

UNIWIN VERSION 10.1.0

ANALYSE SUR TABLEAU DE DISTANCES OU DE DISSIMILARITES

Révision : 23/11/2024

Définition	1
Entrée des données.....	2
Données manquantes	3
Exemple 1 : Tableau de Distances – Fichier VILLES	3
L’option Rapports.....	7
L’option Graphiques.....	8
Exemple 2 : Tableau de Dissimilarités – Fichier ELEVES	11
Exemple 3 : Données d’une enquête – Fichier CEREALES	14
Les variables internes créées par la procédure.....	16

Définition

La méthode d'Analyse sur Tableau de Distances ou de Dissimilarités (ATD) permet d'étudier un tableau de distances euclidiennes (tableau diagonal inférieur) individus x individus contenant à chaque intersection ligne-colonne la distance ou la dissimilarité entre cette ligne et cette colonne.

Voir la procédure 'Créer un tableau de distances euclidiennes' pour plus d'informations sur la façon de construire ce tableau.

Si le tableau n'est pas symétrique, il est symétrisé par la procédure et s'il ne définit pas une "vraie" distance, une technique dite de "la constante additive" est utilisée.

La méthode permet d'obtenir une carte des individus en fonction de leurs proximités ou dissimilarités.

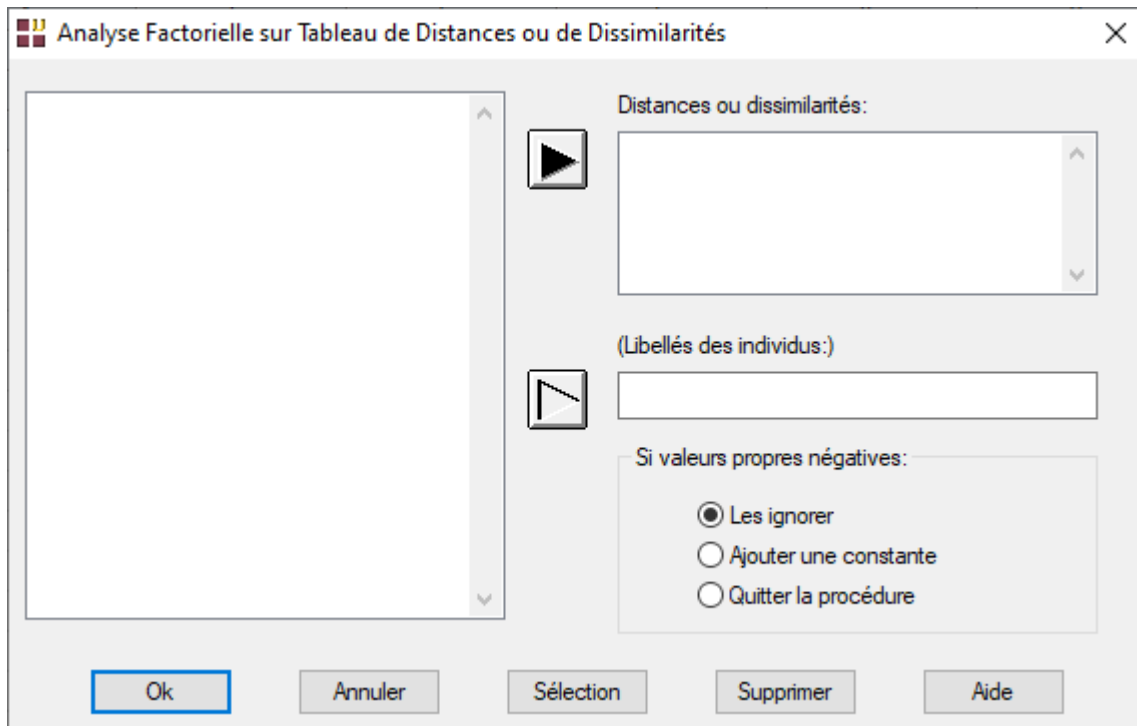
Après affichage du tableau et de l'histogramme des inerties, vous pouvez choisir le nombre d'axes factoriels à extraire et la méthode à utiliser si des valeurs propres sont négatives.

Un rapport général de synthèse est proposé ainsi que les graphiques des plans factoriels des individus.

De nombreux outils d'aide à l'interprétation sont fournis : contributions, cosinus carrés, distances calculées, écarts entre les distances d'origine et celles calculées.

Entrée des données

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



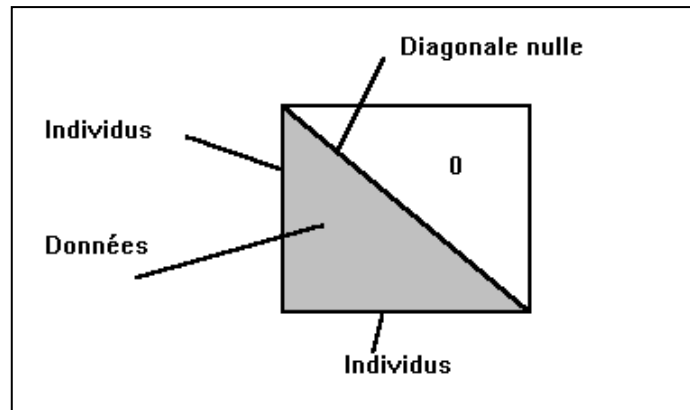
Cette boîte de dialogue permet de préciser les variables qui définissent la matrice de distances ou de dissimilarités entre les individus ainsi que la variable contenant les libellés des individus.

L'ensemble des variables choisies doit définir une matrice triangulaire inférieure à diagonale nulle. L'ordre de sélection des variables est donc important. La matrice est symétrisée automatiquement par la procédure.

La boîte de dialogue permet également d'indiquer comment les valeurs propres négatives sont gérées : on les ignore, on ajoute une constante ou on abandonne le traitement.

Si les variables choisies ne définissent pas une matrice triangulaire inférieure, la procédure propose la transformation de cette matrice par la formule :

$$0,5 \times [(Matrice) + (Matrice transposée)]$$



Données manquantes

Dans cette procédure, les données manquantes ne sont pas autorisées.

Exemple 1 : Tableau de Distances – Fichier VILLES

Dans ce premier exemple, nous utiliserons le fichier VILLES contenant les distances en kilomètres entre 35 villes, principalement françaises:

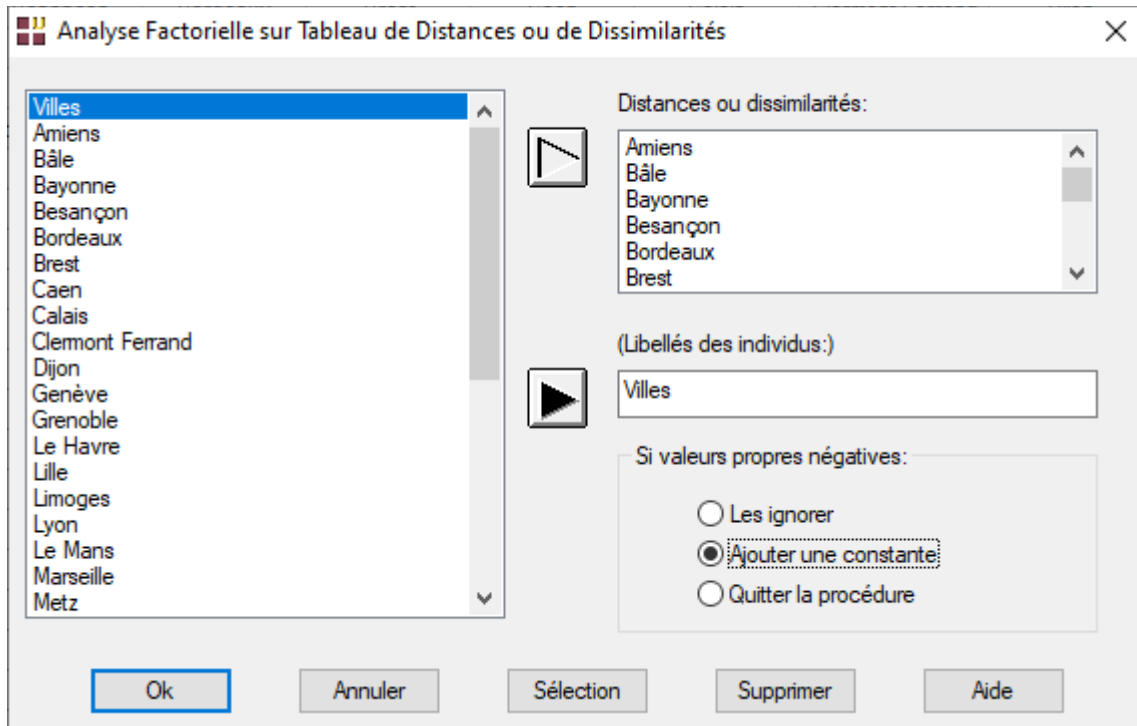
Amiens	Bâle	Bayonne
Besançon	Bordeaux	Brest
Caen	Calais	Clermont-Ferrand
Dijon	Genève	Grenoble
Le Havre	Lille	Limoges
Lyon	Le Mans	Marseille
Metz	Montpellier	Mulhouse
Nancy	Nantes	Nice
Orléans	Paris	Perpignan
Reims	Rennes	Rouen
Saint Etienne	Strasbourg	Toulon
Toulouse	Tours	

Les données proviennent du Guide Michelin (81ème édition). Les distances sont calculées depuis les centres des villes et en utilisant les meilleures routes pour la circulation, pas nécessairement les plus courtes en kilomètres. Si les distances étaient proportionnelles aux distances géographiques, il serait possible de reconstruire la vraie carte par une simple analyse bi-dimensionnelle. Cependant, à cause de l'absence de liaisons directes entre certaines villes, les distances par la route sont parfois très éloignées des distances géographiques. Donc, il ne peut être reconstruit qu'une approximation de la carte géographique.

Cliquons sur l'icône ATD dans le ruban Décrire.

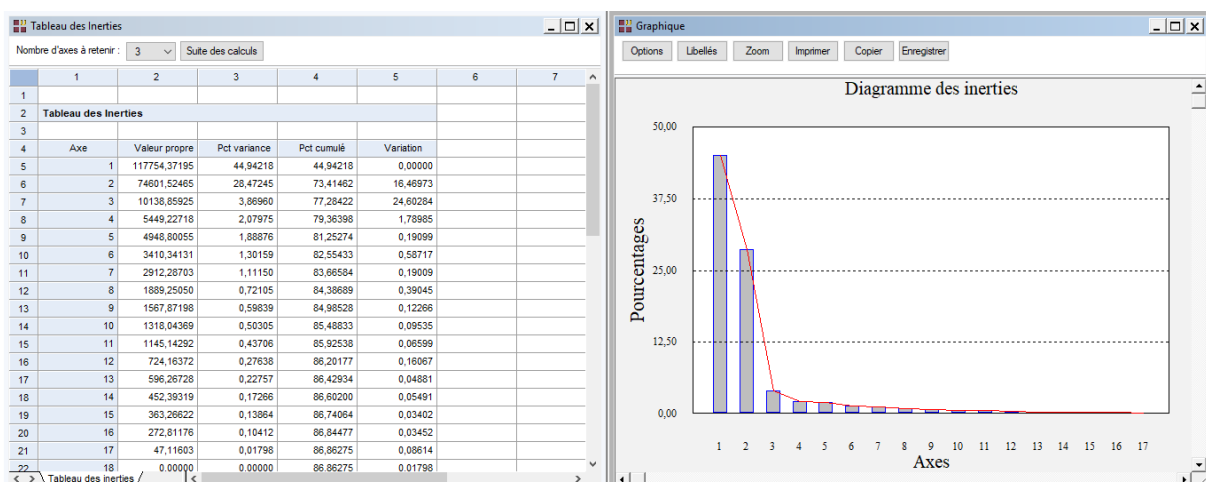
Nous choisissons les variables *Amiens à Tours* (dans cet ordre) comme variables définissant notre tableau de distances, la variable *villes* comme libellés de nos individus et l'option *Ajouter une constante* si il y a des valeurs propres négatives.

La boîte de dialogue renseignée est montrée ci-dessous :



Cliquons sur Ok pour exécuter le traitement.

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.



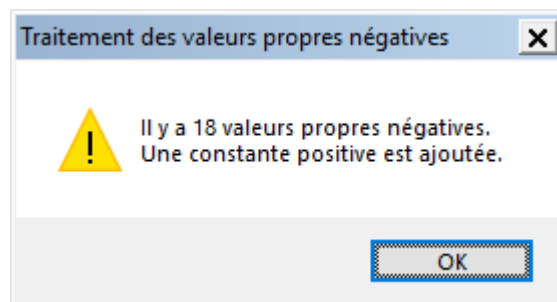
Si des valeurs propres sont négatives, les pourcentages d'inertie expliquée sont à utiliser avec précaution.

Note : Seules les valeurs propres positives sont affichées dans l'histogramme.

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

Cliquons sur le bouton 'Suite des calculs'.

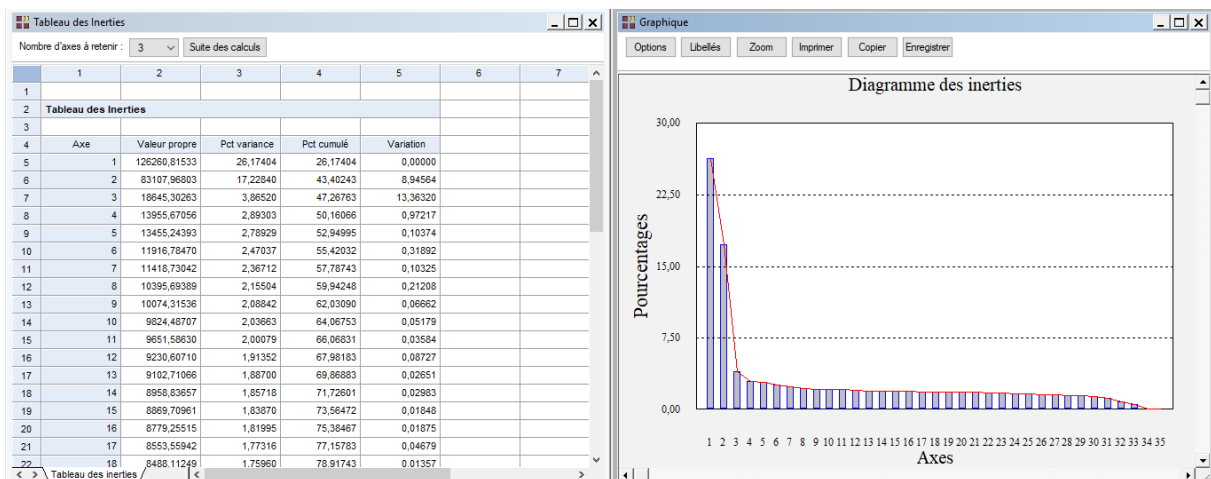
Le message suivant s'affiche, rappelant le choix effectué dans la boîte de dialogue d'entrée des données :



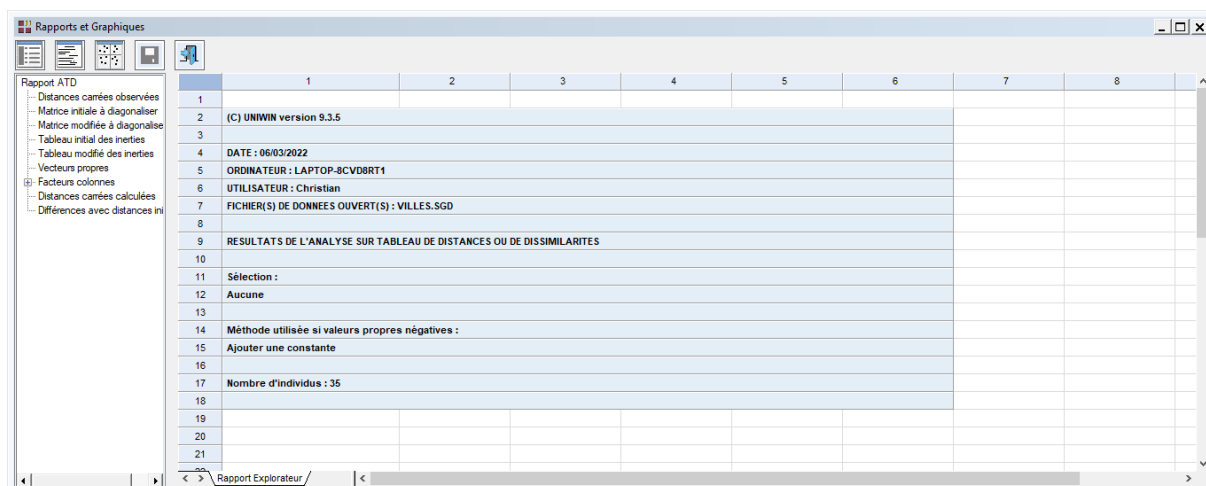
Cette technique consiste à ajouter une constante aux éléments diagonaux de la matrice à diagonaliser de façon à rendre cette matrice semi-définie positive et donc l'ensemble de ses valeurs propres positives.


Nous avons choisi d'ajouter une constante. UNIWIN transforme donc la matrice à diagonaliser et recommence les calculs.

Vous pouvez alors re-visualiser le tableau des inerties, l'histogramme de valeurs propres et choisir le nombre d'axes factoriels à extraire.

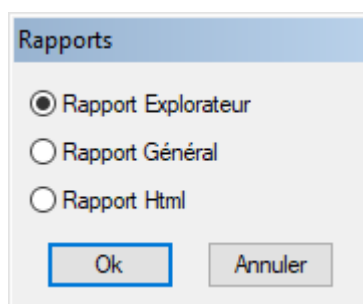



Après quelques instants, l'écran suivant s'affiche :

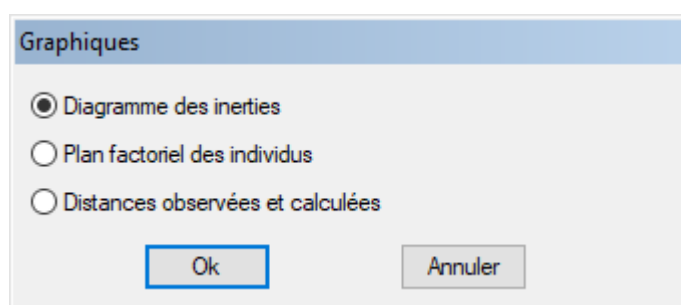



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

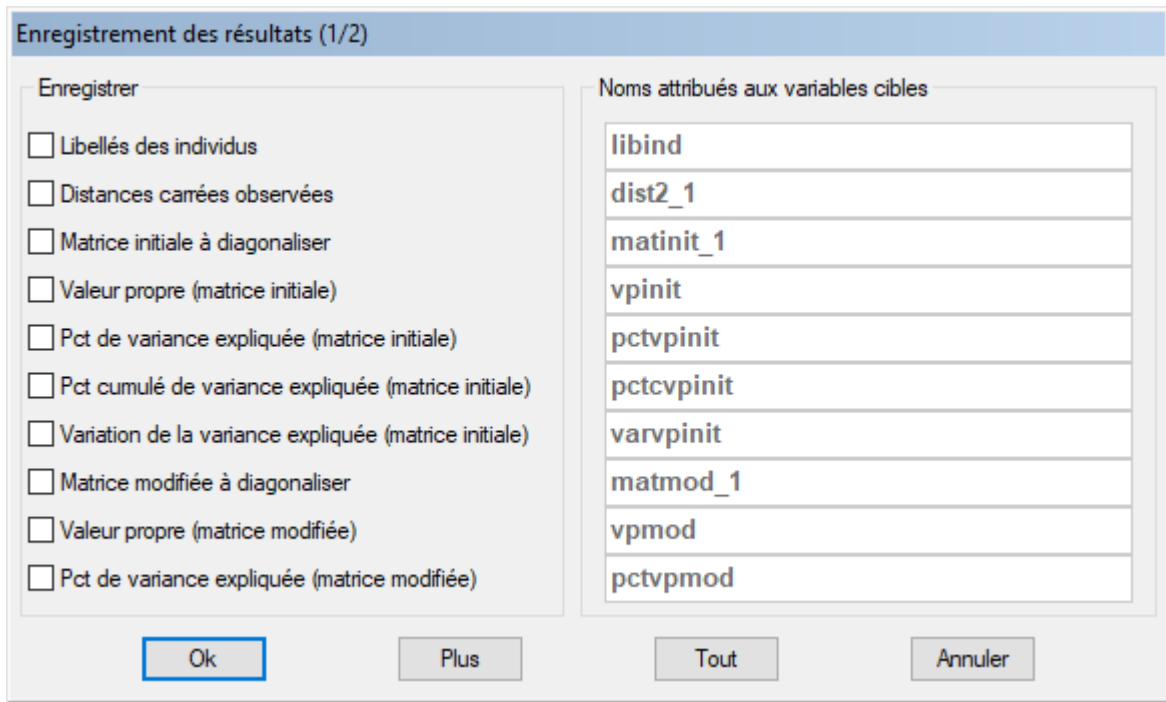
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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 ATD : Explorateur, Général, HTML.

Rapports et Graphiques

Rapport ATD

- Distances carrées observées
- Matrice initiale à diagonaliser
- Matrice modifiée à diagonaliser
- Tableau initial des inerties
- Tableau modifié des inerties
- Vecteurs propres
- Facteurs colonnes
- Axe Factoriel 1**
- Axe Factoriel 2
- Axe Factoriel 3
- Distances carrées calculées
- Différences avec distances in

	1	2	3	4	5	6	7	8
1								
2	FACTEURS COLONNES, COSINUS ET CONTRIBUTIONS POUR L'AXE : 1							
3								
4		Facteur	Cosinus	Contribution				
5	o1	-387,85598	0,33345	3,40412				
6	o2	81,77660	0,01346	0,15133				
7	o3	187,25532	0,05013	0,79347				
8	o4	87,78097	0,02096	0,17437				
9	o5	82,96386	0,01444	0,15575				
10	o6	-575,49611	0,40568	7,49459				
11	o7	-451,37834	0,42054	4,61047				
12	o8	-523,71768	0,46258	6,20666				
13	o9	154,96376	0,07701	0,54341				
14	o10	56,71169	0,01002	0,07278				
15	o11	246,05142	0,14617	1,36998				
16	o12	355,88878	0,29813	2,89610				
17	o13	-443,75492	0,40393	4,45605				
18	o14	-432,75782	0,35849	4,23792				
19	o15	41,01221	0,00494	0,03806				
20	o16	254,65999	0,18536	1,46752				
21	o17	-268,66209	0,19053	1,63334				

Rapport Explorateur /

	1	2	3	4	5	6	7	8	9	10	11	12
21	DISTANCES CARRES OBSERVEES											
22												
23												
24		Amiens	Bâle	Bayonne	Besançon	Bordeaux	Brest	Caen	Calais	Clermont Ferrand	Dijon	Genève
25	Amiens	0	334084	848241	237169	528529	370881	56169	23716	329476	211600	467856
26	Bâle	334084	0	1046529	22801	690561	1201216	613089	467856	228484	59536	67081
27	Bayonne	848241	1046529	0	755161	34225	674041	541696	1136356	292681	687241	904401
28	Besançon	237169	22801	755161	0	456625	933156	416025	352836	102400	6889	31329
29	Bordeaux	528529	690561	34225	456625	0	393129	293764	780384	133225	403225	465124
30	Brest	370881	1201216	674041	933156	393129	0	137641	490000	652864	753424	1168561
31	Caen	56169	613089	541696	416025	293764	137641	0	107584	304704	298116	588289
32	Calais	23716	467856	1136356	352836	760384	490000	107584	0	516961	315844	550564
33	Clermont Ferrand	329476	228484	292681	102400	133225	652864	304704	516961	0	78400	96100
34	Dijon	211600	59536	687241	6889	403225	753424	298116	315844	78400	0	39601
35	Genève	467856	67081	904401	31329	465124	1168561	588289	550564	96100	39601	0
36	Grenoble	509796	162409	688900	79524	422500	1258884	636804	737881	80656	88209	20736
37	Le Havre	31684	556516	646416	370881	372100	231361	11881	72361	338724	260100	532900
38	Lille	13456	373321	986049	271441	638401	512656	118336	12996	416025	238144	447561
39	Limoges	299209	376996	170569	219024	47961	367236	190096	476864	30976	183184	238144
40	Lyon	372100	160801	509796	55225	289444	1036324	484416	570025	29564	37249	22801
41	Le Mans	124609	495616	342225	330625	152881	161604	22801	160000	173889	226576	471969

(C) UNIWIN version 9.3.5

DATE : 06/03/2022
 ORDINATEUR : LAPTOP-8CVD8RT1
 UTILISATEUR : Christian
 FICHIER(S) DE DONNEES OUVERT(S) : VILLES.SGD

RESULTATS DE L'ANALYSE SUR TABLEAU DE DISTANCES OU DE DISSIMILARITES

Sélection :
Aucune

Méthode utilisée si valeurs propres négatives :
Ajouter une constante

Nombre d'individus : 35

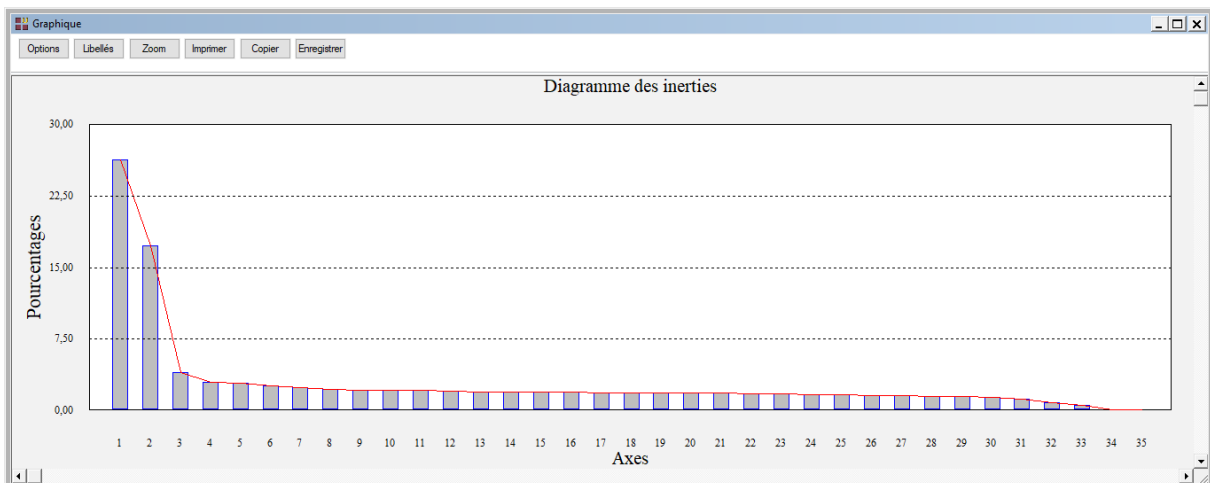
DISTANCES CARRES OBSERVEES

	Amiens	Bâle	Bayonne	Besançon	Bordeaux	Brest	Caen	Calais	Clermont Ferrand	Dijon	Genève	Grenoble	Le Havre	Lille	Limoges	Lyon	Le Mans	Marseille
Amiens	0	334084	848241	237169	528529	370881	56169	23716	329476	211600	467856	509796	31684	13456	299209	372100	124609	8482
Bâle	334084	0	1046529	22801	690561	1201216	613089	467856	228484	59536	67081	162409	556516	373321	376996	160801	495616	5056
Bayonne	848241	1046529	0	755161	34225	674041	541696	1136356	292681	687241	904401	688900	646416	986049	170569	509796	342225	4886

L'option Graphiques

- Diagramme des inerties

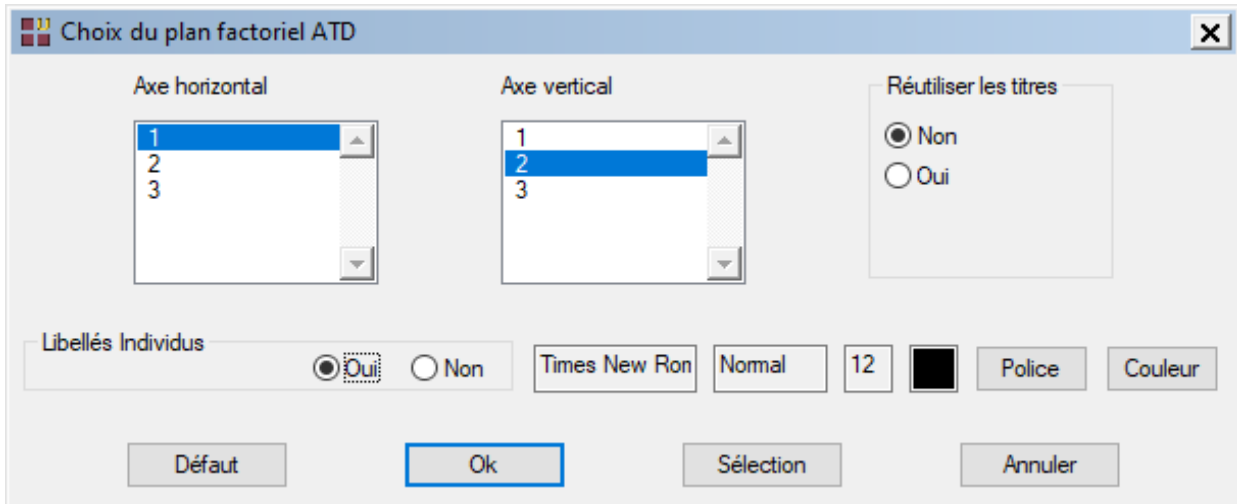
Cette option permet d'afficher le diagramme des inerties.



- Plan factoriel des individus

Cette option permet d'afficher le plan factoriel des villes.

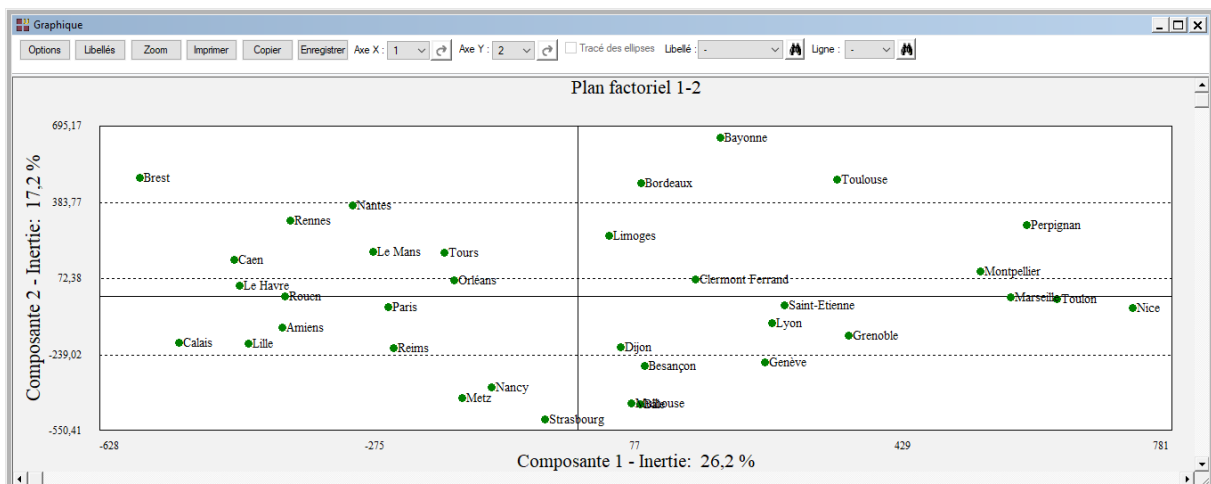
Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.



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

Un exemple du plan factoriel créé est montré ci-après.

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



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

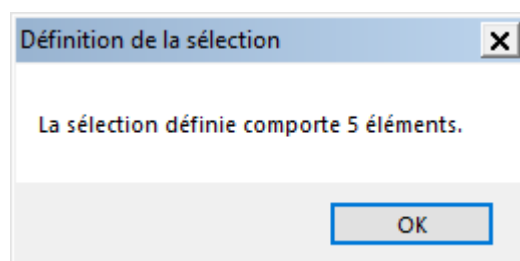
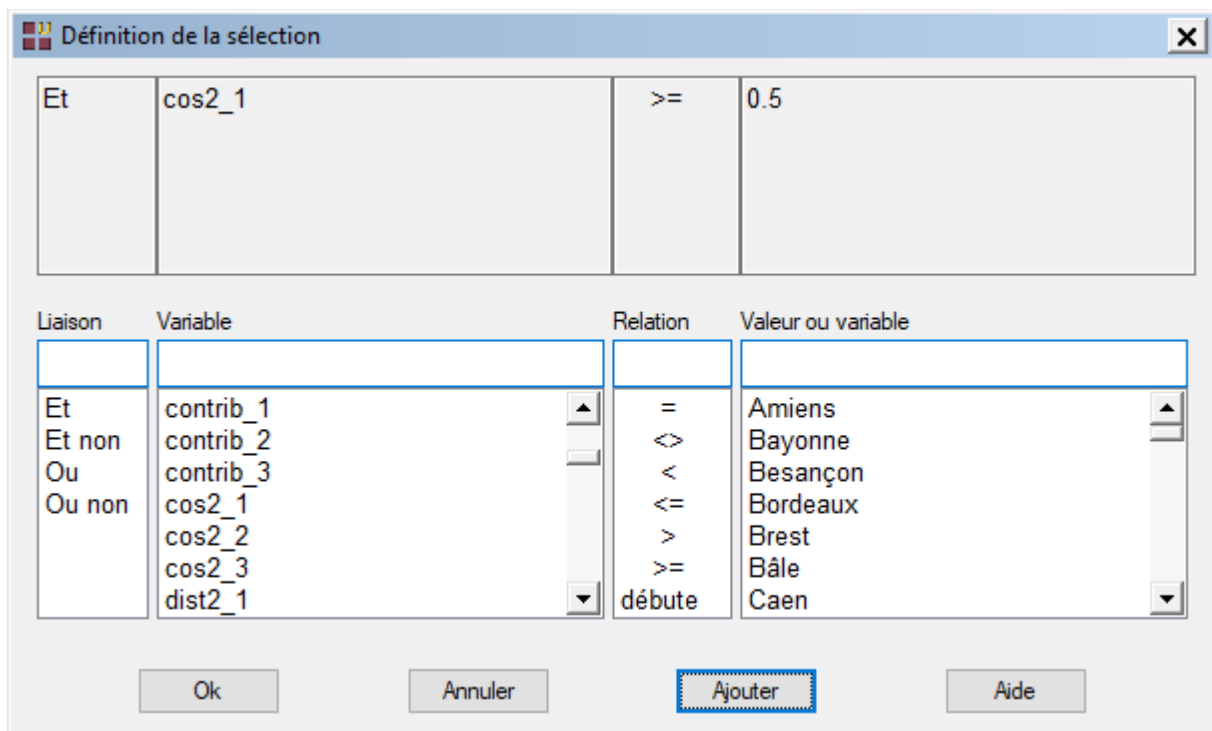
Ceci est très pratique dans le cas d'un nuage comportant beaucoup d'individus : seuls les points significatifs d'un point de vue statistique sont affichés.

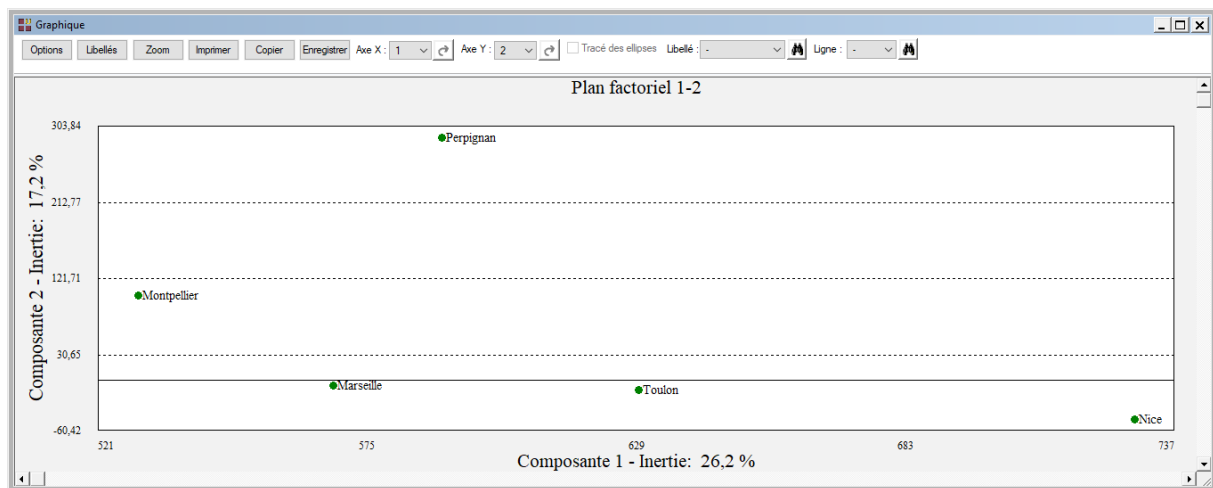
Les principaux résultats de l'analyse sont accessibles dans la boîte de dialogue 'Sélection' en plus des données du fichier VILLES.

Par exemple, sélectionnons les villes dont le cosinus carré sur le premier axe factoriel est supérieur ou égal à 0,5.

La variable cos2_1 contient les cosinus carrés sur le premier axe factoriel extrait.

Cette sélection comporte 5 individus (villes).

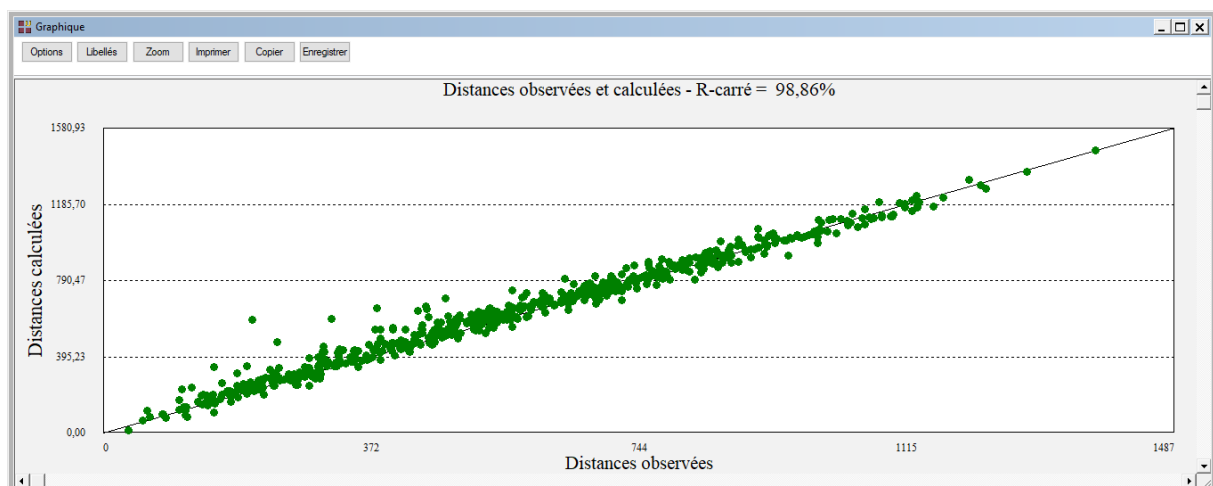




- Distances observées et calculées

Cette option graphique permet d'afficher les distances observées et calculées.

Cette option permet d'afficher un graphique visualisant les distances observées par rapport aux distances calculées. La bissectrice permet de mieux visualiser l'ajustement et le R-carré indique la corrélation entre les valeurs observées et calculées.



Exemple 2 : Tableau de Dissimilarités – Fichier ELEVES

Dans ce deuxième exemple, nous utiliserons le fichier ELEVES contenant les fréquences de communication entre 25 élèves d'une Université.

Cet exemple provient du livre « Eléments d'Analyse des Données » (Diday, Dunod, 1982).

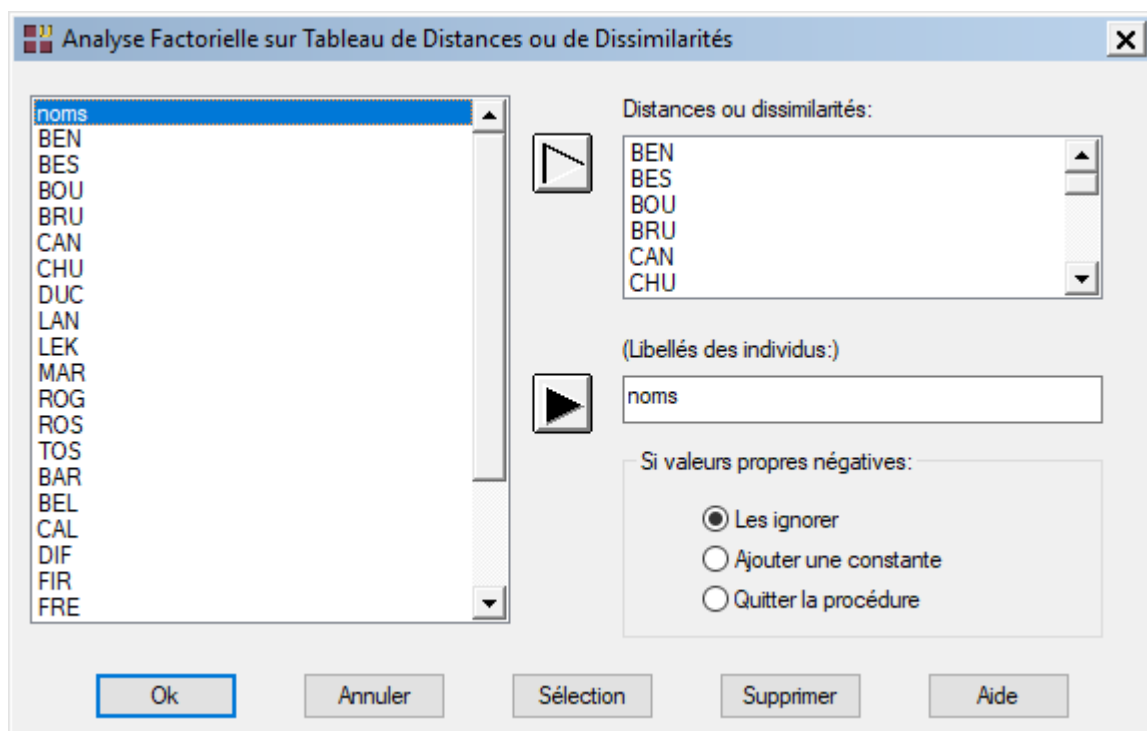
Les notes données par chaque étudiant à un autre étudiant sont :

- 0 pour communication très fréquente
- 1 pour communication fréquente
- 2 pour communication rare
- 3 pour communication très rare
- 4 pour absence de communication

La « socio-matrice » (matrice des dissimilarités) est construite de la façon suivante :

$$d(i,j) = 0,5 \times [(Note\ donnée\ par\ i\ à\ j) + (Note\ donnée\ par\ j\ à\ i)]$$

Cliquons sur l'icône ATD dans le ruban Décrire. Voici la boîte de dialogue renseignée :



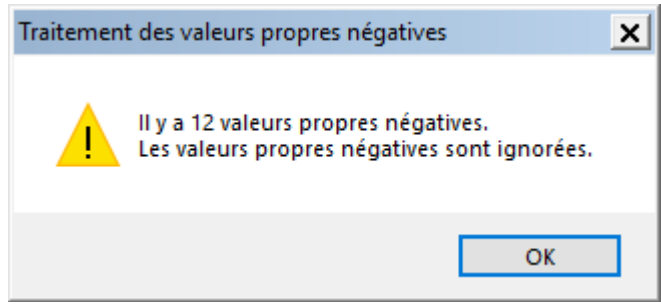
Les variables *BEN* à *VID* contiennent les notes.

Par exemple les éléments de la variable *BEN* sont les notes données par *BEN* à tous les autres étudiants.

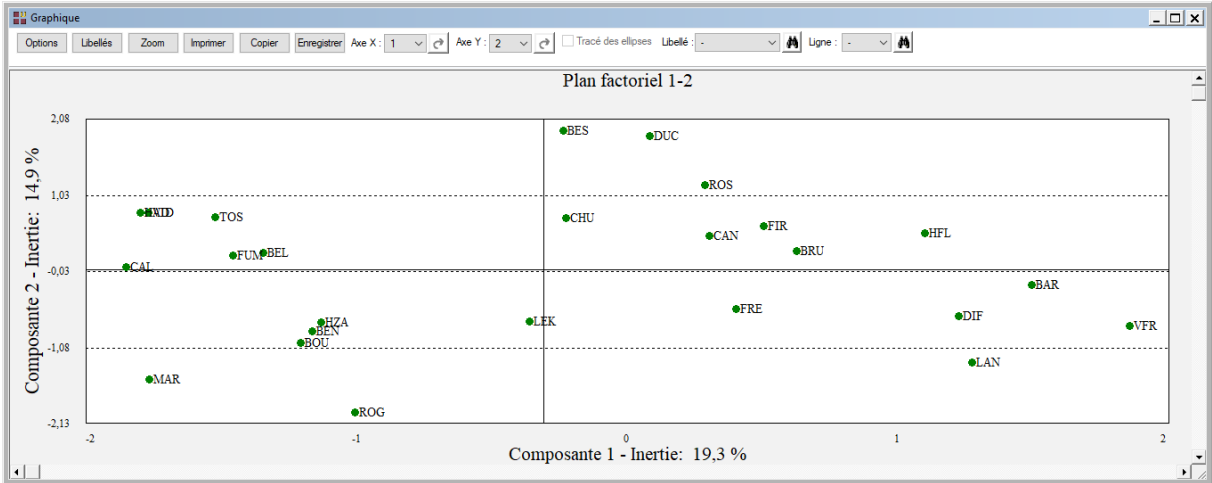
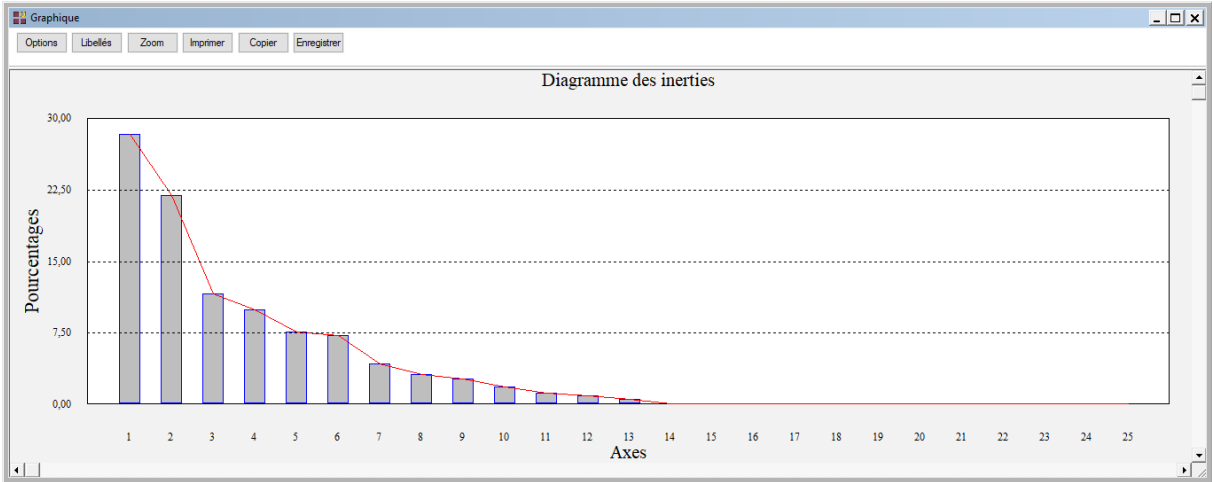
La variable *noms* contient les noms des étudiants.

La procédure se déroule comme montrée dans l'exemple 1.

Dans cet exemple, nous choisissons d'ignorer les valeurs propres négatives c'est-à-dire qu'elles sont mises à 0.



Voici à titre d'exemple le diagramme des inerties et le plan factoriel 1-2.



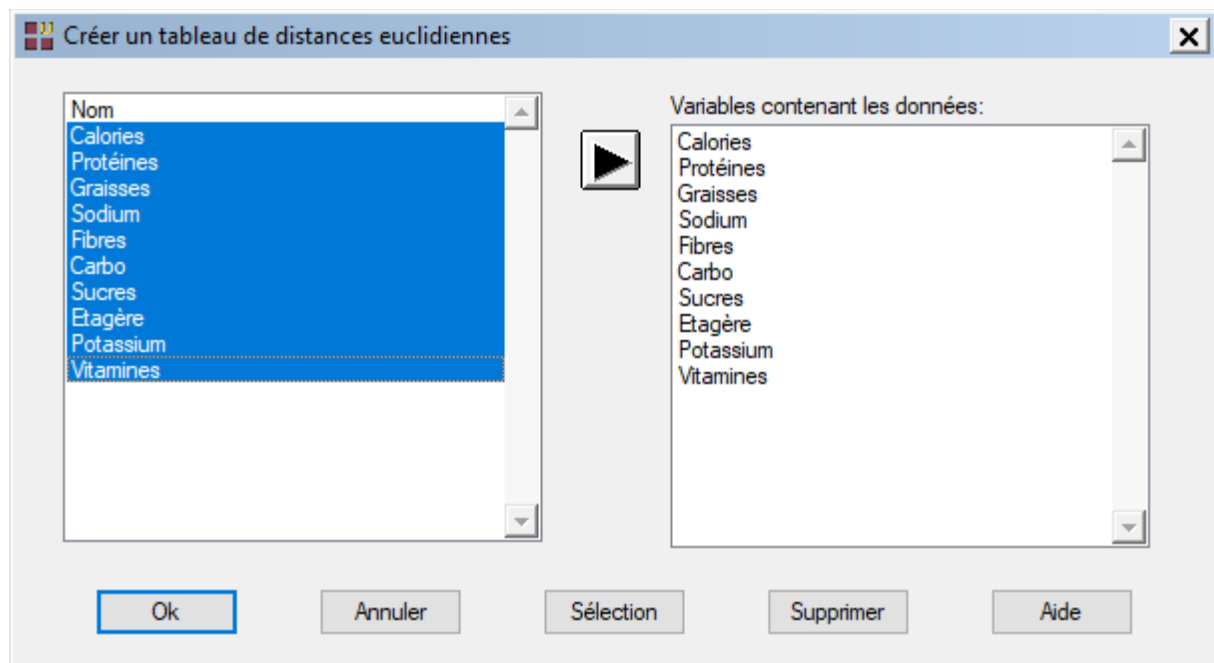
Exemple 3 : Données d'une enquête – Fichier CEREALES

Dans ce troisième exemple, nous utiliserons le fichier CEREALES contenant une partie des résultats issus d'une enquête de comparaison de diverses céréales consommées lors du petit déjeuner (American Statistical Association, 1993).

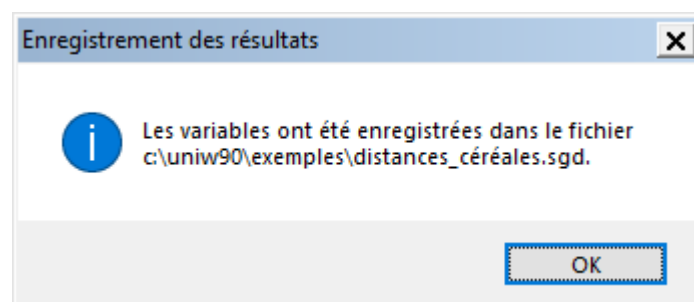
Vingt-trois produits ont été jugés sur dix critères : nombre de calories, teneurs en protéines, en graisses, en sodium, en fibres, en complexes carbohydrates, en sucres, numéro de l'étagère compté à partir du sol, teneurs en potassium et en vitamines.

Notre première étape consiste à transformer ce tableau de données individus x variables en un tableau de distances euclidiennes individus x individus.

Utilisons pour cela la procédure 'Créer un tableau distances euclidiennes' du ruban Décrire.



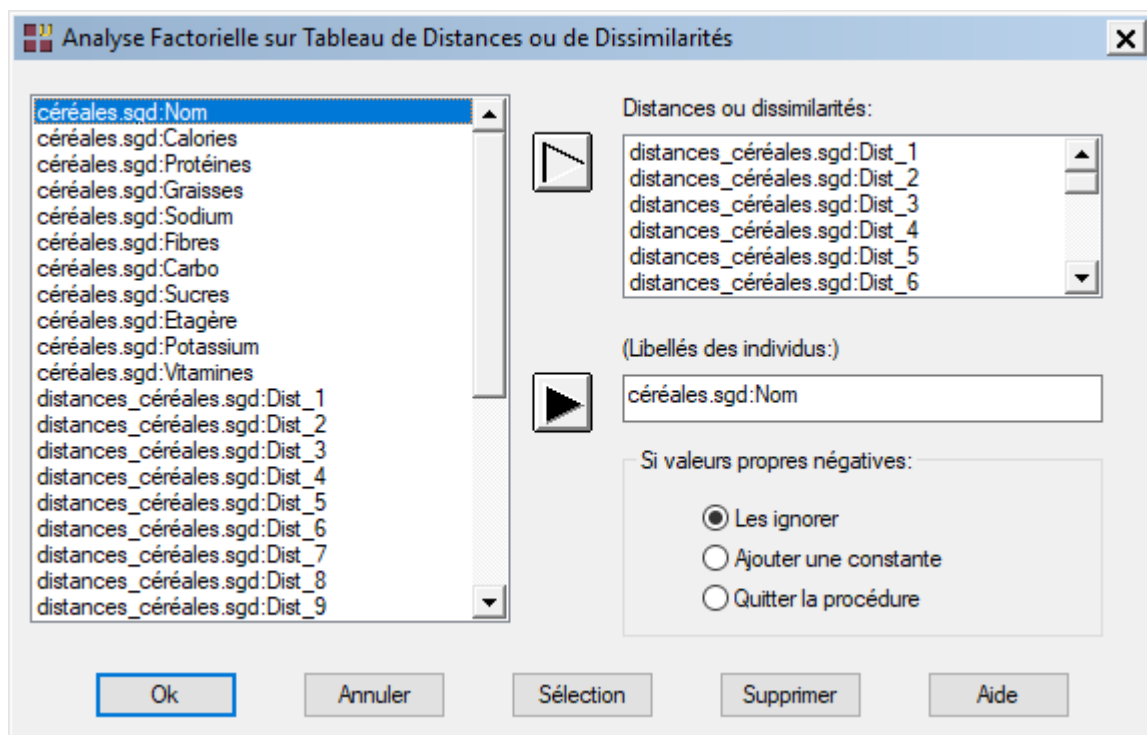
Cliquons sur Ok. Après quelques instants, le message suivant s'affiche :



Ouvrons simultanément ce fichier et le fichier CEREALES.

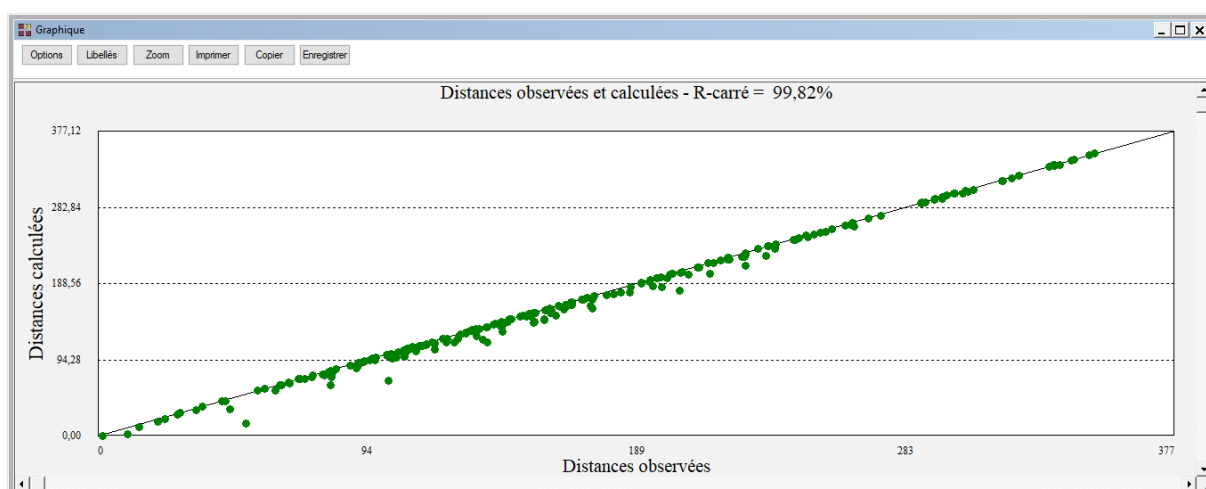
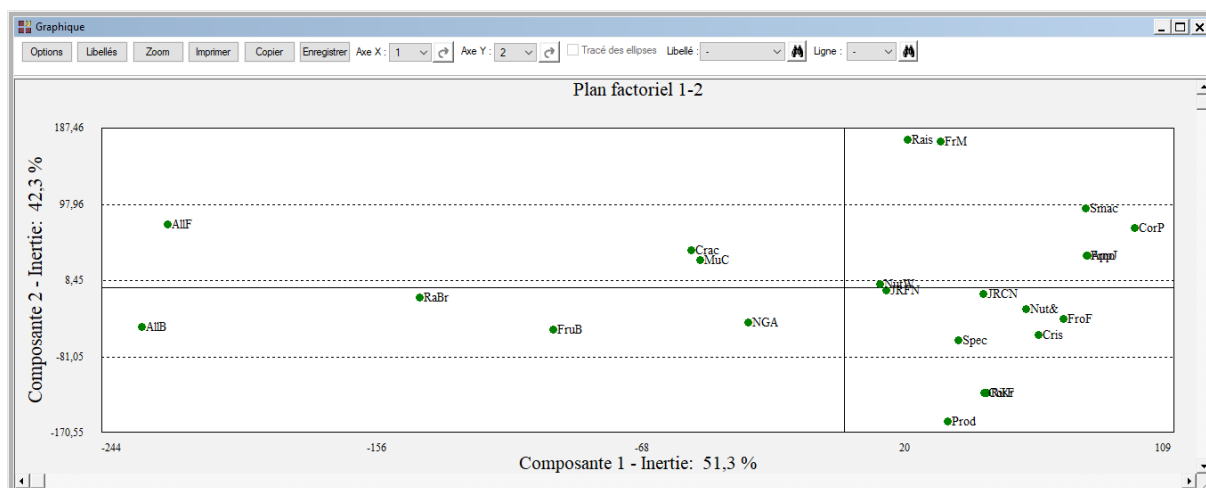
	Dist_1	Dist_2	Dist_3	Dist_4	Dist_5	Dist_6	Dist_7	Dist_8	Dist_9	Dist_10	Dist_11	Dist_12	Dist_13	Dist_14	Dist_15	Dist_16
	Distances 1	Distances 2	Distances 3	Distances 4	Distances 5	Distances 6	Distances 7	Distances 8	Distances 9	Distances 10	Distances 11	Distances 12	Distances 13	Distances 14	Distances 15	Distances 16
	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N
	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23	Longueur = 23
1	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
2	122,27837	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
3	322,83292	306,92870	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
4	288,62259	335,22361	166,11743	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
5	347,36148	320,23117	36,52396	201,22376	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
6	204,08617	180,72908	131,12971	195,94387	148,85228	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
7	295,92060	316,78226	96,16132	70,92249	130,94655	153,15352	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
8	342,04093	274,24077	143,83324	297,32978	120,97107	152,73179	231,23797	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
9	303,95394	317,11827	75,29940	91,82048	110,12720	147,82904	23,30236	213,89250	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
10	322,60502	306,88434	1,41421	166,05120	36,49658	131,06105	96,05727	143,79152	75,27284	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
11	141,12406	186,44570	197,37274	164,60559	226,98458	105,12374	162,00926	257,15754	170,14700	197,36768	0,00000	0,00000	0,00000	0,00000	0,00000	0,00
12	278,33253	278,73105	55,01818	123,13407	89,74965	104,70912	58,53204	174,97714	46,52956	54,87258	148,19919	0,00000	0,00000	0,00000	0,00000	0,00
13	252,70536	254,15743	85,20563	140,20342	113,97368	78,33901	87,54999	174,83707	82,15838	85,14693	119,91664	46,31414	0,00000	0,00000	0,00000	0,00
14	214,48077	203,65657	141,67569	197,42087	160,40574	51,83628	156,28500	172,48188	152,49262	141,86157	103,04368	113,82003	71,07742	0,00000	0,00000	0,00
15	206,97343	234,18796	141,71803	124,75175	173,21085	91,23048	104,52272	225,72993	111,38222	141,86157	66,94774	91,23048	61,09010	79,06959	0,00000	0,00
16	293,19959	303,03795	66,84310	102,53292	102,54755	130,57948	34,27827	200,28979	20,83267	66,76826	158,23400	30,24897	62,06670	132,78554	97,22140	0,00
17	248,12900	245,61352	78,89233	132,44999	108,78419	79,96624	80,71555	170,71028	75,05998	78,75278	126,17448	36,37307	50,79370	101,61693	81,47392	62,20
18	283,48545	341,24478	196,35173	31,73326	231,86203	214,14715	101,62185	324,77993	122,50714	196,30079	167,20646	151,14563	163,21152	214,10745	137,29530	131,87
19	107,31283	134,77388	226,83915	221,34362	250,84258	106,99065	210,82931	253,23902	215,51798	226,83033	58,32667	184,83236	151,92103	107,77755	112,60551	201,08
20	334,99104	264,25934	150,02333	299,80327	129,00775	150,13327	235,04468	14,31782	218,30483	149,97666	254,84701	178,34237	177,93538	173,08668	226,58994	204,73
21	289,90343	336,94807	165,81315	10,14889	200,97264	195,74218	70,22108	297,51639	91,26883	165,75283	163,70706	122,73956	137,65900	194,59702	121,90160	101,05

Renseignons la boîte de dialogue de l'ATD comme montré ci-dessous :



Effectuons ensuite notre analyse sur ce tableau de distances en sélectionnant les 23 colonnes du tableau des distances.

Nous choisissons de conserver 3 axes factoriels.



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>
libind	Libellés des individus
dist2	Distances carrées observées
matinit	Matrice initiale à diagonaliser
vpinit	Valeurs propres (matrice initiale)
pctvpinit	Pourcentages de variance expliquée (matrice initiale)
pctcvpinit	Pct cumulés de variance expliquée (matrice initiale)
varvpinit	Variations de la variance expliquée (matrice initiale)
matmod	Matrice modifiée à diagonaliser
vpmo	Valeurs propres (matrice modifiée)

pctvpmo	Pourcentages de variance expliquée (matrice modifiée)
pctcvpmo	Pct cumulés de variance expliquée (matrice modifiée)
varvpmo	Variations de la variance expliquée (matrice modifiée)
vectpro	Vecteurs propres
cos2	Cosinus carrés
contrib	Contributions à l'inertie
colfact	Facteurs colonnes
distc	Distances carrées calculées
ecdisi	Ecart avec les distances carrées observées