

UNIWIN VERSION 9.7.0

METHODE STATIS DUALE

Révision : 02/09/2023

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichiers CANTONS4.....	3
L'option Rapports	7
L'option Graphiques	8
Exemple 2 : Fichier TELECOM.....	13
Les variables internes créées par la procédure	20

Définition

La méthode Statis duale (Structuration de Tableaux A Trois Indices de la Statistique) permet l'exploration simultanée de plusieurs tableaux de données.

Elle s'applique à des données quantitatives qui sont des tableaux de mesures recueillis en différentes occasions sur les mêmes variables, les individus pouvant être éventuellement différents selon les tableaux.

Si le tableau croise les mêmes variables et les mêmes individus, il peut être intéressant de réaliser également une analyse par la méthode Statis.

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

Un rapport général de synthèse est proposé ainsi que les graphiques des plans factoriels du compromis (variables), de l'interstructure (groupes des variables) et de l'intrastructure (variables et trajectoires des variables).

De nombreux outils d'aide à l'interprétation sont fournis, notamment matrices des corrélations pour les groupes, inerties des groupes, normes des groupes, produits scalaires normés et distances entre les matrices des corrélations, décompositions de ces distances par variable, pondérations des matrices des corrélations, matrice des corrélations du compromis, produits scalaires et distances entre la matrice compromis

et les matrices des groupes, résultats pour les groupes et le compromis (distances à l'origine, contributions, cosinus carrés) ainsi que pour les variables pour les divers groupes.

Entrée des données

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

Méthode STATIS duale

Variables quantitatives:

Codes des groupes de variables:

(Libellés des variables quantitatives:)

(Libellés des individus:)

Codes des variables:

(Libellés des groupes de variables:)

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de définir les variables quantitatives, les codes des groupes de variables et les codes des variables.

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

Les zones de libellés sont optionnelles. Si elles ne sont pas renseignées, UNIWIN génère automatiquement des libellés.

La zone 'Codes des groupes de variables' permet d'indiquer le groupe auquel chaque variable appartient.

La zone optionnelle 'Codes des variables' permet, dans le cas de variables mesurées à différents instants (groupes) d'indiquer les colonnes du fichier qui représentent une même variable. Dans ce cas, le logiciel permettra de visualiser les trajectoires des variables au travers des groupes.

Données manquantes

Dans cette procédure les données manquantes ne sont pas permises.

Exemple 1 : Fichiers CANTONS4

Nous utiliserons le fichier CANTONS4 pour illustrer ce premier exemple.

Ce fichier décrit 30 cantons ruraux de l'Hérault par l'évolution des populations actives entre 1954 et 1982. Voir l'exemple 2 du chapitre 'Analyse Factorielle Multiple' pour une représentation cartographique de ces cantons. Il est issu du livre 'Analyse conjointe de tableaux quantitatifs' de Christine Lavit (Masson).

Quatre recensements de la population ont été effectués en 1954, 1962, 1968 et 1975. Les professions sont réparties en 9 classes de la façon suivante :

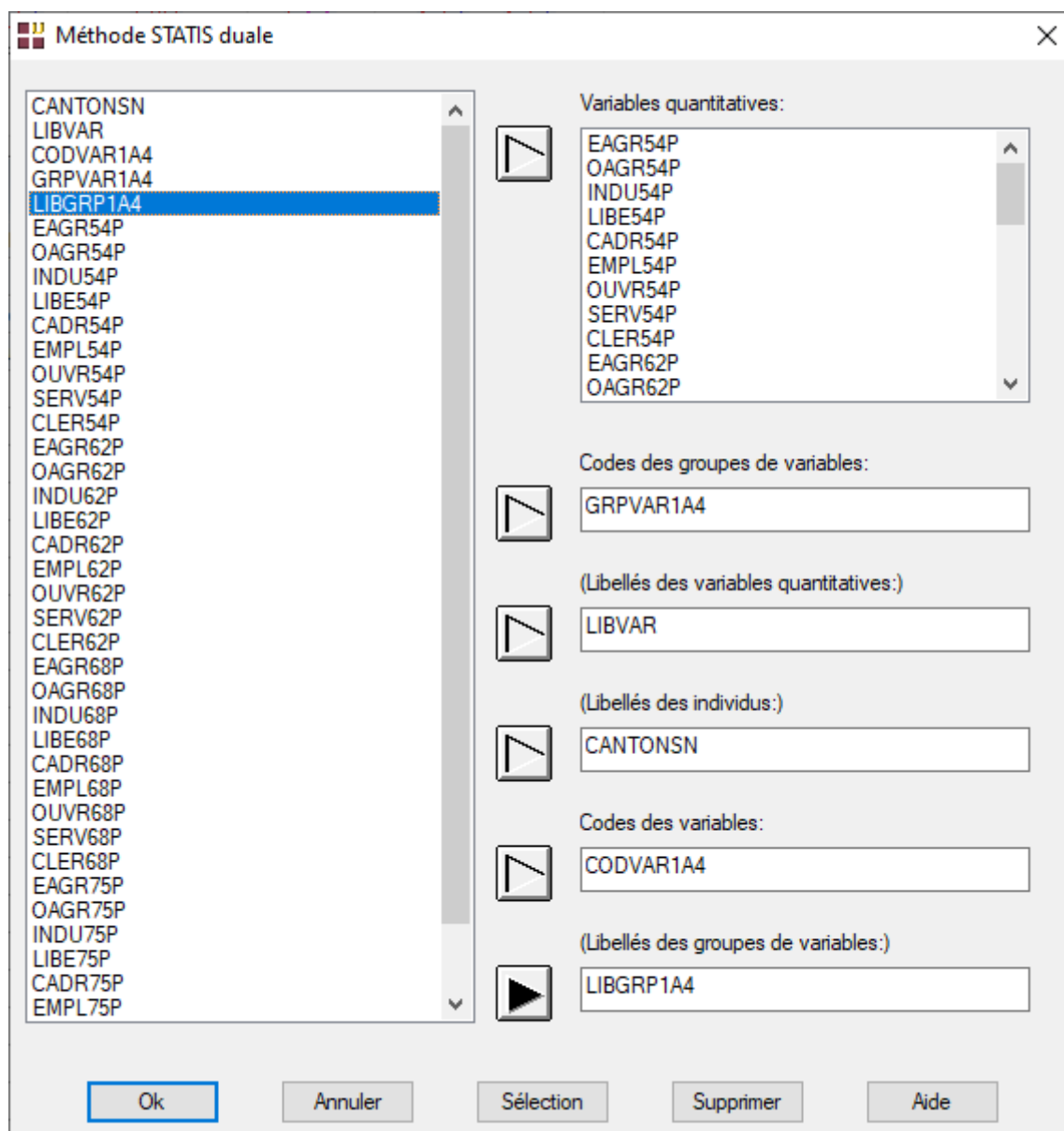
<i>Variable du fichier</i>	<i>Contenu</i>
EAGRnnP	Exploitants agricoles
OAGRnnP	Ouvriers agricoles
INDUnnP	Industriels, commerçants
LIBEnnP	Professions libérales
CADRnnP	Cadres moyens
EMPLnnP	Employés
OUVRnnP	Ouvriers
SERVnnP	Services
CLERnnP	Clergé, Armée, Police

Pour chaque nom de variable, nn indique l'année du recensement. Le P final indique que ces variables contiennent les profils en pourcentages des diverses catégories socioprofessionnelles pour les cantons. Ces pourcentages varient de façon importante en fonction des catégories socioprofessionnelles, les profils ont été centrés et réduits. Par exemple EAGR54P est la variable profil centrée et réduite caractérisant les exploitants agricoles en 1954 dans les 30 cantons.

De plus, dans ce fichier, la variable CANTONSN contient les noms des 30 cantons, la variable LIBVAR contient les noms des 36 variables des 4 recensements, la variable CODVAR1A4 contient le codage des variables en fonction des recensements (un code est ainsi attribué 4 fois à une même variable au travers des 4 tableaux), la variable GRP1A4 définit l'appartenance des variables à chacun des 4 tableaux et enfin la variable LIBGRP1A4 donne les libellés des 4 groupes de variables, ici les années des recensements.

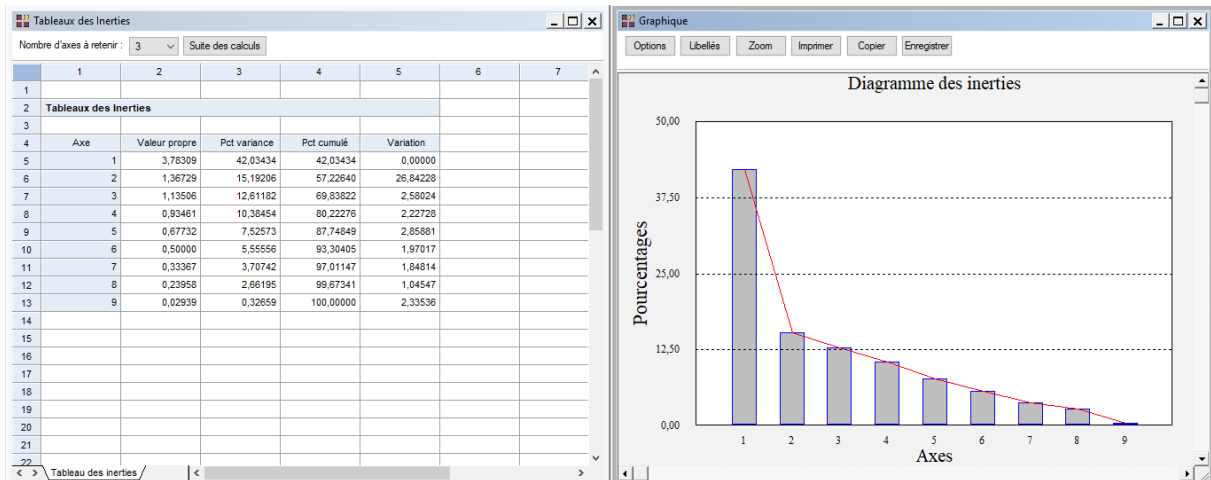
Cliquons sur l'icône STATIS – Statis duale dans le ruban Décrire.

Renseignons la boîte de dialogue comme montré ci-dessous (on sélectionne les variables EAGR54P à CLER75P).



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

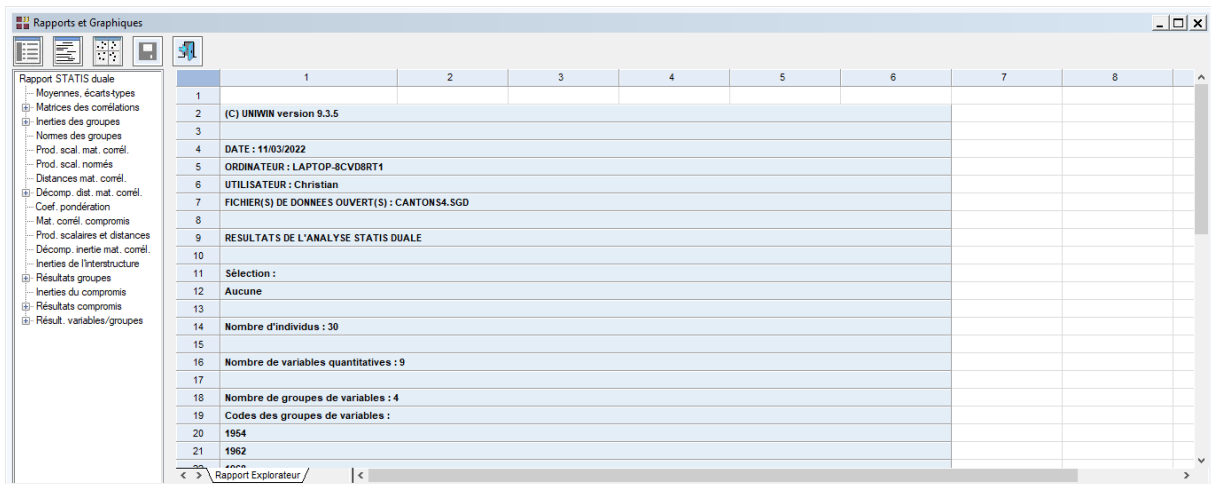
Après quelques instants, un tableau et un diagramme précisant les inerties expliquées par les différents vecteurs propres issus de l'analyse du compromis (infrastructure) apparaît.





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

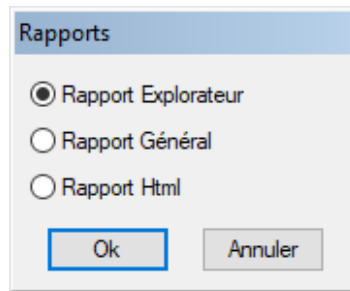
Cliquons sur le bouton 'Suite des calculs'.


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

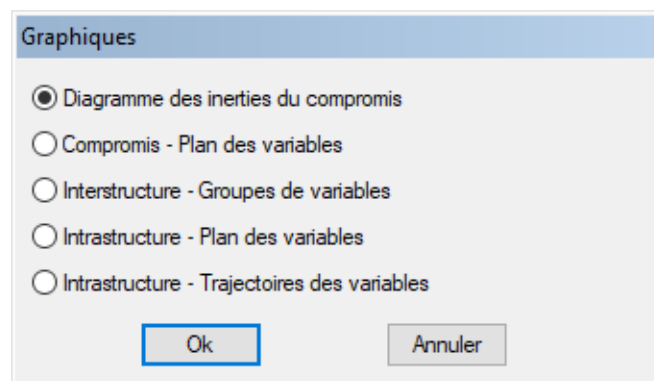



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

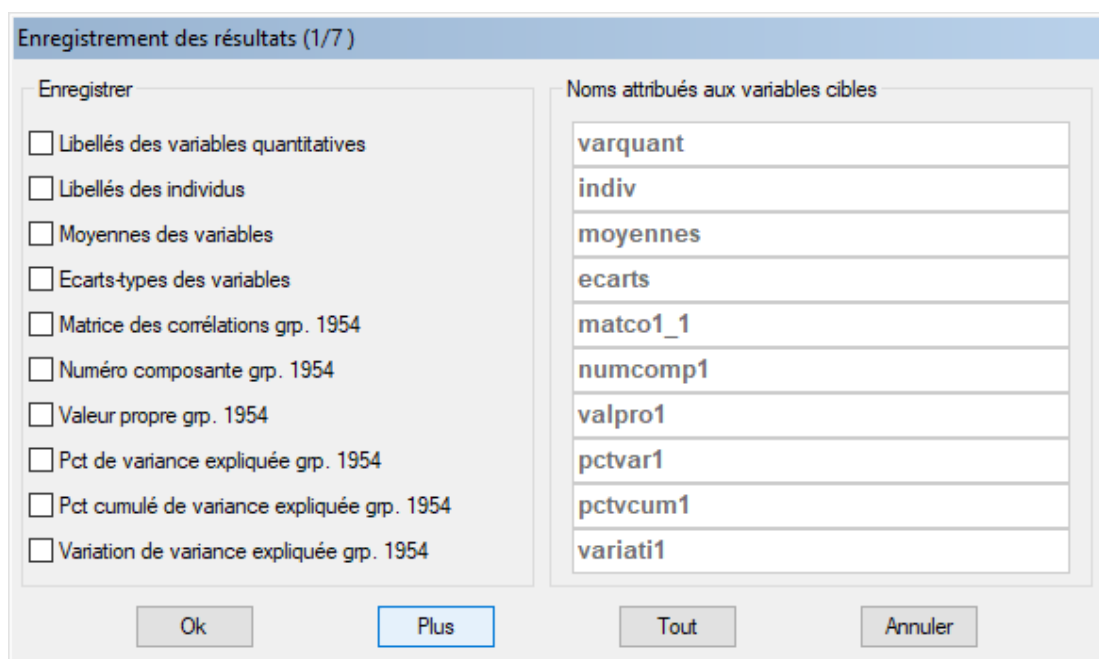
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



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

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

L'option Rapports

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

Voici trois exemples du rapport pour notre méthode Statis duale : Explorateur, Général, HTML.

	1	2	3	4	5	6	7	8
1								
2	RESULTATS COMPROMIS VARIABLES POUR LE FACTEUR : 1							
3								
4								
5		DISTANCE^2	COORDONNEE	CONTRIBUTION	COSINUS^2	COS^2 CUMULE		
6	EAGR	1	-0,77423	15,84522	0,59944	0,59944		
7	OAGR	1	-0,55919	8,26546	0,31269	0,31269		
8	INDU	1	0,39381	4,09943	0,15509	0,15509		
9	LIBE	1	0,74176	14,54379	0,55020	0,55020		
10	CADR	1	0,78220	15,35636	0,58094	0,58094		
11	EMPL	1	0,75854	15,20915	0,57538	0,57538		
12	OUVR	1	0,70994	13,32271	0,50401	0,50401		
13	SERV	1	0,69325	12,70391	0,48060	0,48060		
14	CLER	1	0,15729	0,65398	0,02474	0,02474		
15								
16								
17								
18								
19								
20								
21								

	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 9.3.5							
3								
4	DATE :	11/03/2022						
5	ORDINATEUR :	LAPTOP-8CVD8RT1						
6	UTILISATEUR :	Christian						
7	FICHER(S) DE DONNEES OUVERT(S) :	CANTONS4.SGD						
8								
9	RESULTATS DE L'ANALYSE STATIS DUALE							
10								
11	Sélection :	Aucune						
12								
13								
14	Nombre d'individus :	30						
15								
16	Nombre de variables quantitatives :	9						
17								
18	Nombre de groupes de variables :	4						
19	Codes des groupes de variables :	1954						
20								
21								

	1	2	3	4	5	6	7	8
1	(C) UNIWIN version 9.3.5							
2	DATE :	11/03/2022						
3	ORDINATEUR :	LAPTOP-8CVD8RT1						
4	UTILISATEUR :	Christian						
5	FICHER(S) DE DONNEES OUVERT(S) :	CANTONS4.SGD						
6								
7	RESULTATS DE L'ANALYSE STATIS DUALE							
8	Sélection :	Aucune						
9								
10	Nombre d'individus :	30						
11								
12	Nombre de variables quantitatives :	9						
13								
14	Nombre de groupes de variables :	4						
15	Codes des groupes de variables :	1954						
16								
17								
18								
19								
20								
21								

Ce rapport affiche les informations suivantes :

- ◇ Moyennes et écarts-types des variables
- ◇ Matrices des corrélations des variables pour les groupes
- ◇ Tableaux des inerties des groupes
- ◇ Normes des groupes
- ◇ Produits scalaires entre les matrices des corrélations
- ◇ Produits scalaires normés entre les matrices des corrélations
- ◇ Distances entre les matrices des corrélations
- ◇ Décompositions des distances entre les matrices des corrélations des groupes (contributions des variables en pourcentages)
- ◇ Moyenne générale de ces distances (contributions des variables en pourcentages)
- ◇ Coefficients de pondération des matrices des corrélations
- ◇ Matrice des corrélations du compromis
- ◇ Produits scalaires et distances entre la matrice compromis et les matrices des corrélations des groupes
- ◇ Décomposition de l'inertie selon les matrices des corrélations
- ◇ Tableau de l'inertie de l'interstructure
- ◇ Résultats pour les groupes par facteur (distances carrées à l'origine, coordonnées, contributions, cosinus carrés, cosinus carrés cumulés)
- ◇ Tableau des inerties du compromis
- ◇ Résultats pour le compromis par facteur (distances carrées à l'origine, coordonnées, contributions, cosinus carrés, cosinus carrés cumulés)
- ◇ Résultats pour les variables par groupe et facteur

A la lecture de ces résultats, on remarque que les distances entre les matrices de corrélations augmentent dans le temps.

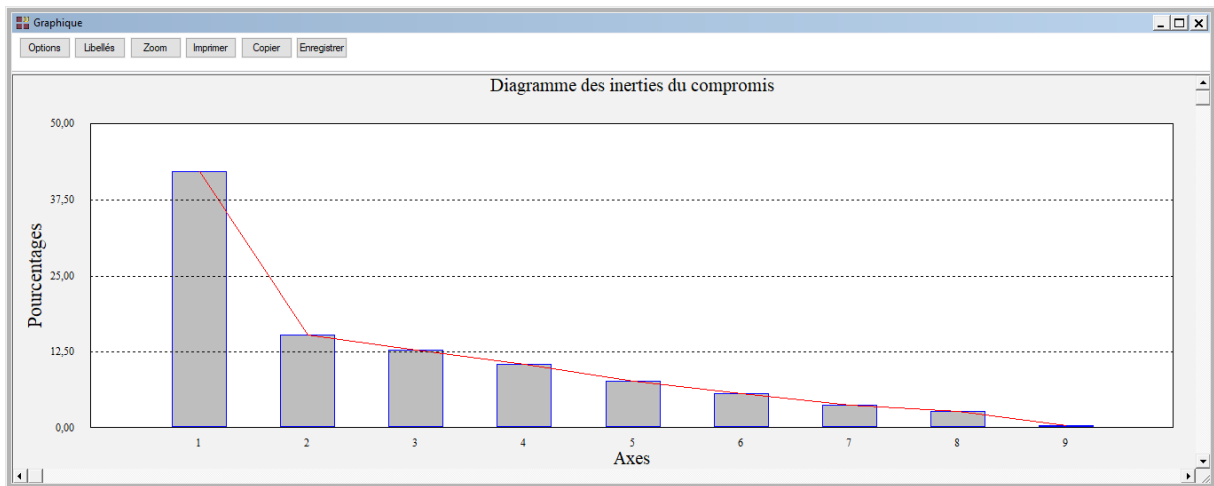
Les tableaux des contributions des variables mettent en évidence que ces distances sont dues principalement aux catégories socioprofessionnelles « professions libérales », « industriels et commerçants » ainsi que dans une moindre mesure aux « ouvriers » et « clergé, armée, police ».

Ces informations sont présentées de façon synthétique dans le tableau des moyennes générales des distances.

L'option Graphiques

- Diagramme des inerties du compromis

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



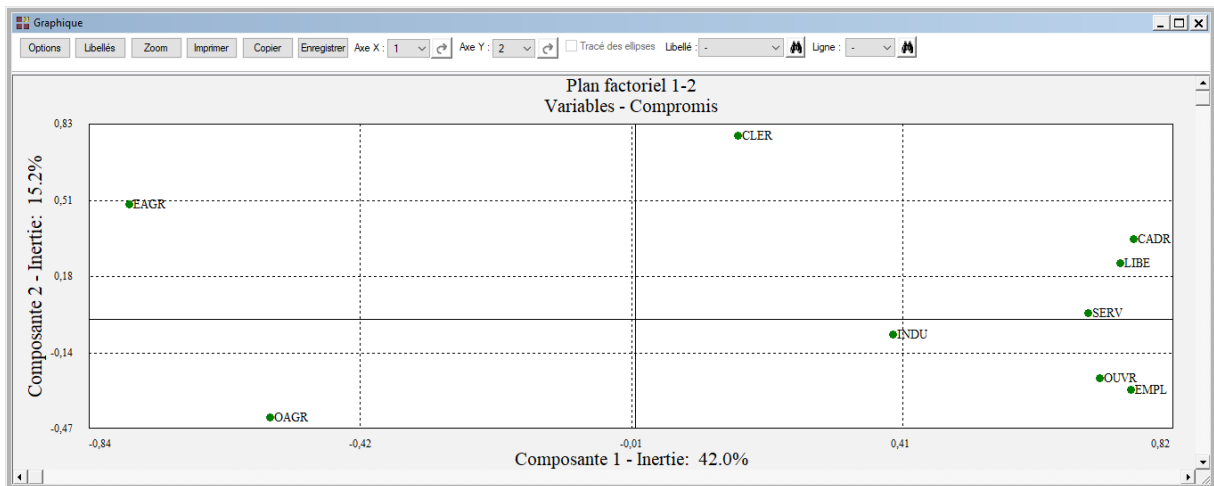
- L'option Compromis – Plan des variables

Cette option permet d'afficher un plan factoriel du compromis des variables. C'est une représentation graphique des 'variables moyennes'. 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.

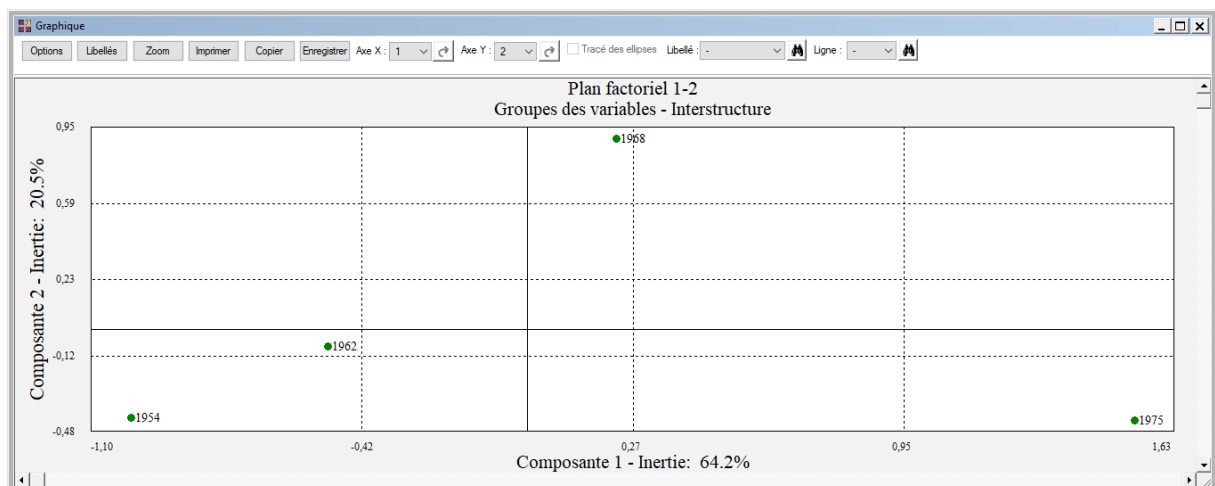
A noter également le bouton Sélection qui permet de n'afficher que les points qui satisfont une condition logique, par exemple une condition basée sur les contributions ou les cosinus carrés.

Ceci est très pratique dans le cas d'un nuage comportant beaucoup d'individus : on n'affiche que ceux qui sont significatifs d'un point de vue statistique.



- L'option Interstructure – Groupes de variables

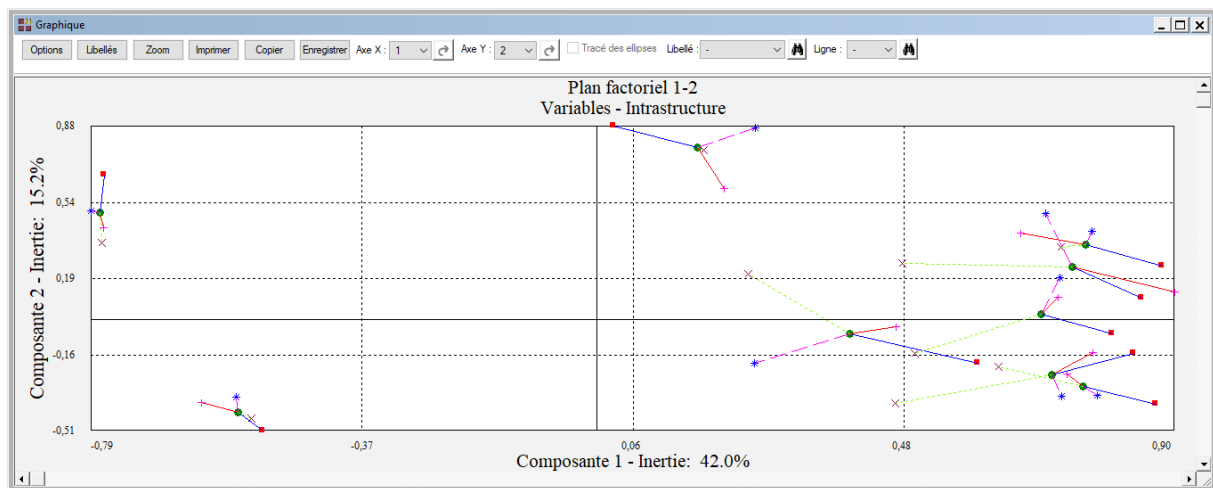
Cette option permet d'afficher le plan factoriel des groupes. La coordonnée d'un groupe s'interprète comme un indice de la liaison entre le groupe et l'axe factoriel. Dans ce graphique, l'origine (0, 0) est le groupe compromis. La boîte de dialogue montrée précédemment s'affiche alors permettant de sélectionner le plan factoriel et d'indiquer si on désire ou non l'affichage des libellés des groupes.



- L'option Intrastructure – Plan des variables

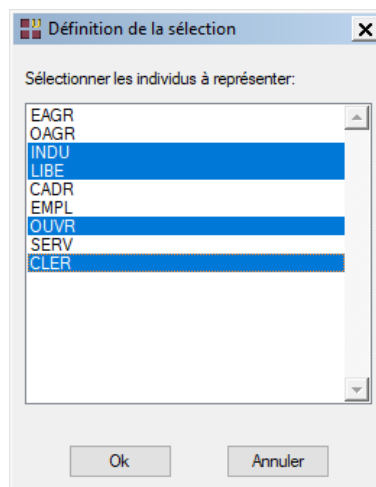
Cette option permet d'afficher simultanément et de façon superposée les plans factoriels des nuages issus des analyses séparées des groupes c'est à dire les positions des variables dans chacune des analyses séparées et les positions des points compromis. La boîte de dialogue montrée précédemment s'affiche alors permettant de sélectionner le plan factoriel et d'indiquer si on désire ou non l'affichage des libellés des variables.

Regardons d'abord le plan sans les libellés des points. A partir de chaque point compromis sont tracées des droites qui relient ce point compromis aux points représentant cet individu dans les analyses séparées des différents groupes.

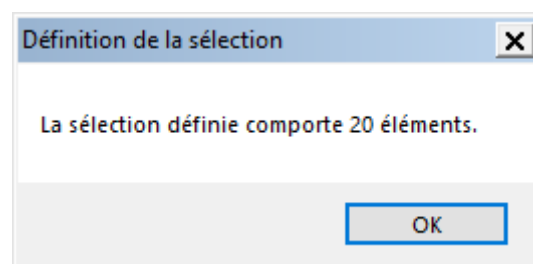


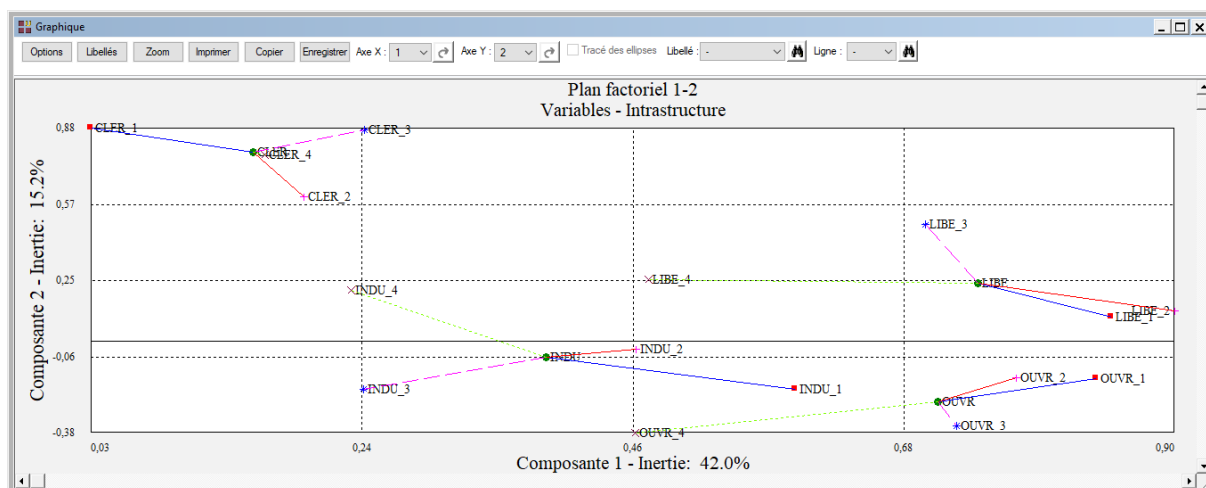
La couleur de ces droites est fonction du groupe de variables. Ainsi toutes les droites reliant les points compromis aux points d'un groupe sont de la même couleur.

Ce graphique pouvant être rapidement peu lisible, utilisons maintenant le bouton Sélection de la boîte de dialogue de choix du plan pour définir les points pour lesquels nous désirons visualiser les points partiels issus des analyses séparées. Sélectionnons les variables LIBE (professions libérales), INDU (industriels et commerçants), OUVR (ouvriers) et CLER (clergé, armée, police).

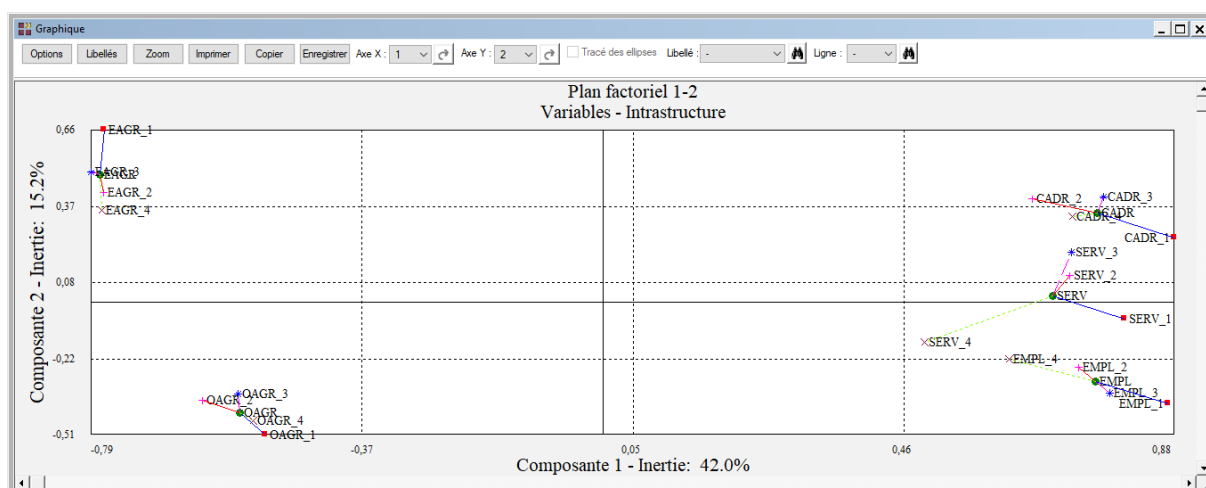


Le logiciel nous indique, à titre informatif, le nombre de points représentés dans le graphique (4 variables au travers de 4 recensements plus le compromis, soit 20 points)





Visualisons également le graphique de EAGR (exploitants agricoles), OAGR (ouvriers agricoles), CADR (cadres moyens), EMPL (employés) et SERV (services).



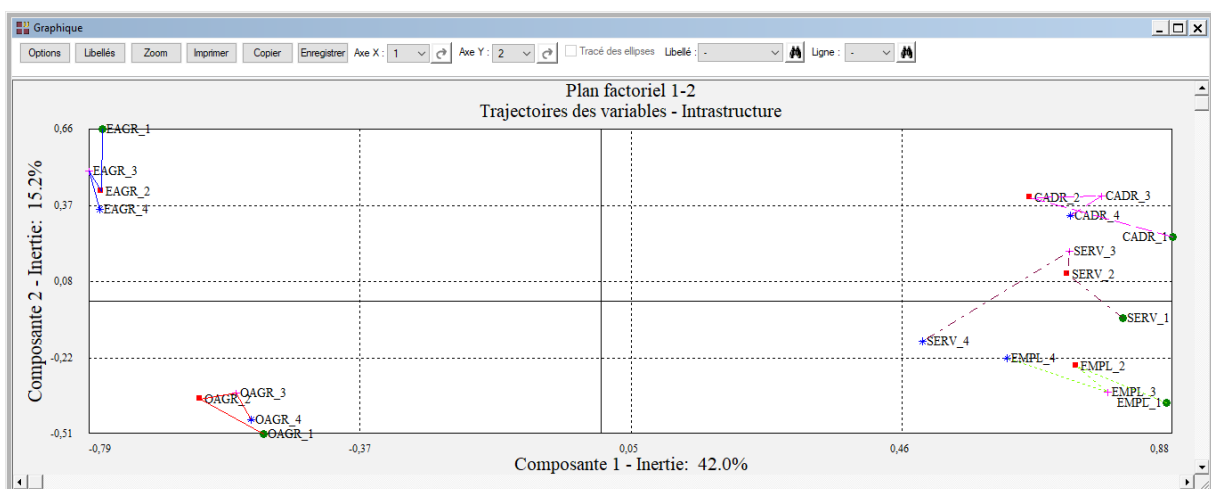
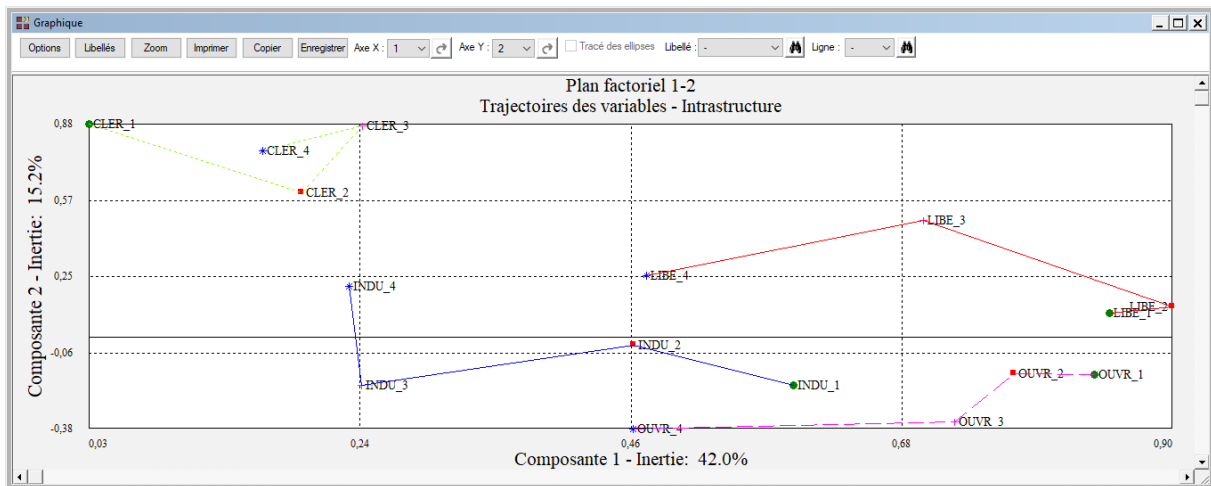
- L'option Infrastructure – Trajectoire des variables

Cette option permet d'afficher les trajectoires des variables au travers des 4 recensements. Les différentes trajectoires des variables nous permettent de les regrouper. Cette trajectoire s'interprète par les écarts de l'évolution de la variable par rapport à l'évolution moyenne. La boîte de dialogue montrée précédemment s'affiche alors permettant de sélectionner le plan factoriel et d'indiquer si on désire ou non l'affichage des libellés des variables.

Ce graphique pouvant être rapidement peu lisible, utilisons maintenant le bouton Sélection de la boîte de dialogue de choix du plan pour définir les points pour lesquels nous désirons visualiser les points partiels issus des analyses séparées.

Sélectionnons les variables LIBE (professions libérales), INDU (industriels et commerçants), OUVR (ouvriers) et CLER (clergé, armée, police).

Visualisons également le graphique de EAGR (exploitants agricoles), OAGR (ouvriers agricoles), CADR (cadres moyens), EMPL (employés) et SERV (services).



Exemple 2 : Fichier TELECOM

Nous utiliserons le fichier TELECOM pour illustrer ce deuxième exemple.

Les tableaux de données traités sont extraits du livre de M. Jambu (Dunod, 1989) « Exploration informatique et statistique de données ». Le lecteur pourra également lire avec intérêt dans le numéro 27 de la revue Modulad (Inria, 2001) l'article « Analyse exploratoire d'une suite de tableaux de données » de Catherine Pardoux.

Les tableaux concernent l'étude des composantes de la qualité de service du réseau téléphonique par région en France métropolitaine. Chaque tableau comporte 22 lignes correspondant aux 22 régions « Télécom » identifiées par le nom de la ville représentative de la région, et sept colonnes correspondant à 7 indicateurs mesurant la qualité de service offerte aux usagers : six indicateurs élémentaires et un indice global de la qualité de service (noté IGQS), combinaison d'indicateurs élémentaires. Ces indicateurs représentant des pourcentages, des taux de signalisation, des vitesses de relevé, définissent un ensemble de variables hétérogènes.

On dispose de trois tableaux de ce type pour trois années successives : 1984, 1985 et 1986 et d'un autre tableau (pour 1984) avec huit indicateurs technico-économiques.

Les trois tableaux de données possèdent les mêmes variables et les mêmes individus.

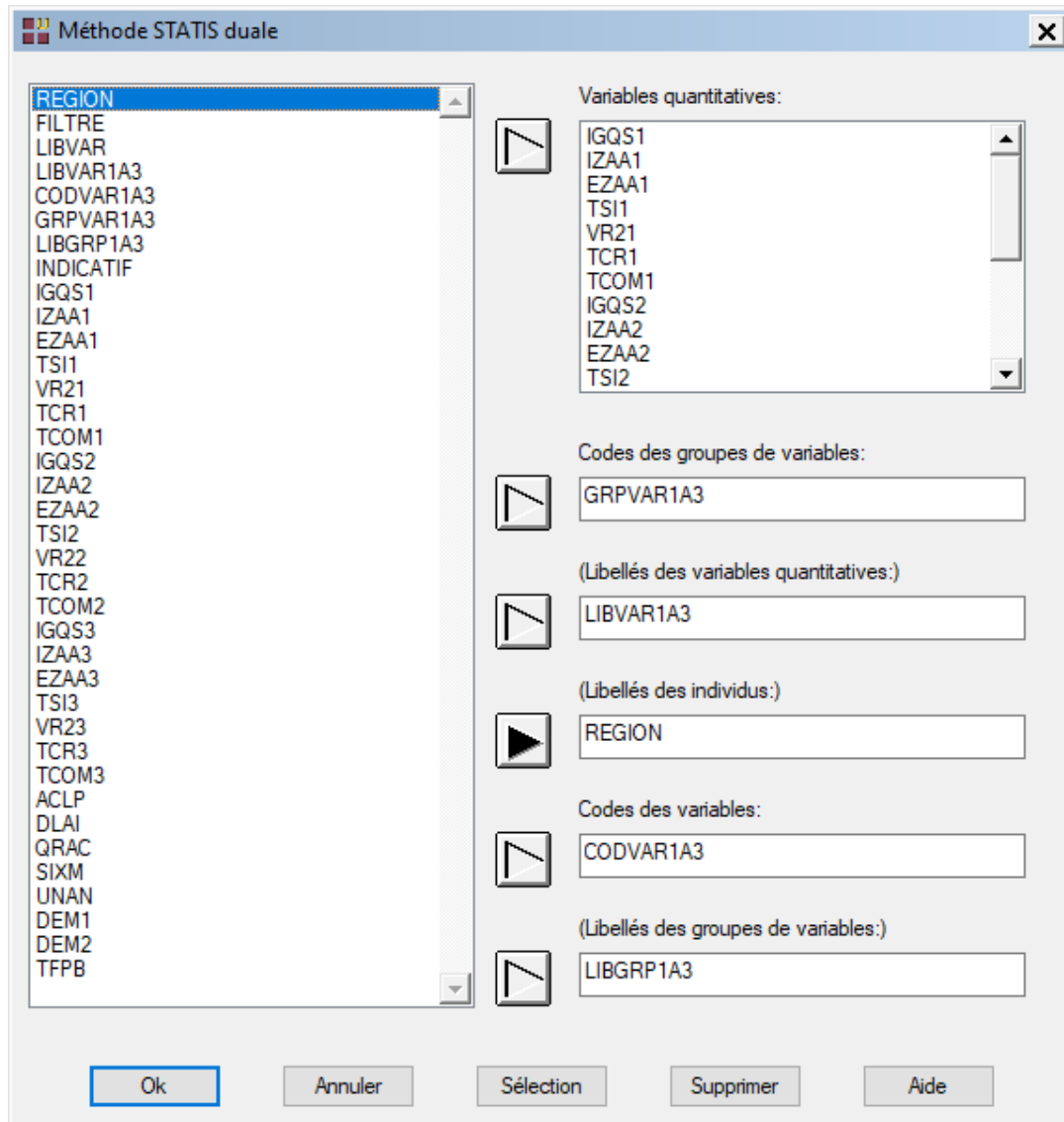
Ainsi nous pourrons mettre en œuvre successivement une analyse Statis duale puis une analyse Statis.

La variable IGQS étant une variable synthétique pour laquelle Ajaccio apparaît comme une valeur extrême pour chacune des trois années, cette région ne sera pas prise en compte dans les analyses.

<i>Variable du fichier</i>	<i>Contenu</i>
IGQS	Indice global de qualité de service (combinaison de 6 autres indicateurs, dont IZAA, EZAA, TSI et VR2)
IZAA	Pourcentage d'appels efficaces par rapport au nombre des appels correctement effectués par les usagers en ne tenant compte que des appels <u>intérieurs</u> à la zone d'autonomie d'acheminement ou à la zone urbaine
EZAA	Pourcentage d'appels efficaces par rapport au nombre des appels correctement effectués par les usagers en ne tenant compte que des appels <u>extérieurs</u> à la zone d'autonomie d'acheminement ou à la zone urbaine
TSI	Taux de signalisation des dérangements (pourcentage du nombre de signalisation de dérangements reçus et traités durant une période donnée par rapport au nombre de lignes principales en service)
VR2	Vitesse de relevé des dérangements dans les 2 jours (pourcentage de dérangements relevés le jour même et le lendemain de la signalisation par rapport au nombre de dérangements qui ont donné lieu à déplacement d'agents)
TCR	Taux d'efficacité à l'arrivée des centres de renseignements (pourcentage d'appels efficaces par rapport au nombre des appels au service des renseignements)
TCOM	Temps d'établissement des communications (délai exprimé en secondes de mise en relation de deux abonnés)

Cliquons sur l'icône STATIS – Statis duale dans le ruban Décrire.

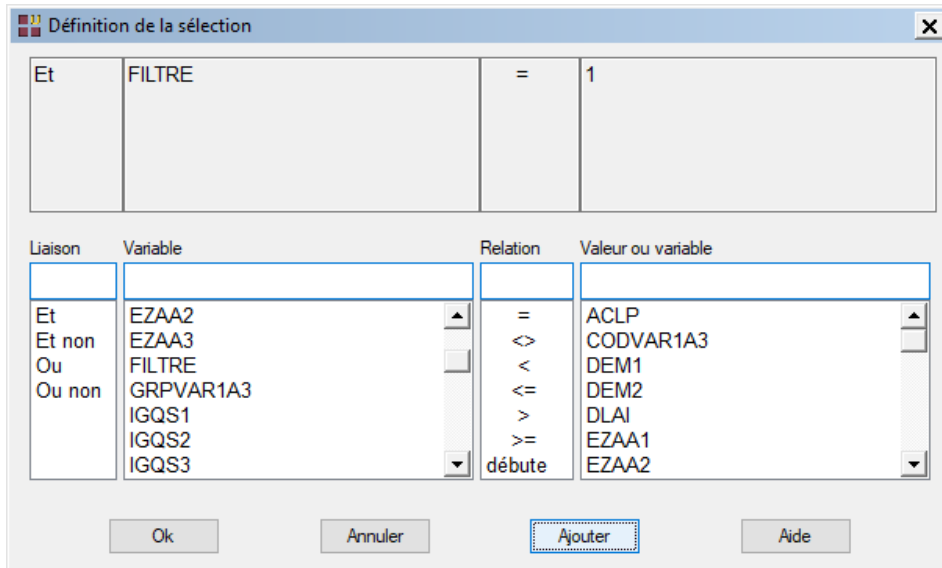
Renseignons la boîte de dialogue comme montré ci-dessous (on sélectionne les variables IGQS1 à TCOM3).



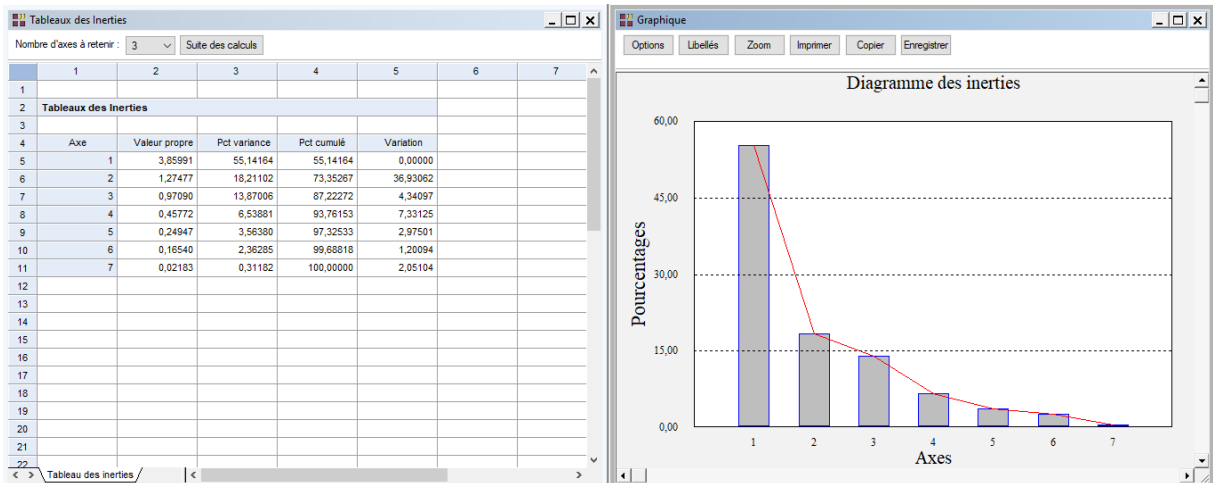
Cliquons sur le bouton Sélection pour sélectionner toutes les régions sauf celle d'Ajaccio (on utilise pour cela la variable FILTRE).

Le logiciel nous indique que 21 régions sont sélectionnées.

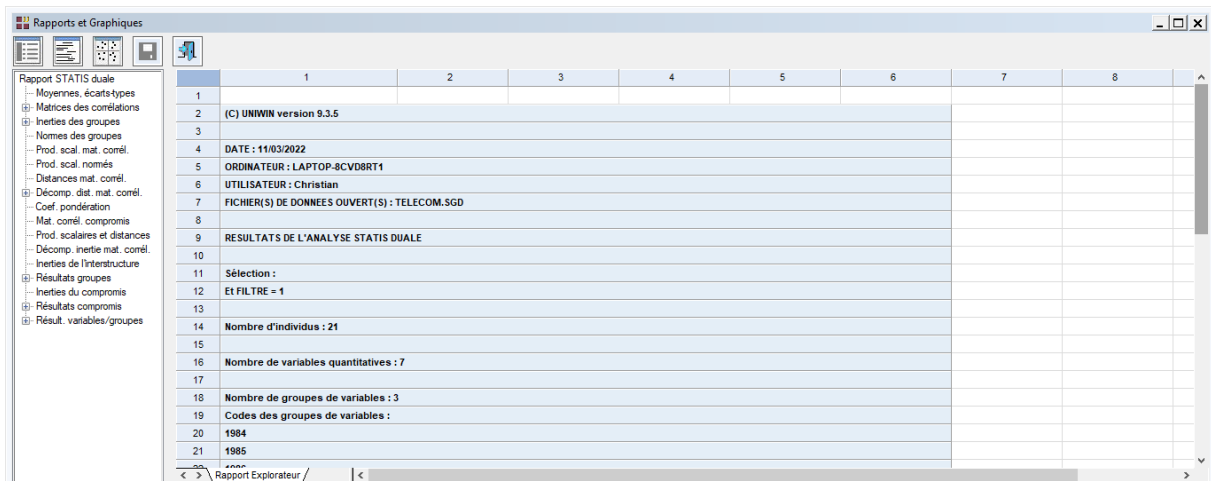
Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.



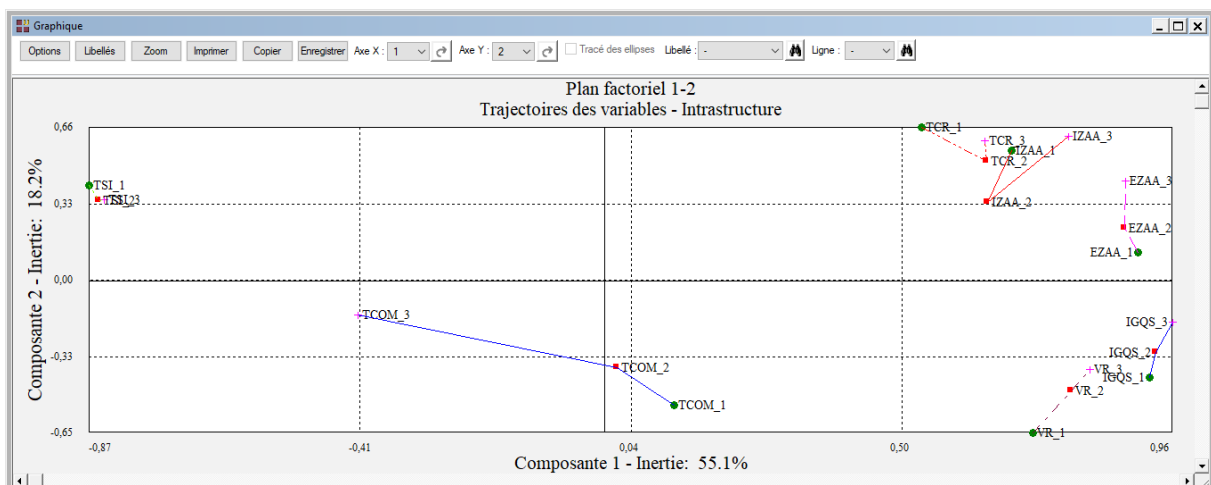
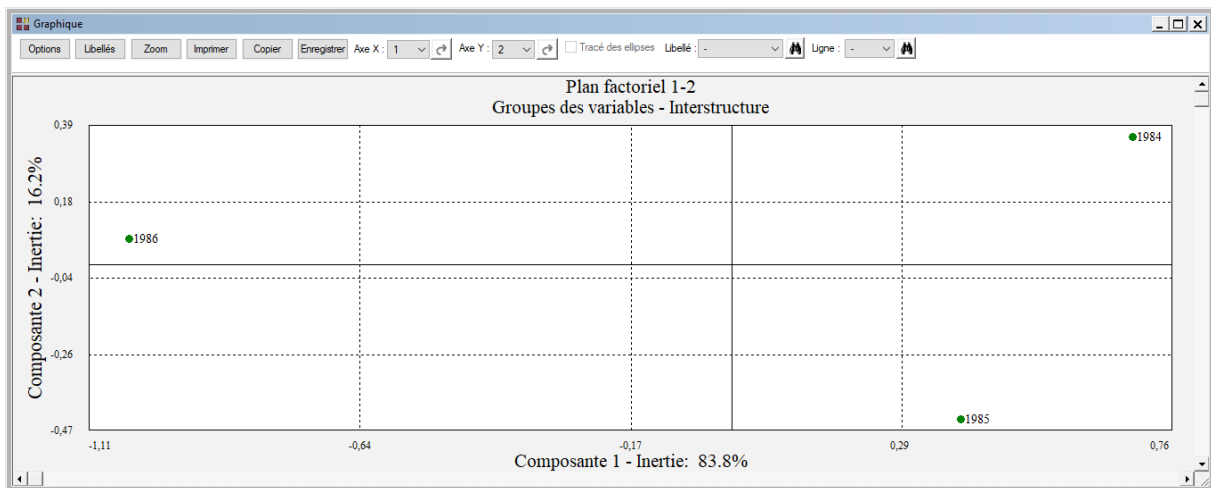
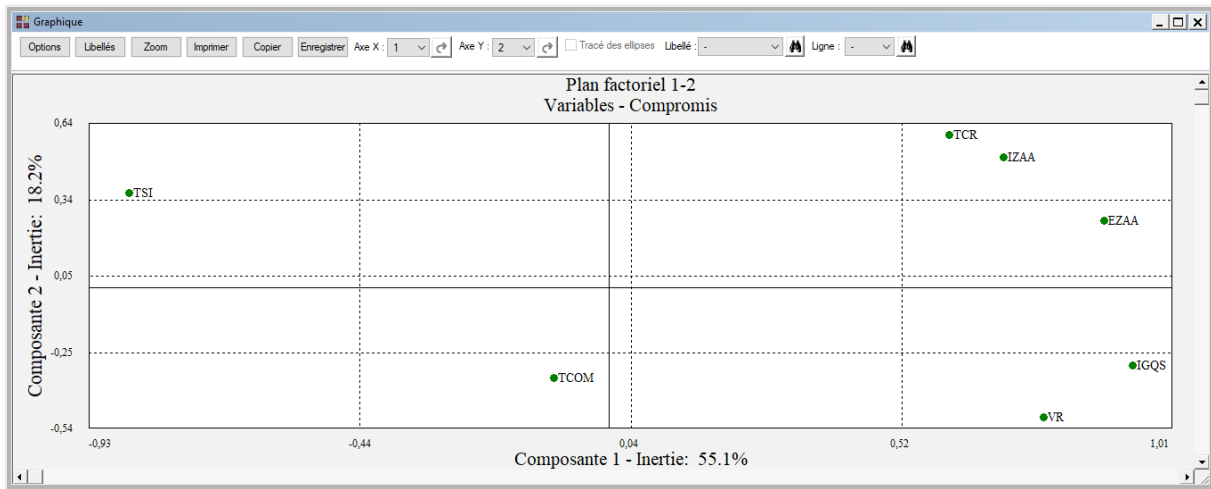
Après quelques instants, un tableau et un diagramme précisant les inerties expliquées par les différents vecteurs propres issus de l'analyse du compromis (infrastructure) apparaît.



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

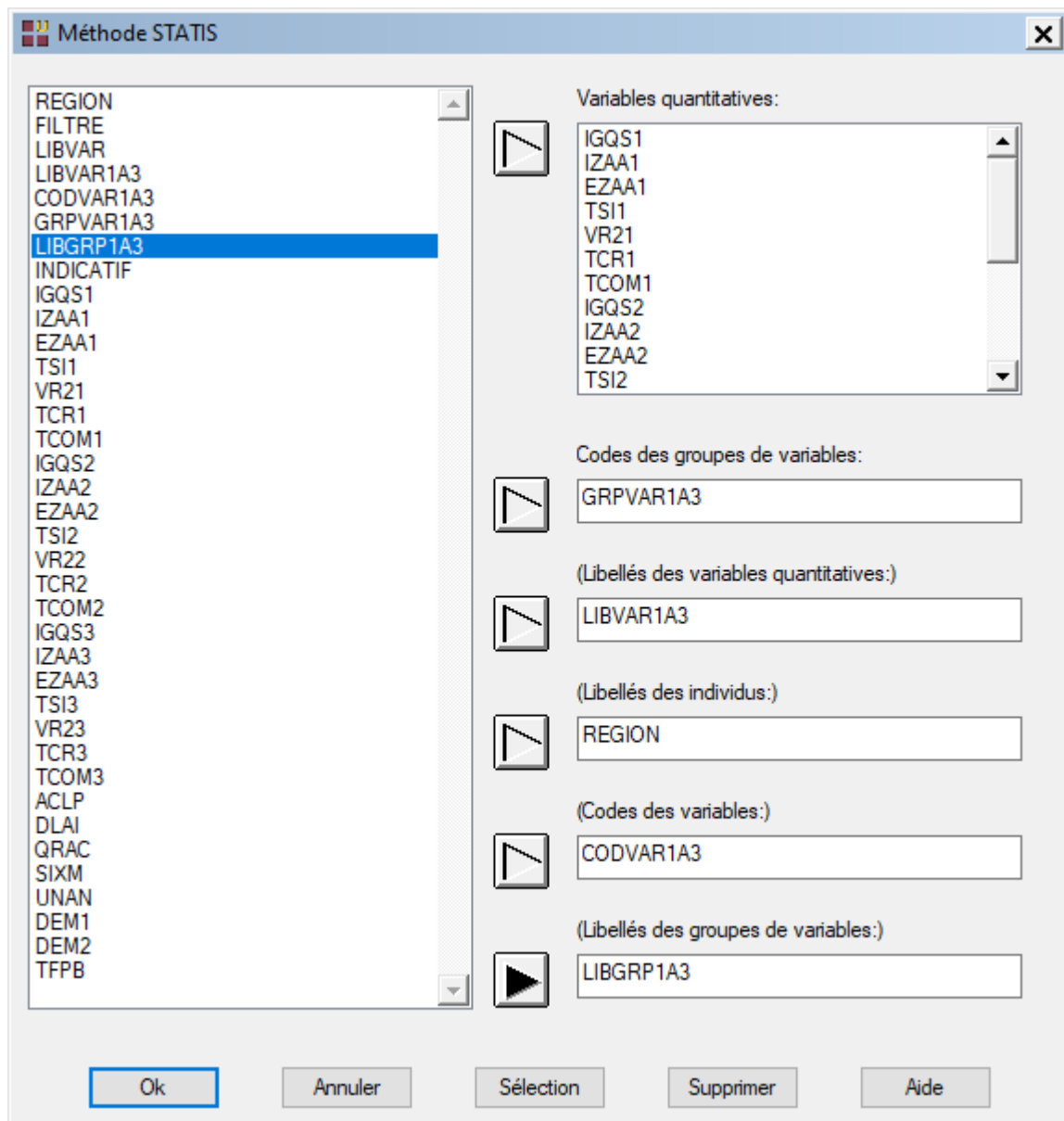


Demandons l'affichage des graphiques du compromis des variables, des groupes de variables (interstructure) et des trajectoires des variables (intrastructure).



Cliquons maintenant sur l'icône STATIS – Statis dans le ruban Décrire.

Renseignons la boîte de dialogue comme montré ci-dessous (on sélectionne les variables IGQS1 à TCOM3).



Cliques sur le bouton Sélection pour sélectionner toutes les régions sauf celle d'Ajaccio (on utilise pour cela la variable FILTRE).

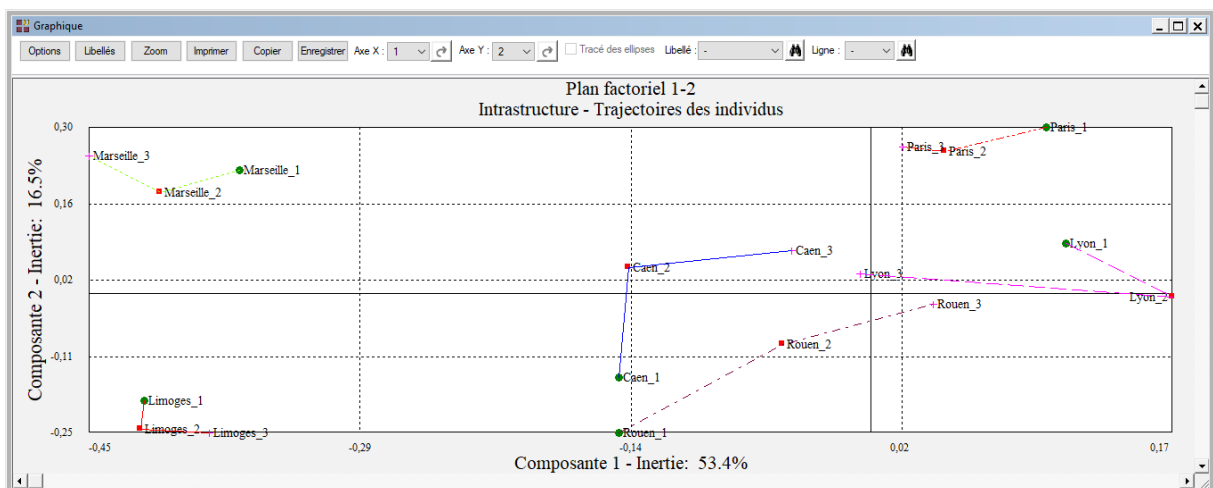
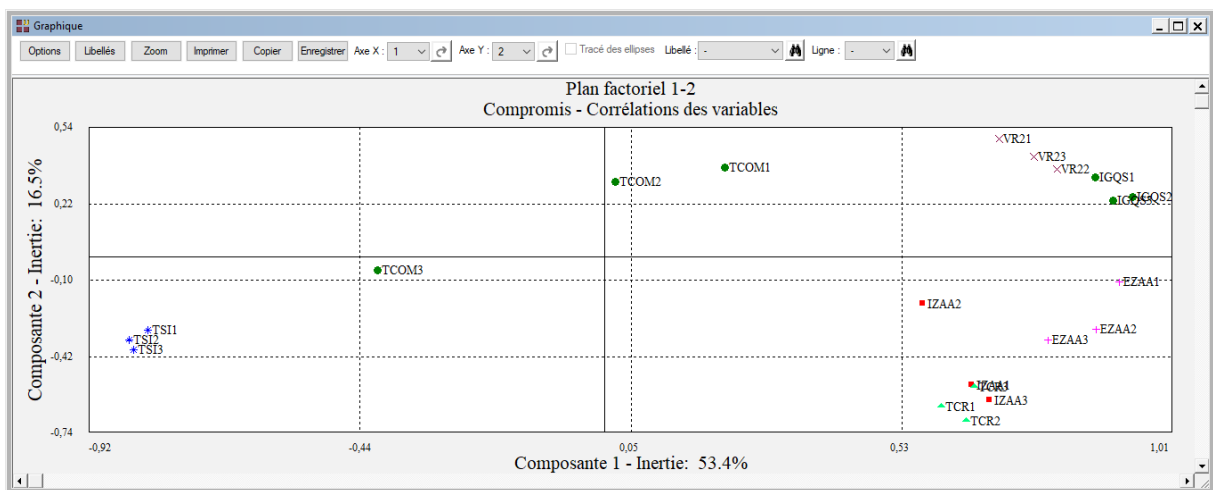
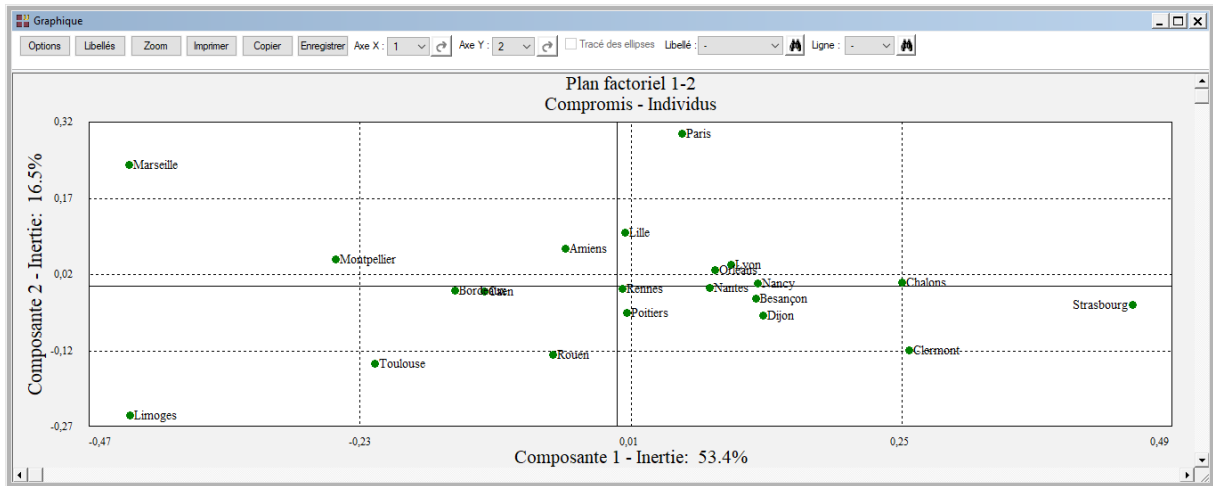
Le logiciel nous indique que 21 régions sont sélectionnées

Cliques alors sur le bouton Ok pour exécuter le traitement de l'analyse.

Après quelques instants, un tableau et un diagramme précisant les inerties expliquées par les différents vecteurs propres issus de l'analyse du compromis (infrastructure) apparaît.

Cliques sur Suite pour sélectionner 3 axes par défaut.

Demandons l'affichage des graphiques du compromis des individus, des corrélations des variables (compromis), et des trajectoires d'individus sélectionnés (infrastructure).



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>
varquant	Libellés des variables quantitatives
indiv	Libellés des individus
moyennes	Moyennes des variables
ecarts	Ecart-types des variables
matco(i)	Matrices des corrélations des variables (groupe i)
numcomp(i)	Numéros des composantes (groupe i)
valpro(i)	Valeurs propres (groupe i)
pctvar(i)	Pourcentages de variance expliquée (groupe i)
pctvcum(i)	Pourcentages cumulés de variance expliquée (groupe i)
variati(i)	Variations de la variance expliquée (groupe i)
norme	Normes des groupes
pscor	Produits scalaires RV
psnor	Produits scalaires RV normés
dmcor	Distances entre matrices des corrélations
dvg(i)	Décompositions des distances entre matrices des corrélations
sdv	Somme des distances entre matrices des corrélations
ponder	Pondérations des matrices des corrélations
compr	Matrice des corrélations du compromis
pscomp	Produits scalaires compromis - matrices des corrélations
dmcomp	Distances compromis - matrices des corrélations
decomposer	Décomposition de l'inertie
numcompg	Numéro des composantes (groupes)
valprog	Valeurs propres (groupes)
pctvarg	Pourcentages de variance expliquée (groupes)
pctvcumg	Pourcentages cumulés de variance expliquée (groupes)
variatiog	Variations de la variance expliquée (groupes)
numcompc	Numéros des composantes (compromis)
valproc	Valeurs propres (compromis)
pctvarc	Pourcentages de variance expliquée (compromis)
pctvcumc	Pourcentages cumulés de variance expliquée (compromis)
variatioc	Variations de la variance expliquée (compromis)
coordgr	Coordonnées des groupes
distacng	Distances carrées à l'origine pour les groupes
contrig	Contributions à l'inertie pour les groupes
cosinug	Cosinus carrés pour les groupes
distances	Distances carrées à l'origine (compromis)
inerlig	Décomposition de l'inertie selon les lignes (compromis)
cosinus	Cosinus carrés (compromis)
contrib	Contributions à l'inertie (compromis)

facteur	Facteurs colonnes (compromis)
pcomp	Composantes principales (compromis)
pcomp(i)	Coordonnées des variables des groupes dans le compromis (groupe i)