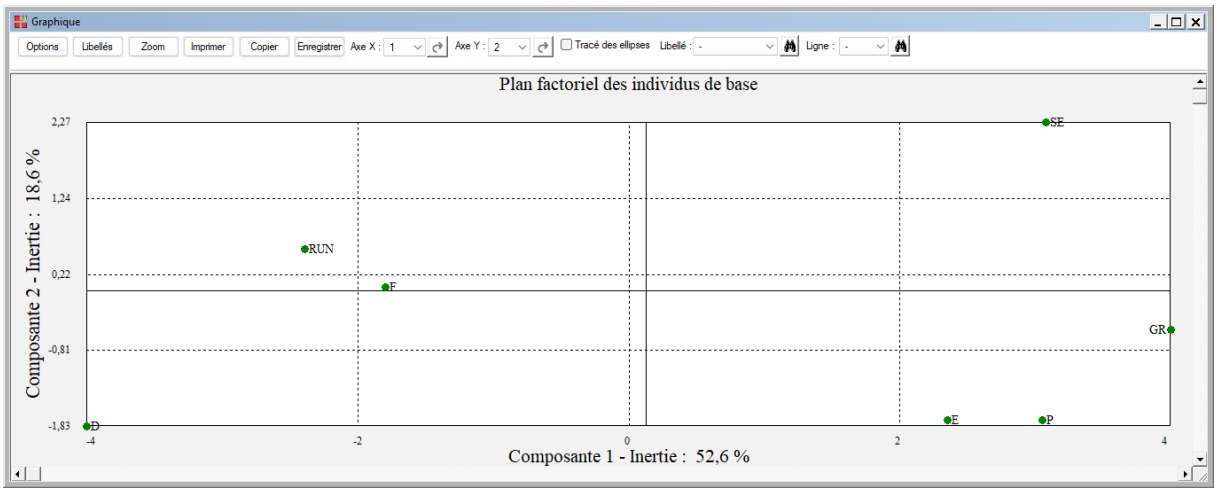


A noter le bouton ‘Sélection’ qui permet de n’afficher que les points qui satisfont une condition logique, par exemple une condition basée sur les contributions ou les cosinus carrés. Ceci est très pratique dans le cas d’un nuage comportant beaucoup d’individus : seuls les points significatifs d’un point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

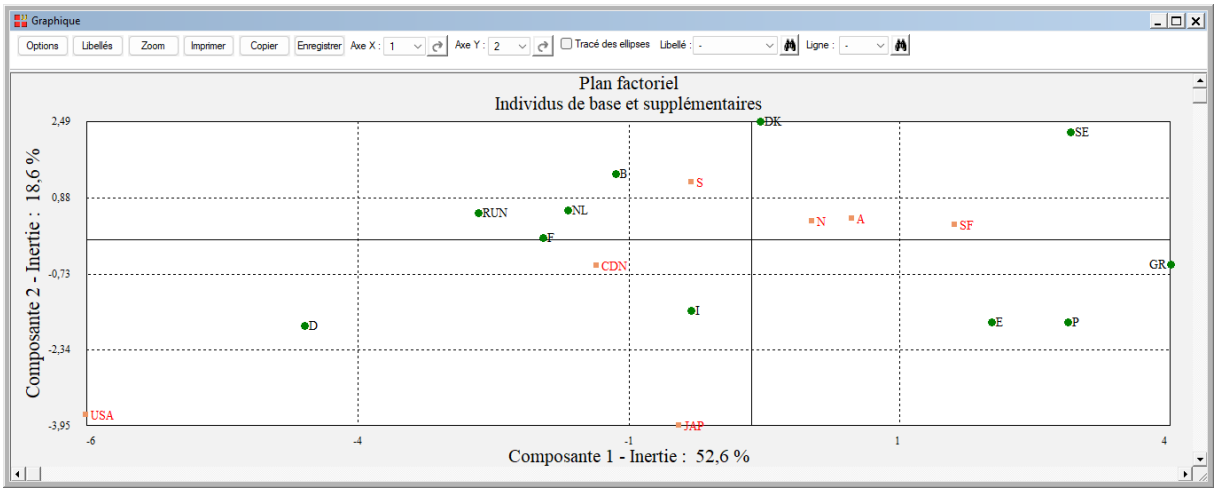
Comme exemple, mettons en œuvre une sélection basée sur la variable ‘cosindivbase_1’ qui contient les cosinus carrés sur le premier axe factoriel.

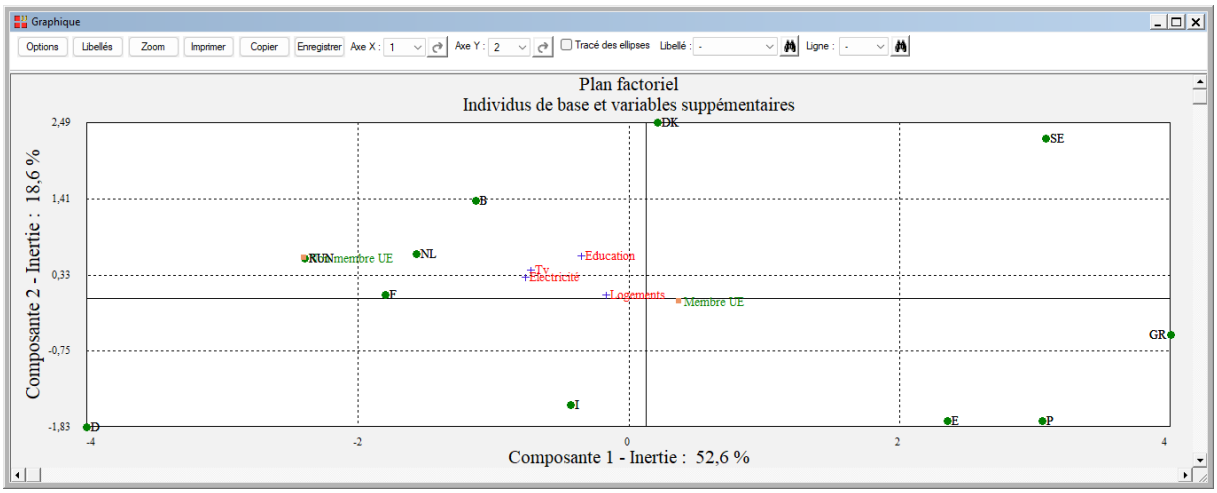




Des exemples des plans factoriels Individus de base et supplémentaires, Individus de base et Variables qualitatives sont également montrés.

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





- Les options Biplot et Biplot (corrélations)

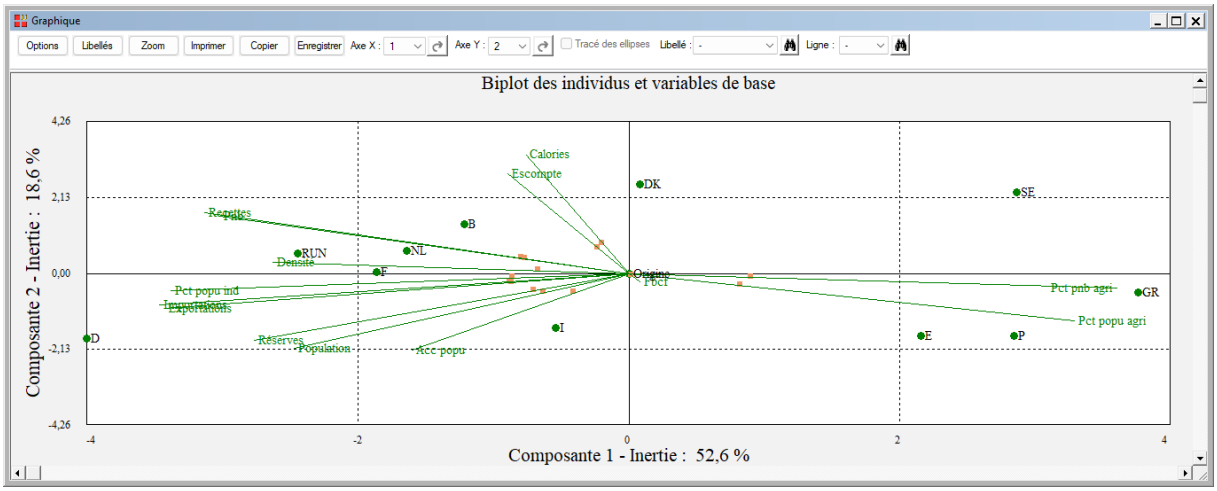
L'option Biplot permet d'obtenir un graphique donnant une représentation simultanée des individus et des directions des variables d'origine dans le plan factoriel choisi.

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.



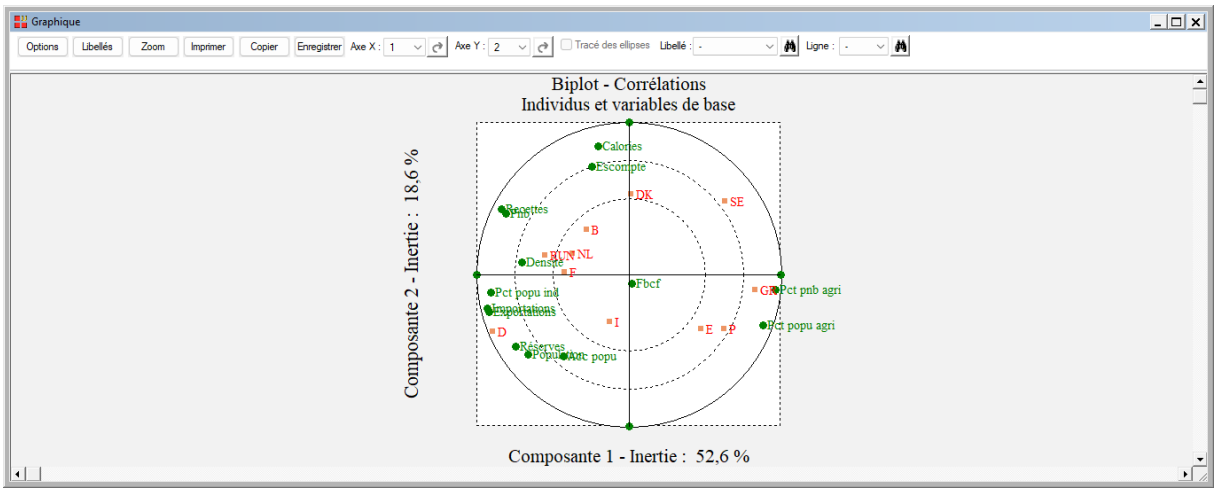
L'option Biplot (corrélations) permet d'obtenir un graphique donnant une représentation simultanée des variables indicatrices des individus et des variables d'origine dans un cercle des corrélations.

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.

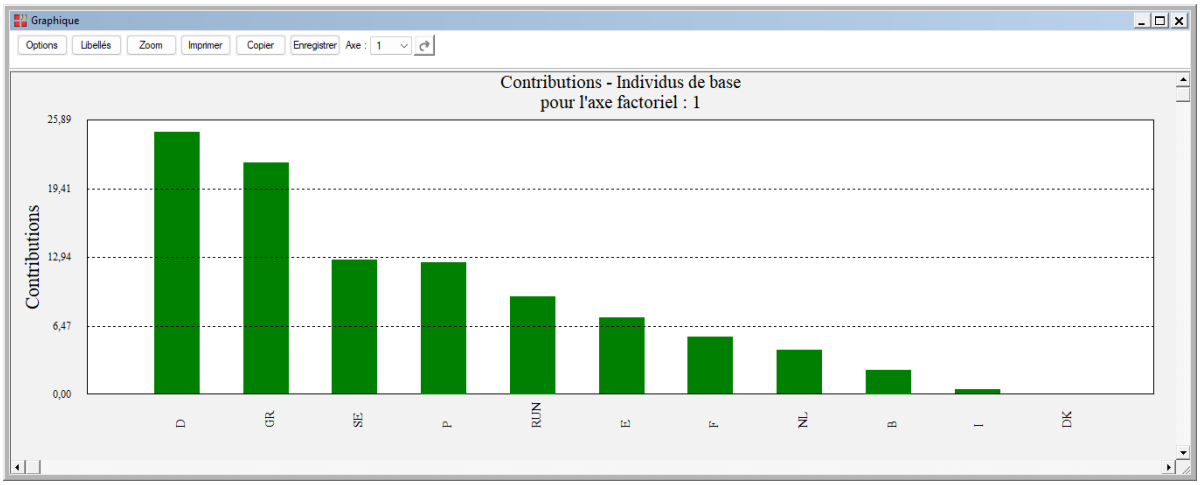
L'option zoom est également disponible.

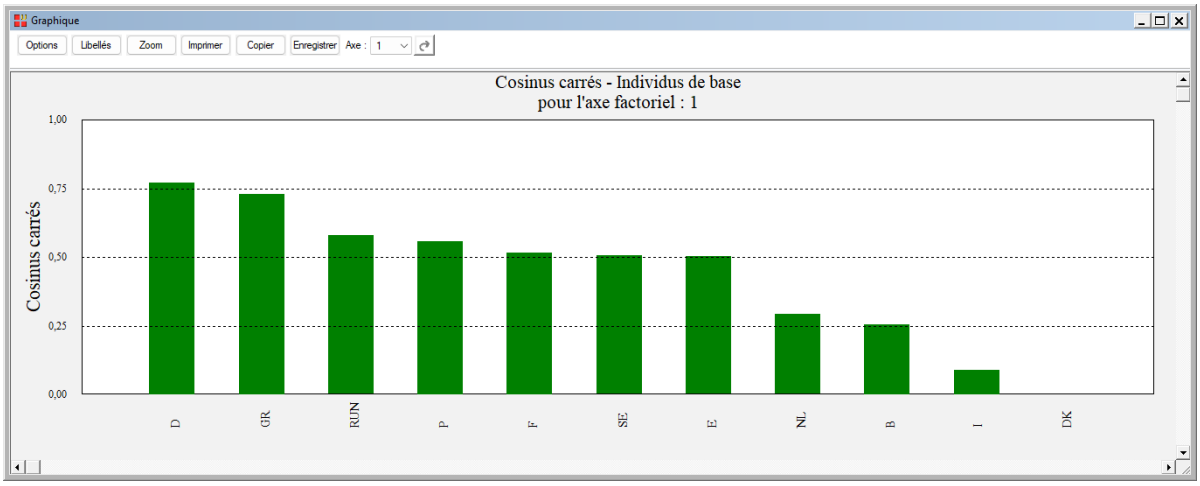


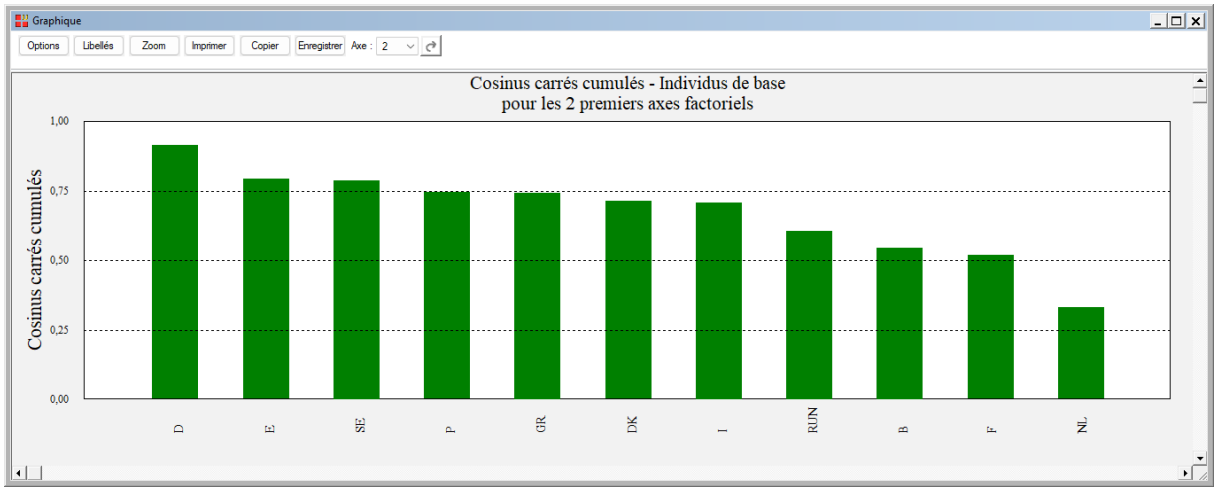
- Les options Contributions, Cosinus carrés, Cosinus carrés cumulés, Distances carrées à l'origine, Contributions à l'inertie totale, T2 de Hotelling

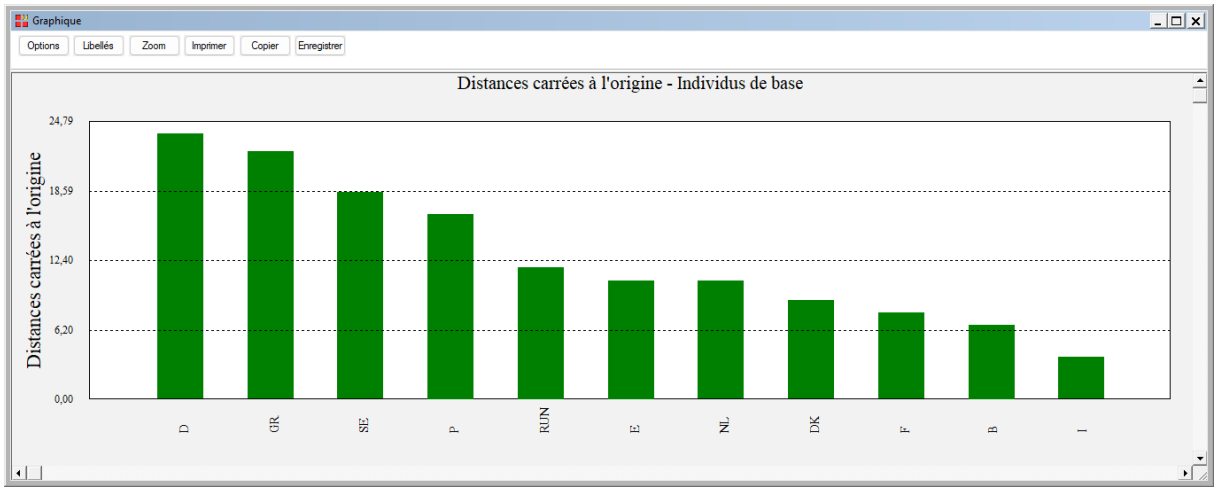
Les cinq premières options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse.

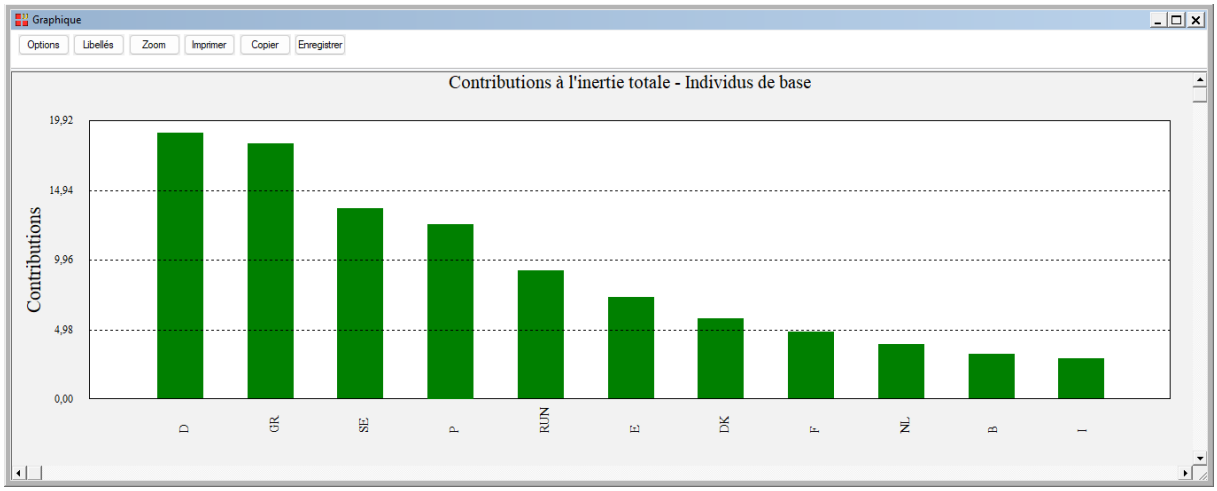
Elles mettent en évidence les individus qui contribuent fortement à la définition des axes factoriels, qui sont bien projetés dans les plans factoriels et qui sont éloignés de l'origine.



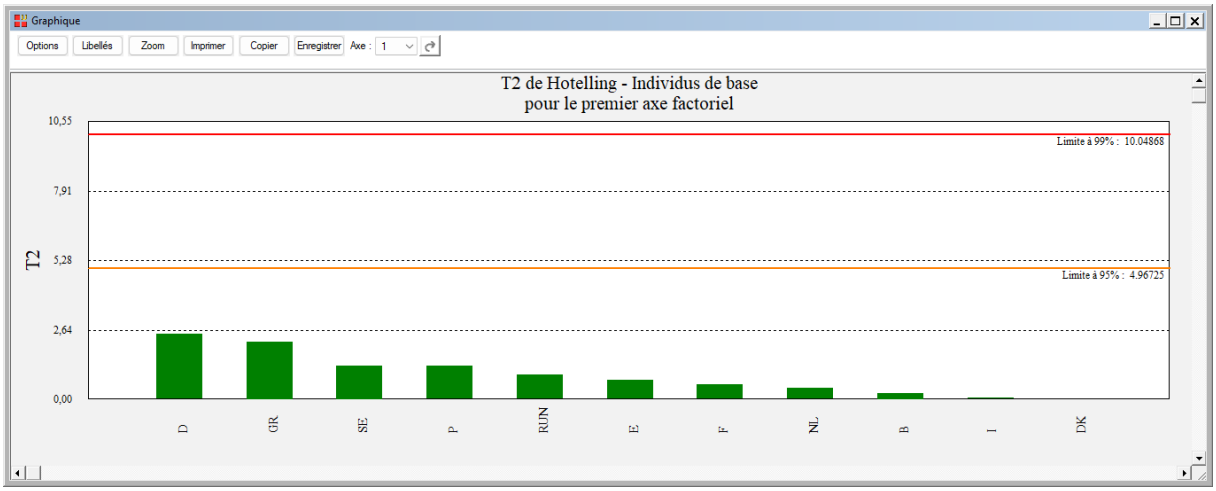


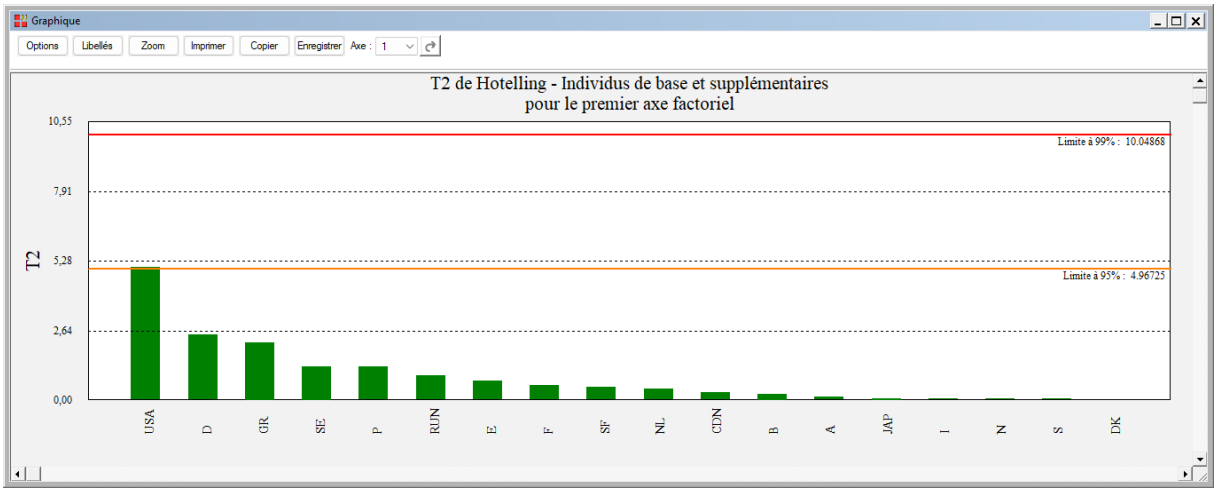






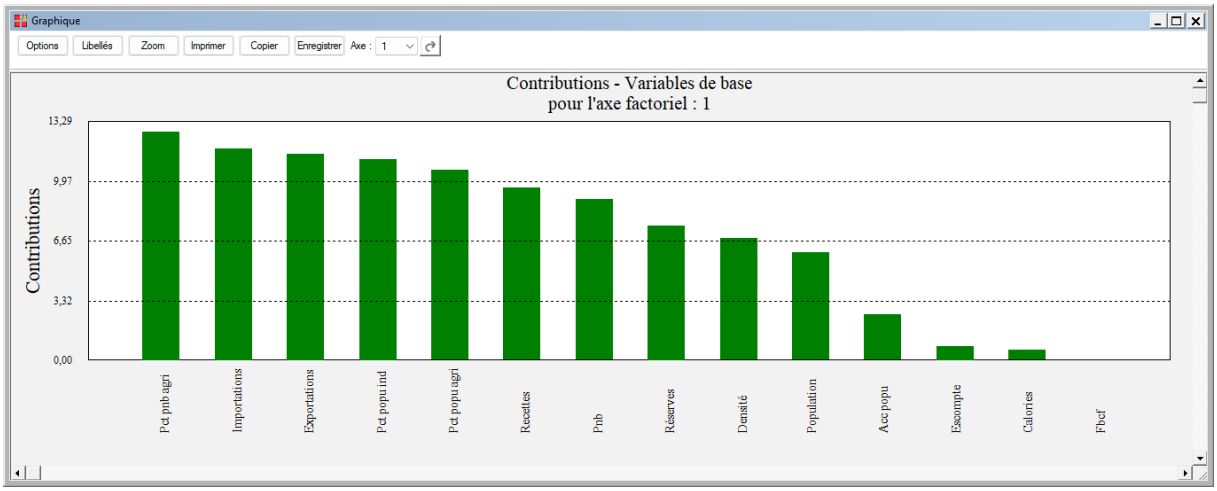
Les sixième et septième options affichent les T2 de Hotelling ainsi que les limites à 95% (individu suspect) et 99% (individu extrême).





- Les options Contributions, Cosinus carrés et Cosinus carrés cumulés pour les variables de base

Ces options permettent de visualiser des diagrammes en bâtons ordonnés de diverses caractéristiques statistiques calculées lors de l'analyse. Elles mettent en évidence les variables qui contribuent fortement à la définition des axes factoriels et qui sont bien projetés dans les plans factoriels.



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 Égyptiens, les Hébreux, les Grecs, et plus tard les Gaulois et les Ibères, fabriquèrent des produits alimentaires fermentés. Le pain était obtenu par des processus empiriques.

Ce n'est qu'au 19e 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 20e 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.

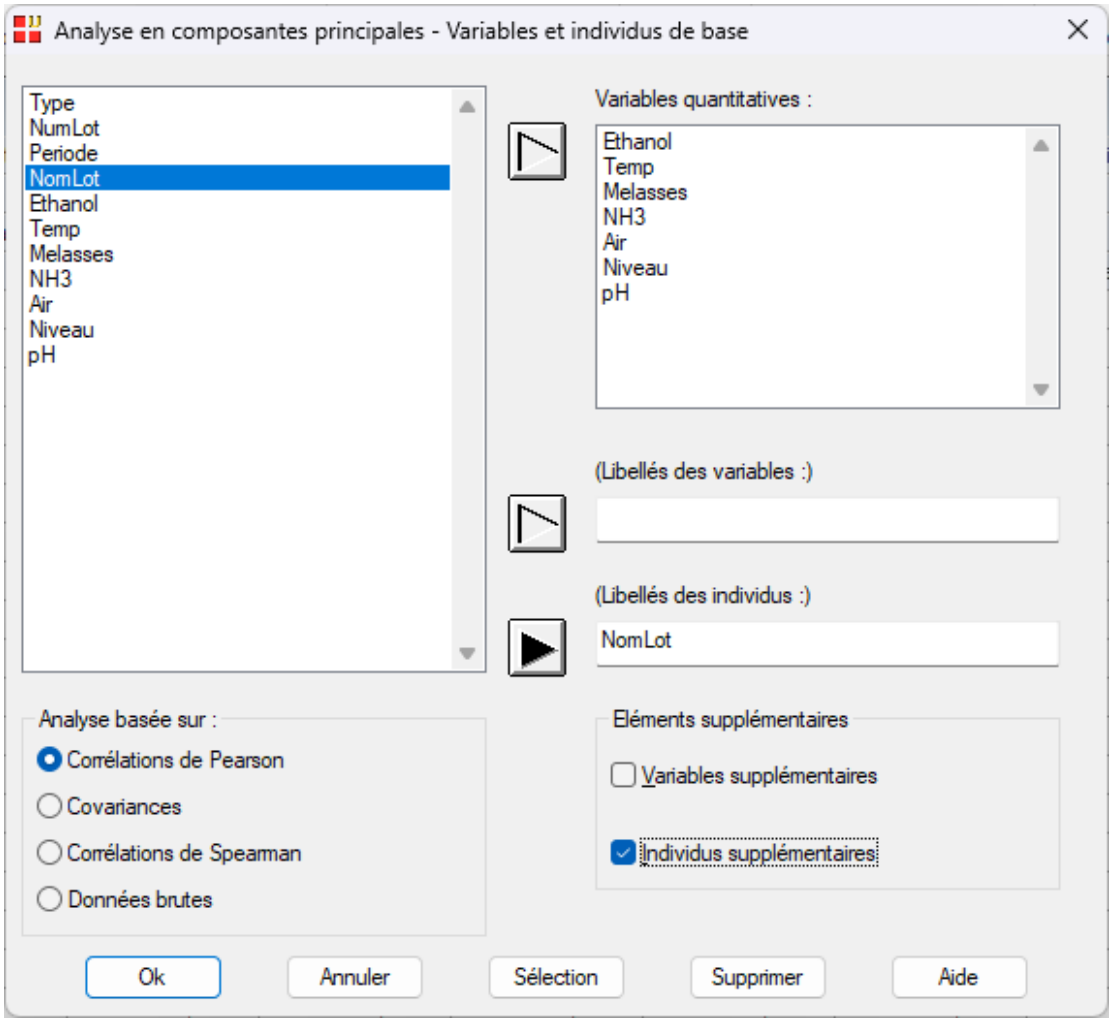
L'objectif est d'aider au pilotage du procédé (MSP) en analysant les trajectoires temporelles des différents lots en mettant en œuvre une analyse multivariée de l'ensemble des caractéristiques (MSP multivariée).

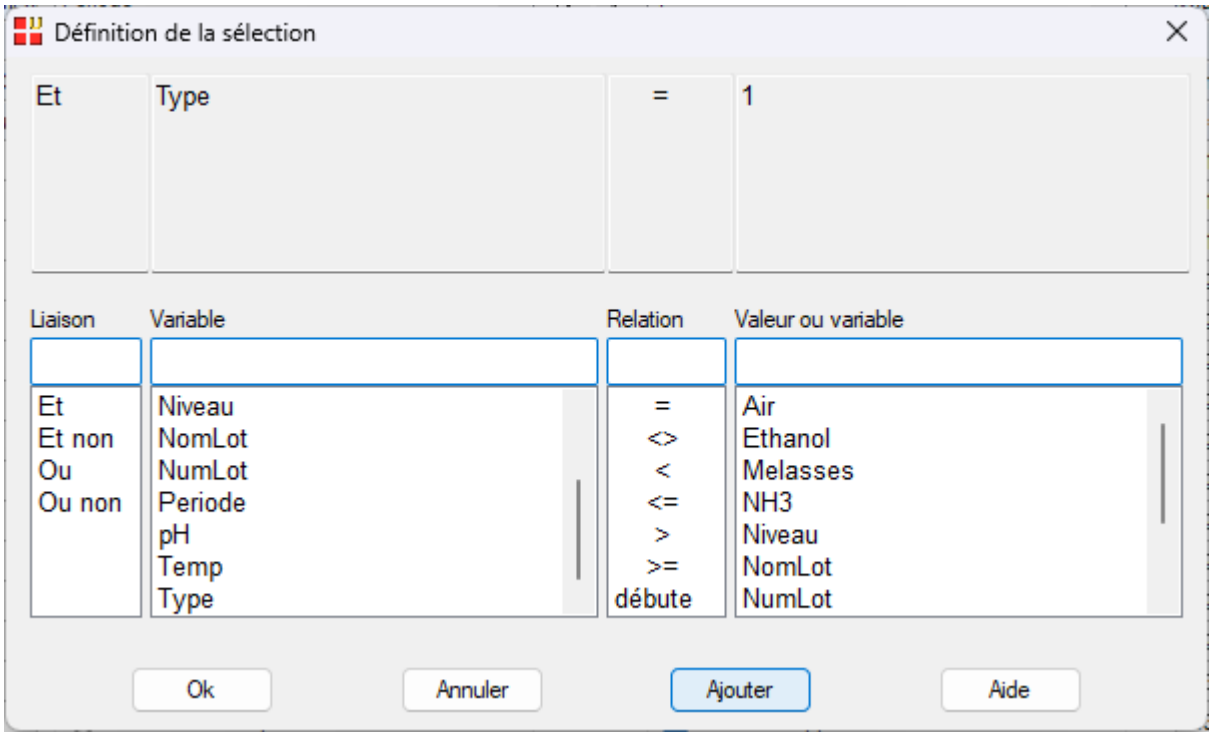
Chaque lot possédant 83 observations, il y a un total de 2.739 individus dans notre analyse.

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).

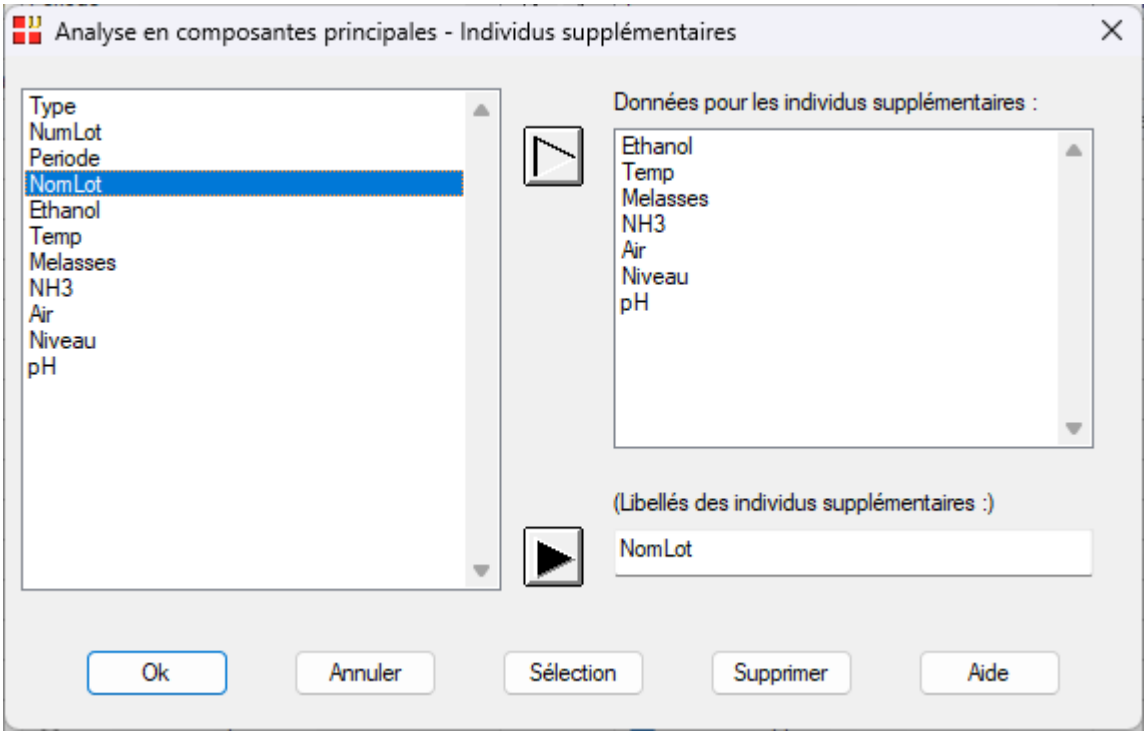
L'analyse initiale porte sur les 20 lots de référence ou « bons lots » (Type =1) parmi les 33 lots.

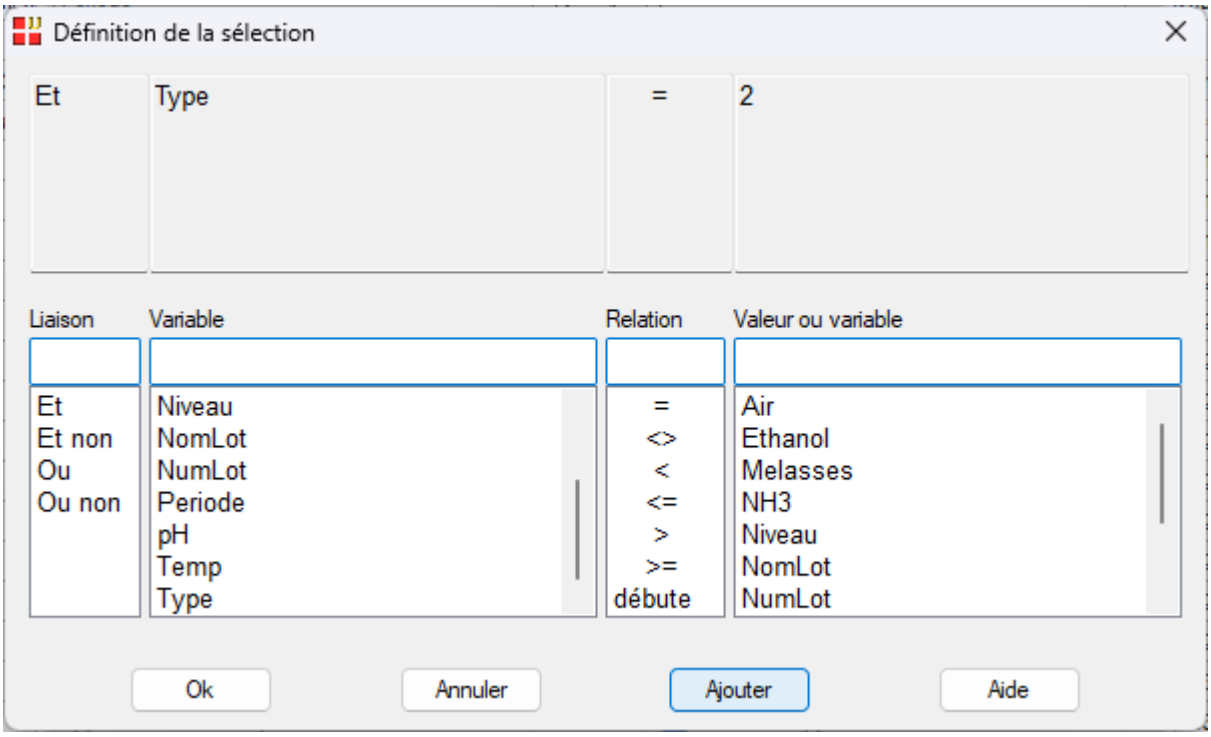
Cliquons sur l'icône ACP ou choisissons Analyse en Composantes Principales dans le ruban Décrire et renseignons la première boîte de dialogue comme montré ci-après.



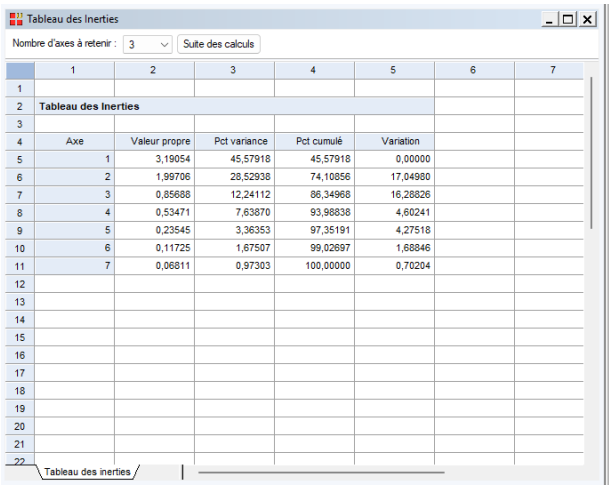


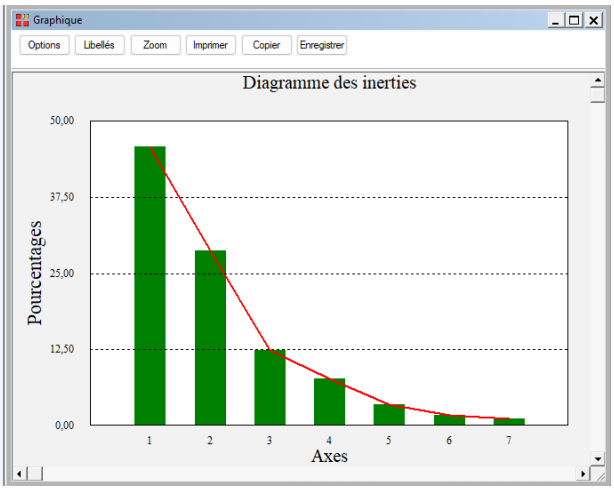
Renseignons la boîte de dialogue des individus supplémentaires (les 13 nouveaux lots).





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



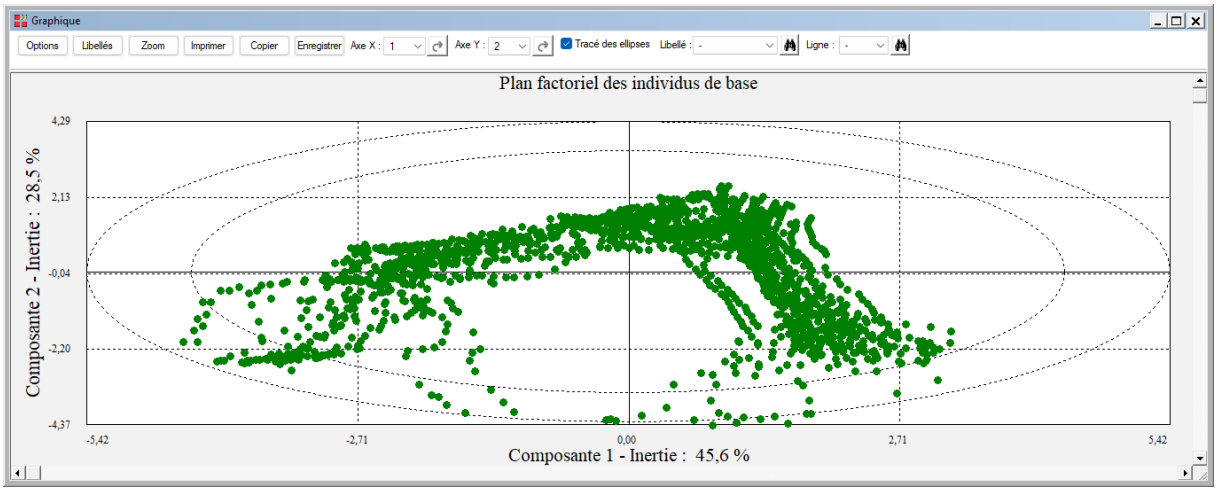


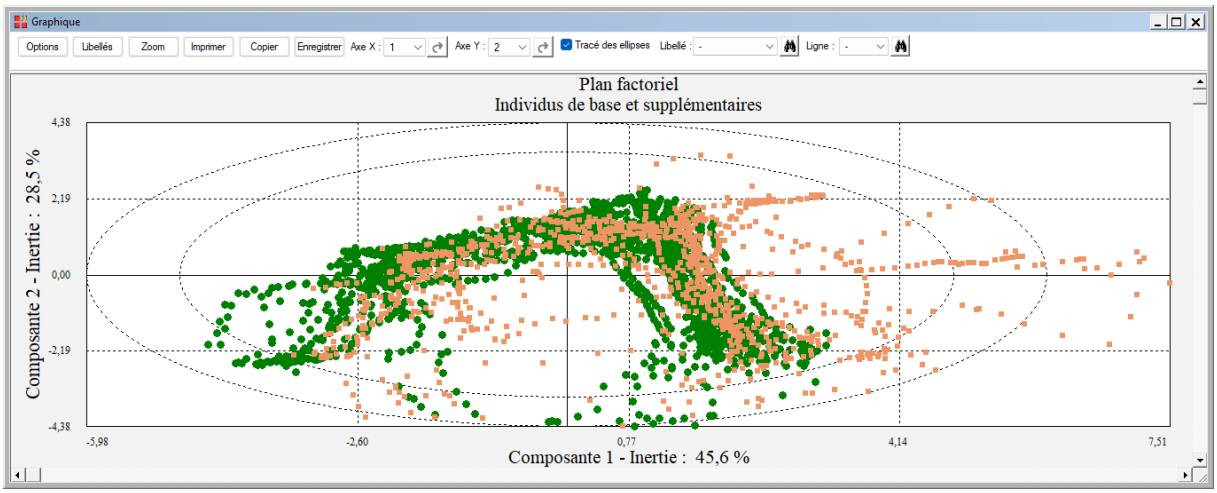
Affichons le plan factoriel 1-2 des individus de base avec les ellipses de confiance.

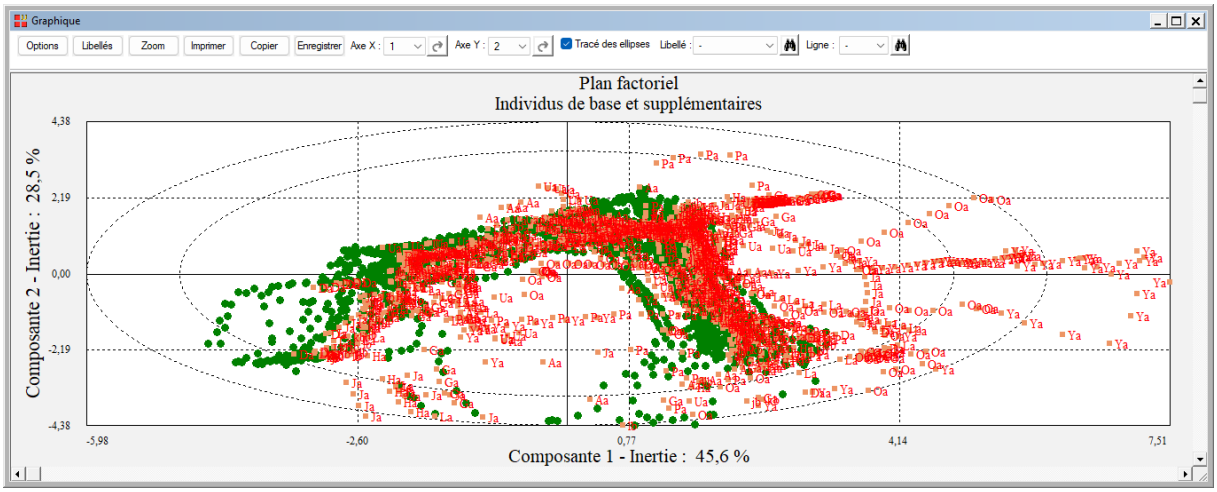
Presque toutes les données de base sont à l'intérieur les ellipses.

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

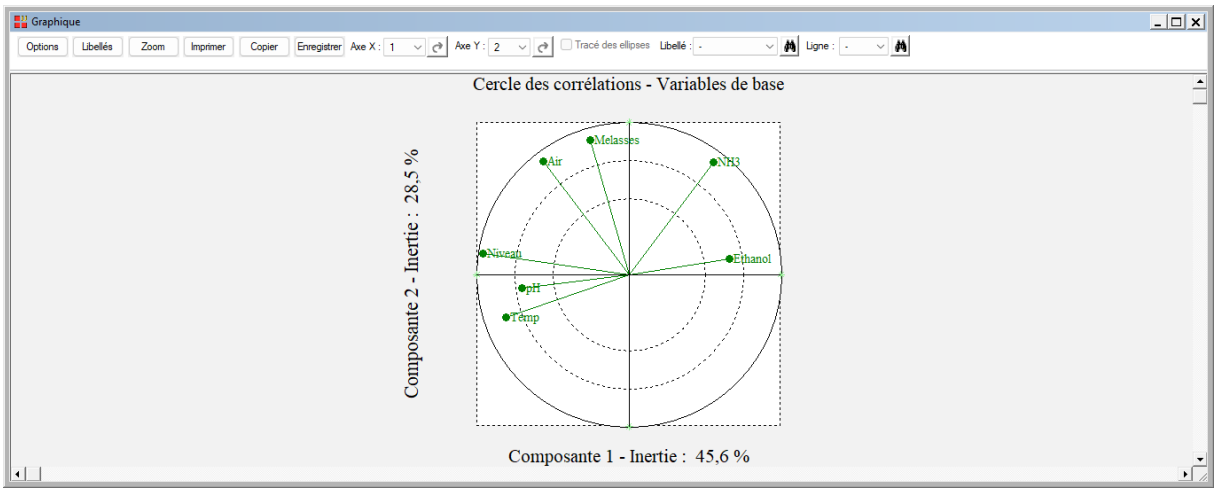
Ce graphique montre que plusieurs des nouveaux lots ne suivent pas les trajectoires des lots de référence. Ce sont principalement les lots Oa, Pa et Ya.





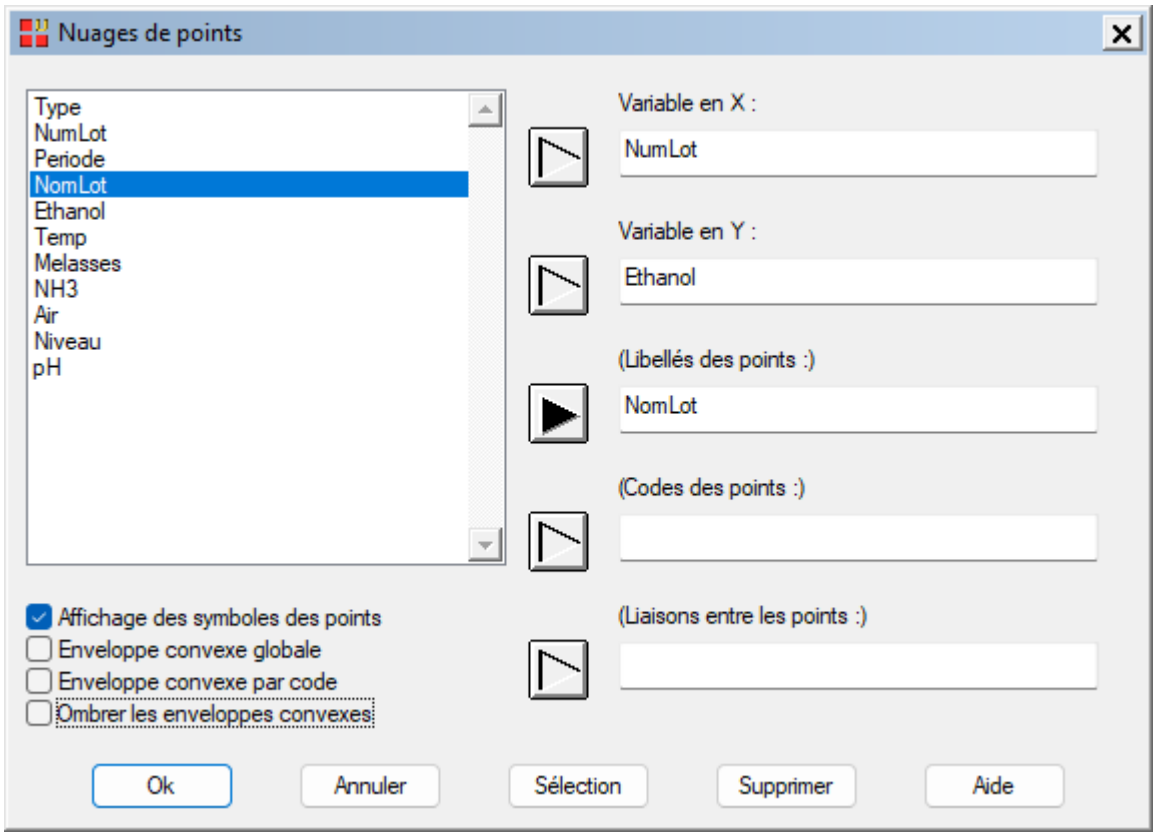


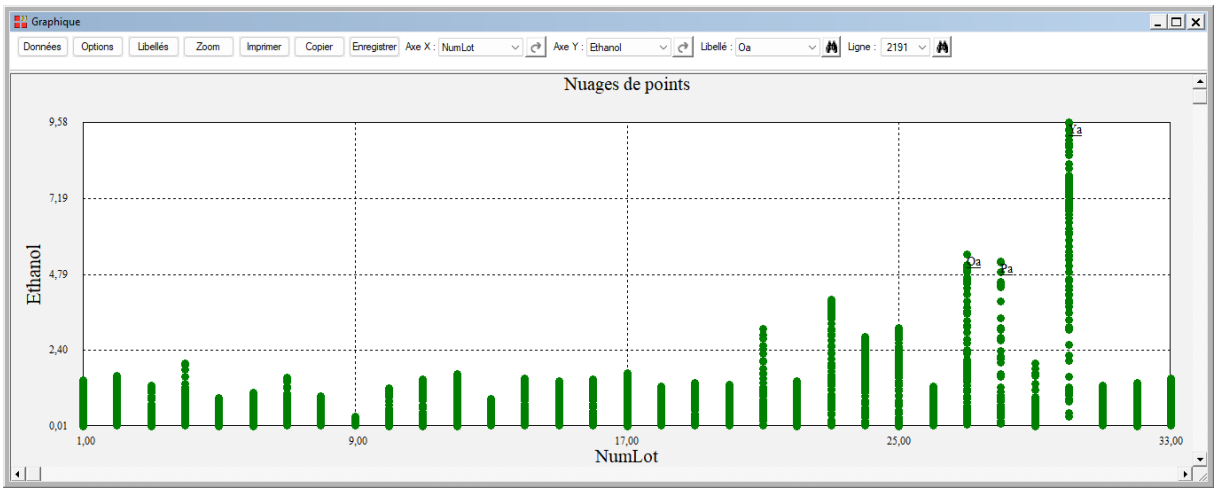
Affichons le cercle des corrélations.



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.

Variable	*Contenu*
composante	Composantes principales
composantesup	Coordonnées des individus supplémentaires
contribindivbase	Contributions à l'inertie (individus de base
contribtotindivbase	Contributions à l'inertie totale (individus de base)
contribvarbase	Contributions à l'inertie (variables de base)
coorvarbase	Coordonnées des variables de base
coorvarqualisup	Coordonnées des variables qualitatives supplémentaires
coorvarquantibase	Coordonnées des variables quantitatives de base
coorvarquantisup	Coordonnées des variables quantitatives supplémentaires
corpearsonbase	Matrice des corrélations de Pearson des variables de base
correlsimplevarbase	Corrélations simples des variables de base
correlsimplevarsup	Corrélations simples des variables supplémentaires
corspearmanbase	Matrice des corrélations des rangs de Spearman des variables de base
cosindivbase	Cosinus carrés (individus de base)
cosvarbase	Cosinus carrés (variables de base)
covarbase	Matrice des covariances des variables de base
cvvarbase	Coefficients de variation des variables quantitatives de base
distindivbase	Distances carrées à l'origine (individus de base)
distvarbase	Distances carrées à l'origine (variables de base)
ectvarbase	Ecarts-types des variables quantitatives de base
facteur	Facteurs colonnes
indivbase	Libellés des individus de base
indivsup	Libellés des individus supplémentaires

modvarqualisup	Libellés des modalités des variables qualitatives supplémentaires
moyvarbase	Moyennes des variables quantitatives de base
t2indivbase	T2 de Hotelling (individus de base)
t2indivsup	T2 de Hotelling (individus supplémentaires)
varqualisup	Libellés des variables qualitatives supplémentaires
varquantibase	Libellés des variables quantitatives de base
varquantisup	Libellés des variables quantitatives supplémentaires