

The logo for Francestat, featuring the word "francestat" in a sans-serif font. "france" is in white and "stat" is in red, all set against a blue rectangular background.

francestat

**UNIWIN version 9.7.0**

**Manuel de l'Utilisateur**

Révision : 07/09/2023

---

**UNIWIN VERSION 9.7.0**

**ANALYSES STATISTIQUES  
MULTIVARIEES POUR  
STATGRAPHICS CENTURION**

**MANUEL DE L'UTILISATEUR**

© 1994-2023 par FRANCESTAT

[www.FRANCESTAT.com](http://www.FRANCESTAT.com)

Tous droits réservés. Aucune partie de ce document ne peut être reproduite, sous toute forme ou par tout moyen, sans l'accord écrit de FRANCESTAT.

Référencé comme : *UNIWIN version 9.7.0 - Manuel de l'Utilisateur*

---

# Table des matières

<b>Table des matières</b> .....	<b>iii</b>
<b>Préface</b> .....	<b>vi</b>
<b>Démarrer</b> .....	<b>7</b>
1.1 Installation du logiciel sur un ordinateur individuel ou sur un serveur .....	7
1.2 Demande du fichier de licence du logiciel .....	12
1.3 Installation nécessaire sur chaque poste client .....	15
1.4 Enregistrement des DLL et OCX.....	16
<b>Rubans et barre d'outils rapide</b> .....	<b>18</b>
2.1 Les rubans.....	18
2.2 La barre d'outils rapide .....	20
<b>Boîtes de dialogue</b> .....	<b>21</b>
3.1 Les éléments d'une boîte de dialogue .....	22
3.2 Les boutons de commande .....	22
3.3 Sélection des données .....	23
<b>Le ruban Fichiers</b> .....	<b>26</b>
4.1 L'option 'Ouvrir'.....	27
4.2 L'option 'Fichiers récents' .....	29
4.3 L'option 'Fermer'.....	29
4.4 L'option 'Supprimer' .....	30
4.5 L'option 'Options du logiciel' .....	31
4.6 L'option 'Installer R et les packages nécessaires' .....	32
4.7 L'option 'Console R' .....	40
4.8 L'option 'Quitter le logiciel' .....	40
<b>Le ruban Accueil</b> .....	<b>41</b>
5.1 Les options 'Copier / Coller' .....	41
5.2 L'option 'Imprimer'.....	41
5.3 L'option 'Calculatrice' .....	46
5.4 L'option 'Lecteur Media' .....	46
5.5 Les options 'Fenêtre' .....	47
5.6 L'option 'Manuel' .....	47
5.6 L'option 'Aide' .....	47
<b>Le ruban Graphique</b> .....	<b>48</b>
6.1 Nuages de points et courbes 2D .....	48

6.2 Bulles 2D .....	60
6.3 Histogramme.....	64
6.4 Courbe ROC.....	67
6.5 Courbe de Gini .....	75
6.6 Cartes géographiques .....	77
6.7 Options graphiques.....	81
6.8 Composition graphique .....	89
6.9 Accès à Microsoft Paint.....	91
<b>Le ruban Décrire .....</b>	<b>93</b>
7.1 Recoder des variables alphabétiques.....	93
7.2 Créer un tableau de distances euclidiennes.....	96
7.3 Tri à plat et tri croisé.....	99
7.4 Analyse en composantes principales .....	105
7.5 Analyse en composantes principales symbolique .....	107
7.6 Analyse NIPALS .....	109
7.7 Analyse factorielle des correspondances.....	111
7.8 Analyse des correspondances multiples .....	113
7.9 Analyse sur tableau de distances ou de dissimilarités.....	115
7.10 Analyse factorielle de données mixtes.....	117
7.11 Analyse en composantes indépendantes.....	119
7.12 Classification ascendante hiérarchique .....	121
7.13 Classification par cartes auto-organisatrices.....	123
7.14 Ellipses de tolérance ou de confiance .....	125
7.15 Analyse factorielle multiple.....	127
7.16 Méthode STATIS.....	129
7.17 Méthode STATIS duale.....	131
7.18 Analyse procrustéenne généralisée .....	133
7.19 Analyse APRIORI.....	135
<b>Le ruban Expliquer .....</b>	<b>137</b>
8.1 Analyse discriminante pas à pas .....	138
8.2 Analyse factorielle discriminante.....	140
8.3 Analyse discriminante bayésienne.....	142
8.4 Analyse discriminante qualitative .....	145
8.5 Scoring .....	148
8.6 Méthode KNN .....	150
8.7 Méthode SIMCA .....	152
8.8 Méthode de classement bayésienne naïve.....	154
8.9 Régression sur composantes principales.....	156
8.10 Régression PLS.....	158

8.11 Réseaux de neurones artificiels.....	160
8.12 Arbres de décision et de régression .....	162
<b>Bibilographie .....</b>	<b>164</b>
9.1 Bibliographie française.....	164
9.2 Bibliographie anglo-saxonne.....	169

# Préface

Ce manuel est conçu pour présenter, aux utilisateurs du logiciel UNIWIN version 9.7.0, les opérations de base du logiciel et son utilisation pour analyser des données. Il donne un aperçu complet des fonctionnalités du logiciel : installation, gestion des données, mise en œuvre d'analyses statistiques, impression et publication des résultats. Comme ce manuel a pour but de permettre aux utilisateurs d'utiliser le logiciel rapidement, il se concentre sur les fonctionnalités les plus importantes du logiciel plutôt que d'essayer de couvrir tous les détails. L'option *Manuel* du ruban *Accueil* dans UNIWIN version 9 donne accès à un grand nombre d'informations additionnelles avec un fichier PDF pour chacune des procédures statistiques.

Ce manuel présente les éléments de base pour l'utilisation du logiciel. Bien que vous puissiez probablement découvrir tout cela par vous-même en utilisant le logiciel, une lecture complète de ces chapitres vous aidera à prendre en main rapidement le logiciel et vous assurera que vous ne passez pas à côté d'importantes fonctionnalités.

NOTE : Les fichiers de données utilisés dans ce manuel sont fournis avec le logiciel.

FRANCESTAT  
Septembre 2023

## Démarrer

*Installer UNIWIN version 9.7.0, utiliser le logiciel et ouvrir un fichier de données.*

### 1.1 Installation du logiciel sur un ordinateur individuel ou sur un serveur

Pour utiliser UNIWIN version 9.7.0, vous devez posséder le logiciel STATGRAPHICS Centurion (Statgraphics Technologies, USA), un ordinateur individuel ou un serveur fonctionnant sous Windows 7, 8, 10, 11, Server 2016, 2019 ou 2022 avec au minimum 1 Go de mémoire vive et environ 130 Mo d'espace libre sur le disque dur.

Microsoft .NET Framework version 4.8 ou supérieure est requis. L'utilitaire 'DotNetVersion.exe' livré avec le logiciel UNIWIN permet d'afficher la ou les versions présentes sur votre ordinateur.

Il faut obligatoirement avoir les droits- administrateur pour installer le logiciel et demander le fichier de licence nécessaire à l'utilisation du logiciel.

UNIWIN version 9.7.0 est livré sous la forme d'un unique fichier à télécharger sur votre ordinateur.

Après avoir téléchargé votre logiciel, suivre les instructions ci-dessous pour l'installer sur votre ordinateur ou serveur et l'activer.

**Etape 1 :** Avec les droits administrateur, localiser le fichier téléchargé et double-cliquer sur son nom pour débiter le processus d'installation.

**Etape 2 :** UNIWIN version 9.7.0 utilise l'installateur standard de Windows pour installer le logiciel sur votre ordinateur. L'installateur pilote l'installation via un ensemble de boîtes de dialogue. La première boîte de dialogue vous accueille dans UNIWIN version 9.7.0 :



Figure 1-1. Boîte de dialogue d'accueil

Cliquer sur le bouton *Suivant*.

**NOTE :** Pour installer et activer UNIWIN 9.7.0 vous devez avoir le droit administrateur sur votre ordinateur. Si une personne du service informatique doit intervenir pour cette installation, nous vous recommandons d'installer et de faire la demande du fichier de licence en sa présence.

**Etape 3** : La deuxième boîte de dialogue affiche le contrat de licence du logiciel :

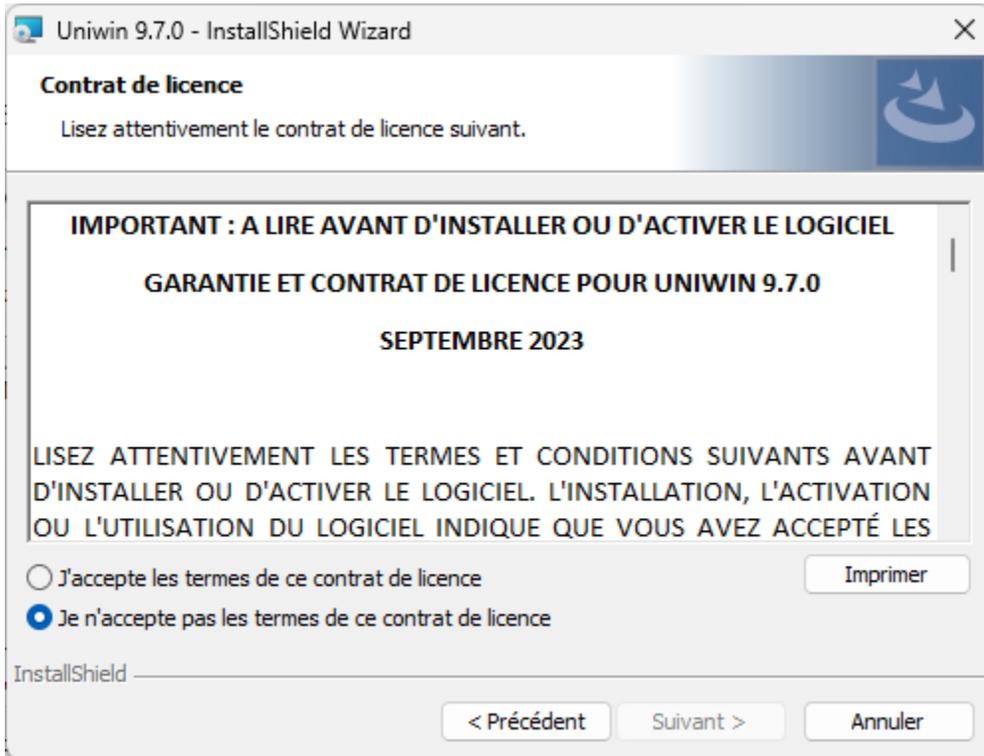
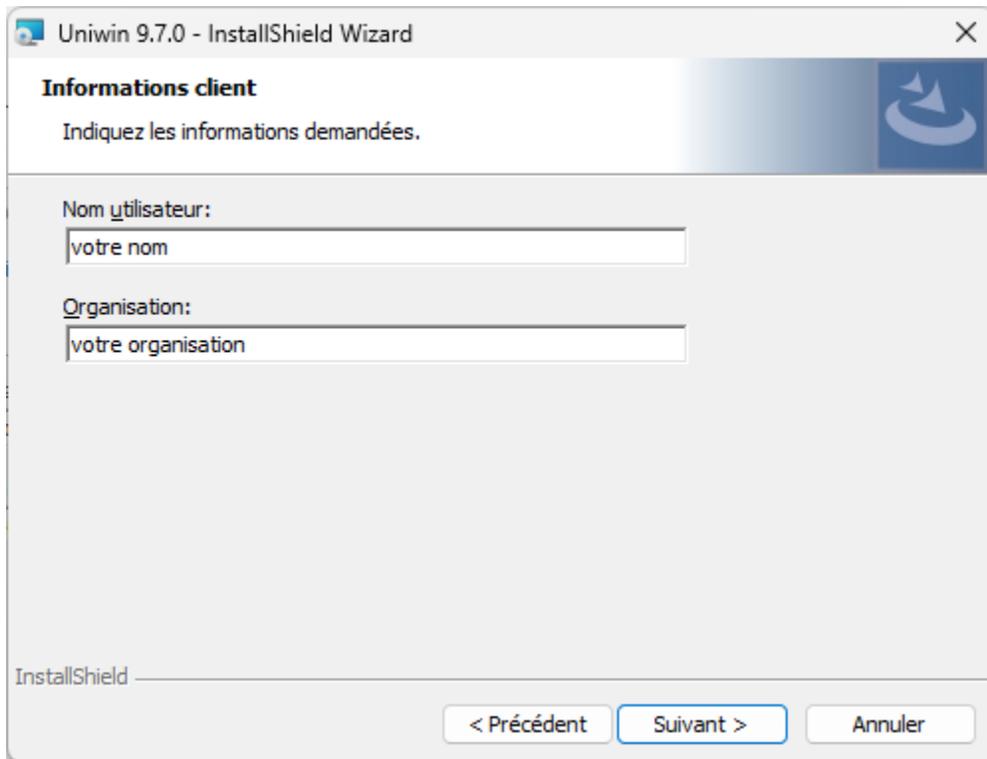


Figure 1-2. Boîte de dialogue du contrat de licence

Lire attentivement le contrat de licence. Si vous en acceptez les termes, cliquez sur le bouton radio approprié puis cliquez sur *Suivant* pour continuer. Si vous ne l'acceptez pas, cliquez sur *Annuler*. Si vous n'en acceptez pas les termes, vous ne devez pas utiliser le logiciel.

**Etape 4** : La boîte de dialogue suivante vous demande des informations identifiant la personne qui utilisera le logiciel :



Uniwin 9.7.0 - InstallShield Wizard

**Informations client**

Indiquez les informations demandées.

Nom utilisateur:  
votre nom

Organisation:  
votre organisation

InstallShield

< Précédent   Suivant >   Annuler

Figure 1-3. Boîte de dialogue d'identification de l'utilisateur

**Etape 5 :** La boîte de dialogue suivante indique le répertoire dans lequel le logiciel sera installé par défaut :

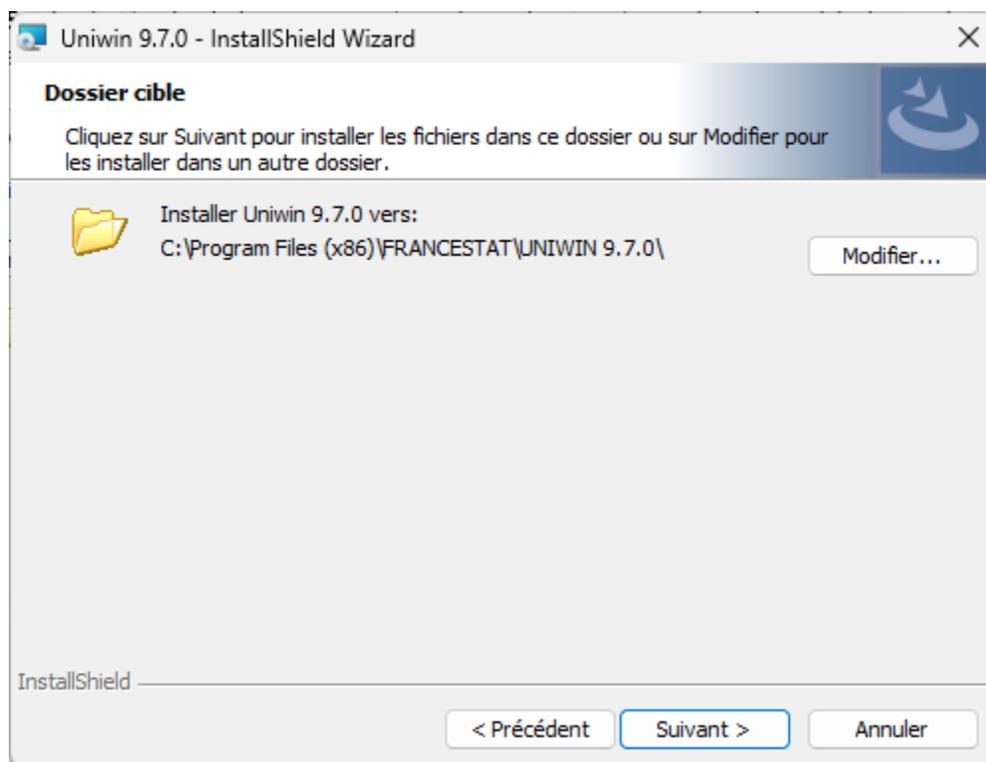


Figure 1-4. Boîte de dialogue pour le répertoire de destination

Note : Si vous installez le logiciel sur un serveur, installez-le dans un répertoire auquel tous les utilisateurs potentiels auront accès en lecture. L'accès en écriture par les utilisateurs n'est pas nécessaire.

**Etape 6 :** Suivre ensuite les instructions pour terminer l'installation. Lorsque l'installation est terminée, une boîte de dialogue finale s'affiche :

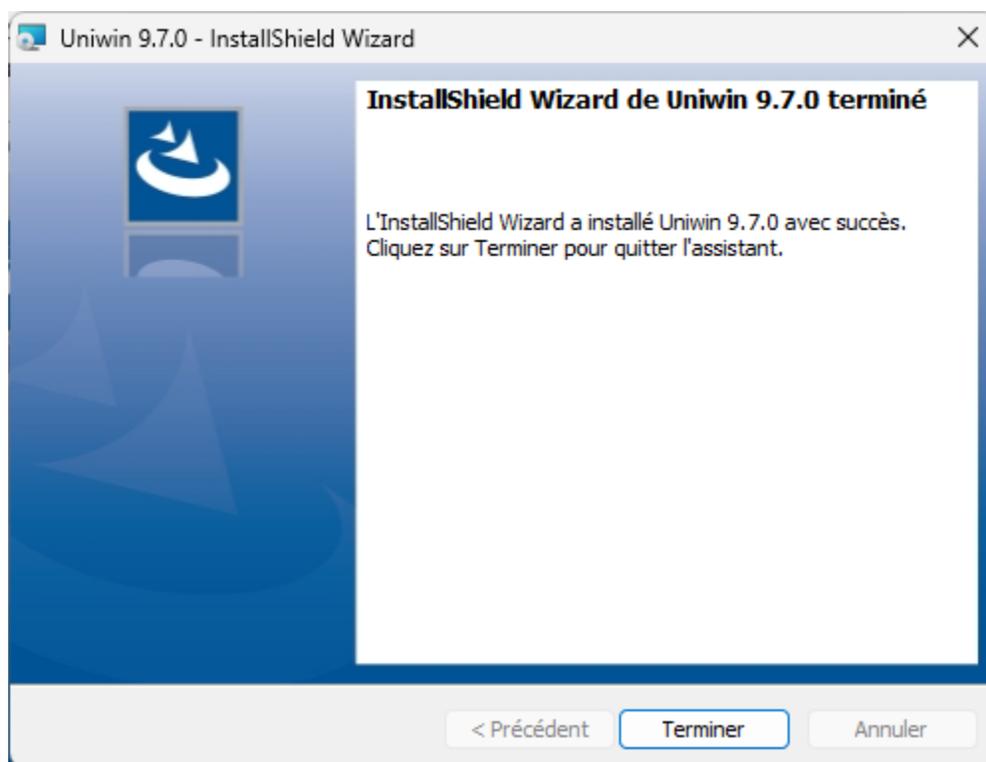


Figure 1-5. Boîte de dialogue finale de l'installation

Cliquez sur *Terminer* pour terminer l'installation.

## 1.2 Demande du fichier de licence du logiciel

Dans le cadre de la procédure d'installation, un raccourci vers UNIWIN 9.7.0 sera ajouté dans le menu *Démarrer* de Windows ainsi que sur votre bureau. Pour utiliser le logiciel :

**Etape 1 :** Dans le menu *Démarrer* de Windows, rechercher l'application *Francestat – Uniwin*, faire un clic droit, sélectionner *Plus puis Exécuter en tant qu'administrateur* comme montré dans la figure ci-dessous.

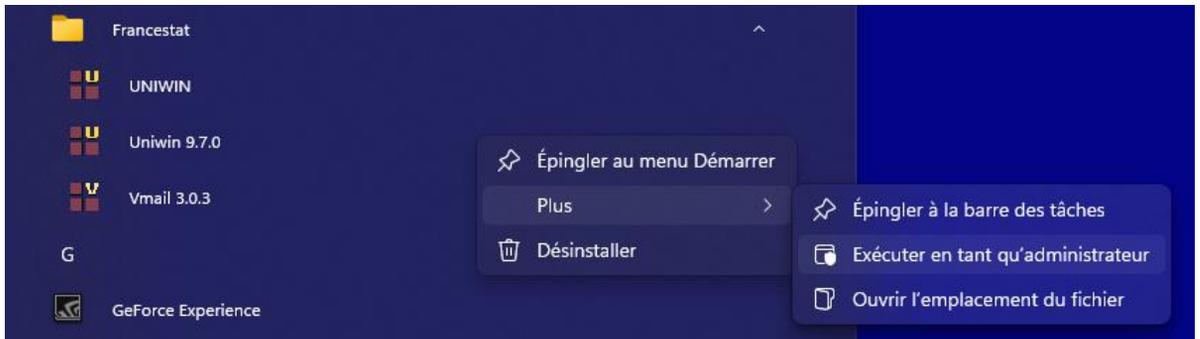


Figure 1-6. Utiliser le logiciel

Vous pouvez également localiser le fichier *Uniw.in.exe* dans le répertoire d'installation du logiciel, sélectionner ce fichier, faire un clic droit et choisir *Exécuter en tant qu'administrateur* dans le menu popup.

**Etape 2 :** Lorsque UNIWIN 9.7.0 est chargé pour la première fois, il ouvre une fenêtre d'enregistrement :

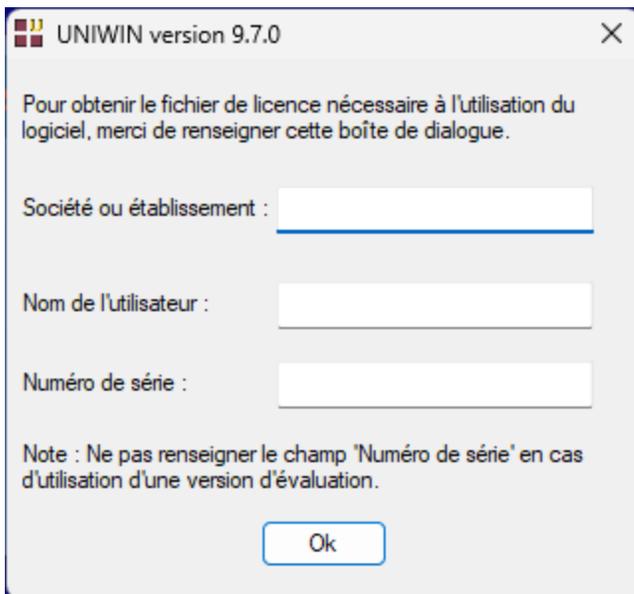


Figure 1-7. Fenêtre d'enregistrement

Entrer le nom de votre société ou établissement, votre nom d'utilisateur et le numéro de série du logiciel (sauf en cas d'utilisation d'une version d'évaluation).

Après avoir cliqué sur le bouton *Ok*, une nouvelle fenêtre s'affiche contenant les informations permettant de demander le fichier de licence nécessaire pour l'utilisation du logiciel.

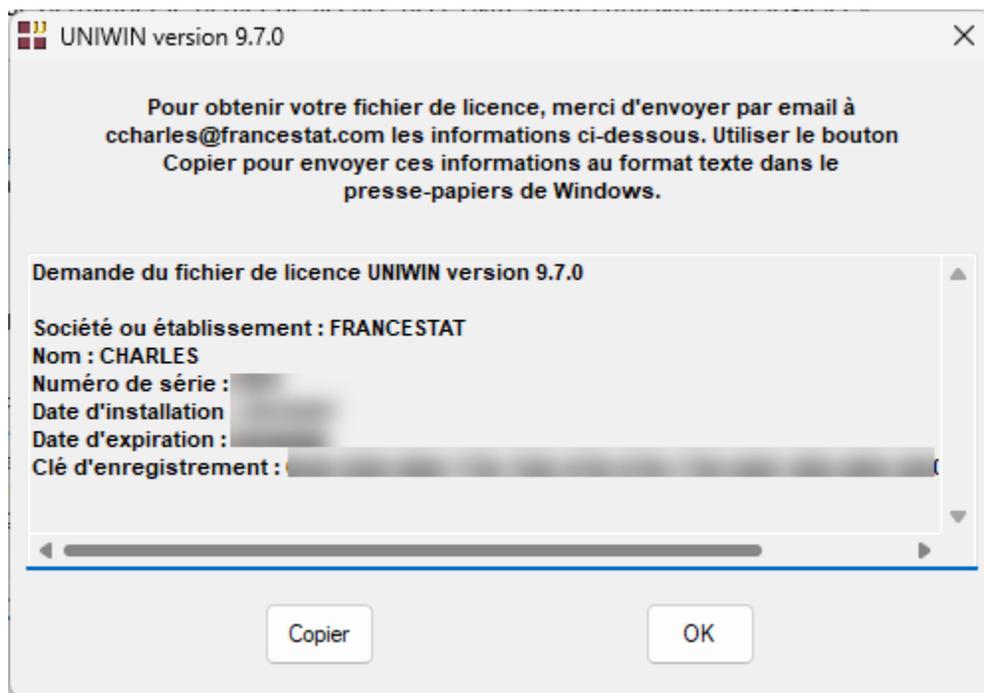


Figure 1-8. Demande du fichier de licence

Utiliser le bouton *Copier* de cette boîte de dialogue pour copier ces informations dans le presse-papiers de Windows puis dans votre message électronique à envoyer à l'adresse de messagerie indiquée en haut de cette boîte de dialogue. Vous recevrez en retour un fichier de licence *Uninwin90.lic*. Ce fichier devra être copié dans le répertoire d'installation du logiciel qui est par défaut :

*C:\Program Files (x86)\Francestat\Uninwin 9.7.0*

**Etape 3 :** Une fois le fichier de licence reçu et copié dans le répertoire du logiciel, le logiciel se charge et la fenêtre principale s'affiche :

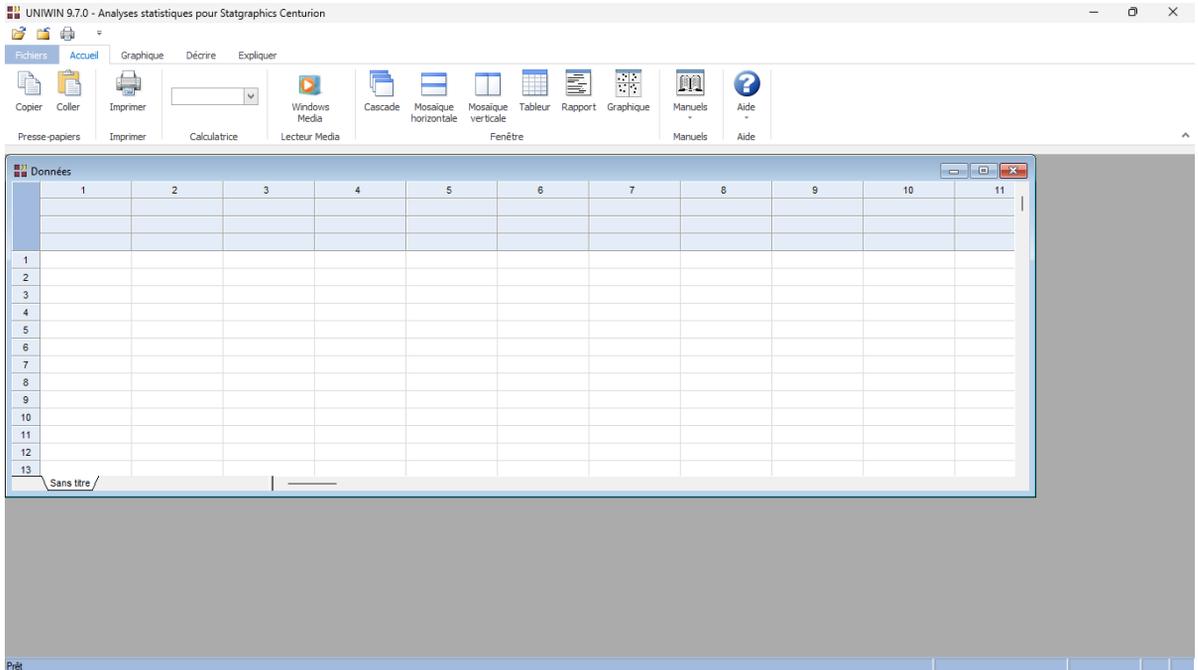


Figure 1-9. Fenêtre principale du logiciel UNIWIN version 9.7.0.

## 1.3 Installation nécessaire sur chaque poste client

Dans le cas d'une installation en réseau (et uniquement dans ce cas), il faut en plus de l'installation sur le serveur sur chaque poste client connecté au serveur et souhaitant utiliser le logiciel :

- Déclarer un raccourci pointant vers le fichier *Unwin.exe* présent dans le répertoire d'installation du logiciel sur le serveur.
- Créer un répertoire et copier depuis le serveur dans ce répertoire les quatre fichiers suivants :

LC.Objects.RibbonDX.dll  
 LC.Objects.RibbonDX.tlb  
 RegAsmInvoker.exe  
 RegisterRibbonDX.bat

- Exécuter dans une fenêtre DOS depuis le répertoire créé sur le poste client et avec les droits administrateur les trois commandes suivantes :

```

Regsvr32.exe APLGrid.dll
Regsvr32.exe APLDraw.ocx
RegisterRibbonDX.bat

```

## 1.4 Enregistrement des DLL et OCX

Si les enregistrements nécessaires des fichiers APLGrid.dll, APLDraw.ocx et RibbonDX.dll ne se font pas automatiquement lors de la première session du logiciel sur le poste client, il faut dans le menu *Démarrer* de Windows, rechercher l'application *Francestat – UniwIn*, faire un clic droit, sélectionner *Plus* puis *Exécuter en tant qu'administrateur* comme montré ci-dessous.

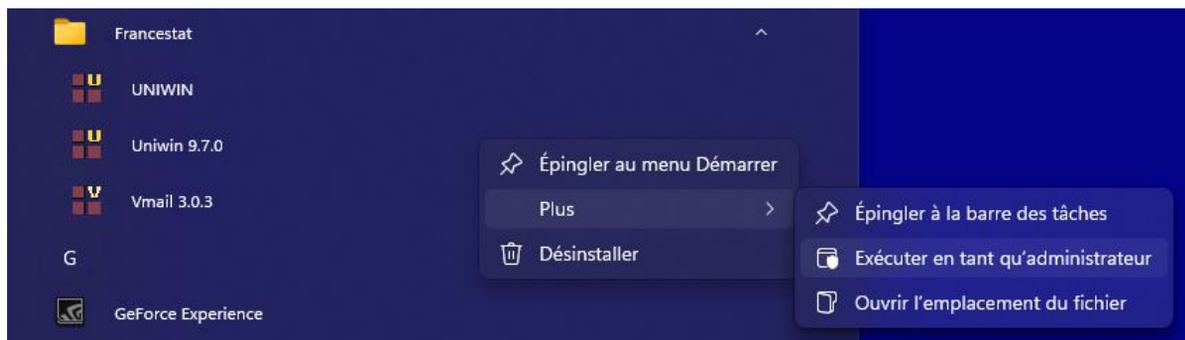


Figure 1-10. Utiliser le logiciel

Vous pouvez également localiser le fichier *UniwIn.exe* dans le répertoire d'installation du logiciel sur le poste client, sélectionner ce fichier, faire un clic droit et choisir « *Exécuter en tant qu'administrateur* » dans le menu popup.

Une autre solution consiste à ouvrir sur le poste client avec les droits administrateur une fenêtre DOS et à faire les actions suivantes :

```

chemin\regsvr32.exe chemin\APLGrid.dll
chemin\regsvr32.exe chemin\APLDraw.ocx
chemin\RegisterRibbonDX.bat

```

où *chemin* indique le chemin d'installation du logiciel, par défaut :

*C:\Program Files (x86)\Francestat\Uniwinn 9.7.0*

### Comment ouvrir une fenêtre DOS sous Windows ?

L'accès à la fenêtre DOS se fait par le menu *Démarrer* (en bas à gauche). Un clic droit dans ce menu vous affiche la commande *Exécuter* qui ouvre une fenêtre dans laquelle il faudra taper *cmd.exe*.

## Rubans et barre d'outils rapide

### 2.1 Les rubans

- Le ruban **Fichiers** propose les options suivantes :

Ouvrir	Ouvrir un ou plusieurs fichiers de données
Fichiers récents	Afficher les fichiers récemment ouverts
Fermer	Fermer le ou les fichiers de données ouverts
Supprimer	Supprimer un ou plusieurs fichiers de données
Options du logiciel	Permet de gérer diverses options du logiciel
Installer R et les packages	Permet l'installation de R et des packages nécessaires
Console R	Permet d'accéder à la console R
Quitter le logiciel	Après demande de confirmation, quitte le logiciel

- Le ruban **Accueil** propose les options suivantes :

Copier	Copier usuel de Windows
Coller	Coller usuel de Windows
Imprimer	Configuration de l'impression ou aperçu avant impression
Calculatrice	Affiche une calculatrice
Lecteur Media	Permet de charger le lecteur Windows Media
Gestion des fenêtres	Cascade, mosaïque, tableur, rapport, graphique
Manuel	Accès à ce manuel et aux documentations des procédures
Aide	Aide en ligne, Site web, A propos de

- Le ruban **Graphiques** propose les options suivantes :

Nuages et courbes 2D	Nuages de points ou courbes à 2 dimensions
Bulles 2D	Graphique en bulles à 2 dimensions
Histogramme	Histogramme à classes d'amplitudes inégales
Courbe ROC	Courbe ROC (Receiver Operating Characteristic)
Courbe de Gini	Courbe de concentration de Gini
Cartes géographiques	Cartographie thématique
Composition graphique	Mise en page de graphiques
Microsoft Paint	Accès à Microsoft Paint

- Le ruban **Décrire** propose les options suivantes :

Recodification de variables alphabétiques  
 Calcul d'un tableau de distances euclidiennes  
 Tris à plat et tris croisés  
 Analyse en composantes principales  
 Analyse en composantes principales symbolique  
 Analyse NIPALS  
 Analyse des correspondances simples  
 Analyse des correspondances multiples  
 Analyse d'un tableau de distances ou de dissimilarités  
 Analyse factorielle de données mixtes  
 Analyse en composantes indépendantes (via R)  
 Classification ascendante hiérarchique  
 Classification par cartes auto-organisatrices (via R)  
 Ellipse de tolérance ou de confiance  
 Analyse factorielle multiple  
 Méthode Statis  
 Méthode Statis duale  
 Analyse GPA (via R)  
 Analyse Apriori (via R)

- Le ruban **Expliquer** propose les options suivantes :

Analyse discriminante pas à pas  
 Analyse factorielle discriminante  
 Analyse discriminante bayésienne

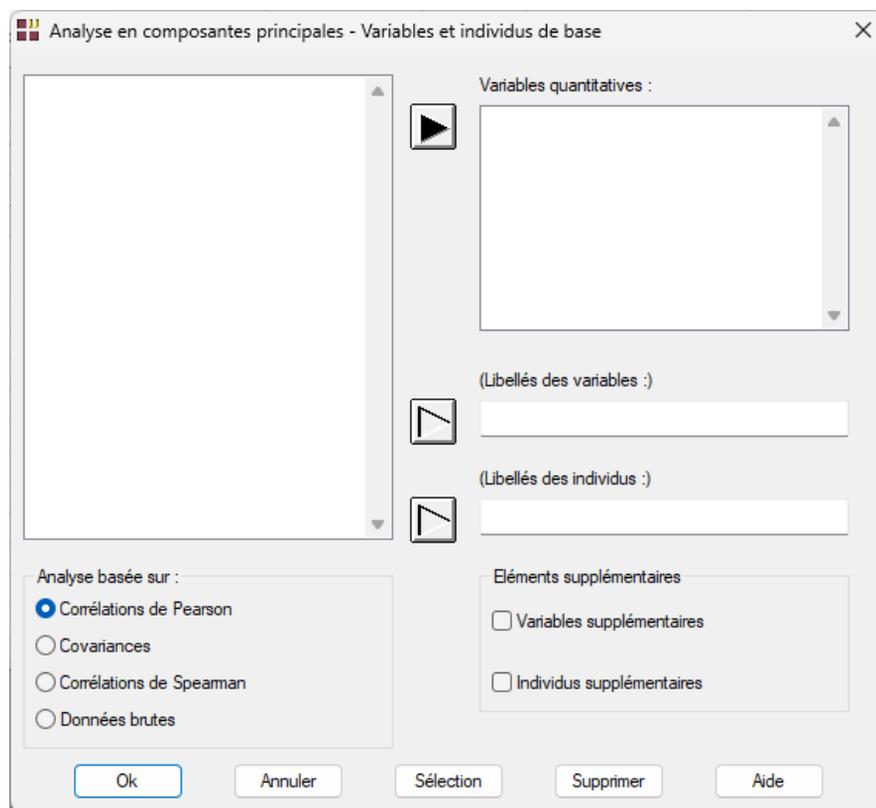
Analyse discriminante qualitative  
Scoring  
Méthode KNN (via R)  
Méthode bayésienne naïve (via R)  
Méthode SIMCA (via R)  
Régression sur composantes principales  
Régression PLS (via R)  
Réseaux de neurones artificiels (via R)  
Arbres de décision et de régression (via R)

## 2.2 La barre d'outils rapide

Cette barre, placée au-dessus des titres des rubans, permet un accès rapide à certaines fonctionnalités. Dans cette version du logiciel, la gestion de ses éventuelles modifications n'est pas gérée.

## Boîtes de dialogue

Dans ce chapitre, nous présentons les différents éléments d'une boîte de dialogue type d'UNIWIN. Comme exemple, affichons la boîte de dialogue de l'Analyse en Composantes Principales (ACP) utilisant les variables du fichier OCDE :



## 3.1 Les éléments d'une boîte de dialogue

Cette boîte de dialogue contient :

- une zone liste à gauche affichant la liste des variables disponibles
- une zone liste en haut à droite qui contiendra les variables à utiliser dans l'analyse
- deux zones de type liste optionnelles (titres entre parenthèses) pour préciser les noms des variables contenant les libellés des variables et des individus
- une zone 'Transformation des variables' composée de trois boutons d'options exclusives
- une zone 'Éléments supplémentaires' composée de deux cases à cocher
- cinq boutons de commande : Ok, Annuler, Sélection, Supprimer et Aide

Il y a deux façons de renseigner les champs de droite :

1. Sélectionner les variables dans la zone liste de gauche puis cliquer sur le bouton  du champ de droite à renseigner qui deviendra . Les variables s'afficheront dans ce champ.
2. Choisir le champ à renseigner en cliquant dans le champ. Le bouton associé  deviendra  puis double-cliquer sur les noms des variables sélectionnées. Ces variables s'afficheront dans ce champ.

## 3.2 Les boutons de commande

### ◇ **Bouton Ok**

Cliquer sur ce bouton permet de prendre en compte les réponses apportées dans la boîte de dialogue et de démarrer le traitement.

### ◇ **Bouton Annuler**

Cliquer sur ce bouton permet d'abandonner le traitement en cours.

#### ◇ **Bouton Sélection**

Cliquer sur ce bouton permet de définir une sélection des données à traiter (voir le paragraphe ‘Sélection des données’ plus loin dans ce chapitre).

#### ◇ **Bouton Supprimer**

Cliquer sur ce bouton permet de supprimer une variable affichée dans un des champs de droite.

#### ◇ **Bouton Aide**

Cliquer sur ce bouton permet d’afficher l’aide en ligne contextuelle et d’accéder à la documentation au format PDF.

### 3.3 Sélection des données

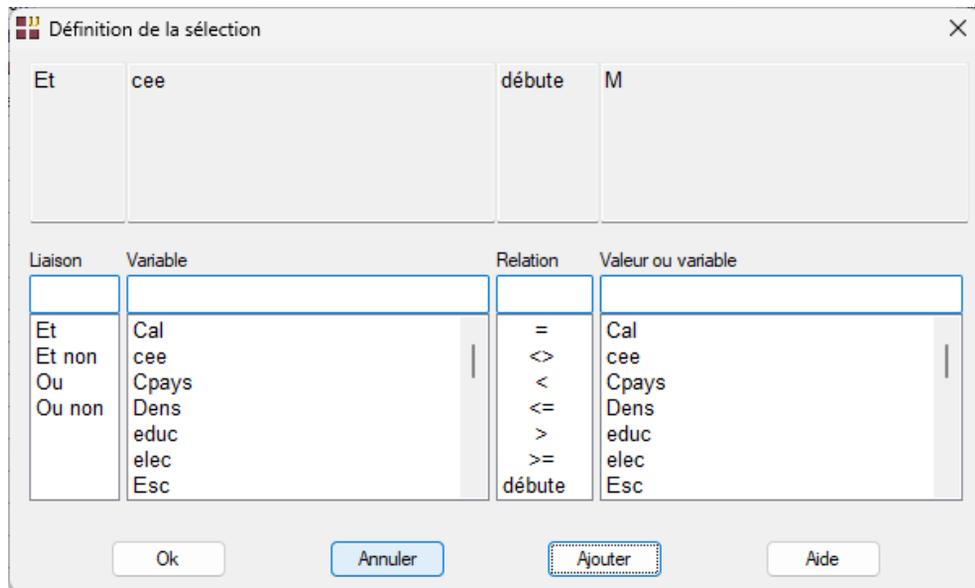
Cliquer sur le bouton de commande ‘Sélection’ permet de définir une sélection qui s’applique à toutes les variables choisies dans la boîte de dialogue qui dépendent du nombre d’individus (observations, sujets, lignes du tableur).

La sélection doit obligatoirement commencer par l’opérateur de liaison Et. et peut être constituée de un à cinq filtres successifs. Les variables utilisées dans les filtres peuvent appartenir à un même fichier ou à des fichiers différents. La syntaxe et la validité des filtres construits sont validées par UNIWIN.

Un filtre est composé de quatre éléments :

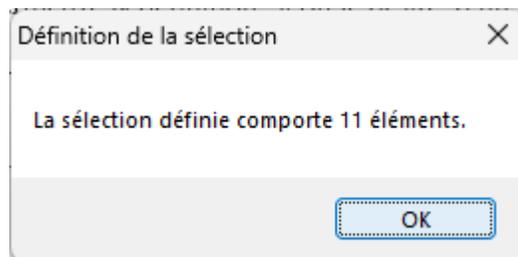
- |                          |   |
|--------------------------|---|
| 1. Opérateur de liaison  | Et, Et non, Ou, Ou non  |
| 2. Variable              | Nom d’une variable d’un des fichiers ouverts  |
| 3. Opérateur de relation | Egal (=), différent de (<>), inférieur à (<), inférieur ou égal à (<=), supérieur à (>), supérieur ou égal à (>=), débute |
| 4. Valeur ou variable    | Valeur numérique, texte ou nom d’une variable d’un des fichiers ouverts   |

Voici un exemple avec le fichier OCDE : il s’agit de sélectionner uniquement les pays (individus) qui appartiennent à la CEE (la variable ncee est codée ‘Membre CEE’ et ‘Non membre CEE’).



Cliquer sur le bouton ‘Ajouter’ a pour effet de mettre cette phrase de sélection dans la partie haute de la boîte de dialogue et de permettre la définition, si on le désire, d’un filtre complémentaire.

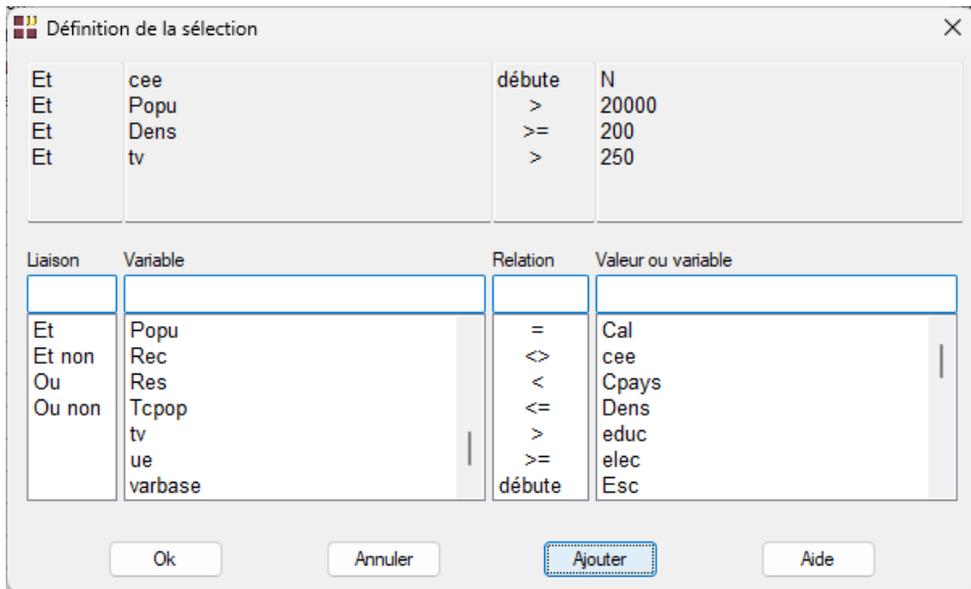
Cliquer sur le bouton ‘Ok’ nous retourne à la boîte de dialogue de l’analyse et affiche le nombre d’individus sélectionnés.



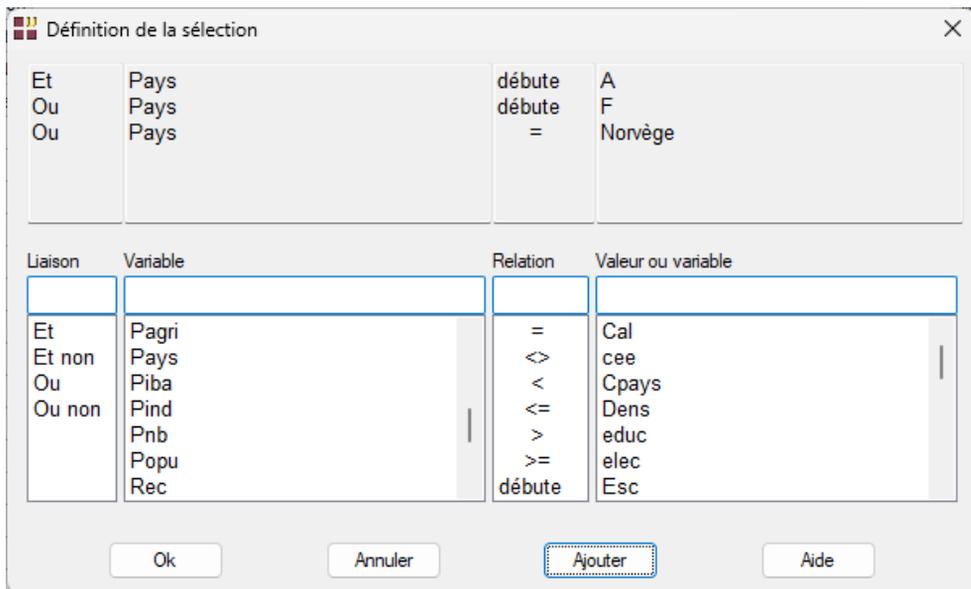
Cliquer sur le bouton ‘Annuler’ annule la sélection définie ou active.

A titre illustratif, voici deux autres exemples de sélections.

Le premier exemple montre une sélection contenant quatre filtres définis à partir des variables du fichier OCDE.

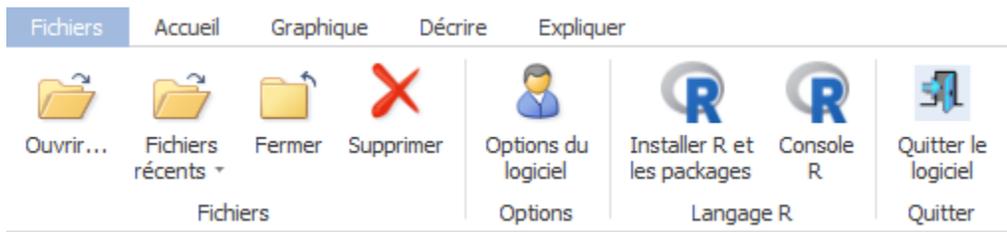


Le deuxième exemple utilise l'opérateur 'débute'. Il permet de sélectionner les pays dont les noms commencent par 'A' ou par 'F' ainsi que le pays 'Norvège'.



## Le ruban Fichiers

Le ruban 'Fichiers' permet d'ouvrir, de fermer et de supprimer un ou plusieurs fichiers :



Il permet également :

- de définir certaines options du logiciel
- d'installer R et les packages nécessaires aux procédures
- d'accéder à la console R
- de quitter UNIWIN

Les fichiers en entrée pour UNIWIN sont des fichiers au format .SGD provenant du logiciel STATGRAPHICS Centurion.

Les fichiers en sortie d'UNIWIN sont aux formats .SGD pour les données et HTML pour les rapports.

A noter que ce format HTML permet notamment la reprise des rapports issus d'UNIWIN dans Microsoft Excel.

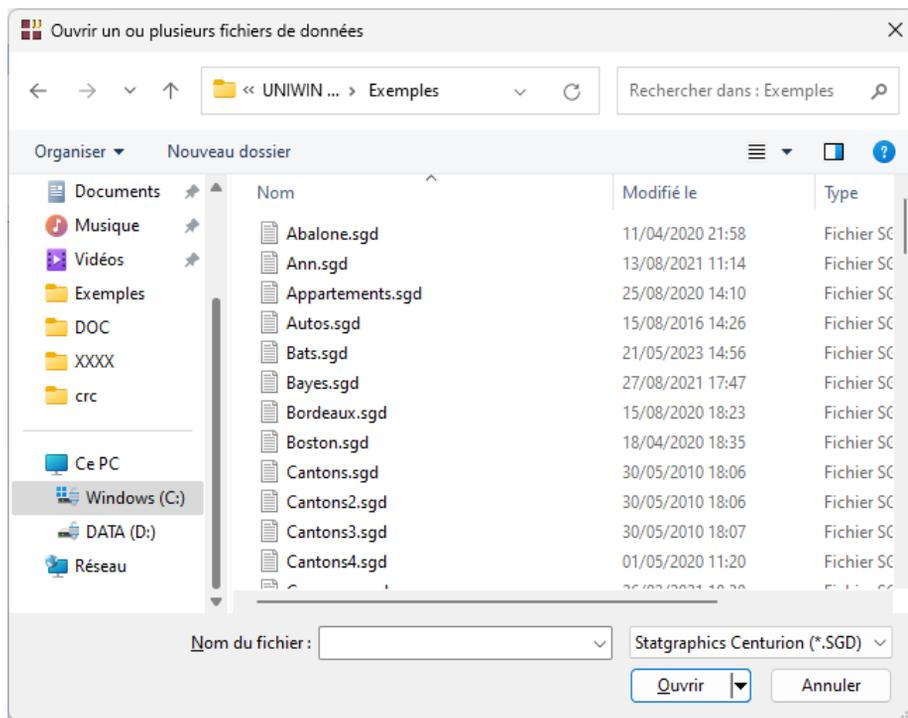
## 4.1 L'option 'Ouvrir'

L'option 'Ouvrir' affiche la boîte de dialogue permettant de choisir le ou les fichiers à ouvrir (contenus dans un même répertoire), c'est-à-dire le ou les fichiers à utiliser dans les procédures d'UNIWIN.

A noter que si plusieurs fichiers de données sont ouverts simultanément, les variables utilisées dans les procédures n'ont pas à appartenir obligatoirement à un seul de ces fichiers.

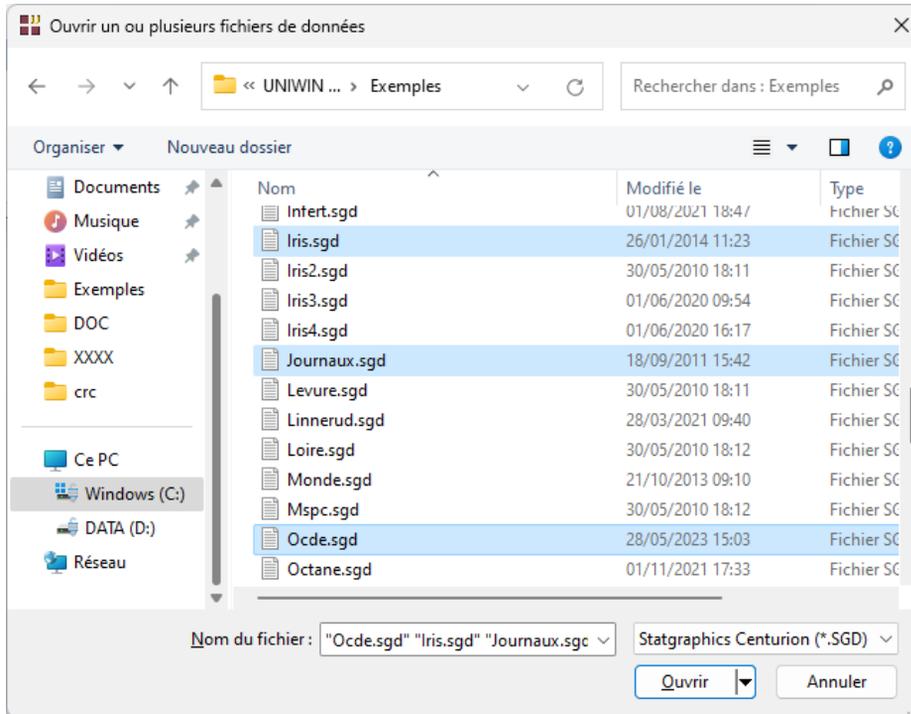
Les noms des fichiers UNIWIN doivent respecter les contraintes pour les noms des fichiers sous Windows.

Voici la boîte de dialogue 'Ouvrir' telle qu'elle s'affiche pour les fichiers STATGRAPHICS Centurion et le tableur des données associé :



	Pays	Cpays	varbase	varsup	Popu	Dens	Tctop	Pagri	Pind	Pnb	Piba	Fbcf	Rec
	Pays	Code pays	Variables de base	Variables supp.	Population	Densité de population	Taux croissance popu	Pct popu agriculture	Pct popu industrie	Produit national brut	Pct pnb agri	Form. brute capital fixe	Recettes
	Type = C	Type = C	Type = C	Type = C	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N
	Longueur = 18	Longueur = 18	Longueur = 14	Longueur = 4	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18	Longueur = 18
1	Allemagne	D	Population	Logements	60348	245	1,05	9,6	49,1	2520	3,6	24,4	37,9
2	Autriche	A	Densité	Electricité	7373	88	0,50	19,1	39,9	1690	7,0	23,2	37,5
3	Belgique	B	Acc popu	Education	9984	332	0,60	5,4	44,8	2352	5,4	23,1	35,1
4	Canada	CDN	Pct popu agri	Tv	21089	2	1,85	8,2	32,3	3460	5,9	21,7	35,2
5	Danemark	DK	Pct popu ind		4893	114	0,75	11,9	38,5	2860	8,9	22,0	37,1
6	Espagne	E	Pnb		32949	65	0,95	30,7	37,1	870	15,0	22,0	22,4
7	Usa	USA	Pct pnb agri		203213	22	1,35	4,6	33,7	4660	2,9	16,7	31,5
8	Finlande	SF	Fbcf		4706	14	0,70	24,5	34,6	1940	14,7	23,8	35,9
9	France	F	Recettes		50325	91	1,05	15,1	40,6	2770	6,0	25,4	38,1
10	Grèce	GR	Réserves		8666	67	0,70	48,2	22,5	950	20,3	29,7	26,9
11	Irlande	SE	Esompte		2921	42	0,25	28,4	29,7	1040	19,7	19,9	30,7
12	Italie	I	Importations		54120	180	0,85	21,5	43,1	1520	11,3	20,5	33,3
13	Japon	JAP	Exportations		102380	277	1,05	18,8	35,0	1630	8,7	35,2	21,2
14	Norvège	N	Calories		3851	12	0,80	14,7	36,8	2530	6,5	25,3	43,4
15	Pays Bas	NL			12873	352	1,25	7,5	41,3	2190	7,0	25,5	41,9
16	Portugal	P			9583	105	0,90	31,5	35,5	600	17,7	18,4	24,0
17	Grande Bretagne	RUN			55643	228	0,85	2,9	46,8	1970	3,0	17,3	39,0
18	Suède	S			7989	18	0,70	8,8	40,4	3230	5,9	23,6	48,1
19													
20													
21													

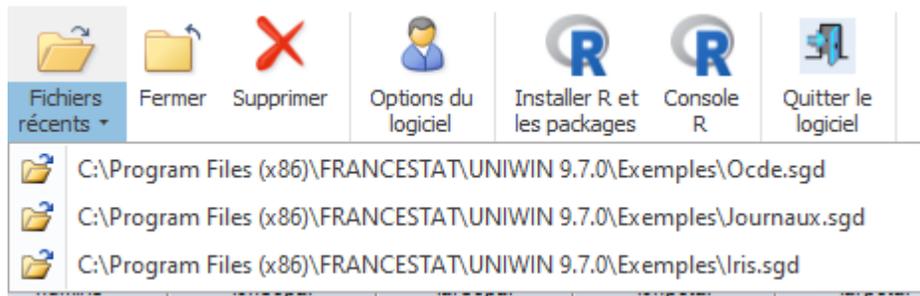
Si plusieurs fichiers sont ouverts simultanément, le tableau des données affiche alors plusieurs feuilles :



Données															
mesures	nomgroupe	bfrcar	bincar	binsys	bprdia	bpapul	bpvent	brepul	bgroupe	blabobs	efcar	enincar	einsys	eprdia	
Noms des mesures	Noms des groupes	frcar base	incarc base	insys base	prdia base	papul base	pvent base	repul base	Groupes de base	Labels obs. base	frcar essais	incarc essais	insys essais	prdia essais	
Type = C	Type = C	Type = N	Type = C	Type = N	Type = N	Type = N	Type = N								
Longueur = 7	Longueur = 2	Longueur = 81	Longueur = 81	Longueur = 20	Longueur = 20	Longueur = 20	Longueur = 20								
1	frcar	Décédé	105	1,84	17,5	18	22,0	10,0	957	1	Obs011	90	1,68	18,7	24,0
2	incarc	Vivant	87	1,20	13,8	34	41,0	20,0	2733	1	Obs012	120	1,40	11,7	23,0
3	insys		85	1,19	18,3	15	18,0	13,0	1210	1	Obs013	80	1,58	19,7	21,0
4	prdia		103	0,91	8,6	30	33,5	10,0	2945	1	Obs014	80	1,13	14,1	18,0
5	papul		90	0,95	10,6	20	24,0	6,0	2021	1	Obs015	110	1,86	15,1	23,0
6	pvent		95	0,99	10,4	20	27,5	8,0	2222	1	Obs016	80	1,50	18,7	13,0
7	repul		95	0,85	8,9	19	22,0	15,5	2071	1	Obs017	118	1,03	8,7	19,0
8			70	1,44	20,6	19	26,5	11,0	1472	1	Obs018	95	1,89	19,9	25,0
9			94	1,31	13,9	26	40,0	15,0	2443	1	Obs019	80	1,45	18,1	19,0
10			79	1,29	16,3	24	31,0	10,0	1922	1	Obs020	85	1,30	15,1	13,0
11			75	1,21	16,1	19	24,0	4,0	1587	1	Obs021	80	2,85	35,6	25,0
12			110	1,24	11,3	22	27,5	11,0	1774	1	Obs022	100	2,31	23,1	8,0
13			116	1,85	15,9	33	42,0	13,0	1816	1	Obs023	101	2,55	25,2	23,2
14			92	1,97	21,4	18	27,0	3,0	1096	1	Obs024	92	2,17	23,5	19,0
15			102	1,60	15,7	24	31,0	16,0	1550	1	Obs025	80	1,59	19,9	13,0
16			80	0,82	13,7	22	32,0	13,0	3122	1	Obs026	90	1,45	16,1	17,0
17			92	1,37	14,8	25	46,0	11,0	2686	1	Obs027	87	2,37	27,2	15,0
18			100	1,38	13,8	20	31,0	11,0	1797	1	Obs028	108	2,40	22,2	26,0
19			87	2,51	28,8	16	24,0	20,0	765	1	Obs029	86	2,36	27,4	24,0
20			120	1,18	9,9	25	36,0	8,0	2441	1	Obs030	90	3,03	33,6	17,0
21			115	1,83	15,9	25	30,0	8,0	1311	1	Obs031				
22			87	1,42	16,1	20	26,0	10,0	1485	1	Obs032				

## 4.2 L'option 'Fichiers récents'

L'option 'Fichier récents' permet d'ouvrir un des fichiers récemment utilisés.



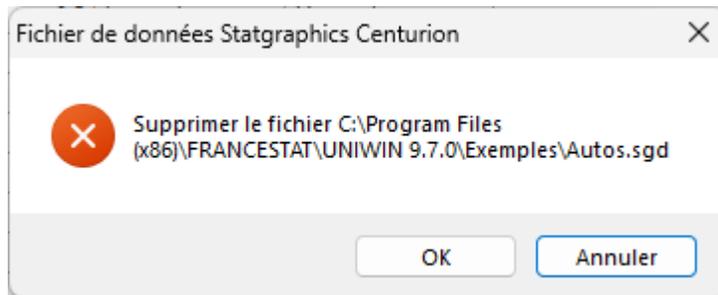
## 4.3 L'option 'Fermer'

L'option 'Fermer' fait apparaître la boîte de dialogue montrée ci-après.

Cette boîte permet de fermer le ou les fichiers de données ouverts.

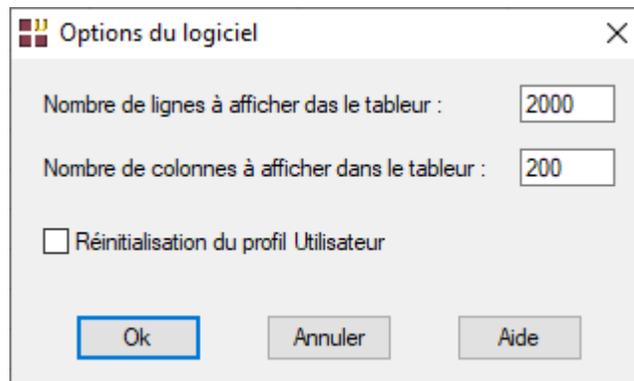


Cette option permet d'effacer un fichier. Attention, par cette option vous effacez définitivement le fichier sélectionné. UNIWIN vous demande de confirmer la suppression du fichier.



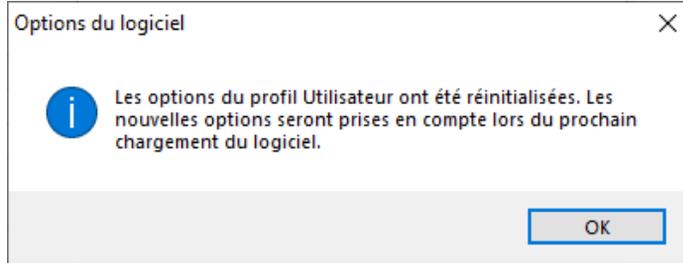
## 4.5 L'option 'Options du logiciel'

Cette option permet de définir les nombres de lignes et de colonnes affichées dans le tableur et de réinitialiser toutes les options graphiques de toutes les procédures d'UNIWIN aux valeurs par défaut, c'est-à-dire aux valeurs initiales lors de l'installation du logiciel. Elle réinitialise également la liste des derniers fichiers de données ouverts. UNIWIN vous demande de confirmer cette réinitialisation :



Note : Les nombres de lignes et de colonnes indiqués concerne uniquement l'affichage des données dans le tableur et ne représentent pas les limites des volumes de données pouvant être traités par le logiciel.

Cocher la case 'Réinitialisation du profil Utilisateur' puis cliquer sur le bouton 'Ok', affiche la boîte de dialogue suivante :

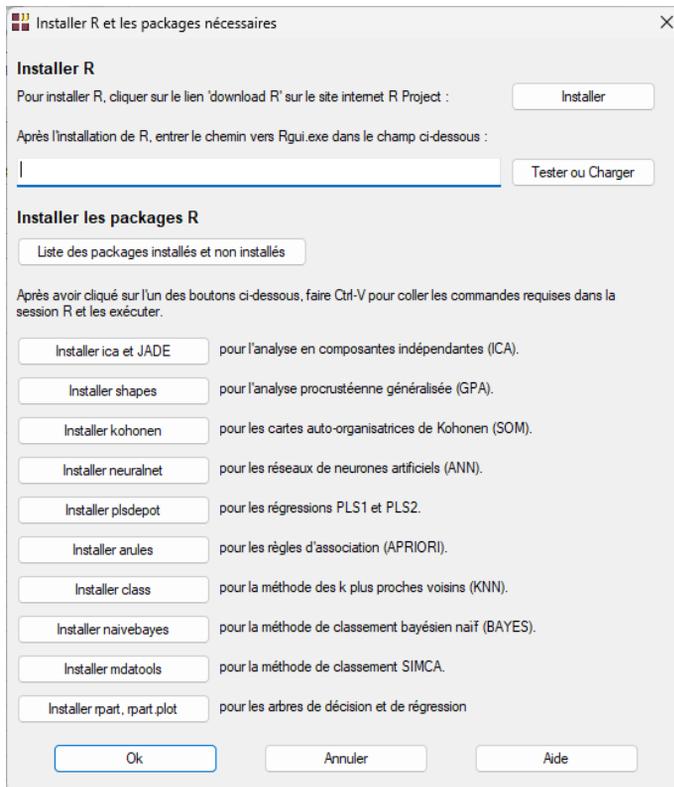


Si vous ne confirmez pas, UNIWIN revient au menu général sans faire la réinitialisation.

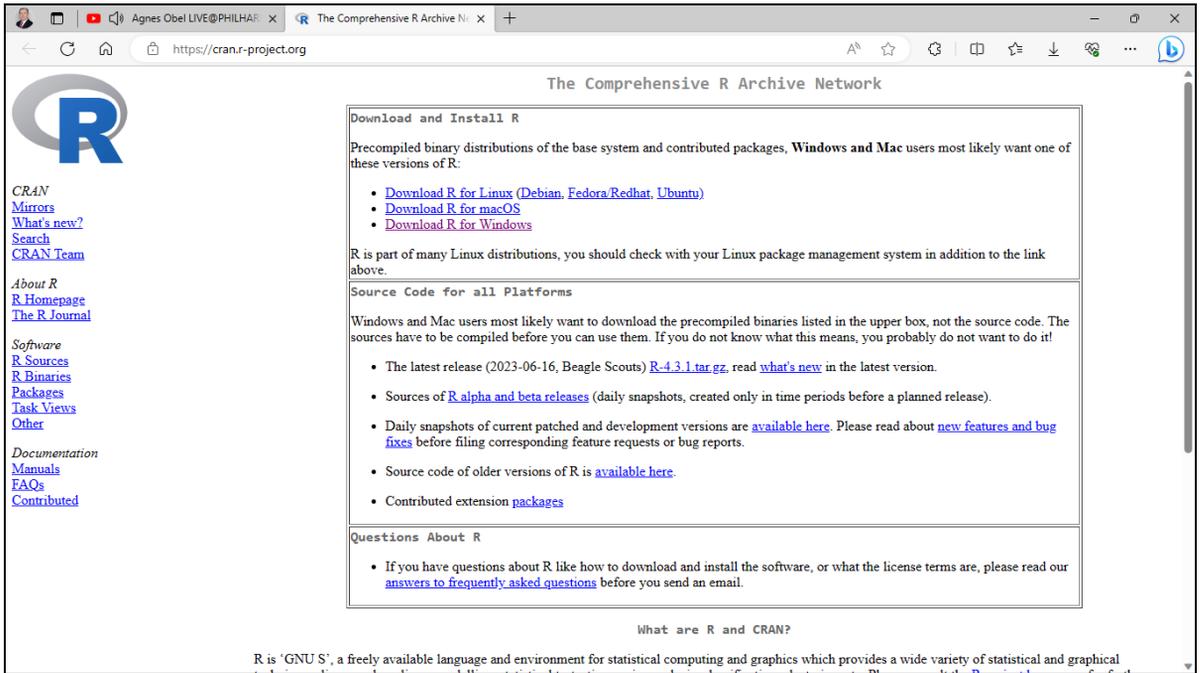
## 4.6 L'option 'Installer R et les packages nécessaires'



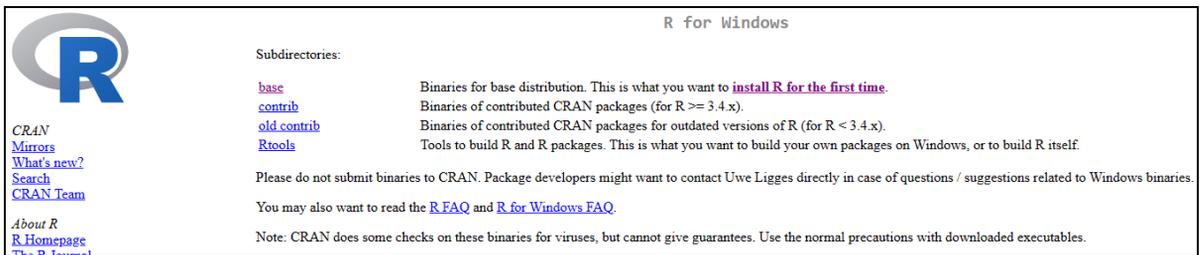
Dans le ruban Fichiers, cliquer sur l'icône . La fenêtre suivante s'affiche alors :



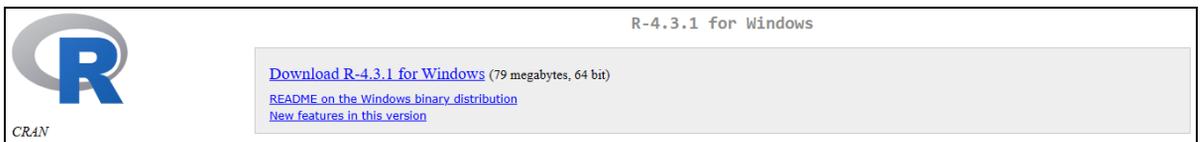
Cliquer sur le bouton 'Installer' pour installer le langage R. La page Internet suivante s'affiche :



Cliquer sur le lien 'Download R for Windows'.



Cliquer sur 'Install R for the first time'.

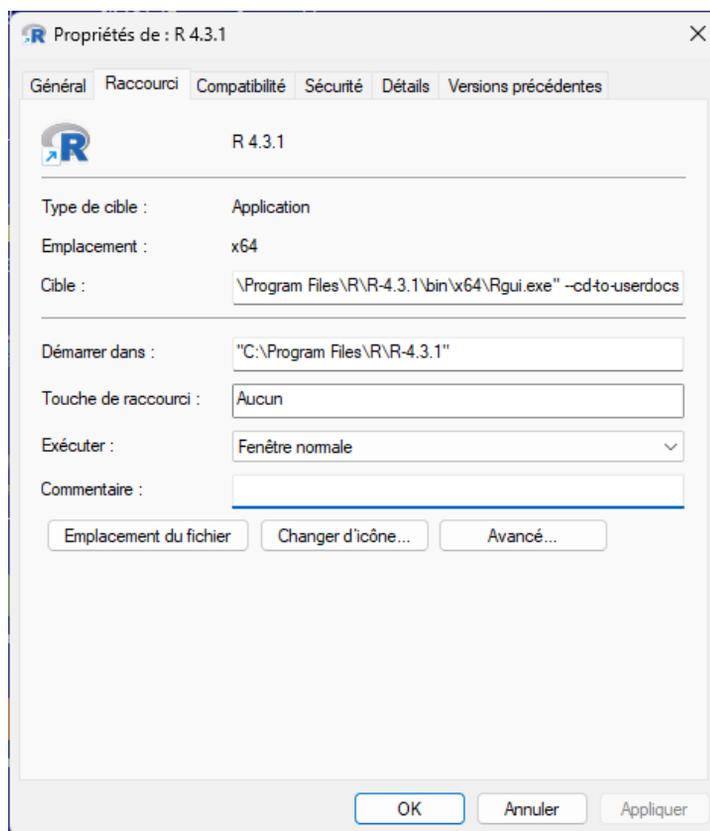


Cliquer sur 'Download R 4.3.1 for Windows' (ou version supérieure).

Enregistrer le fichier téléchargé sous le nom 'R-4.3.1-win.exe'.

Exécuter ce fichier pour installer R et demander la création d'icônes sur le bureau.

Cliquer avec le bouton droit sur l'icône créé sur le bureau et sélectionner 'Propriétés' dans le menu popup. Une fenêtre similaire à celle montrée ci-après s'affiche.



Sélectionner le début du contenu du champ 'Cible', soit dans notre cas :

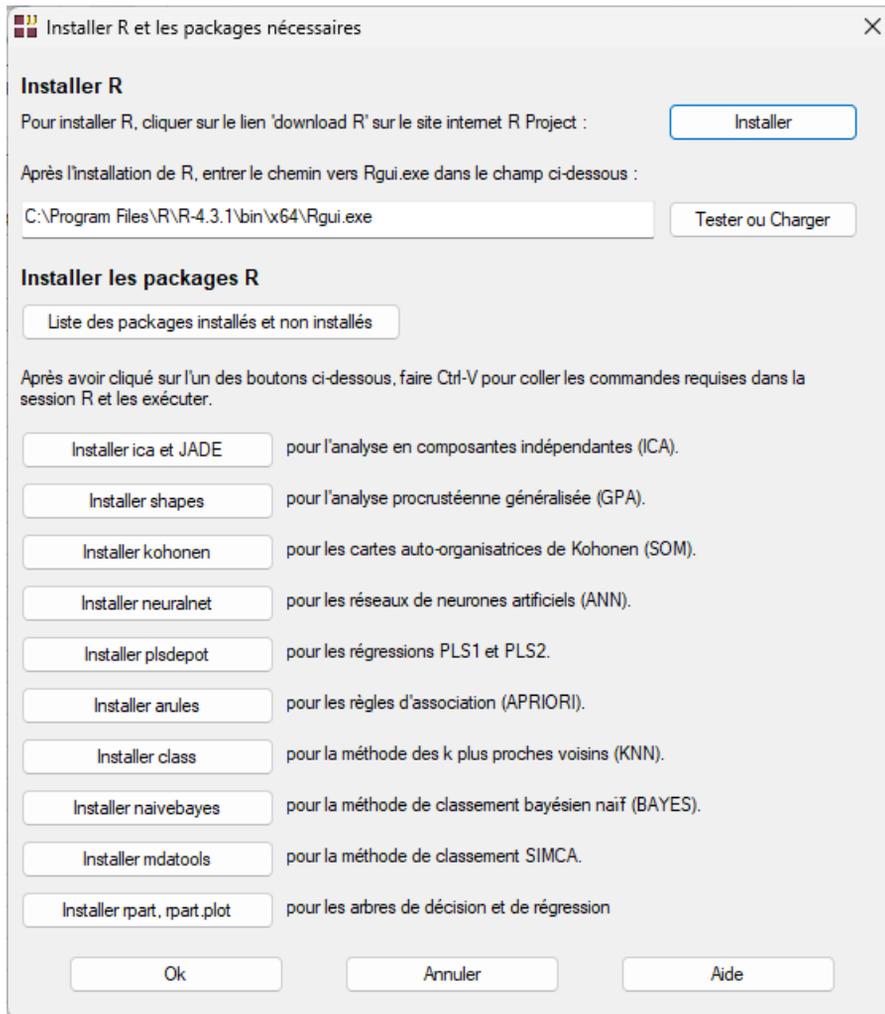
C:\Program Files\R\R-4.3.1\bin\x64\Rgui.exe

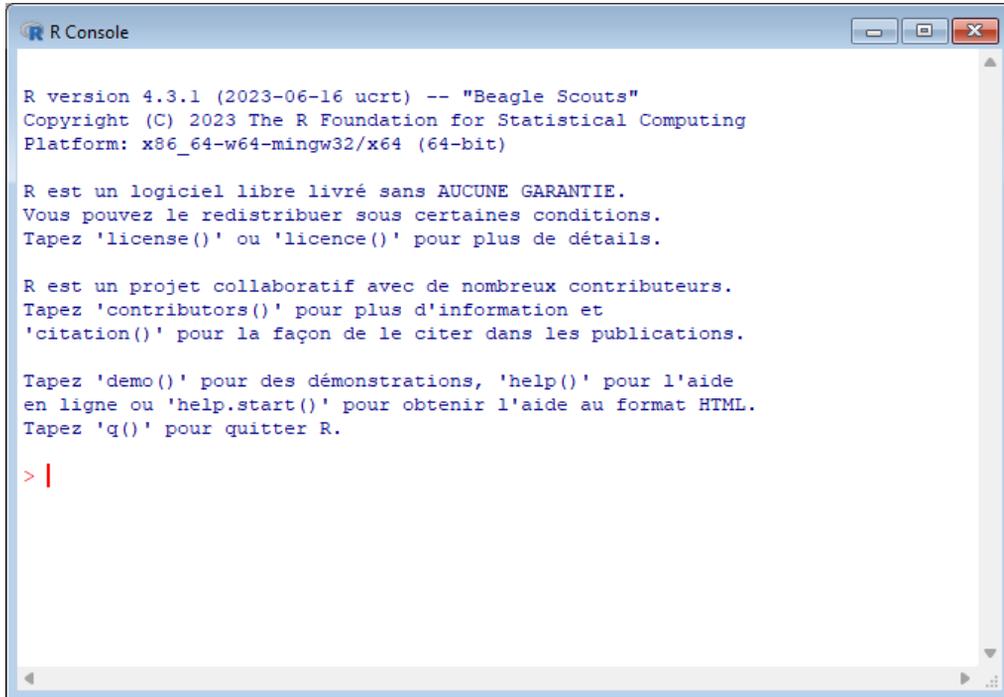
et copier ce texte dans le presse-papiers de Windows.

Dans la fenêtre UNIWIN d'installation de R, coller ce contenu dans le champ précisant le chemin vers le fichier Rgui.exe.

Cliquer ensuite sur tester pour vérifier que le lancement de R s'effectue correctement. Pour quitter R, entrer 'q()' dans la fenêtre (console) puis cliquer sur 'Non' pour ne pas sauvegarder de session.

Cliquer ensuite sur 'Ok' dans la fenêtre d'UNIWIN. R est maintenant installé et son chemin d'accès est enregistré dans le logiciel.





```
R version 4.3.1 (2023-06-16 ucrt) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

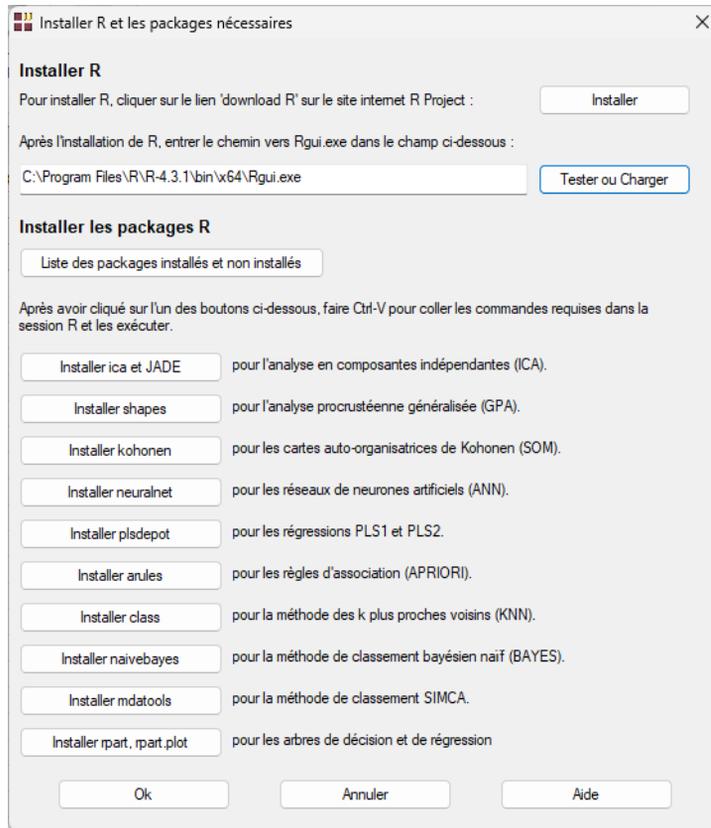
## Installer les packages nécessaires



Dans le ruban 'Fichiers', cliquer sur l'icône . La fenêtre montrée ci-après s'affiche alors.

Pour chacune des options concernant les packages R, cliquer sur le bouton pour placer dans le presse-papiers de Windows la commande R qui a été générée.

Il suffit ensuite de charger R (par exemple en cliquant sur le bouton "Tester ou Charger") et de coller le contenu du presse-papiers dans la fenêtre R (console).



## Liste des packages installés et non installés

Le bouton 'Liste des packages installés et non installés' permet de visualiser les noms des packages déjà installés et ceux non encore installés.

## Installer le package 'ica'

Si vous souhaitez utiliser la procédure 'ICA', il faut installer le package 'ica'.

Pour cela cliquer sur le bouton 'Installer ica' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package 'JADE'**

Si vous souhaitez utiliser l'algorithme 'SOBI' dans procédure 'ICA', il faut, en plus du package 'ica', installer le package 'JADE'.

Pour cela cliquer sur le bouton 'Installer JADE' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package 'shapes'**

Si vous souhaitez utiliser la procédure 'GPA', il faut installer le package 'shapes'.

Pour cela cliquer sur le bouton 'Installer shapes' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package 'kohonen'**

Si vous souhaitez utiliser la procédure 'SOM', il faut installer le package 'kohonen'.

Pour cela cliquer sur le bouton 'Installer kohonen' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package 'neuralnet'**

Si vous souhaitez utiliser la procédure 'ANN', il faut installer le package 'neuralnet'.

Pour cela cliquer sur le bouton 'Installer neuralnet' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package 'plsdepot'**

Si vous souhaitez utiliser les procédures 'PLS', il faut installer le package 'plsdepot'.

Pour cela cliquer sur le bouton 'Installer plsdepot' pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package ‘arules’**

Si vous souhaitez utiliser la procédure ‘APRIORI’, il faut installer le package ‘rules’.

Pour cela cliquer sur le bouton ‘Installer rules’ pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package ‘class’**

Si vous souhaitez utiliser la procédure ‘KNN’, il faut installer le package ‘class’.

Pour cela cliquer sur le bouton ‘Installer class’ pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package ‘naivebayes’**

Si vous souhaitez utiliser la procédure ‘BAYES’, il faut installer le package ‘naivebayes’.

Pour cela cliquer sur le bouton ‘Installer naivebayes’ pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

### **Installer le package ‘mdatools’**

Si vous souhaitez utiliser la procédure ‘SIMCA’, il faut installer le package ‘mdatools’.

Pour cela cliquer sur le bouton ‘Installer mdatools’ pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

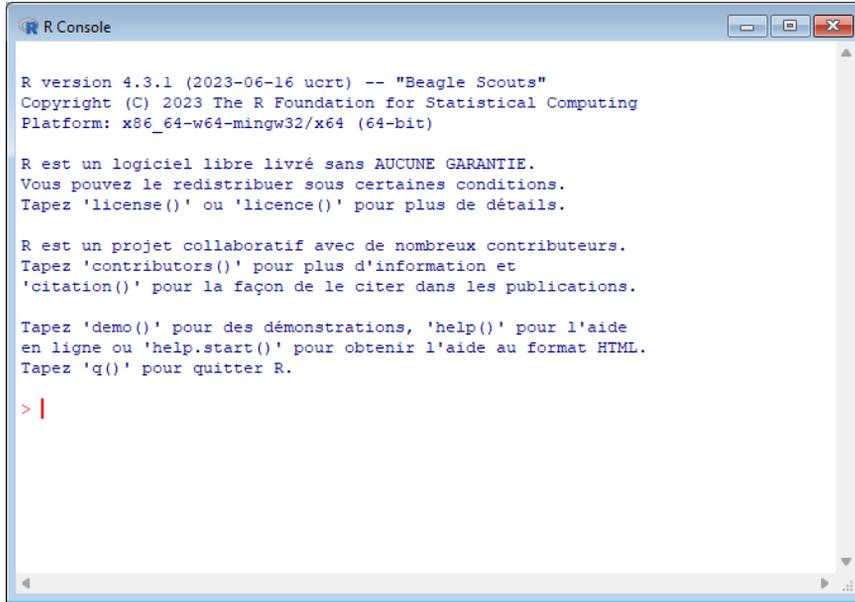
### **Installer les packages ‘rpart’ et ‘rpart.plot’**

Si vous souhaitez utiliser la procédure ‘ARBRE’, il faut installer les packages ‘rpart’ et ‘rpart.plot’.

Pour cela cliquer sur le bouton ‘Installer rpart, rpart.plot’ pour générer la commande R pour cette installation, coller cette commande dans la console R et faire Entrée.

## 4.7 L'option 'Console R'

Cette option permet de charger la console R.



```
R Console

R version 4.3.1 (2023-06-16 ucrt) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

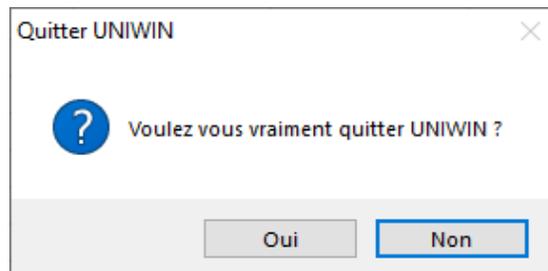
R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

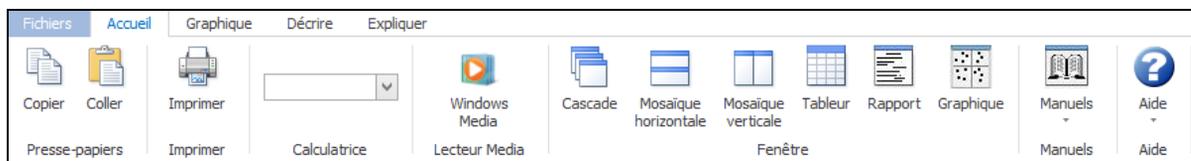
## 4.8 L'option 'Quitter le logiciel'

Cette option permet de quitter le logiciel UNIWIN. Elle affiche la boîte de dialogue montrée ci-dessous. Si vous cliquez sur 'Oui', vous quittez l'application UNIWIN.



## Le ruban Accueil

Le ruban ‘Accueil’ permet de copier/coller, d’imprimer, d’afficher une calculatrice, de charger le lecteur Windows Media, de gérer les fenêtres, d’accéder au manuel de l’utilisateur et aux documentations des procédures ainsi qu’au système d’aide en ligne :



### 5.1 Les options ‘Copier / Coller’

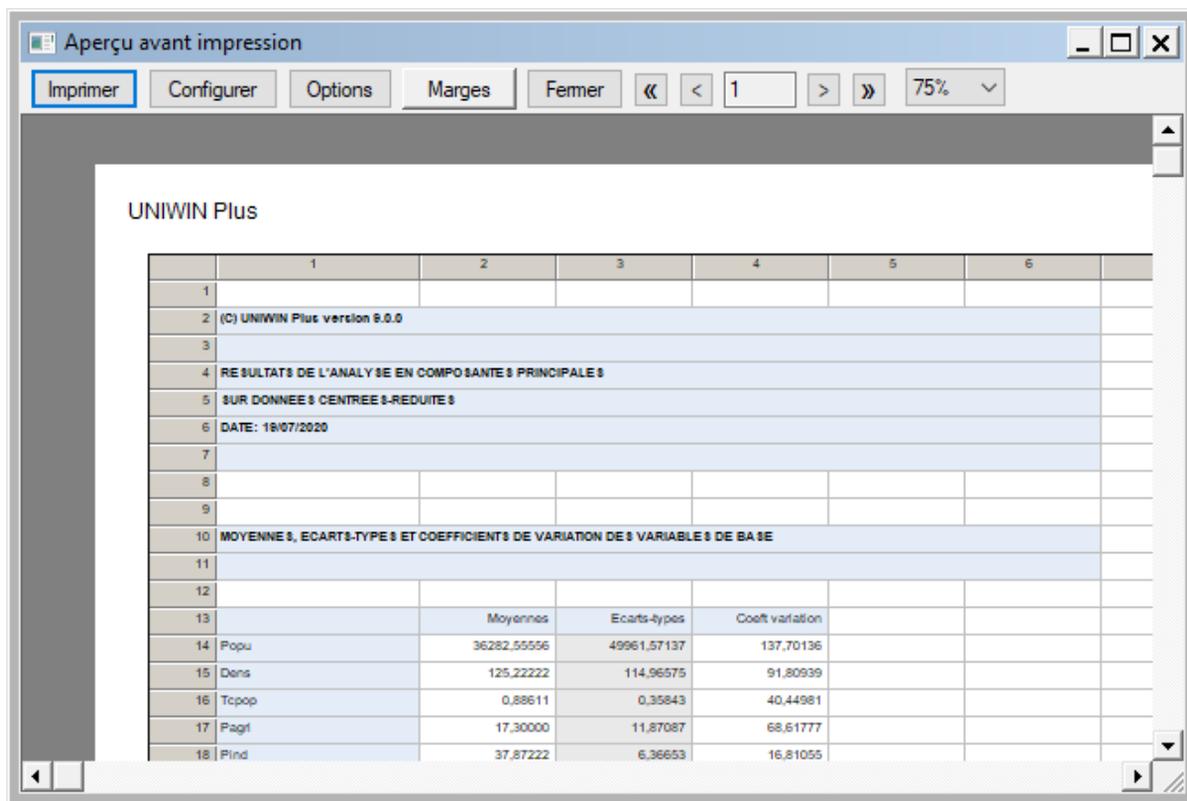
L’option ‘Copier’ permet de copier des cellules du tableau affichant les données ou les rapports des diverses analyses dans une autre application Windows.

Pour un graphique, il est possible d’utiliser cette option ou l’option ‘Copier’ affichée dans la barre d’outils du graphique.

Note : L’option ‘Coller’ est inactive dans cette version du logiciel.

### 5.2 L’option ‘Imprimer’

L’option ‘Imprimer’ permet d’afficher un aperçu avant impression pour un rapport ou la boîte de configuration de l’impression pour un graphique.



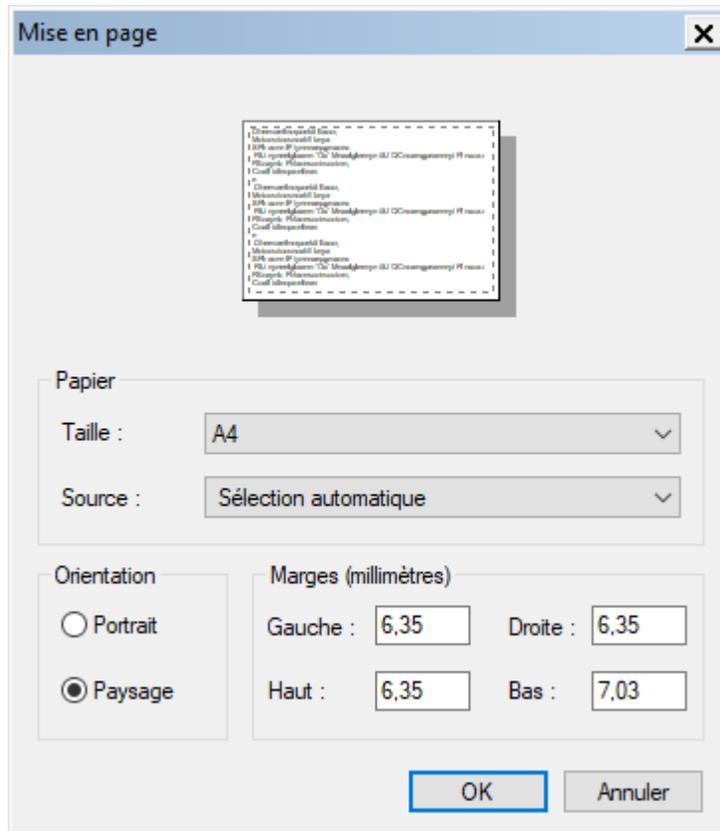
Voici une description des options offertes par l'aperçu avant impression.

### Bouton Imprimer

Cliquer sur ce bouton provoque l'impression du document. Une règle indique la progression de l'impression.

### Bouton Configurer

Cliquer sur ce bouton affiche la boîte de dialogue 'Mise en page', dont un exemple est montré ci-dessous :



### Bouton Options

Cliquer sur ce bouton affiche la boîte de dialogue ‘Options de la page’ montrée ci-après.

Cette boîte de dialogue permet de définir les espaces, en pixels, entre l’entête du rapport et le tableur, entre le pied de page du rapport et le tableur ainsi qu’entre le tableur et la marge.

Elle permet également de préciser si les entêtes des colonnes et des lignes sont imprimées et répétées sur chaque page.

De même, il est possible de définir les types et couleurs des traits définissant les lignes et colonnes ainsi que la couleur de la bordure.

**Options de la page**

Position		Impression	
Espace entêtes - tableur	15 Px.	Entêtes des colonnes	Imprimer <input checked="" type="checkbox"/> Répéter <input checked="" type="checkbox"/>
Espace pied de page - tableur	30 Px.	Entêtes des lignes	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Espace tableur - marge	10 Px.	Entêtes des colonnes fixes	<input checked="" type="checkbox"/>
		Entêtes des lignes fixes	<input checked="" type="checkbox"/>

**Lignes et couleurs**

<input checked="" type="checkbox"/> Séparateurs des lignes	<input type="text"/>	Couleur	Plein	1 Px.
<input checked="" type="checkbox"/> Séparateurs des colonnes	<input type="text"/>	Couleur	Plein	1 Px.
<input checked="" type="checkbox"/> Couleur de la bordure	<input type="text"/>	Couleur	Plein	2 Px.

Entête et pied de page      OK      Annuler

Cliquer sur le bouton ‘Entête et pied de page’ permet de définir un entête et un pied de page personnalisés.

**Entête et pied de page**

Police      OK      Annuler

**Entête**

Section gauche:	Section centrale:	Section droite:
UNIWIN		

**Pied de page:**

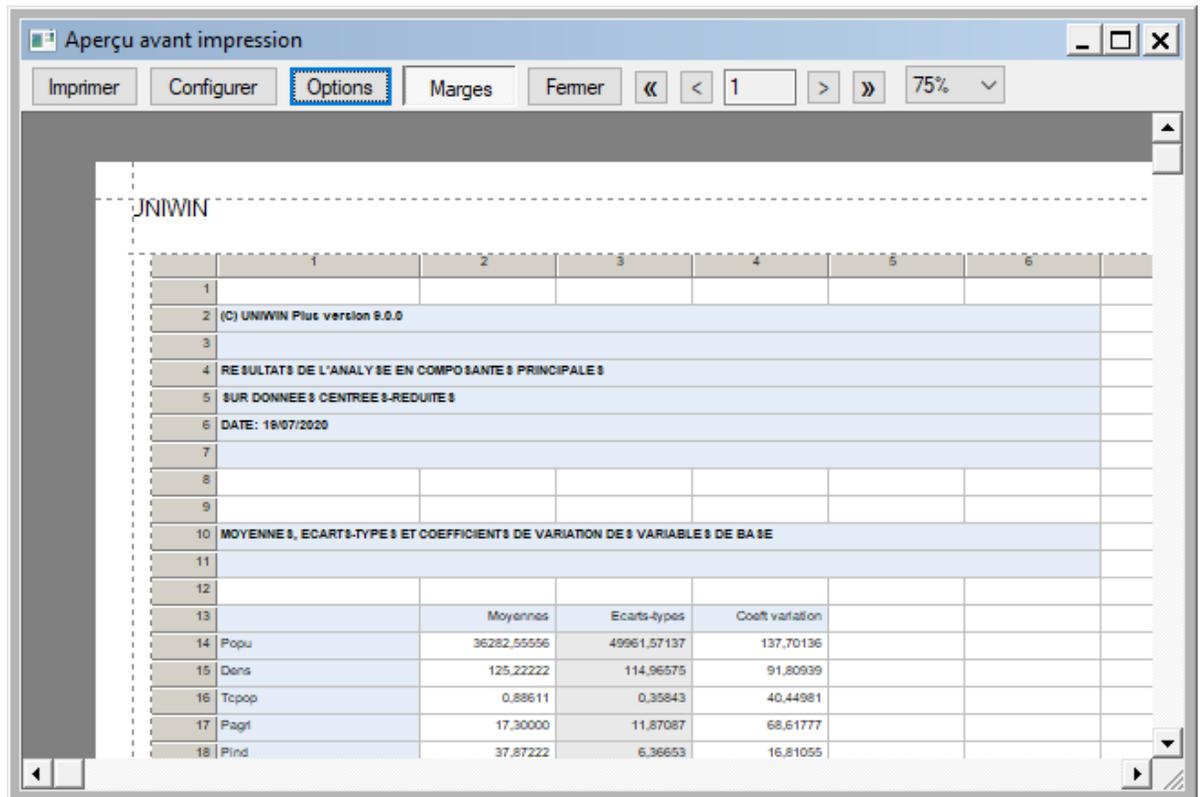
Section gauche:	Section centrale:	Section droite:
Créé le &D à &T		Page &P de &N

Note : &D indique la date, &T l'heure, &P le numéro de page, &N le nombre total de pages.

### Bouton Marges

Cliquer sur le bouton 'Marges' affiche un ensemble de lignes en pointillés permettant de modifier directement les différentes marges du rapport par simple déplacement de ces lignes.

Après avoir défini les marges souhaitées, un nouveau clic sur le bouton 'Marges' permet de revenir au rapport. C'est un équivalent des fonctions proposées dans le cadre 'Position' du menu 'Options'.



### Bouton Fermer

Cliquer sur ce bouton permet de quitter l'aperçu avant impression.

### Boutons <<, <, >, >>

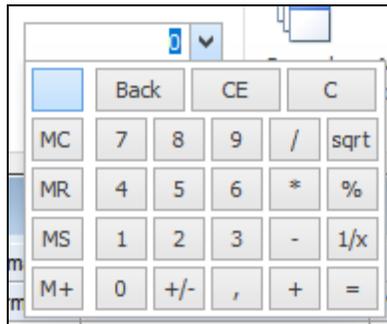
Ces boutons de navigation permettent de revenir à la première page du rapport, de revenir en arrière d'une page, d'avancer d'une page ou d'aller à la fin du rapport.

### Bouton %

Ce bouton permet de choisir le format de l'affichage de l'aperçu : 30%, 50%, 75%, 100%, 150%. Ils permettent d'obtenir une vue globale ou détaillée de chaque page du rapport.

## 5.3 L'option 'Calculatrice'

Cette option affiche une calculatrice :



## 5.4 L'option 'Lecteur Media'



Windows  
Media

Lecteur Media

Cette option charge le lecteur Windows Media permettant d'écouter des fichiers .wav. Cela est pratique dans le cadre d'un des exemples illustrant la procédure ICA.

## 5.5 Les options ‘Fenêtre’

Elles permettent de gérer l’arrangement des fenêtres (cascade, mosaïque horizontale, mosaïque verticale) et d’afficher la fenêtre ‘Tableur’, la fenêtre ‘Rapport’, la fenêtre ‘Graphique’ si elles sont minimisées.

## 5.6 L’option ‘Manuel’

Elle permet d’accéder au manuel de l’Utilisateur au format PDF ainsi qu’à chacune des documentations des procédures au format PDF.

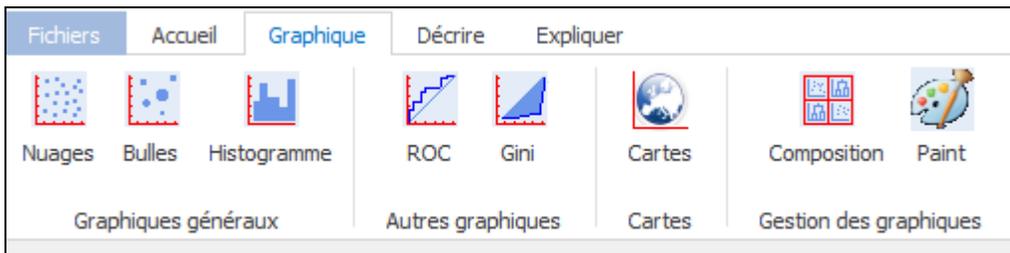
## 5.6 L’option ‘Aide’

Elle permet d’accéder à l’aide en ligne du logiciel, au site Internet pour le logiciel et aux informations concernant votre licence.



## Le ruban Graphique

Le ruban ‘Graphique’ permet de créer des nuages de points 2D, des courbes 2D, des graphiques en bulles 2D et des histogrammes. Il permet également de créer des courbes ROC et de Gini ainsi que des cartes géographiques thématiques. L’outil de composition graphique permet de créer une présentation graphique. Enfin, ce ruban donne un accès direct à Microsoft Paint.

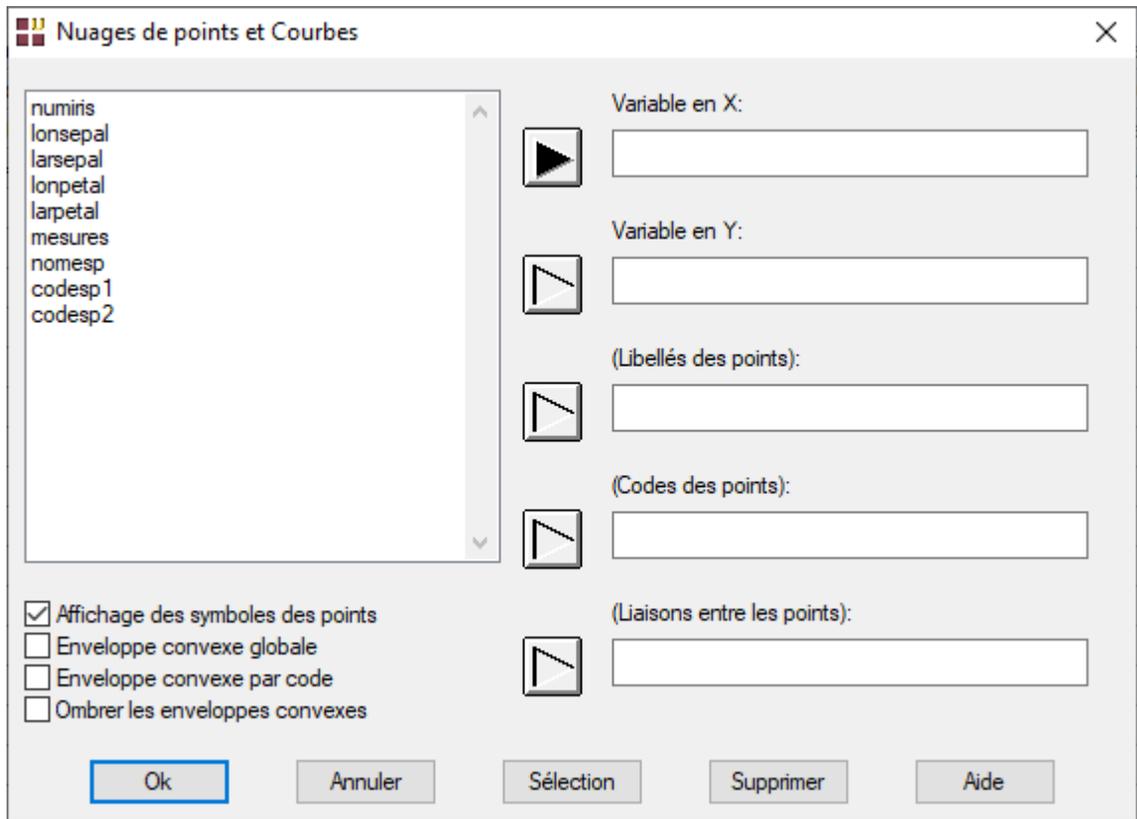


### 6.1 Nuages de points et courbes 2D

Ces procédures vous permettent de créer des nuages de points et des courbes à deux dimensions utilisant des variables provenant de vos fichiers de données et/ou simultanément des variables provenant des fichiers de résultats créés par UNIWIN lors des procédures statistiques.

Ces procédures sont utiles pour explorer vos données et découvrir les relations bivariées entre les variables avant de mettre en œuvre des analyses multivariées comportant un grand nombre de variables. Elles sont également utiles pour poursuivre ou reprendre l’analyse statistique sur les résultats obtenus suite à la mise en œuvre d’une des procédures d’UNIWIN.

Lorsque vous choisissez cette option, la boîte de dialogue suivante s'affiche (on suppose ici le fichier IRIS ouvert) :



Les champs 'Variable en X' et 'Variable en Y' permettent de préciser les variables numériques à représenter sur les axes X et Y respectivement.

Le champ optionnel 'Libellés des points' permet d'indiquer le nom de la variable contenant les libellés des points.

Le champ optionnel 'Codes des points' permet de préciser le nom de la variable numérique ou non numérique contenant les codes des différents nuages de points à représenter simultanément.

Le champ optionnel 'Liaisons entre les points' permet de préciser le nom de la variable numérique contenant les codes d'appartenance des points à des sous-ensembles de points. Pour chaque sous-ensemble, les points seront reliés pour former une courbe dans l'ordre des lignes dans le tableur (voir un exemple plus loin).

La case à cocher 'Affichage des symboles des points' permet d'afficher ou non les symboles des points.

Les cases à cocher 'Enveloppe convexe globale' et 'Enveloppe convexe par code' permettent d'afficher la ou les enveloppes convexes du ou des nuages de points.

La case à cocher 'Ombre les enveloppes convexes' permet d'ajouter une trame de fond dans l'enveloppe globale ou dans chacune des enveloppes associées aux différents codes.

Le fichier IRIS contient des données relatives à 150 iris issus de trois espèces : Setosa (codé 1), Versicolor (codé 2) et Virginica (codé 3).

Les mesures effectuées sur chaque iris sont : longueur du sépale (lonsepal), longueur du pétale (lonpetal), largeur du sépale (larsepal), largeur du pétale (larpetal).

**Iris Setosa (1)**



**Iris Versicolor (2)**

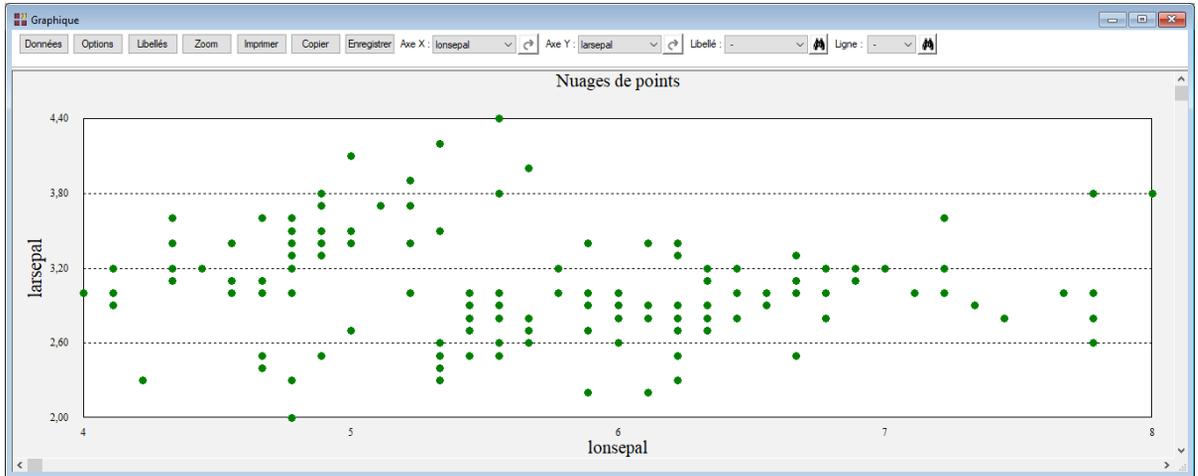


**Iris Virginica (3)**



Choisir les variables pour l'axe des X (axe horizontal) et l'axe des Y (axe vertical), par exemple lonsepal en X et larsepal en Y.

Cliquer sur le bouton Ok et le graphique montré ci-dessous s'affiche.



La barre d'outils au-dessus du graphique affiche les possibilités suivantes : Options, Libellés, Zoom, Imprimer, Copier, Enregistrer, Axe X, Axe Y, Libellé et Ligne.

Note : Un clic droit dans le graphique affiche également un menu popup.

- Données

Cliquer sur ce bouton permet de réafficher la boîte de dialogue d'entrée des données.

- Options

Toutes les options graphiques sont les options par défaut ou les options définies par l'utilisateur (voir le paragraphe 'Options graphiques' pour une description détaillée de ces options) :

- Titres et libellés des axes (police, taille, couleur, ...)
- Points, lignes, remplissages (style, taille, couleur, remplissage, ...)
- Echelles des axes (minimum, maximum, pas, nombre de décimales, ...)
- Couleur de fond
- Couleur de la bordure
- Grille et cadre (type, tracé, couleur)
- Toutes options par défaut

## Libellés

Permet d'indiquer si on désire ou non afficher les libellés des points. Permet également de choisir la police et la couleur des libellés.

- Zoom

Permet d'effectuer un zoom manuel ou automatique en X et Y (150%, 200% et 250%) ou uniquement en X ou uniquement en Y (300%, 350%, 400%) sur le graphique et de le désactiver. Si le graphique est plus grand que la taille de l'écran, un ascenseur horizontal et un ascenseur vertical s'activent. Ils permettent de faire défiler le graphique.

- Imprimer

Permet de définir les paramètres de l'impression et d'imprimer le graphique complet (notamment s'il a été zoomé et qu'il n'apparaît plus entièrement à l'écran) ou uniquement la partie affichée à l'écran.

- Copier

Permet d'envoyer le graphique dans le presse-papiers. Ceci permet de le récupérer dans d'autres applications sous Windows.

- Enregistrer

Permet d'enregistrer le graphique dans le format vectoriel EMF (Enhanced Metafile Format) ou dans d'autres formats tels que BMP, GIF, JPEG, TIFF, PNG via Paint de Microsoft, si ces types de fichiers sont associés à Microsoft Paint.

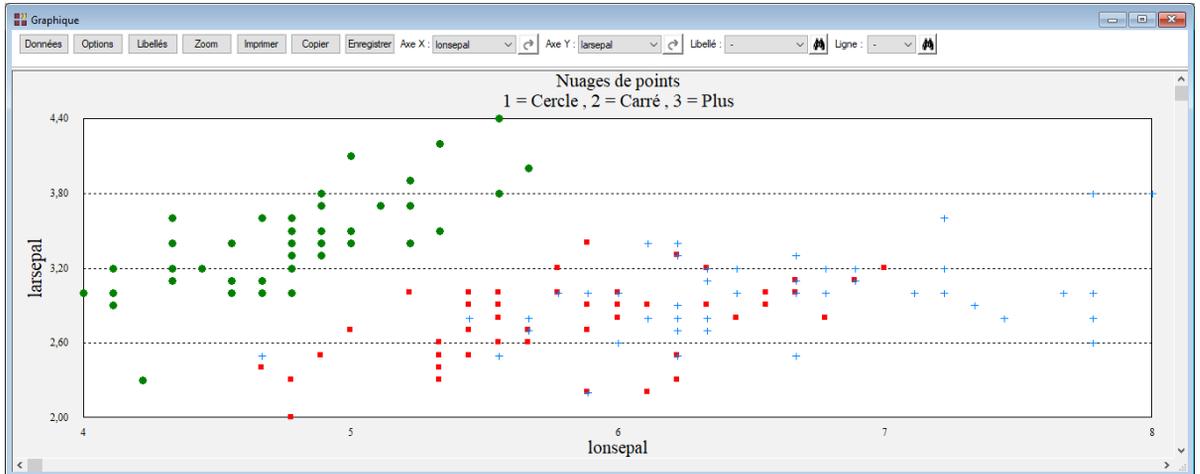
- Axe X et Axe Y

Permet de choisir dans les listes déroulantes les variables à représenter sur les axes X et Y.

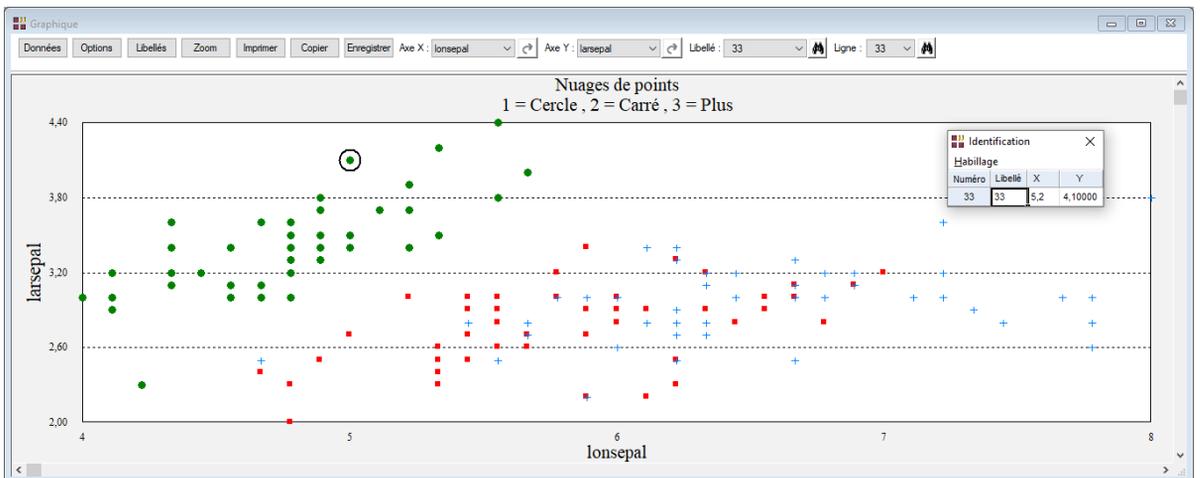
- Libellé et Ligne

Permet de localiser un point dans le graphique en choisissant son libellé ou son numéro de ligne dans le tableur puis en cliquant sur l'icône .

Rappelons la boîte de dialogue en utilisant le bouton Données pour définir des codes pour les points. Pour les mêmes variables en X et en Y, définissons la variable codesp1 comme variable des Codes pour représenter par trois symboles différents nos trois espèces d'iris.

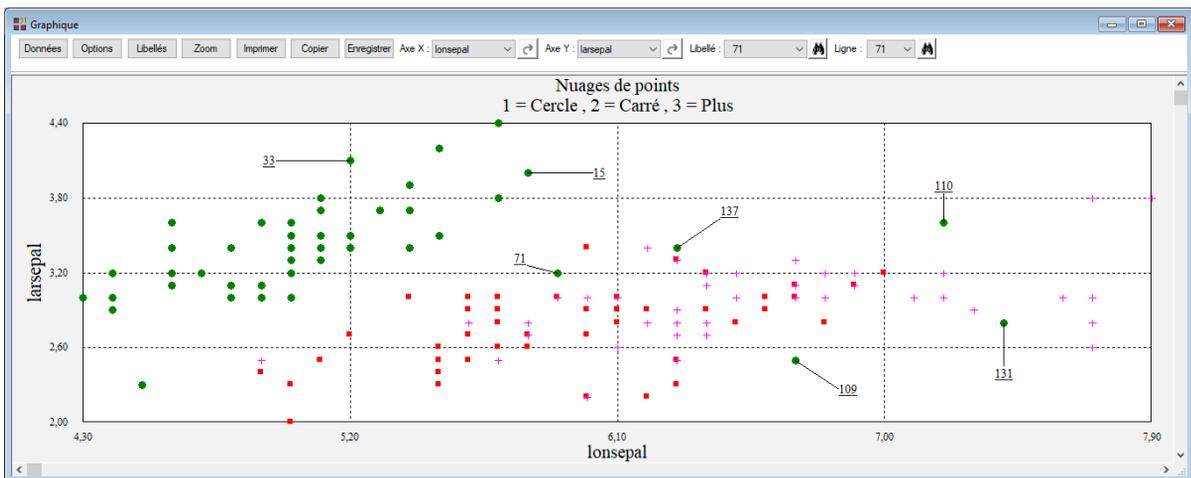
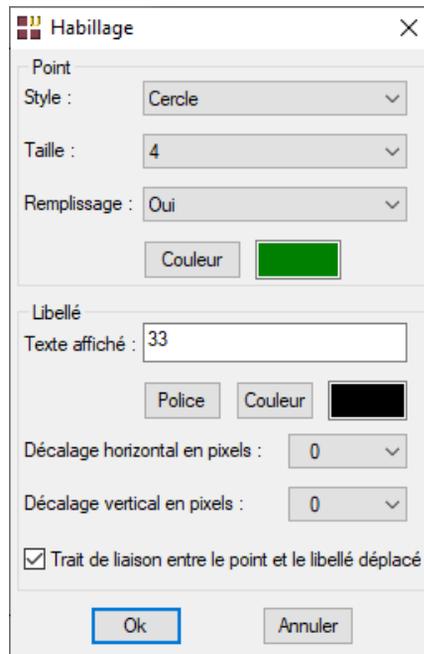


UNIWIN a utilisé un symbole différent pour chaque espèce d'iris, permettant ainsi de visualiser les trois groupes de données. Ce graphique comporte 150 points. Si nous désirons connaître les coordonnées en X, Y d'un point précis ainsi que son numéro d'ordre dans le tableau et son éventuel libellé, il suffit de mettre le curseur sur ce point (ou près de ce point, s'il n'y a pas d'ambiguïté) et de cliquer sur le bouton gauche de la souris. UNIWIN affiche alors une fenêtre donnant les informations concernant ce point.

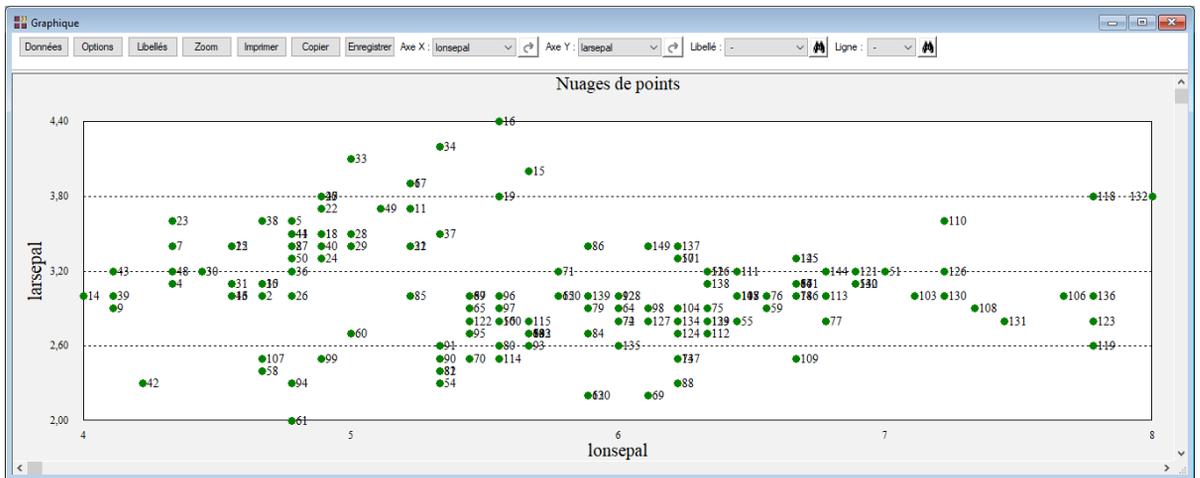
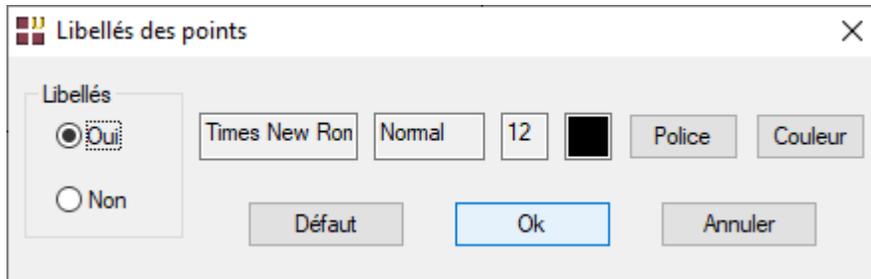


Les champs 'Libellé' et 'Ligne' de la barre d'outils sont également automatiquement renseignés.

Le menu 'Habillage' de la fenêtre d'identification permet d'habiller le point sélectionné. Il permet de sélectionner un symbole spécifique pour le point, sa taille, son remplissage éventuel, sa couleur, d'y associer un libellé pour lequel on peut choisir la police, la couleur, les décalages en pixels par rapport au point et préciser si un trait est tracé entre le point et le libellé décalé.



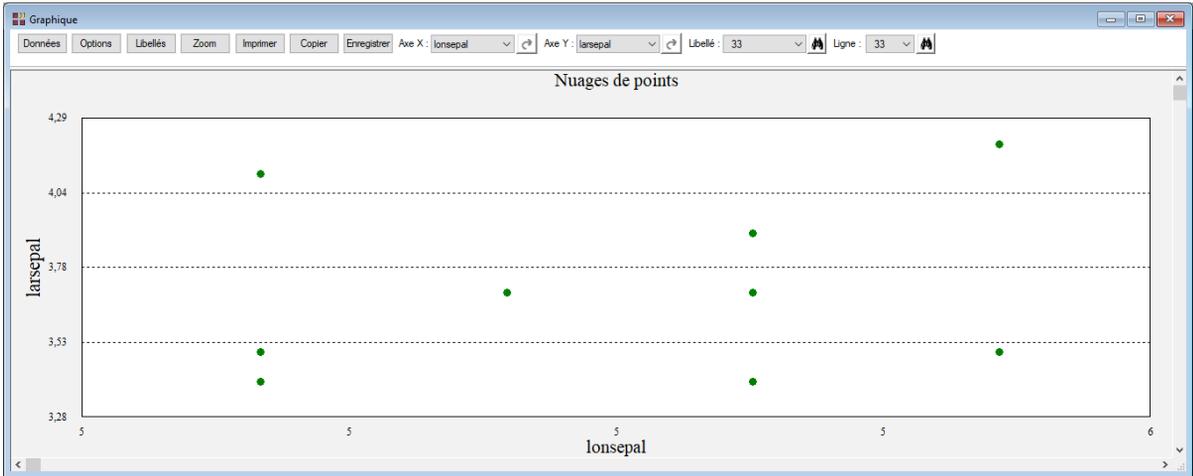
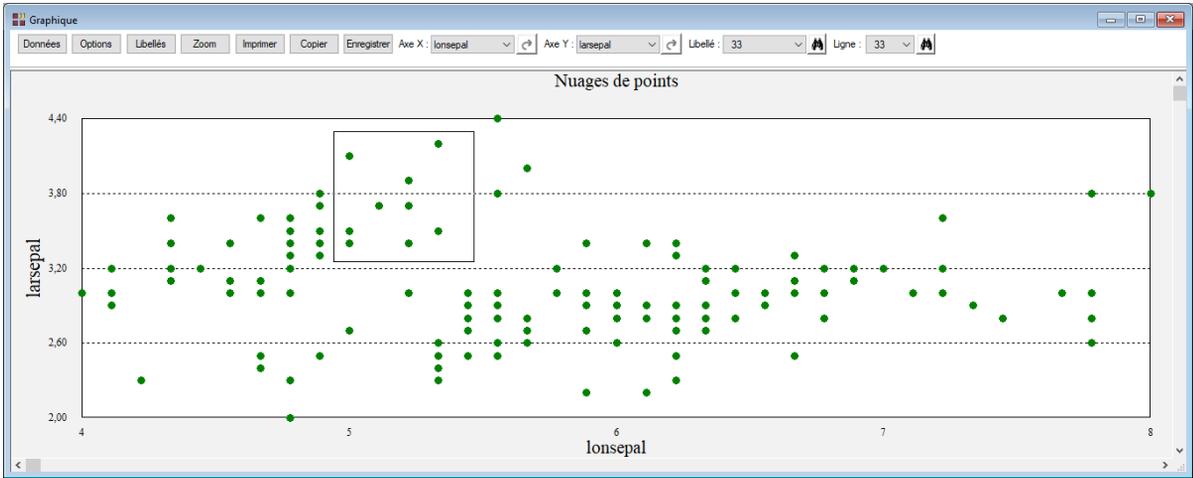
Pour faire apparaître des libellés pour tous les points, après avoir défini une variable de libellés dans la boîte de dialogue principale (par exemple numiris), cliquer sur le menu Libellés et cliquer sur le bouton d'option Oui dans la boîte montrée ci-après et choisir éventuellement la police et sa couleur.



Par défaut, UNIWIN n'affiche pas les libellés des points.

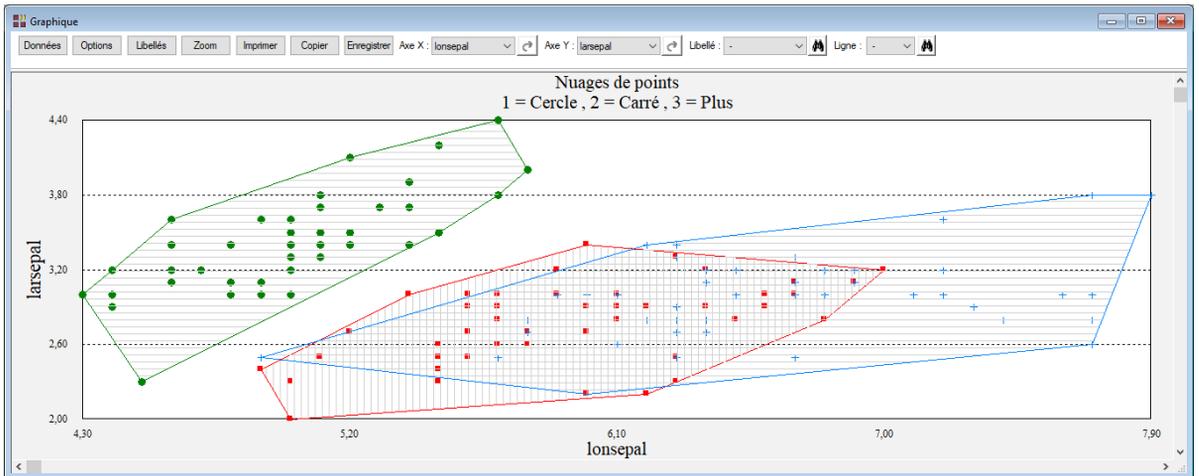
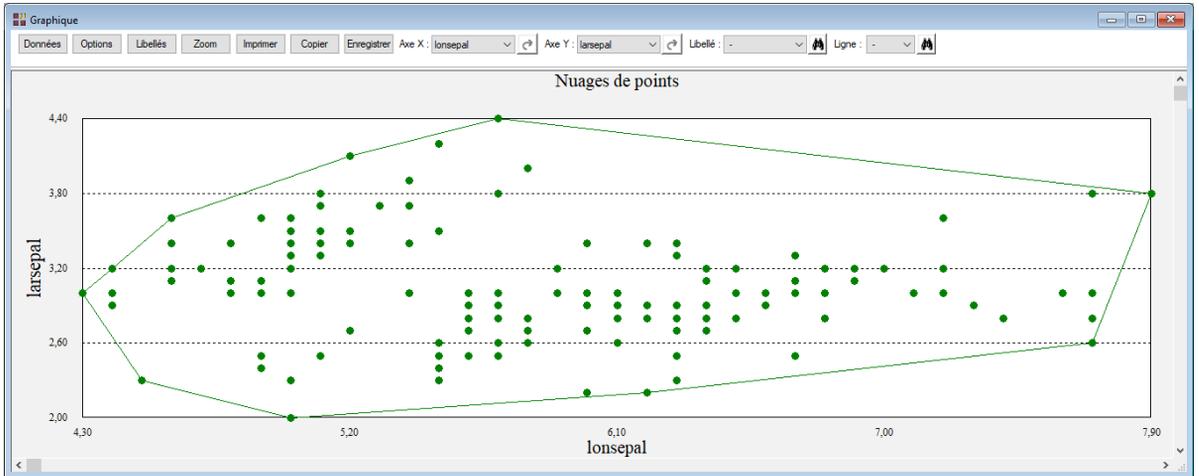
De même, pour mieux analyser un groupe de points important, il est possible de zoomer de façon directe dans le graphique.

Choisir dans le menu Zoom l'option 'Manuel' puis positionner le curseur à l'endroit définissant le début de la zone à zoomer, cliquer sur le bouton gauche de la souris et glisser. Un rectangle se dessine montrant la zone qui sera zoomée.



Illustrons maintenant l'ajout d'une enveloppe convexe pour le nuage global et d'une enveloppe convexe pour chaque espèce d'iris.

Note : les codes des lignes des enveloppes sont 1 pour l'enveloppe globale et 2, 3 et 4 pour les 3 espèces d'iris.



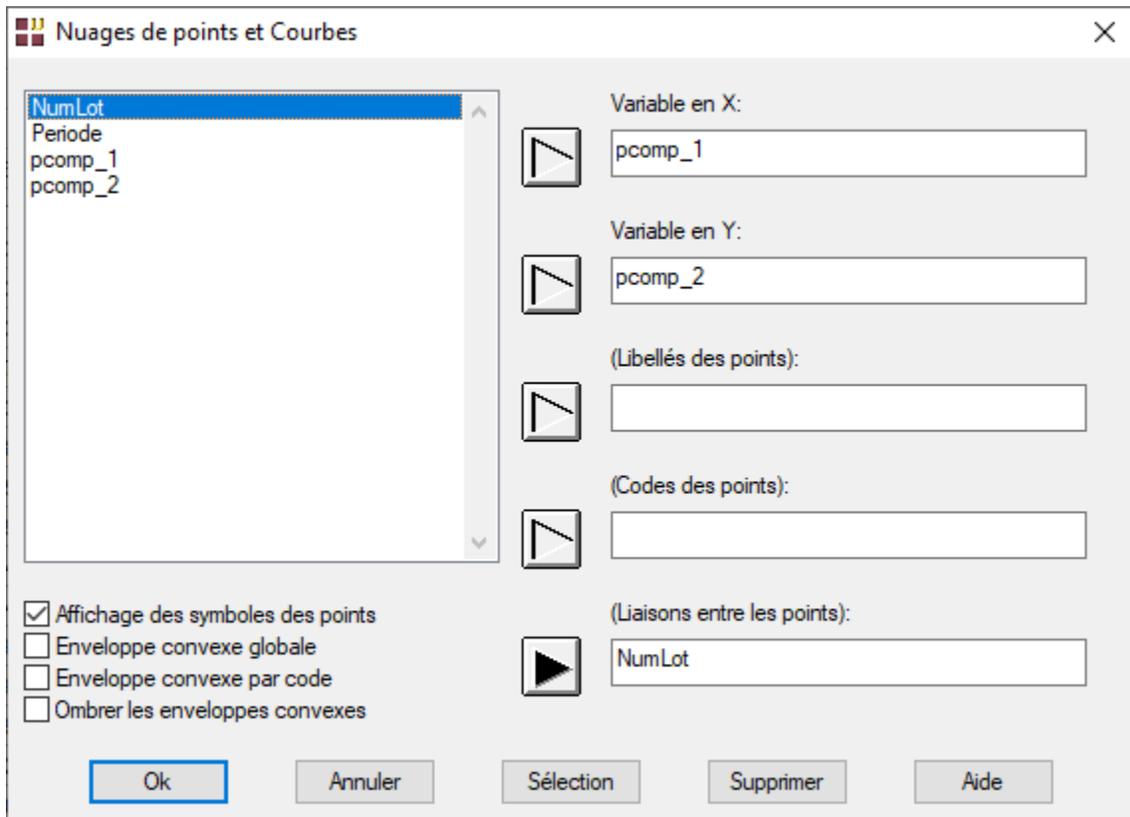
Illustrons enfin le tracé de courbes avec cette procédure. Pour cela ouvrons le fichier MSPC. Il contient les résultats d'une analyse en composantes principales (ACP) effectuée sur des données de production de levure industrielle par lots (fichier LEVURE).

Les variables pcomp\_1 et pcomp\_2 sont respectivement les première et deuxième composantes principales et définissent nos axes des X et des Y.

Le champ 'Liaisons entre les points' permet de préciser le nom de la variable numérique contenant les codes d'appartenance des points à des sous-ensembles de points. Dans notre exemple, nous utiliserons la variable NumLot qui représente les différents lots de la production de levure.

Pour chaque lot, les points seront reliés pour former une courbe dans l'ordre des lignes dans le tableau (évolution du lot pendant sa fabrication).

Renseignons la boîte de dialogue comme montré ci-dessous et demandons le tracé des courbes uniquement sans l'affichage des symboles des points. Il suffit pour cela dans la boîte de dialogue de décocher 'Affichage des symboles des points'.



Visualisons uniquement les 5 premiers lots en fonction des périodes.

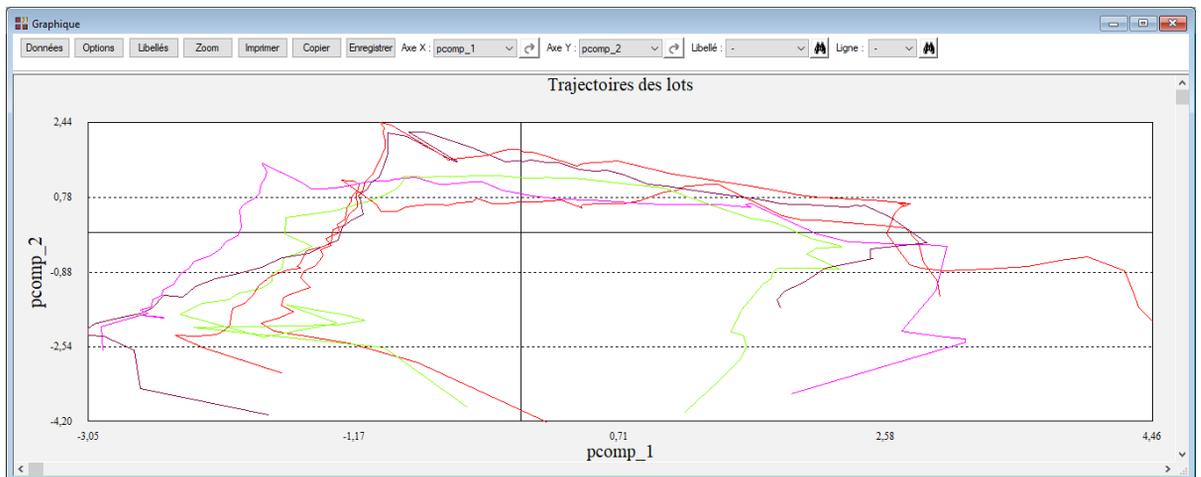
**Définition de la sélection** [X]

Et	NumLot	<=	5
----	--------	----	---

Liaison	Variable	Relation	Valeur ou variable
Et	NumLot	=	NumLot
Et non	Periode	<>	Periode
Ou	pcomp_1	<	pcomp_1
Ou non	pcomp_2	<=	pcomp_2
		>	
		>=	
		début	

Ok      Annuler      **Ajouter**      Aide



## 6.2 Bulles 2D

Cette procédure permet de créer un nuage de points à deux dimensions dans lequel chaque point est représenté par un symbole dont la taille est proportionnelle à la valeur associée dans une troisième variable.

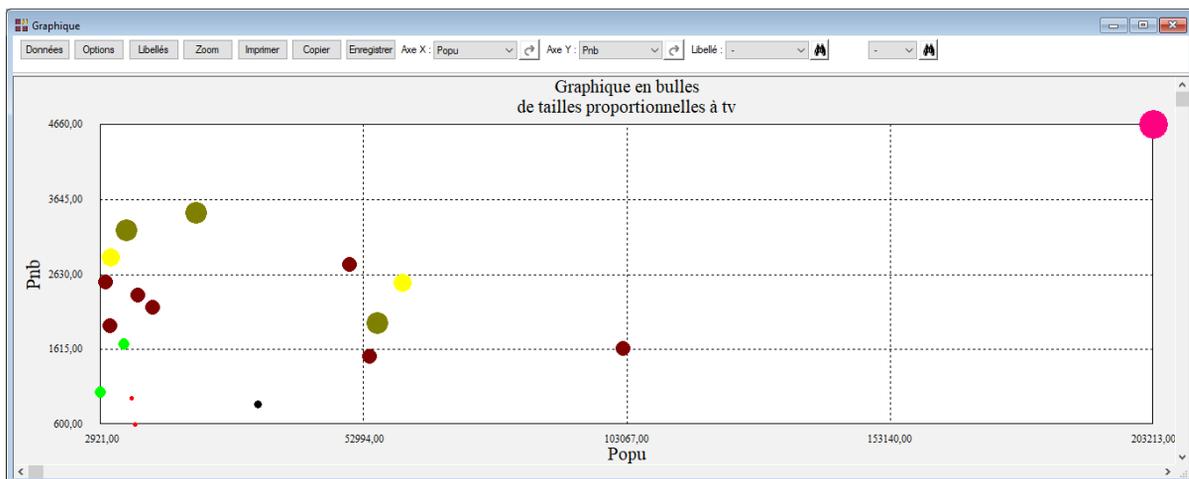
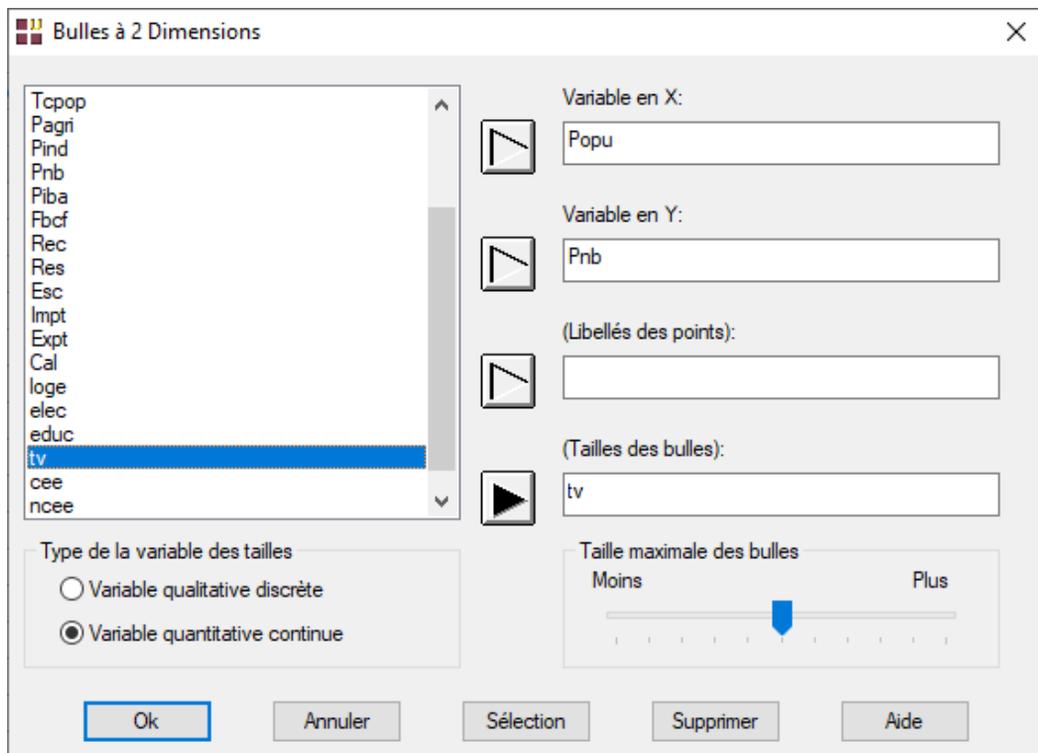
Cette procédure est utile pour poursuivre ou reprendre l'analyse statistique sur les résultats obtenus suite à la mise en œuvre d'une des procédures d'UNIWIN, par exemple une analyse factorielle.

Dans ce cas, vous pouvez tracer un plan factoriel dans lequel chaque point a une taille proportionnelle à ses cosinus carrés ou à ses contributions. Ceci permet une lecture rapide du plan factoriel grâce à la mise en évidence des points les plus significatifs.

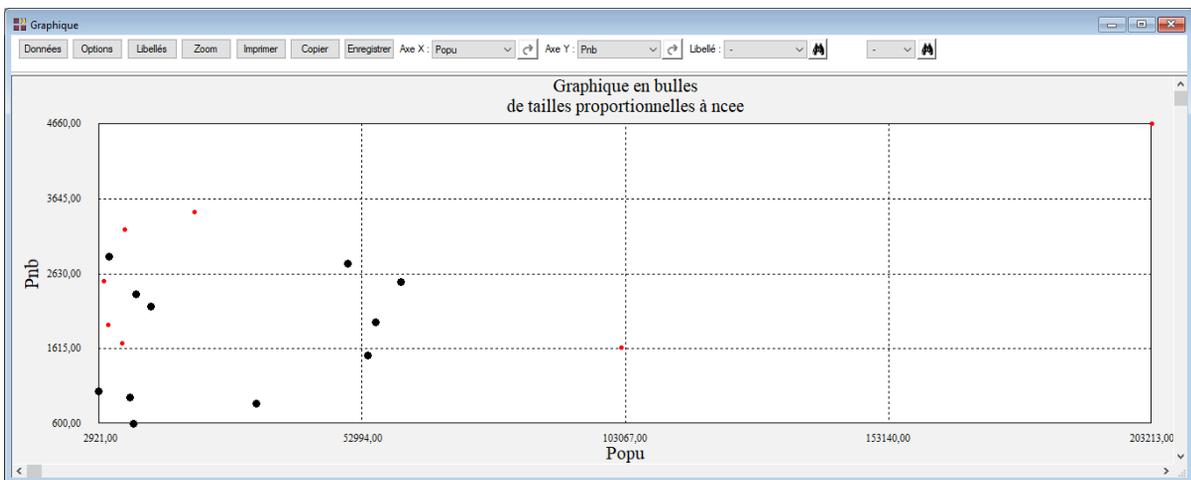
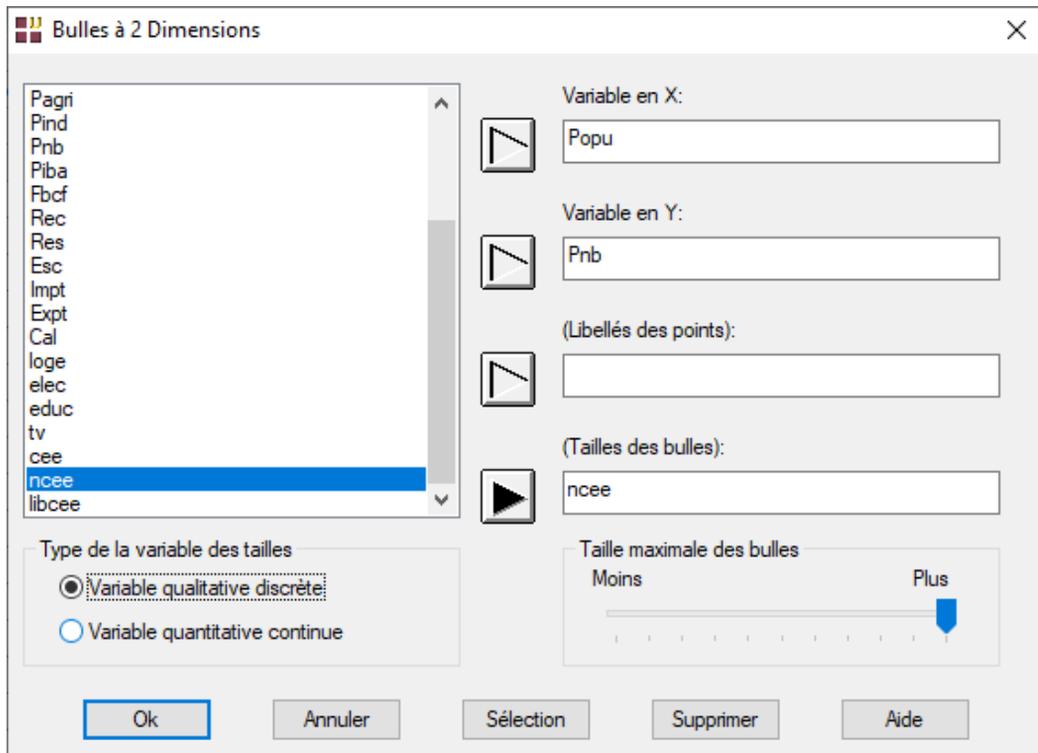
Choisissons l'option 'Bulles'. Elle permet de préciser :

- la variable pour l'axe des X
- la variable pour l'axe des Y
- la variable optionnelle contenant les libellés des points
- la variable optionnelle définissant les tailles des bulles
- le type de la variable des tailles (qualitative ou quantitative)
- la taille maximale des bulles

Ouvrons le fichier OCDE et renseignons la boîte de dialogue comme montré ci-dessous :



Prenons un deuxième exemple, avec la variable ncee pour variable des tailles. Cette variable codée indique l'appartenance ou non à la CEE.

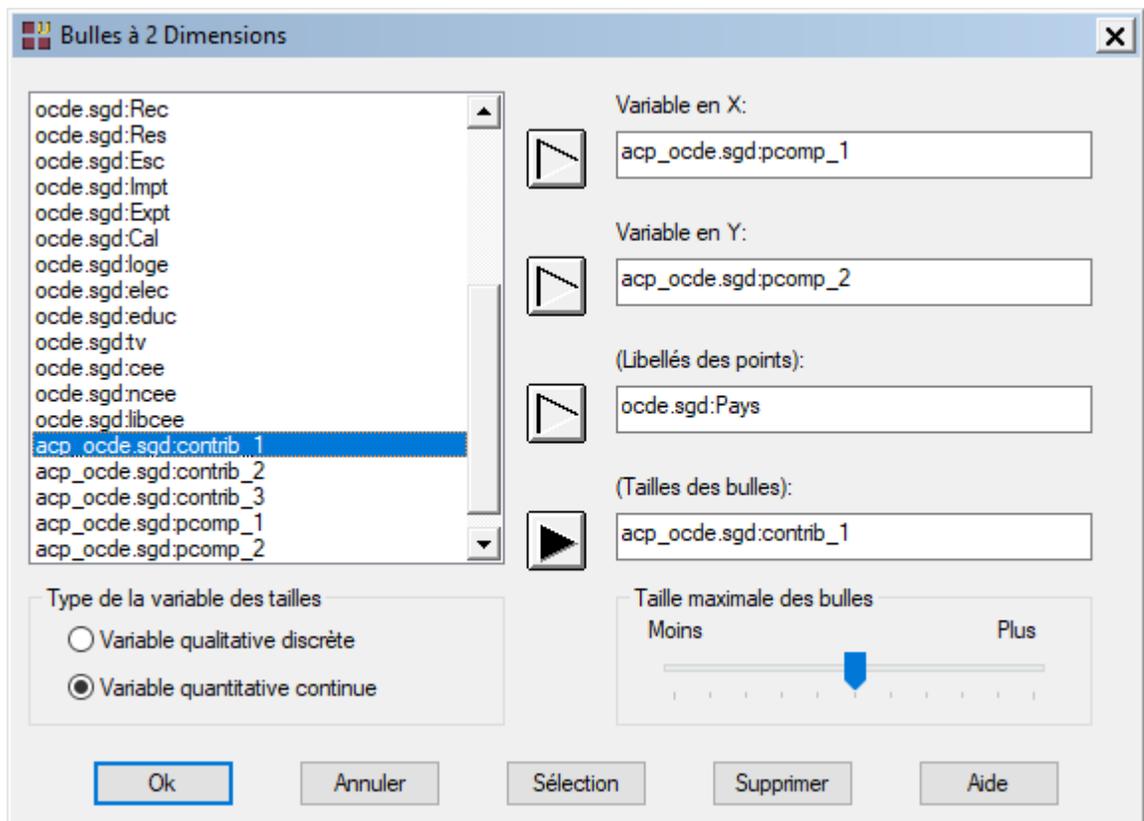


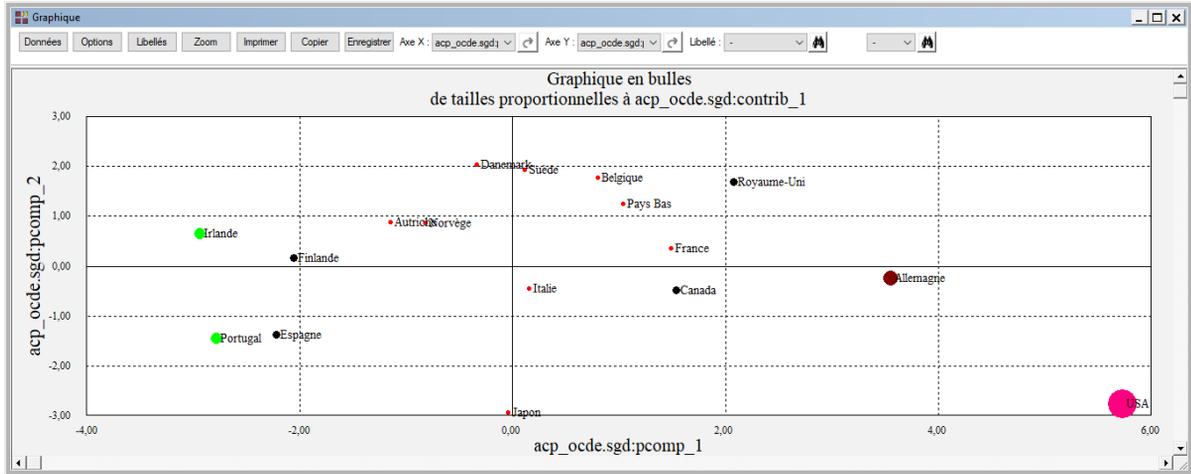
Prenons un troisième exemple illustrant cette procédure sur les résultats d'une analyse en composantes principales (ACP) sur les données de ce fichier OCDE (Popu à Cal).

On supposera les première et deuxième composantes principales enregistrées dans les variables pcomp\_1 et pcomp\_2 et les contributions au premier axe factoriel enregistrées dans la variable contrib\_1.

Nous renseignons la boîte de dialogue comme montré ci-après (le fichier ACP\_OCDE contient les résultats de l'ACP sur le fichier OCDE).

Le graphique en bulles sur les données du premier plan factoriel permet de rapidement connaître les points contribuant le plus à la définition de l'axe 1.





## 6.3 Histogramme

Cette procédure décrit comment tracer un histogramme d'une variable continue répartie en classes d'amplitudes inégales.

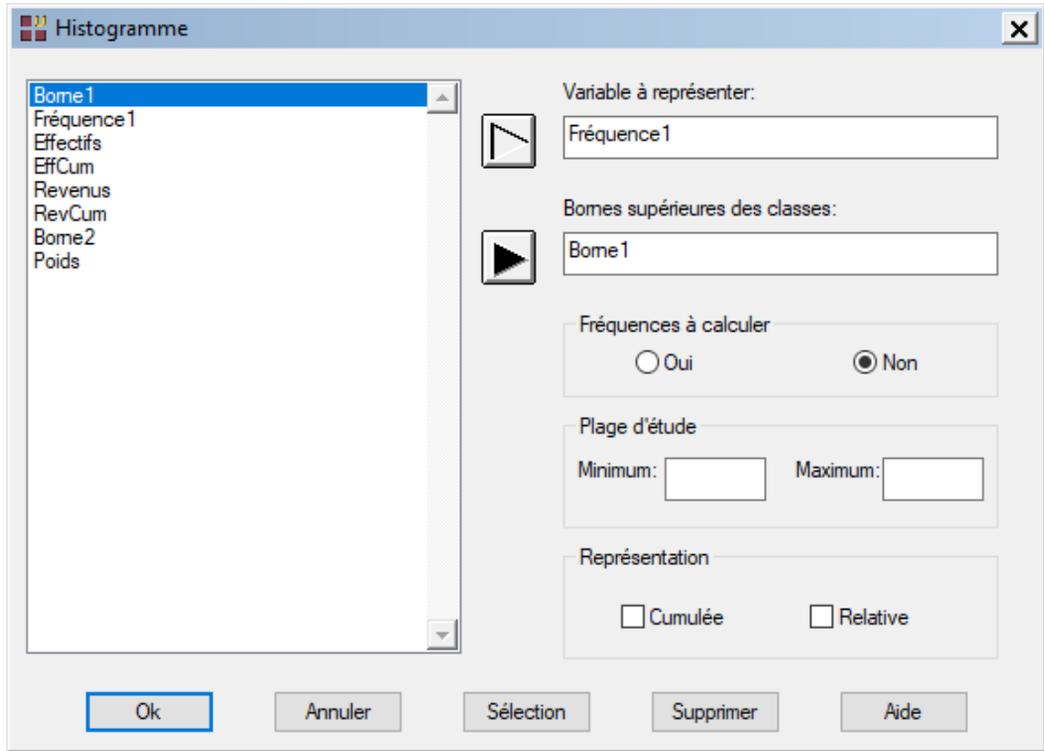
Dans ce cas, le rectangle construit pour chaque classe a une surface égale à la fréquence de la classe.

Le cas de classes d'amplitudes égales peut également être traité, bien entendu.

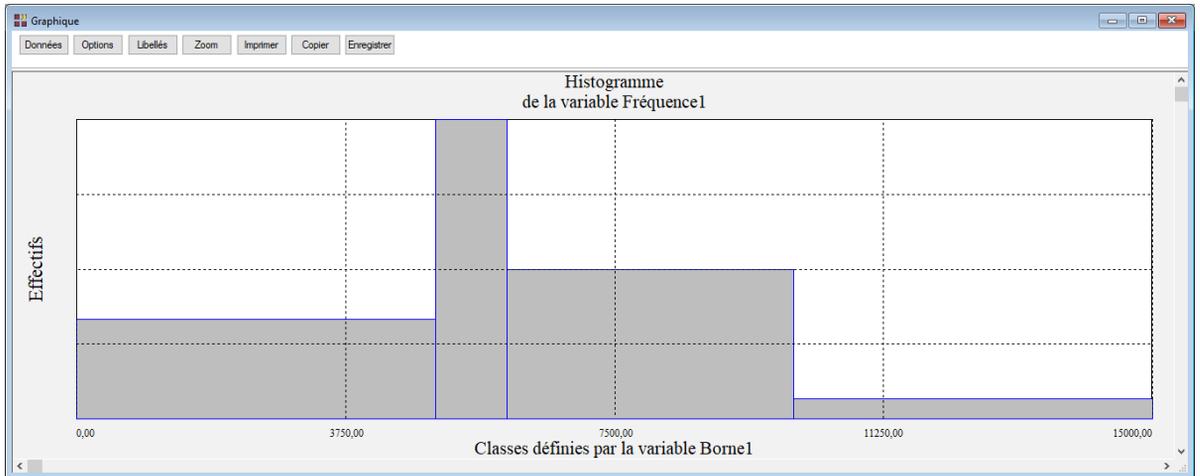
Choisissons 'Histogramme'. La boîte de dialogue qui s'affiche permet de préciser :

- la variable numérique à représenter (variable continue ou fréquences)
- les bornes supérieures des classes à former
- s'il faut ou non calculer les fréquences (en fonction du type de la variable à représenter)
- la plage d'étude de la variable (minimum et maximum pour le graphique)
- le type de représentation (cumulée, relative, cumulée et relative)

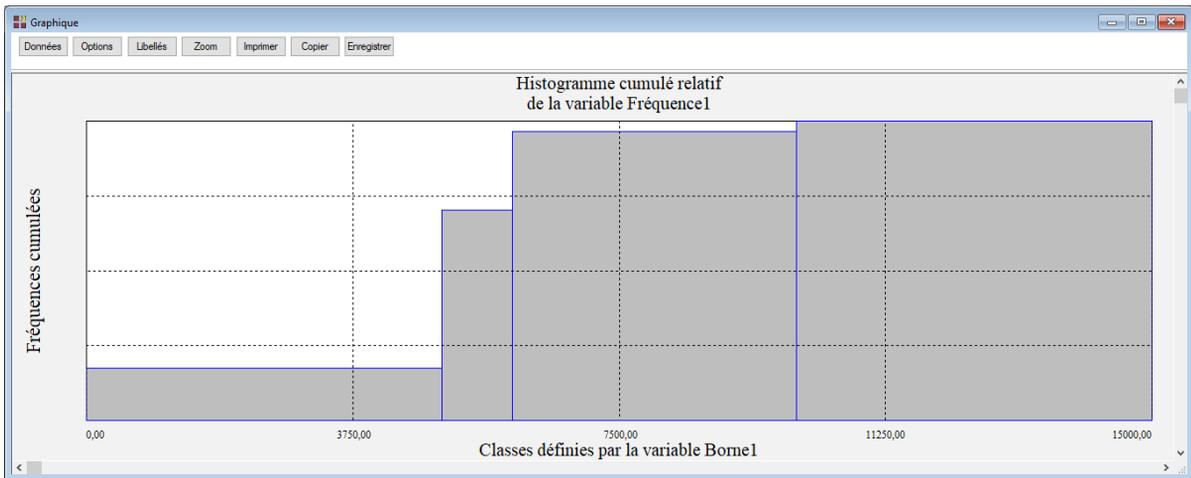
Ouvrons le fichier DESCR1 et renseignons la boîte de dialogue de la façon suivante :



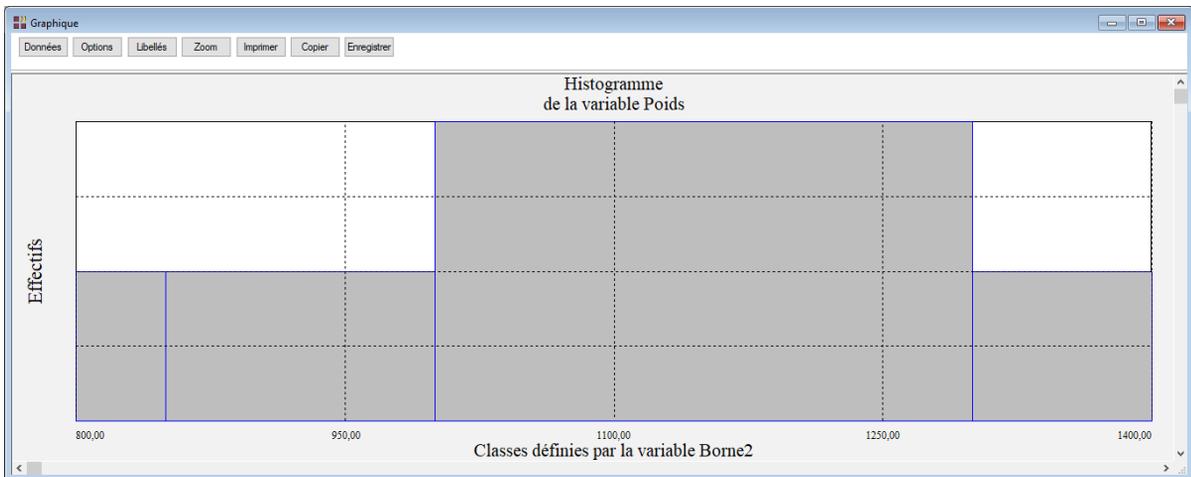
La variable à étudier Fréquence1 est une variable numérique contenant les fréquences pour chacune des classes dont les bornes sont définies par la variables Borne1.



Demandons maintenant une représentation cumulée et relative.



Choisissons ensuite, dans le fichier DESCR1, les variables 'Poids' (variable numérique à étudier) et 'Borne2' (bornes supérieures des classes) et sélectionnons 'Oui' pour l'option 'Fréquences à calculer', car la variable 'Poids' est une variable continue.



## 6.4 Courbe ROC

Ce paragraphe décrit comment tracer une courbe ROC (Receiver Operating Characteristic) et calculer l'aire sous la courbe (AUC) ainsi que d'autres statistiques. Cette courbe est très utilisée en médecine pour étudier un événement du type présence ou absence d'une maladie, mais également pour le scoring (voir la documentation concernant la méthode de scoring proposée par UNIWIN).

Choisissons 'Courbe ROC'. La boîte de dialogue qui s'affiche permet de préciser :

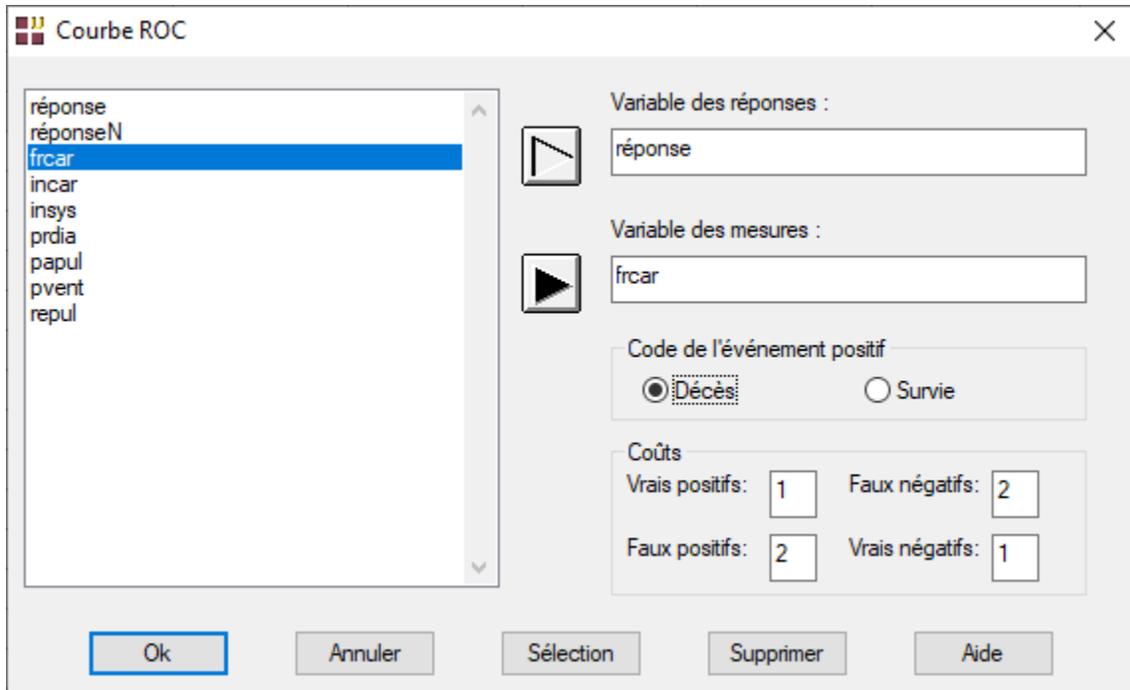
- la variable des réponses
- la variable des mesures
- le code de l'événement considéré positif pour l'analyse
- les coûts associés aux différents cas (VP, FN, FP, VN)

### Exemple 1 : Fichier ROC1

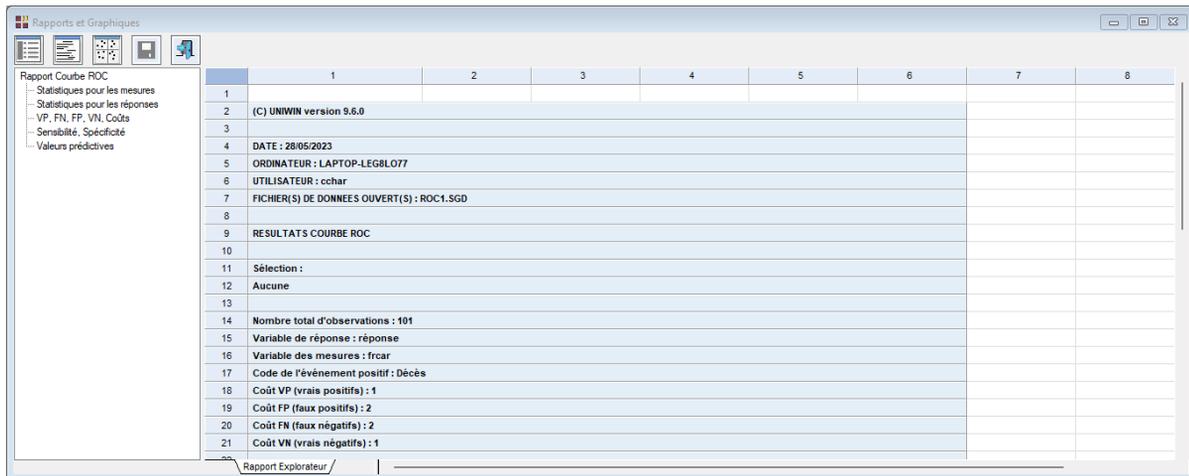
Le fichier ROC1 contient des informations concernant 101 victimes d'un infarctus du myocarde. Les variables collectées sont :

- réponse 'Décès' ou 'Survie'
- frcar fréquence cardiaque
- incar index cardiaque
- insys index systolique
- prdia pression diastolique
- papul pression artérielle pulmonaire
- pvent pression ventriculaire
- repul résistance pulmonaire

Renseignons la boîte de dialogue comme montré ci-dessous :



Après avoir cliqué sur Ok, la fenêtre suivante s'affiche :



La barre d'outils contient cinq icônes permettant de rappeler la boîte de dialogue d'entrée des données, de visualiser le rapport sous différents formats, d'afficher divers graphiques, d'enregistrer des résultats calculés et de quitter cette procédure.

L'explorateur à gauche dans cette fenêtre permet de visualiser les différents éléments calculés. Il rappelle dans son premier item les variables utilisées et indique l'aire sous la courbe (AUC) et l'indice de Gini ( $2 \times \text{AUC} - 1$ ).

Les deuxième et troisième items affichent des statistiques descriptives pour la variable des  $n$  mesures et pour la variable des réponses et notamment le calcul de la prévalence  $(\text{VP} + \text{FN})/n$ .

Le quatrième item affiche, pour chacune des valeurs uniques de la variable des mesures (seuil), les nombres de vrais positifs (VP), de faux négatifs (FN), de faux positifs (FP), de vrais négatifs (VN) et les coûts.

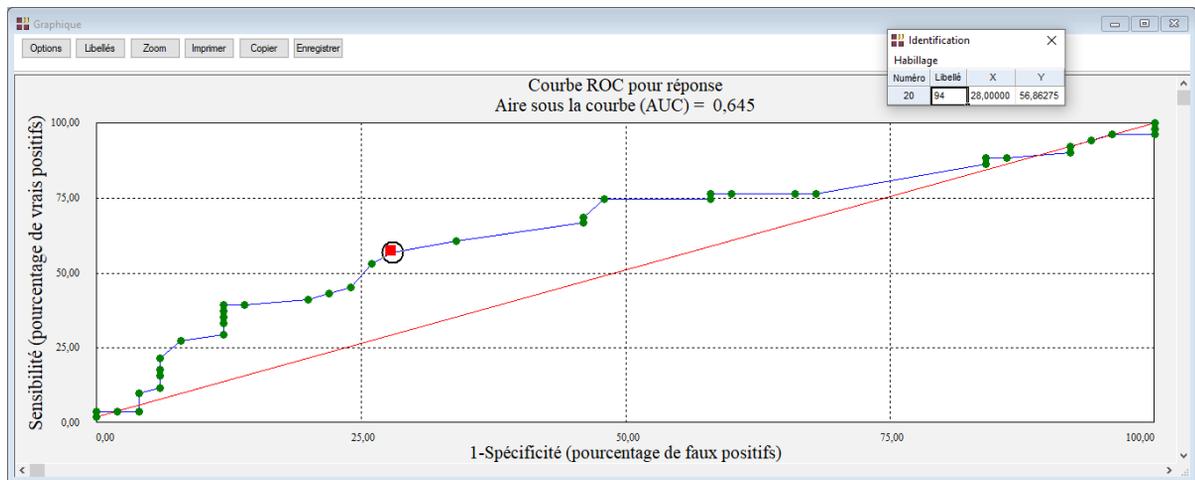
Le cinquième item affiche, pour chacune des valeurs uniques de la variable des mesures (seuil), la sensibilité, la spécificité et l'indice J de Youden (sensibilité + spécificité - 1).

Enfin le sixième item affiche :

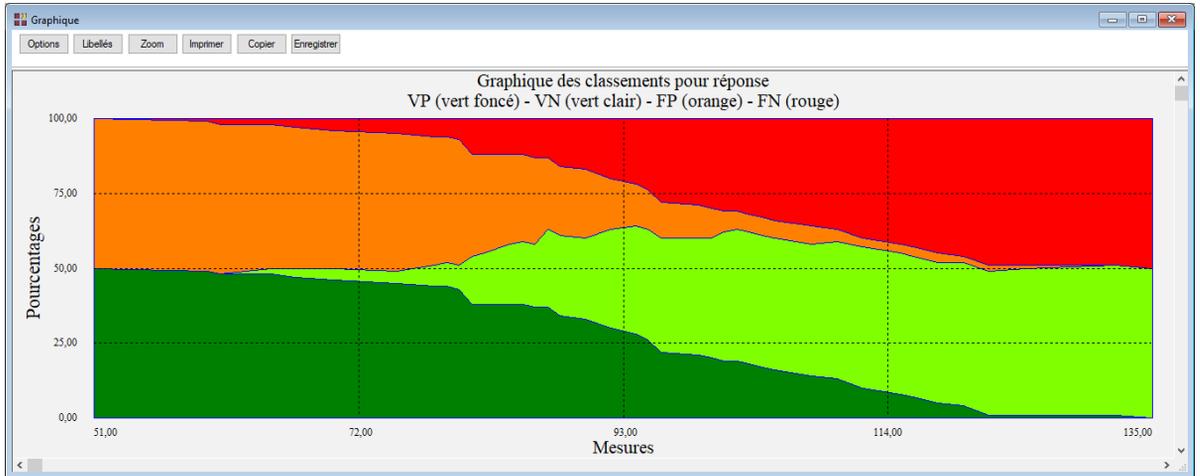
- les valeurs prédictives positives :  $\text{VPP} = \text{VP} / (\text{VP} + \text{FP})$
- les valeurs prédictives négatives :  $\text{VPN} = \text{VN} / (\text{VN} + \text{FN})$
- les taux de fausses alarmes :  $1 - \text{VPP}$
- les taux de fausses assurances :  $1 - \text{VPN}$
- les rapports de vraisemblance positifs :  $\text{RVP} = (\text{sensibilité} / (1 - \text{spécificité}))$
- les rapports de vraisemblance négatifs :  $\text{RVN} = ((1 - \text{sensibilité}) / \text{spécificité})$

	Prévu Décès	Prévu Survie	Total	% correct	
Observé Décès	$VP$	$FN$	$VP + FN$	$\frac{100 * VP}{(VP + FN)}$	Sensibilité (ou Rappel)
Observé Survie	$FP$	$VN$	$FP + VN$	$\frac{100 * VN}{(VN + FP)}$	Spécificité
Total	$VP + FP$	$FN + VN$			
% correct	$\frac{100 * VP}{(VP + FP)}$	$\frac{100 * VN}{(FN + VN)}$		$\frac{100 * (VP + VN)}{(VP + VN + FP + FN)}$	
	Précision			% total correctement prévu	

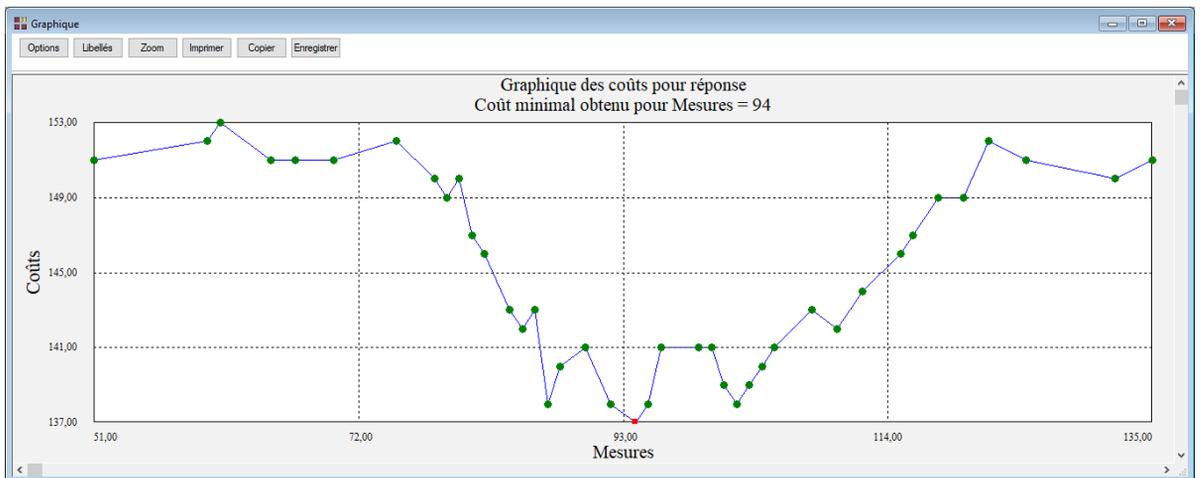
Le premier graphique proposé est le graphique de la courbe ROC qui rappelle la valeur de l'AUC. Le point coloré par défaut en rouge est celui donnant la valeur optimale à l'indice J de Youden. Un clic sur ce point permet d'obtenir les valeurs de la sensibilité et de 1- la spécificité :



Le deuxième graphique représente les évolutions des classements en fonction des valeurs de la variable des mesures :



Le troisième graphique affiche les évolutions des coûts en fonction des valeurs de la variable des mesures :



Il indique que le coût minimal (137) est obtenu pour la valeur des mesures égale à 94. Cette valeur de Mesures est celle qui donne la valeur maximale à l'indice J de Youden.

L'option d'enregistrement des résultats permet de sauvegarder dans le fichier des résultats les éléments affichés dans le rapport, dont VP, FN, FP, VN, VPP, VPN, indice J de Youden, RVP, RVN, spécificité, sensibilité et coûts.

## Exemple 2 : Fichier ROC2

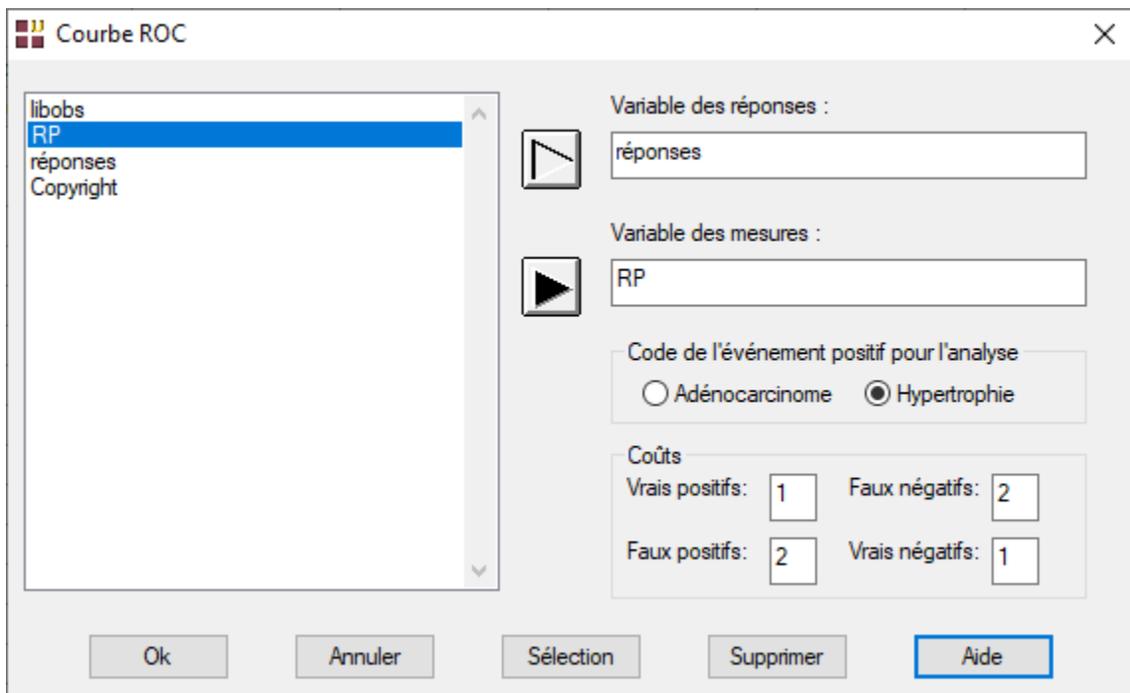
Ce fichier est issu d'une étude des caractéristiques analytiques et cliniques des dosages de PSA total et de PSA libre par électroluminescence. Au total 104 prélèvements ont été réalisés, 52 appartenant à des sujets présentant un adénocarcinome et 52 une hypertrophie de la prostate.

Il provient du Ann Biol Clin (ABC), volume 63, numéro 2, pages 145-154, mars-avril 2005.

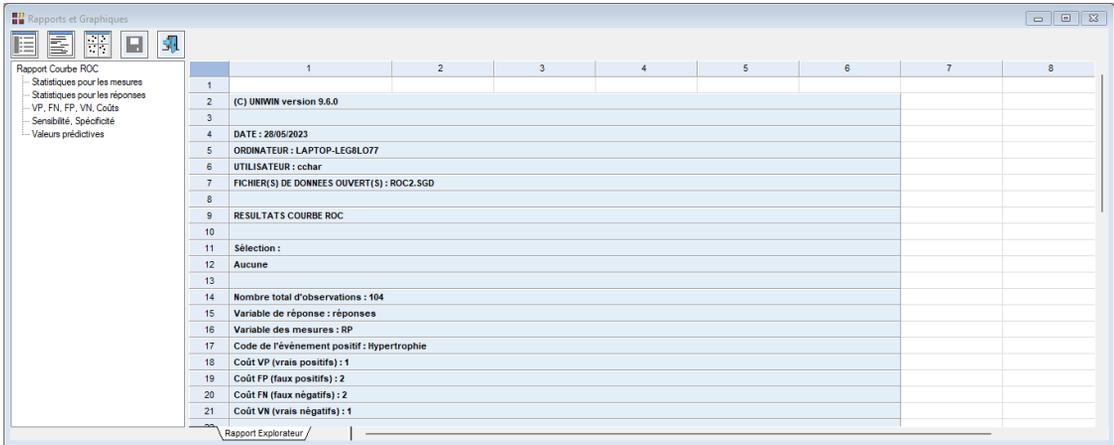
Il contient les variables suivantes :

- RP                      Rapport PSA libre sur PSA total (%)
- réponses              Hypertrophie, Adénocarcinome

Renseignons la boîte de dialogue comme montré ci-dessous :



Cliquons sur Ok. Le rapport s'affiche alors. Visualisons également les graphiques obtenus.



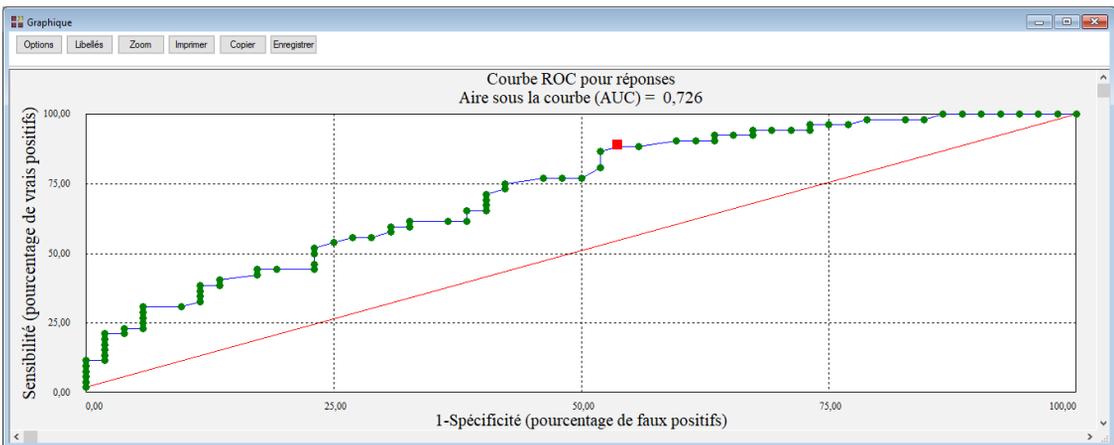
Rapports et Graphiques

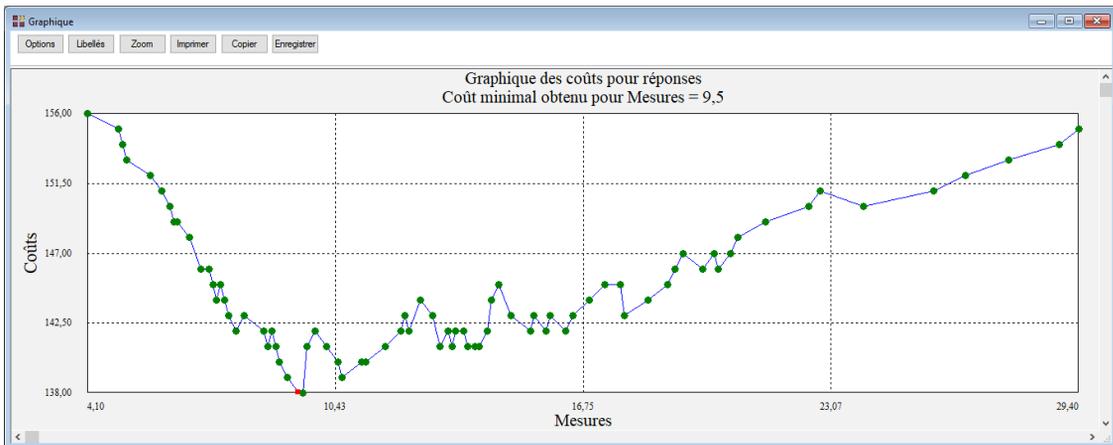
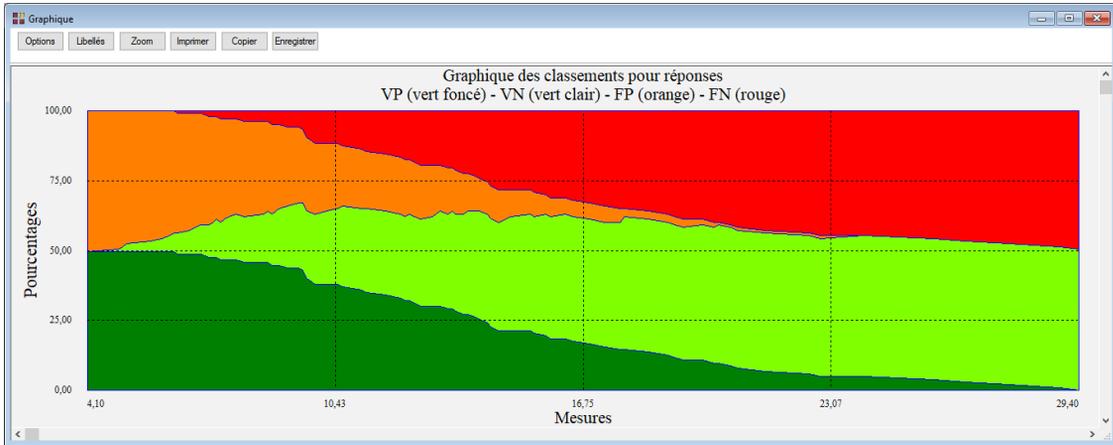
Rapport Courbe ROC

- Statistiques pour les mesures
- Statistiques pour les réponses
- VP, FN, FP, VN, Coûts
- Sensibilité, Spécificité**
- Valeurs prédictives

1								
2	TABLEAU 2 DES RESULTATS							
3								
4	Mesures = Mesures uniques							
5	Sensibilité = Sensibilité en %							
6	Spécificité = Spécificité en %							
7	Indice J de Youden = $(0,01 \times \text{Sensibilité} + 0,01 \times \text{Spécificité}) - 1$							
8	Maximum de l'indice J de Youden = 0,346 pour Mesures = 9,5							
9								
10								
11		Mesures	Sensibilité	Spécificité	Indice J de Youden			
12	1	4,1	100,00000	0,00000	0,00000			
13	2	4,9	100,00000	1,92308	0,01923			
14	3	5,0	100,00000	3,84615	0,03846			
15	4	5,1	100,00000	5,76923	0,05769			
16	5	5,7	100,00000	7,69231	0,07692			
17	6	6,0	100,00000	9,61538	0,09615			
18	7	6,2	100,00000	11,53846	0,11538			
19	8	6,3	100,00000	13,46154	0,13462			
20	9	6,4	98,07692	15,38462	0,13462			
21	10	6,7	98,07692	17,30769	0,15385			

Rapport Explorateur /





Classiquement l'aire sous la courbe est jugée d'apport nul si  $AUC = 0,5$ , peu informative si  $AUC$  est compris entre  $0,5$  et  $0,7$ , moyennement informative si  $AUC$  est compris entre  $0,7$  et  $0,9$  et très informative si  $AUC$  est supérieur à  $0,9$ . Dans cet exemple, l'aire sous la courbe est égale à  $0,726$  ce qui est donc moyennement informatif. Cela signifie qu'un sujet malade aura un résultat pour le test supérieur à celui d'un sujet sain dans  $72,6\%$  des cas.

Le graphique des coûts indique que le coût minimal ou seuil optimal (138) est obtenu pour la valeur des mesures égale à  $9,5$ .

## 6.5 Courbe de Gini

Ce paragraphe décrit comment tracer une courbe de concentration de Gini et calculer l'indice de Gini. Cette courbe est très utilisée, principalement en statistique économique pour étudier les inégalités de répartition d'une grandeur positive.

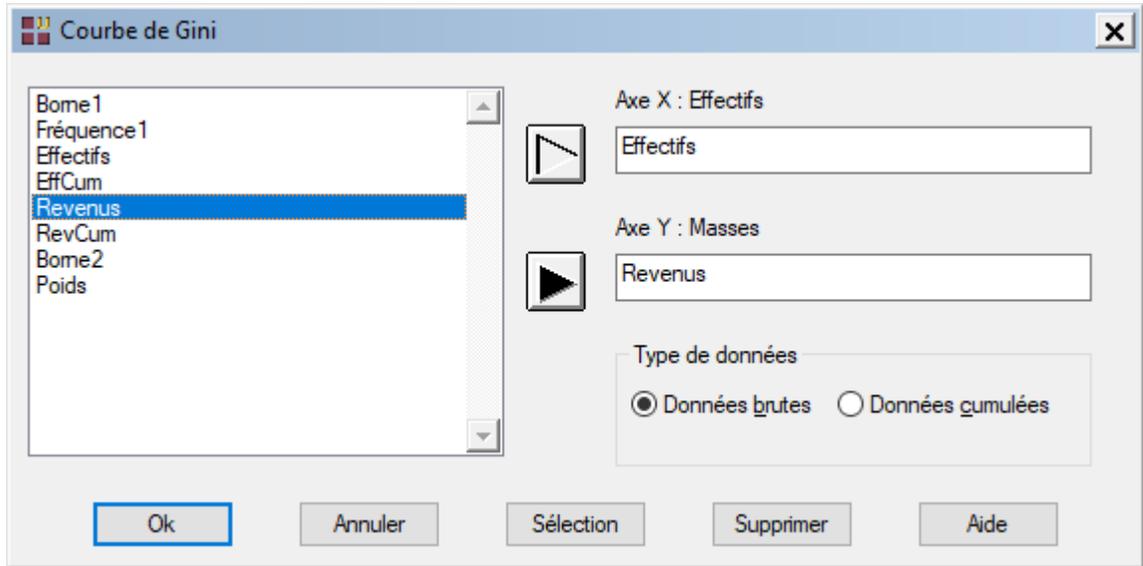
La boîte de dialogue 'Courbe de Gini' permet de préciser :

- la variable pour l'axe des X (effectifs)
- la variable pour l'axe des Y (masses)
- le type de données : données brutes ou données cumulées

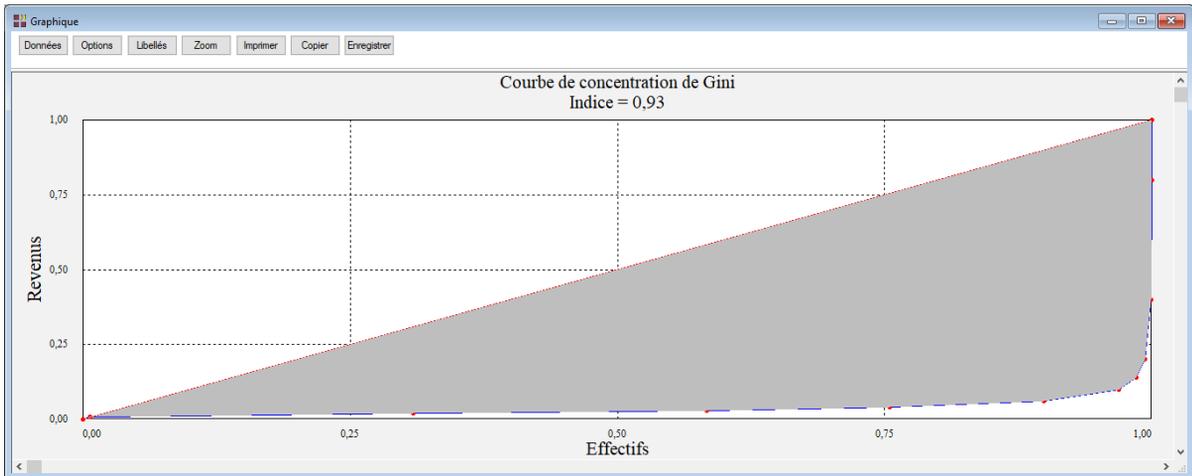
Pour illustrer cette procédure, nous utiliserons un fichier donnant la distribution des revenus imposables des français en 1970 (voir la référence bibliographique : Saporta). Voici les données :

Tranches de revenus en francs	Pourcentages du nombre total de contribuables	Pourcentages cumulés
0 à 5000	0,67	0,67
5000 à 10000	30,18	30,85
10000 à 15000	27,5	58,35
15000 à 20000	17,09	75,44
20000 à 30000	14,45	89,89
30000 à 50000	7,01	96,90
50000 à 70000	1,66	98,56
70000 à 100000	0,81	99,37
100000 à 200000	0,51	99,88
200000 à 400000	0,10	99,98
400000 et plus	0,02	100

Ouvrons le fichier DESCR1 et renseignons la boîte de dialogue comme montré ci-dessous :



En cliquant sur le bouton Ok, on obtient le graphique de la courbe de Gini.



Les variables utilisées 'Effectifs' et 'Revenus' sont des variables contenant les données brutes. Les variables 'EffCum' et 'RevCum' du fichier DESCR1 contiennent ces mêmes données, mais cumulées. Ces données sont à utiliser avec l'option 'Données cumulées'.

## 6.6 Cartes géographiques

Ce paragraphe décrit l'outil de cartographie thématique permettant de faire des représentations cartographiques dans lesquelles les zones géographiques sont colorées en fonction des valeurs d'une variable sélectionnée.

Dans cette version du logiciel, les cartes de France par régions et de France par départements ainsi qu'une carte d'Europe sont disponibles.

Note 1 : Francestat peut mettre à disposition d'autres fonds de cartes. Nous contacter pour préciser votre demande.

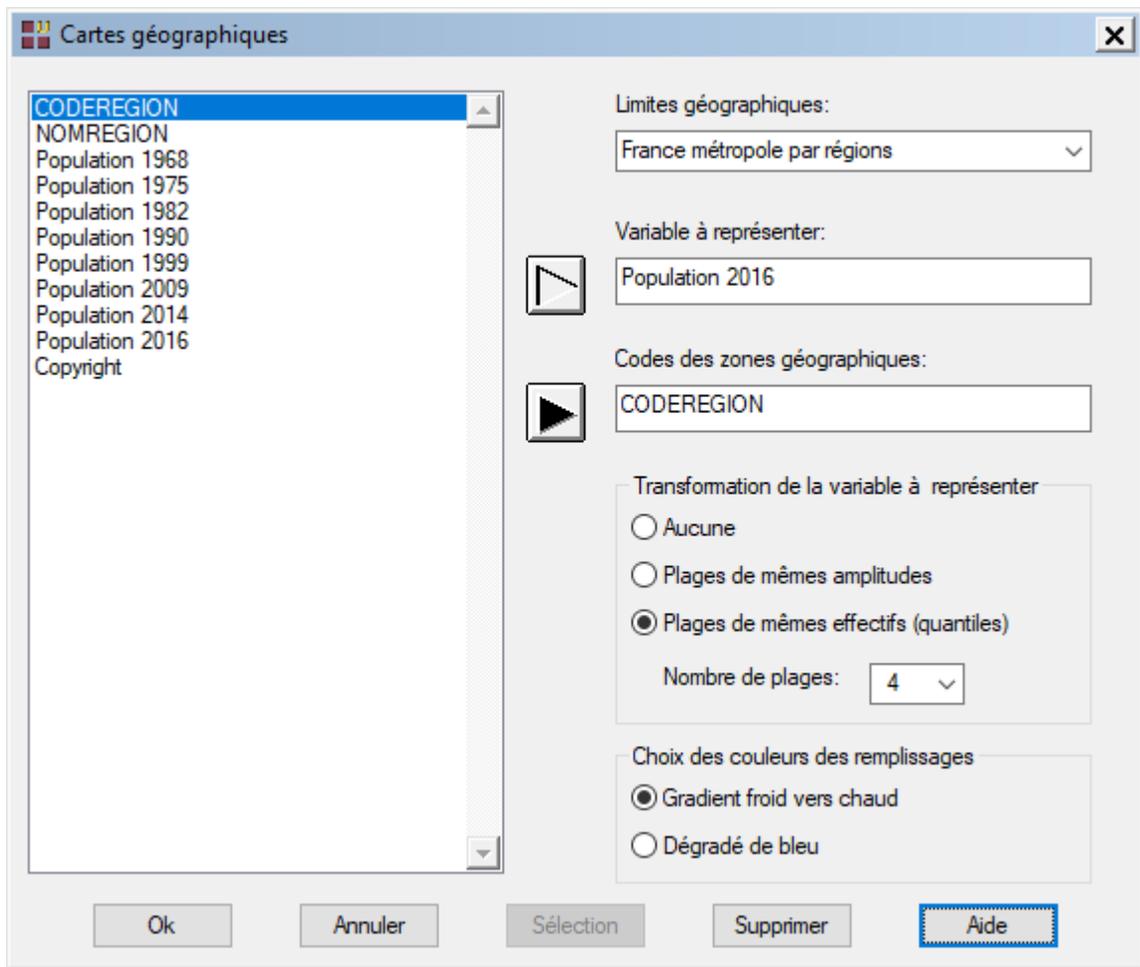
Note 2 : Les fichiers pour la France proviennent de © IGN - GEOFLA ([www.ign.fr](http://www.ign.fr)) et sont distribués avec la permission de l'IGN.

Note 3 : Les fichiers de données proviennent de © INSEE ([www.insee.fr](http://www.insee.fr)) et sont distribués avec la permission de l'INSEE.

Pour illustrer cette procédure, ouvrons le fichier France\_Métropole\_Régions. Il contient pour chacune des 13 régions françaises les évolutions des populations entre 1968 et 2016.

	CODEREGION	NOMREGION	Population 1968	Population 1975	Population 1982	Population 1990	Population 1999	Population 2009	Population 2014	Population 2016	Copyright
	Codes des régions										
	Type = N	Type = C	Type = N	Type = C							
	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 13	Longueur = 2
1	44	Grand-Est	5734998	6111163	6348625	6671915	6954285	7518004	7820966	7940652	© INSEE
2	75	Nouvelle-Aquitaine	2495379	2631260	2680103	2706929	2727126	2810648	2820623	2820150	<a href="http://www.insee.fr">www.insee.fr</a>
3	84	Auvergne-Rhône-Alpes	2468227	2595431	2707886	2795638	2906197	3175064	3276543	3310341	
4	27	Bourgogne-Franche-Comté	1990238	2152500	2264164	2371036	2440329	2538590	2577435	2587004	
5	53	Bretagne	205268	225562	240178	250371	260196	305674	324212	330354	
6	24	Centre-Val-de-Loire	4966097	5184984	5231888	5277946	5386884	5531118	5554645	5558304	
7	94	Corse	5394454	5592417	5673260	5775745	5854069	5944354	6006156	6030309	
8	76	Occitanie	9248631	9878565	10073059	10660554	10952011	11728240	12027565	12142802	
9	32	Hauts-de-France	2757520	2901847	3006341	3128565	3202385	3303822	3335645	3343247	
10	28	Normandie	4676995	4817190	4961927	5113789	5259366	5708497	5879144	5943096	
11	52	Pays-de-la-Loire	3892344	4057772	4251833	4545648	4847335	5473597	5730753	5830166	
12	93	Provence-Alpes-Côte-d'Azur	2582866	2767163	2930398	3059112	3222061	3539048	3690659	3743982	
13	11	Île-de-France	3298836	3675730	3965209	4257907	4506151	4889053	4983438	5024192	

Précisons les données à représenter dans la boîte de dialogue :

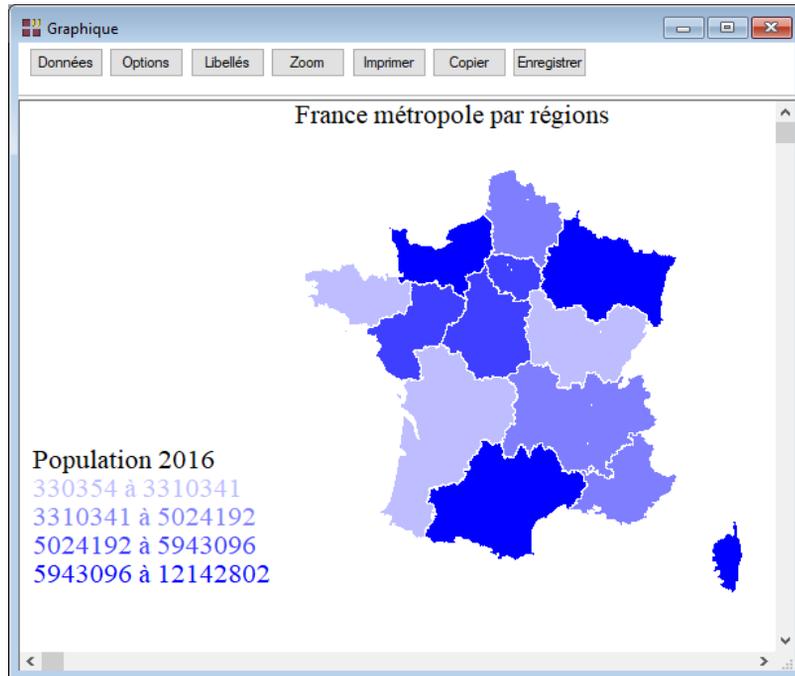
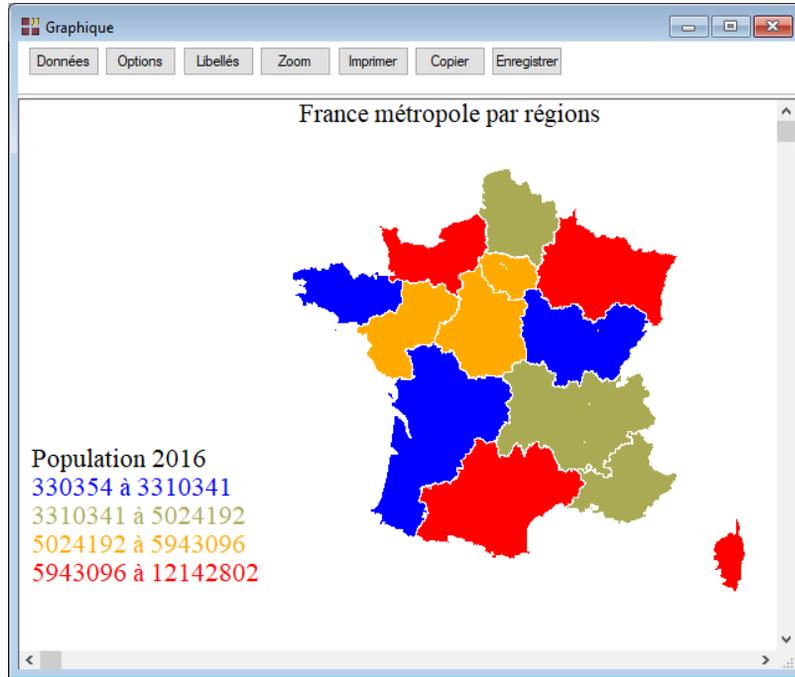


Note : Les codes des zones géographiques ne doivent pas être modifiés.

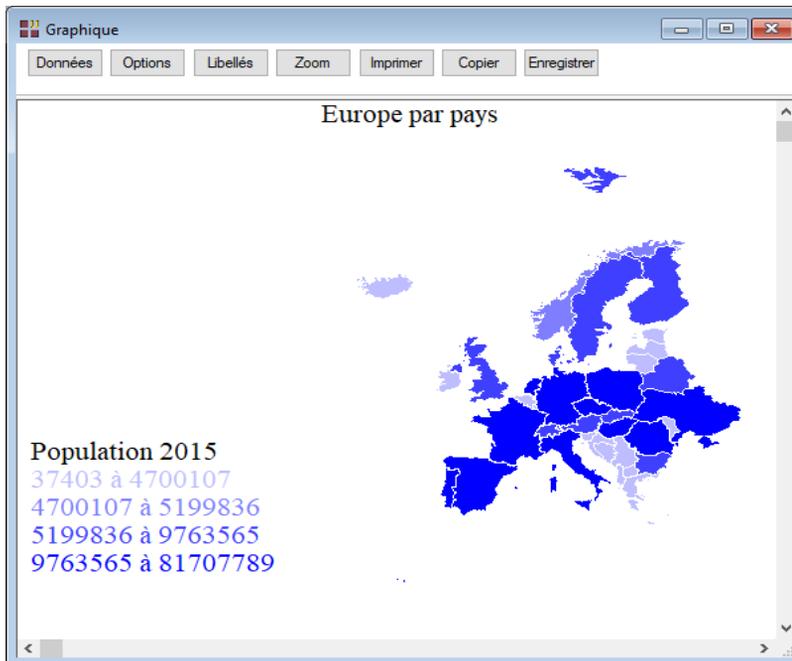
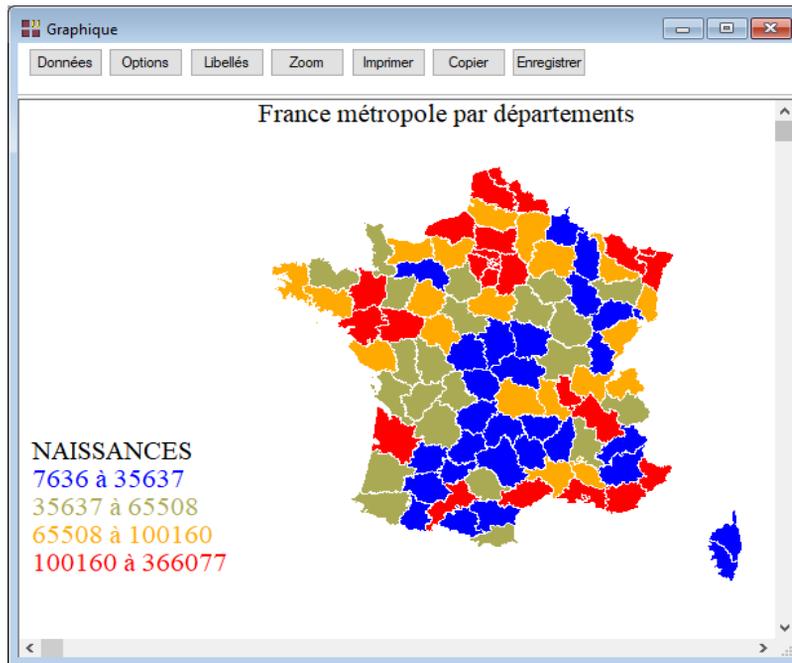
Plusieurs transformations de la variable à représenter sont proposées : aucune (à utiliser si la variable est qualitative) et découpages de l'étendue des données en plages de mêmes amplitudes ou de mêmes effectifs (quantiles).

Les couleurs de remplissages utilisent un gradient froid vers chaud ou un dégradé de bleu.

Note : Pour choisir la police utilisée pour les légendes, modifier dans la boîte de dialogue 'Titres et libellés des axes' la police utilisée pour afficher le titre de l'axe Z.



Voici des exemples utilisant les données France\_Métropole\_Départements et Europe\_Pays :



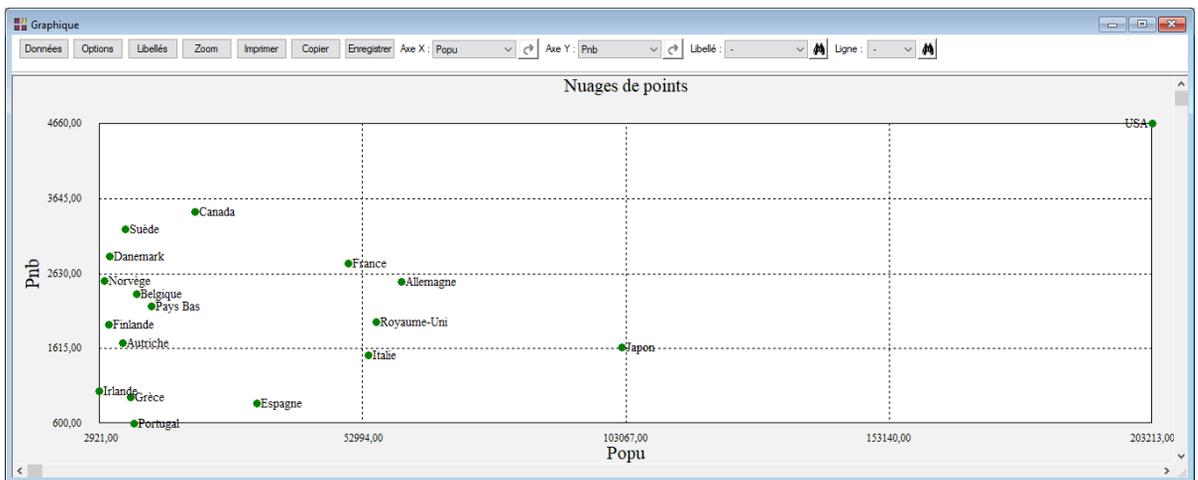
## 6.7 Options graphiques

Ce paragraphe décrit les différentes options disponibles lorsqu'un graphique est à l'écran. Ces options apparaissent dans la barre de menu au-dessus du graphique.

Le paragraphe 6-1 a montré comment on pouvait définir des symboles et des libellés. Il a également montré les possibilités d'interrogation des points du graphique et de zoom manuel.

Ce paragraphe montre comment modifier les caractéristiques d'un graphique (titre, libellés des axes, valeurs sur les axes, grille, police, couleur, couleur de fond, couleur de la bordure, type de symbole, type de ligne ou de remplissage), comment modifier les échelles, comment inverser les axes, comment imprimer le graphique, comment l'enregistrer dans des formats tels que EMF, BMP, GIF, JPEG, TIFF, PNG ou le mettre dans le presse-papiers pour l'envoyer vers d'autres applications Windows.

Créons un graphique à partir du fichier OCDE montrant la relation entre la population (Popu) et le produit national brut (Pnb) et affichons les libellés des pays (Pays).

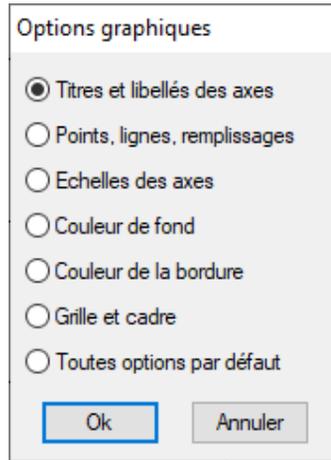


Examinons les possibilités proposées par la barre d'outils :

### Données

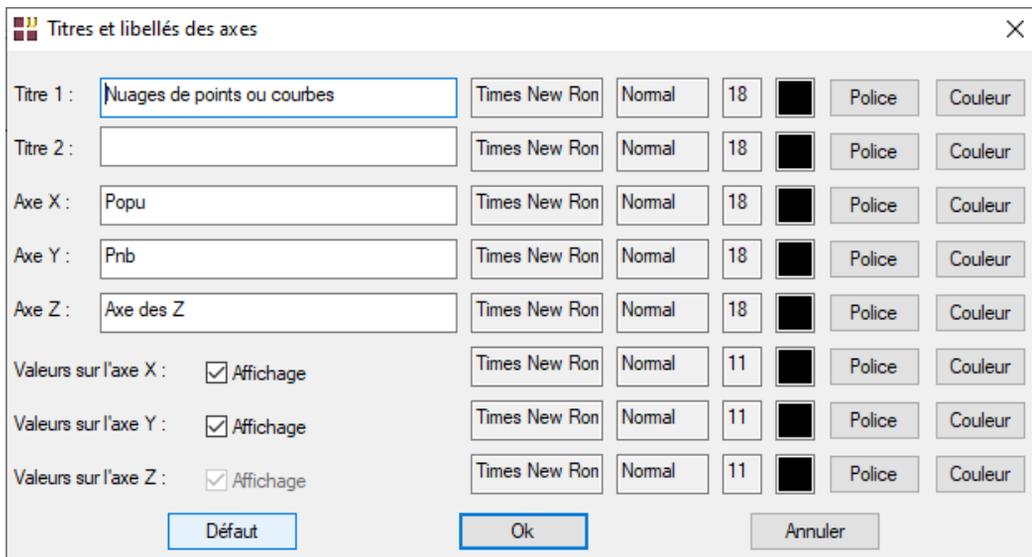
Permet de rappeler la boîte de dialogue d'entrée des données.

## Options



### ◇ Titres et libellés des axes

Le choix de cette option affiche la boîte de dialogue montrée ci-après. Cette boîte de dialogue permet de définir le titre, le sous-titre, les libellés des axes X, Y et Z, les polices et couleurs associées. En cliquant sur l'un quelconque des boutons Police, la boîte de dialogue Police apparaît. Vous pouvez choisir le type de police, son style, sa taille. Les options Affichage permettent d'afficher ou non les valeurs sur les axes.

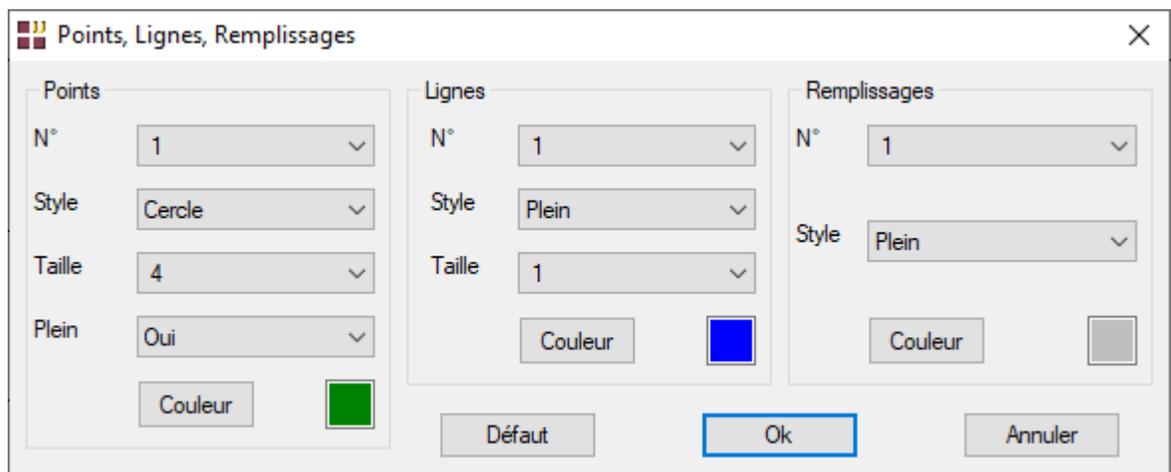


En cliquant sur l'un quelconque des boutons 'Police' ou 'Couleur', les boîtes de dialogue 'Police' et 'Couleurs' apparaissent.

En cliquant sur le bouton Défaut, UNIWIN redéfinit toutes les options du graphique à leurs valeurs par défaut. En cliquant sur Ok, le graphique modifié apparaît. En cliquant sur le bouton Annuler, on revient au graphique inchangé.

◇ Points, Lignes, Remplissages

Le choix de cette option affiche la boîte de dialogue suivante :



Elle permet de choisir les styles, tailles et couleurs pour les points, les lignes et les remplissages.

Vingt types et huit styles de points sont possibles. Les styles sont : carré, cercle, plus, étoile, croix, triangle haut, triangle bas et losange. La taille varie de 1 à 20 pour chacun d'eux. La zone Plein permet de préciser si le symbole utilisé doit être peint intérieurement ou non.

Vingt types et cinq styles de lignes sont possibles. Les styles sont: : plein, plein intérieur, tiret, pointillé, tiret+point. La taille varie de 1 à 20 pour chacun d'eux.

Vingt types et sept styles de remplissages sont possibles. Les styles sont : plein, hachures horizontales, hachures verticales, hachures \\, hachures ///, croix, croix diagonales.

En cliquant sur l'un quelconque des boutons Couleur, la boîte de dialogue Couleurs apparaît.

En cliquant sur le bouton Défaut, UNIWIN redéfinit toutes les options à leurs valeurs par défaut. En cliquant sur Ok, le graphique modifié apparaît. En cliquant sur le bouton Annuler, on revient au graphique inchangé.

◇ Echelles des axes

Le choix de cette option affiche la boîte de dialogue suivante :

Axe X	Axe Y	Axe Z
Minimum : -5090,68	Minimum : 437,6	Minimum : -1,25971200
Maximum : 211224,68	Maximum : 4822,4	Maximum : 1,25971200
Pas : 54078,84	Pas : 1096,19999	Pas : 1
Décimales : 2	Décimales : 2	Décimales : 2
Echelle inversée : <input type="checkbox"/>	Echelle inversée : <input type="checkbox"/>	Echelle inversée : <input type="checkbox"/>
Graphique orthonomé en X et Y (ratio 1/1) : <input type="checkbox"/>		

Elle permet de définir les valeurs minimales et maximales sur chacun des axes X, Y et Z ainsi que les pas et nombres de décimales désirés.

Il est aussi possible de définir une échelle inversée pour un axe, c'est-à-dire d'afficher les valeurs les plus grandes à gauche (en bas) et les plus petites à droite (en haut) ainsi que de tracer un graphique orthonomé (ratio 1/1 pour les axes X et Y).

Par défaut, UNIWIN définit des échelles permettant de visualiser toutes les données et deux décimales pour les graduations. Vous pouvez grâce à cette boîte de dialogue zoomer de façon précise dans votre graphique. Rappelons que cela est directement possible dans le graphique par le clic sur le bouton gauche de la souris (zoom manuel) et par l'option Zoom mais que ces façons de faire sont moins précises.

Après avoir défini vos nouvelles valeurs, en cliquant sur le bouton Ok le nouveau graphique s'affiche.

En cliquant sur le bouton Défaut, on revient aux valeurs initiales. En cliquant sur le bouton Annuler, on revient au graphique inchangé.

#### ◇ Couleur de fond

Cette option permet de choisir la couleur de fond de votre graphique, c'est-à-dire la partie de l'écran intérieure au cadre formé par les axes X et Y. Le choix de cette option fait apparaître la boîte de dialogue Couleurs.

#### ◇ Couleur de la bordure

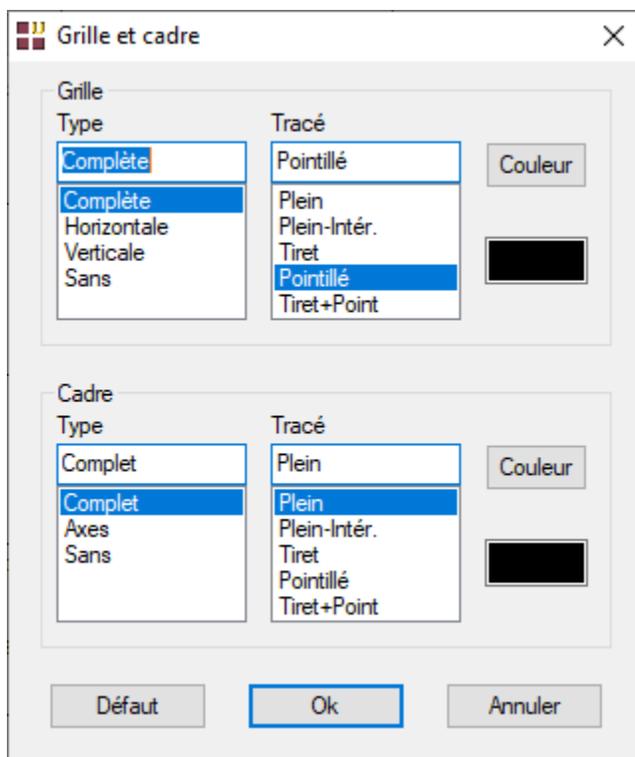
Cette option permet de choisir la couleur de la bordure de votre graphique, c'est-à-dire la partie de l'écran extérieure au cadre formé par les axes X et Y. Le choix de cette option fait apparaître la boîte de dialogue Couleurs.

#### ◇ Grille et cadre

Le choix de cette option affiche la boîte de dialogue montrée ci-après.

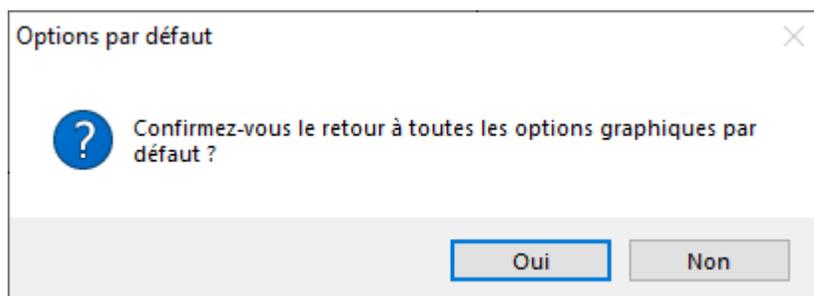
Elle permet de préciser le type de grille désiré (complète (horizontale et verticale), horizontale, verticale ou sans) et le type du cadre (complet, axes, sans). Elle permet également de choisir les types de tracé et les couleurs de la grille et du cadre.

En cliquant sur le bouton Ok on fait apparaître le nouveau graphique. En cliquant sur le bouton Défaut, on revient à la grille par défaut (complète) et aux polices et couleurs par défaut. En cliquant sur le bouton Annuler, on revient au graphique inchangé.



◇ Toutes options par défaut

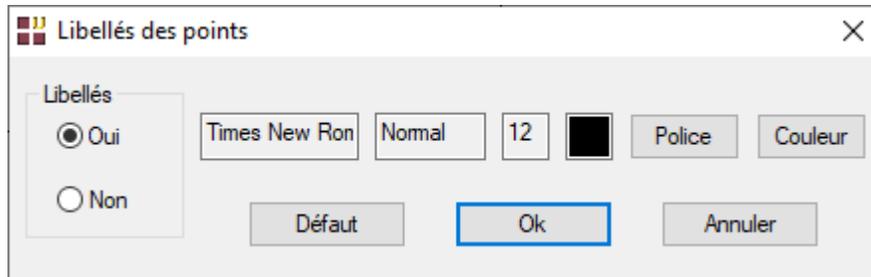
Cette option permet de revenir aux options graphiques par défaut du logiciel UNIWIN pour la procédure en cours. Par sécurité, une confirmation est demandée.



Les paramètres spécifiques définis par l'utilisateur pour la procédure en cours sont alors annulés.

## Libellés

En cliquant sur 'Libellés' dans la barre d'outils, la boîte de dialogue ci-dessous s'affiche :



Cette option permet de préciser si on désire ou non l'affichage des libellés des points ainsi que de choisir la police et la couleur de ces libellés.

Les libellés des points n'apparaîtront que s'ils ont été préalablement définis dans la boîte de dialogue de la procédure. Sinon, les numéros des points seront affichés.

En cliquant sur le bouton Ok, on fait apparaître le nouveau graphique. En cliquant sur le bouton Défaut, on revient à la police et à la couleur par défaut. En cliquant sur le bouton Annuler, on revient au graphique inchangé.

## Zoom

Le zoom permet d'agrandir le graphique de façon manuelle (Manuel), en X & Y (150%, 200% et 250%), en X ou Y (300%, 350%, 400%), de désactiver l'usage du zoom et de revenir à la taille initiale du graphique (Sans). La barre de statut en pied d'écran indique l'option en cours.

Le choix de l'option Manuel permet de définir un zoom personnalisé par le clic gauche de la souris puis en glissant.

Une taille supérieure à la taille de l'écran est très utile lorsque le graphique est dense : elle permet de mieux éparpiller les points et facilite l'analyse exploratoire. Un ascenseur horizontal et un ascenseur vertical permettent de faire défiler le graphique.

Ces options Zoom s'ajoutent à la possibilité du menu Options (zoom par modification des échelles des axes).

## Imprimer

Le bouton Imprimer permet d'imprimer le graphique complet ou seulement sa partie affichée (en cas de zoom) sur votre imprimante, de choisir et de configurer votre imprimante.

En cliquant sur Imprimer, un message apparaît indiquant que l'impression est en cours.

En cliquant sur Configuration de l'impression, la boîte de dialogue de configuration s'affiche.

Vous pouvez notamment choisir l'orientation de votre graphique : portrait ou paysage.

Dans les graphiques d'analyse des données, l'axe des X représente fréquemment le premier axe factoriel. Il est donc judicieux d'imprimer le graphique en mode Paysage.

A noter qu'à l'impression, en fonction de la résolution de votre imprimante, il peut être nécessaire de modifier les tailles des polices, même si l'affichage à l'écran est jugé correct.

## Copier

Le bouton Copier permet d'envoyer l'image graphique à l'écran dans le presse-papiers de Windows. Un message s'affiche indiquant que l'image a été copiée dans le presse-papiers.

## Enregistrer

Le bouton Enregistrer permet d'enregistrer le graphique au format vectoriel EMF (Enhanced Metafile Format) et dans d'autres formats tels que BMP, GIF, JPEG, TIFF et PNG.

L'enregistrement au format EMF permet la récupération de l'image pour la composition graphique dans UNIWIN et dans tous les logiciels qui relisent le format vectoriel EMF. Ce format autorise la modification (police, couleur...) et le déplacement dans le logiciel d'arrivée des objets de l'image, par exemple les libellés des points.

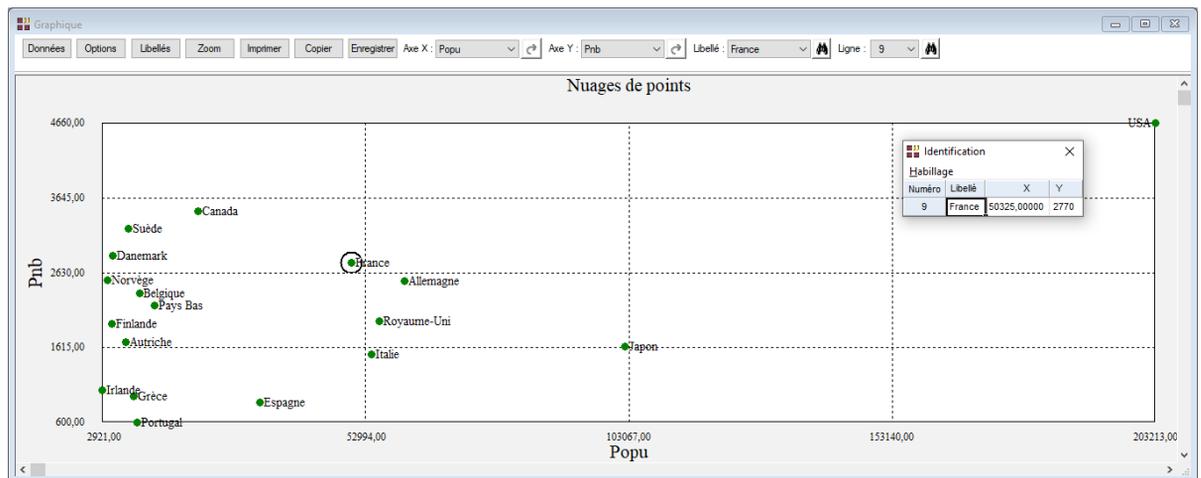
## Axe X, Axe Y, Libellé, Ligne



Les deux premiers boutons permettent de choisir les variables à afficher sur les axes X et Y.

Après avoir choisi les variables, cliquer sur l'une de deux icônes  associées à ces boutons pour activer vos choix.

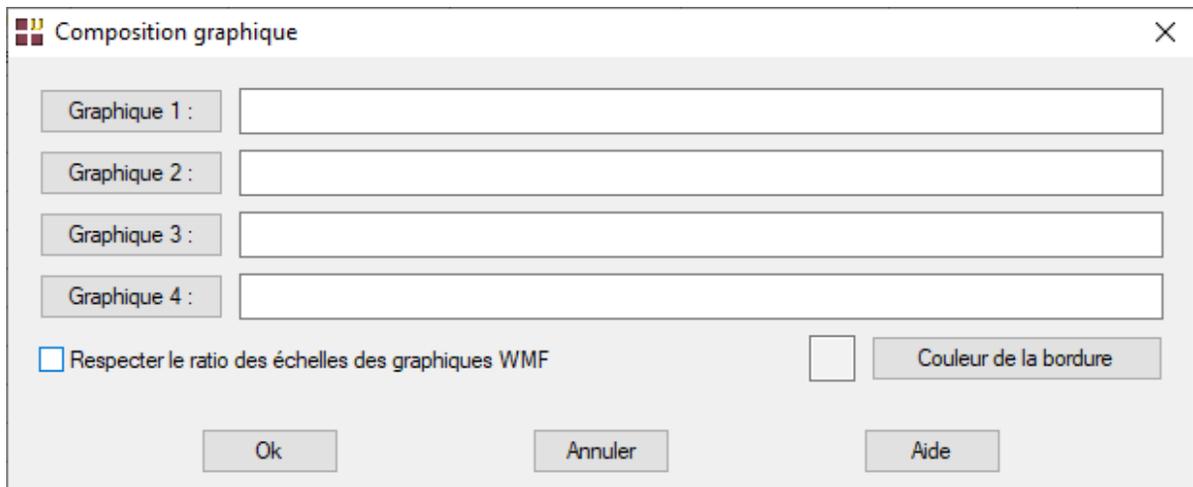
Les boutons Libellé et Ligne affichent les listes des libellés et des numéros des points représentés dans le graphique. En sélectionnant un libellé ou un numéro de ligne puis en cliquant sur l'icône , le point sélectionné est mis en évidence dans le graphique :



## 6.8 Composition graphique

Cette option permet la mise en page de 1, 2 ou 4 graphiques issus d'UNIWIN ou d'autres applications créant des fichiers vectoriels EMF (Enhanced Metafile Format) ou WMF (Windows Metafile Format).

Les clics sur les boutons Graphique 1, 2, 3 et 4 permettent la sélection des fichiers à afficher.



La case à cocher 'Respecter le ratio des échelles des graphiques WMF' permet de ne pas déformer le graphique WMF en préservant ses tailles initiales.

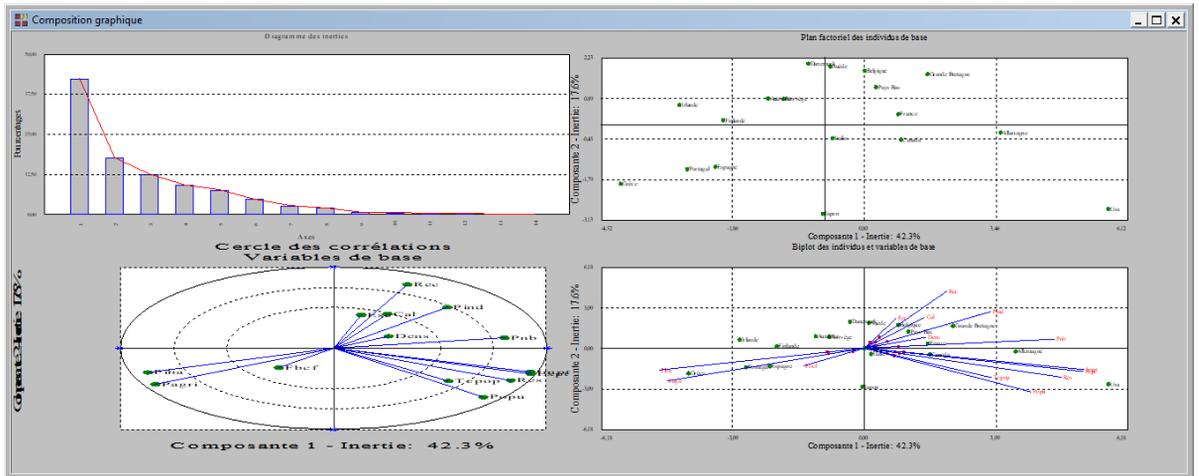
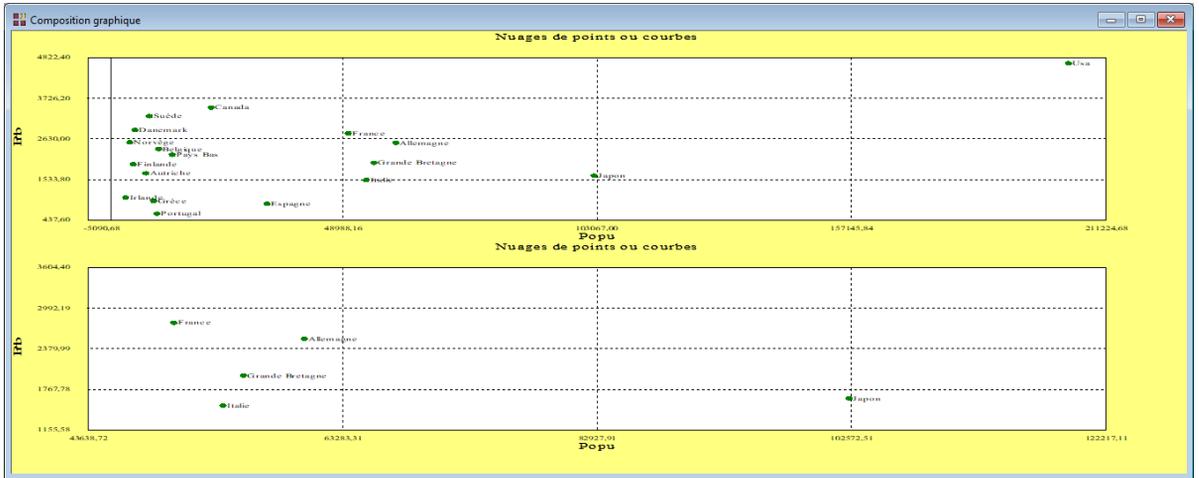
La couleur de la bordure de la composition graphique (partie extérieure à la composition) peut également être choisie.

Pour copier cette composition graphique dans le presse-papiers, utiliser l'option Coller du ruban Accueil.

Pour imprimer cette composition graphique, utiliser l'option Imprimer du ruban Accueil.

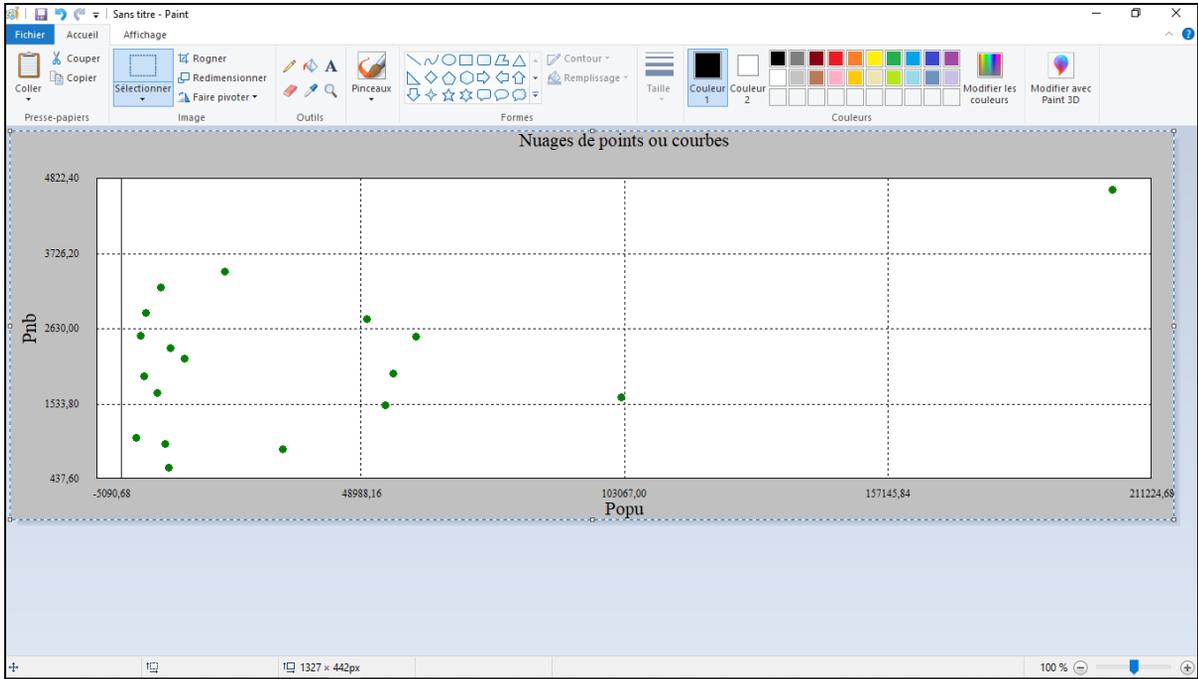
Voici des exemples avec 2 ou 4 graphiques issus d'analyses d'UNIWIN :

1. Exemple utilisant le nuage de points OCDE avant et après zoom.
2. Exemple issu d'une analyse en composantes principales (ACP) sur le fichier OCDE et affichant le diagramme des inerties, le cercle des corrélations, le premier plan factoriel et le Biplot.



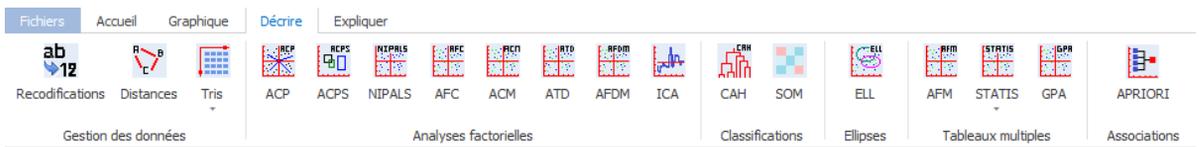
## 6.9 Accès à Microsoft Paint

Ce bouton charge le logiciel Microsoft Paint. Si un graphique d'UNIWIN a été copié dans le presse-papiers, il peut être collé dans cette application pour le modifier et l'enrichir.



## Le ruban Décrire

Le ruban 'Décrire' permet de recoder des variables alphabétiques, de calculer des tableaux de distances euclidiennes, de faire des tris à plat, des tris croisés et de mettre en œuvre des analyses de données multivariées (ACP, ACPS, NIPALS, AFC, ACM, ATD, AFDM, ICA, CAH, SOM, ELL, AFM, STATIS et STATIS duale, GPA, APRIORI).



### 7.1 Recoder des variables alphabétiques

L'option Recodifications permet de recoder des variables alphabétiques en variables numériques qualitatives. Elle ajoute dans le fichier des résultats deux nouvelles variables pour chacune des variables recodées :

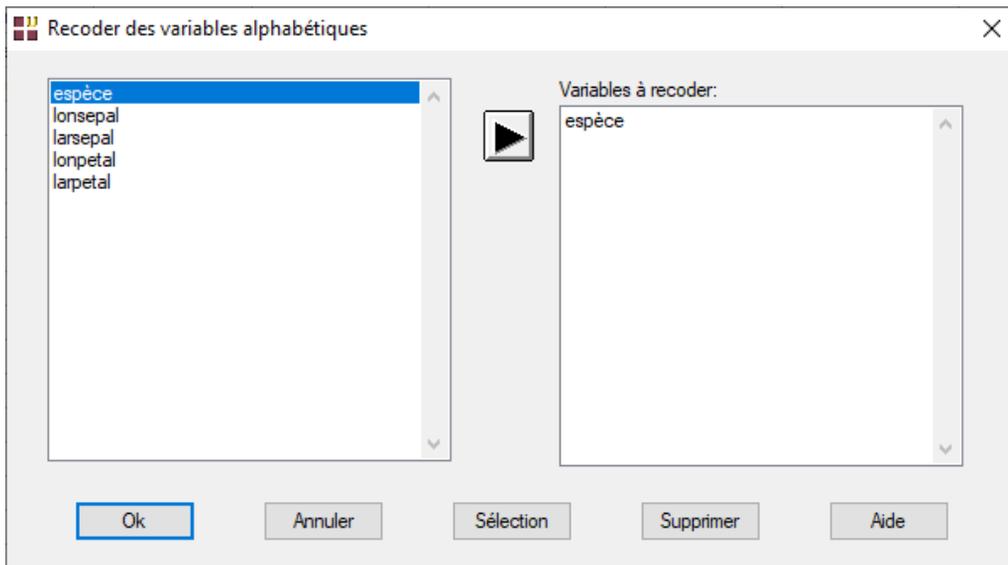
1. Une variable nommée `cod_var` (où `var` est le nom de la variable alphabétique) contenant les codages numériques des données alphabétiques
2. Une variable nommée `lib_var` (où `var` est le nom de la variable alphabétique) contenant les libellés uniques de la variable `var`

Prenons pour exemple le fichier IRIS2 dont un extrait est montré ci-dessous :

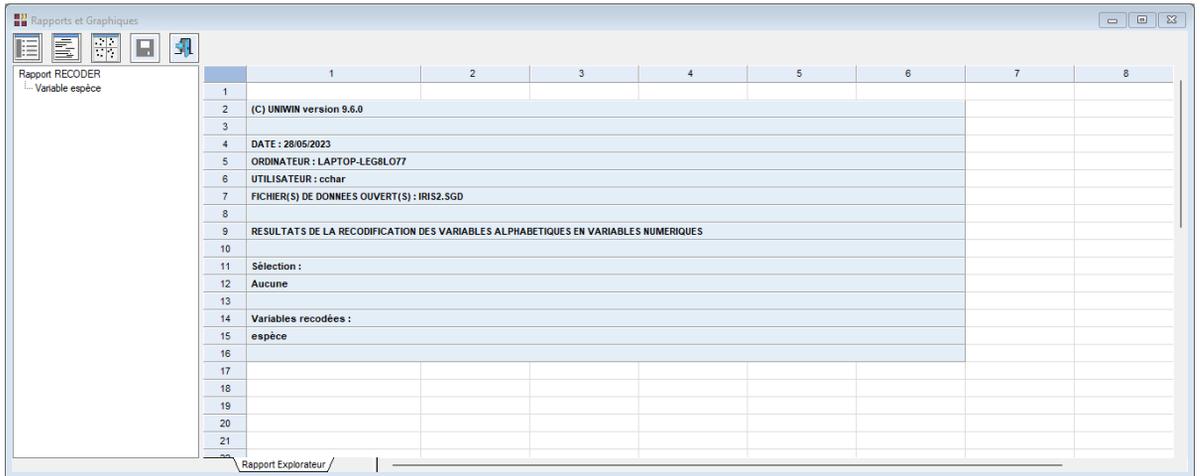
Données					
	espèce	lonsepal	larsepal	lonpetal	larpetal
		Longueurs sépales	Largeurs sépales	Longueurs pétales	Largeurs pétales
	Type = C	Type = N	Type = N	Type = N	Type = N
	Longueur = 150	Longueur = 150	Longueur = 150	Longueur = 150	Longueur = 150
1	Setosa	5,1	3,5	1,4	0,2
2	Setosa	4,9	3,0	1,4	0,2
3	Setosa	4,7	3,2	1,3	0,2
4	Setosa	4,6	3,1	1,5	0,2
5	Setosa	5,0	3,6	1,4	0,2
6	Setosa	5,4	3,9	1,7	0,4
7	Setosa	4,6	3,4	1,4	0,3
8	Setosa	5,0	3,4	1,5	0,2
9	Setosa	4,4	2,9	1,4	0,2
10	Setosa	4,9	3,1	1,5	0,1
11	Setosa	5,4	3,7	1,5	0,2
12	Setosa	4,8	3,4	1,6	0,2

Ce fichier contient une colonne nommée « espèce » indiquant pour chacun des 150 iris étudiés son espèce sous la forme d'un libellé alphabétique. Nous souhaitons recoder cette variable de façon numérique.

La boîte de dialogue de recodification nous permet de sélectionner une ou plusieurs variables à recoder. Sélectionnons la variable « espèce » puis cliquons sur Ok.



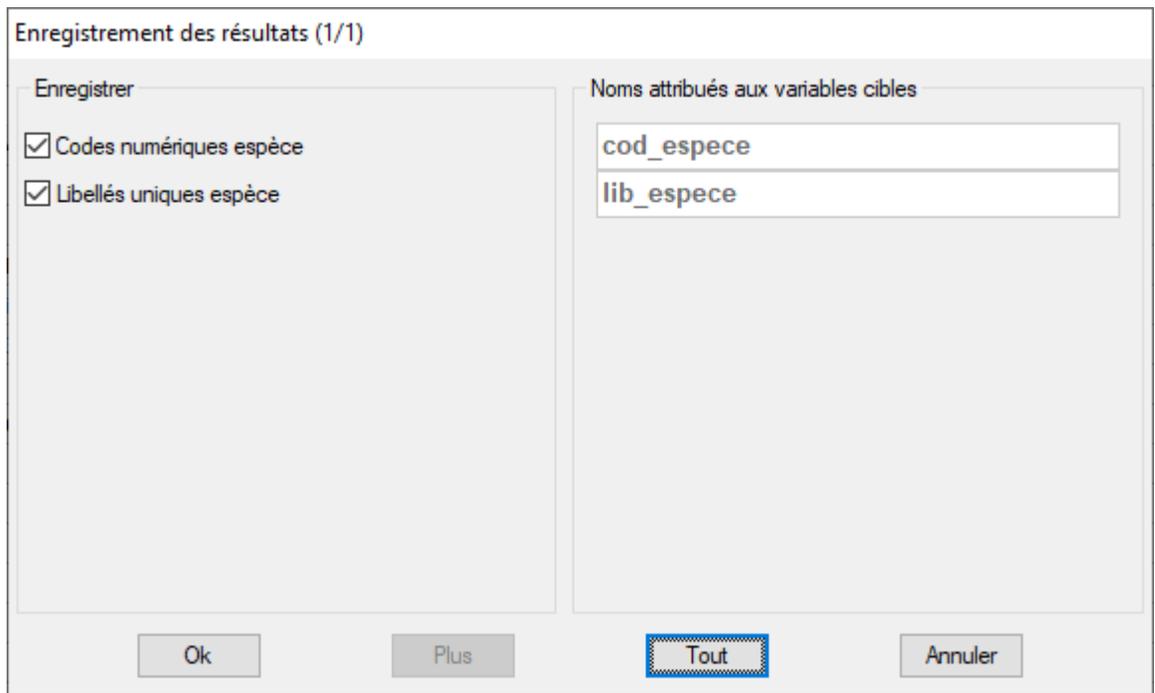
Après quelques instants, l'écran suivant s'affiche :



The screenshot shows a software window titled 'Rapports et Graphiques'. On the left, there is a sidebar with 'Rapport RECODER' and 'Variable espèce'. The main area is a table with 8 columns and 21 rows. The content of the table is as follows:

	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 9.6.0							
3								
4	DATE : 28/05/2023							
5	ORDINATEUR : LAPTOP-LEG8L077							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : IRIS2.SGD							
8								
9	RESULTATS DE LA RECODIFICATION DES VARIABLES ALPHABETIQUES EN VARIABLES NUMERIQUES							
10								
11	Sélection :							
12	Aucune							
13								
14	Variables recodées :							
15	espèce							
16								
17								
18								
19								
20								
21								

Clignons sur l'icône  pour enregistrer les variables créées :



The dialog box is titled 'Enregistrement des résultats (1/1)'. It has two main sections:

- Enregistrer**: Contains two checked checkboxes: 'Codes numériques espèce' and 'Libellés uniques espèce'.
- Noms attribués aux variables cibles**: A text area containing 'cod\_espece' and 'lib\_espece' on separate lines.

At the bottom, there are four buttons: 'Ok', 'Plus', 'Tout' (highlighted with a blue dashed border), and 'Annuler'.

Cliquons sur Tout puis sur Ok. Après avoir précisé le nom du fichier dans lequel ces variables sont enregistrées, visualisons ce fichier :

	cod_espece	lib_espece
	Codes numériques espèce	Libellés uniques espèce
	Type = Numérique	Type = Caractère
	Longueur = 150	Longueur = 3
1	1	Setosa => 1
2	1	Versicolor => 2
3	1	Virginica => 3
4	1	
5	1	
6	1	
7	1	
8	1	
9	1	
10	1	

La variable « lib\_espece » contient les informations relatives à la recodification et la variable « cod\_espece » indique le numéro du libellé unique pour chacun des iris.

## 7.2 Créer un tableau de distances euclidiennes

Cette option permet de créer un tableau de distances euclidiennes (matrice triangulaire inférieure) entre des observations (individus) observées sur un ensemble de variables.

La distance  $d$  entre deux individus  $i$  et  $j$  observés sur  $m$  variables est calculée par :

$$d(i, j) = \sqrt{\sum_{k=1}^m (x_i^k - x_j^k)^2}$$

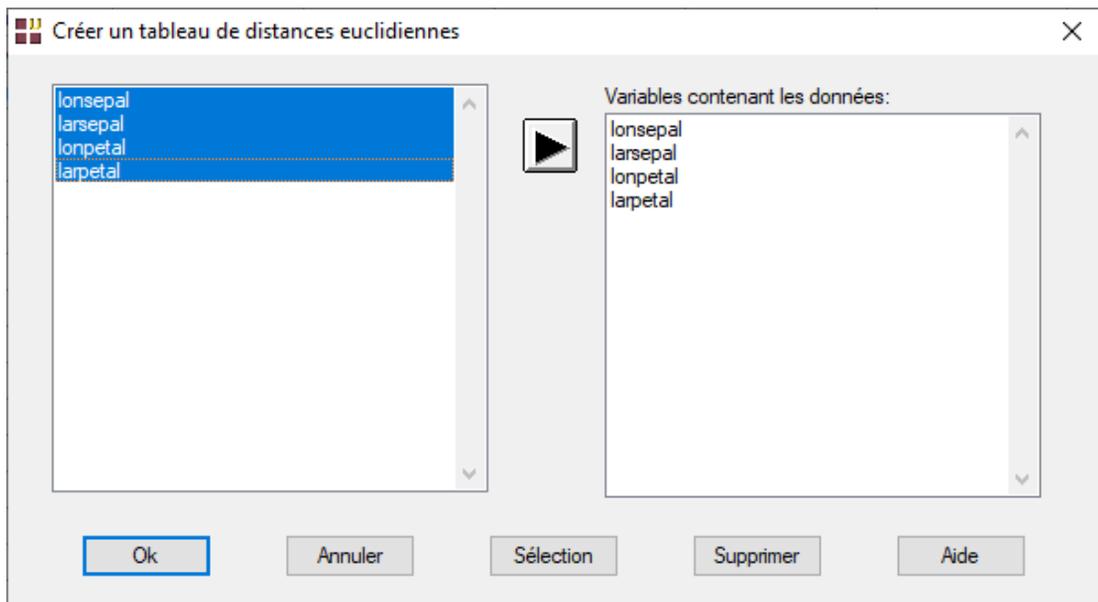
Le tableau résultant de cette procédure est directement utilisable dans l'Analyse sur Tableau de Distances ou de Dissimilarités (ATD).

Une variable (comportant autant de colonnes que d'observations) est créée à l'issue de cette procédure.

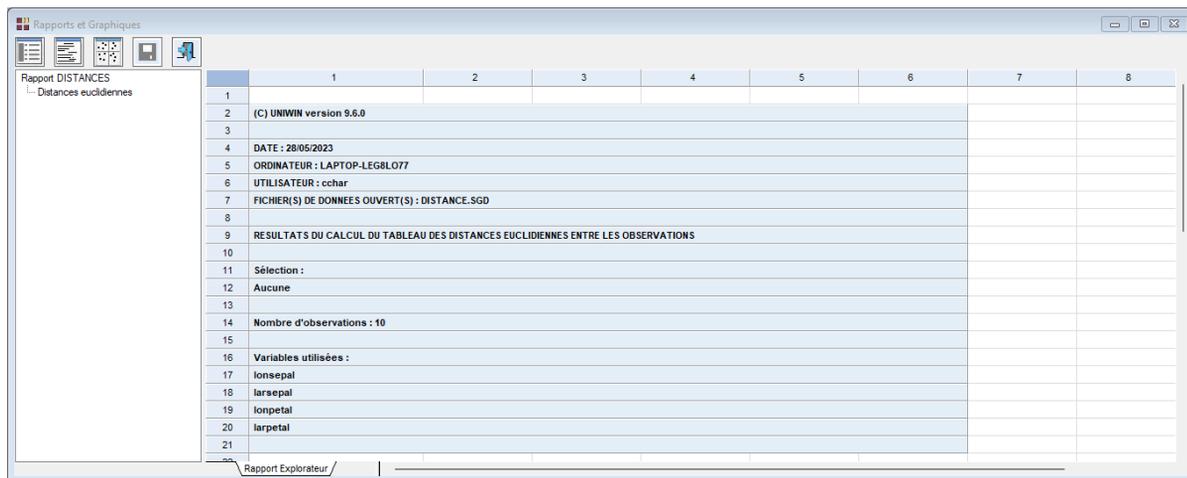
Prenons un exemple en utilisant le fichier DISTANCE constitué des 10 premières observations du fichier IRIS. Quatre caractéristiques sont mesurées sur ces dix iris : longueurs des sépales, largeurs des sépales, longueurs des pétales et largeurs des pétales.

Données				
	lonsepal	larsepal	lonpetal	larpetal
	Longueurs sépales	Largeurs sépales	Longueurs pétales	Largeurs pétales
	Type = N	Type = N	Type = N	Type = N
	Longueur = 10	Longueur = 10	Longueur = 10	Longueur = 10
1	5,1	3,5	1,4	0,2
2	4,9	3,0	1,4	0,2
3	4,7	3,2	1,3	0,2
4	4,6	3,1	1,5	0,2
5	5,0	3,6	1,4	0,2
6	5,4	3,9	1,7	0,4
7	4,6	3,4	1,4	0,3
8	5,0	3,4	1,5	0,2
9	4,4	2,9	1,4	0,2
10	4,9	3,1	1,5	0,1

Renseignons la boîte de dialogue comme montré ci-dessous puis cliquons sur Ok.



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

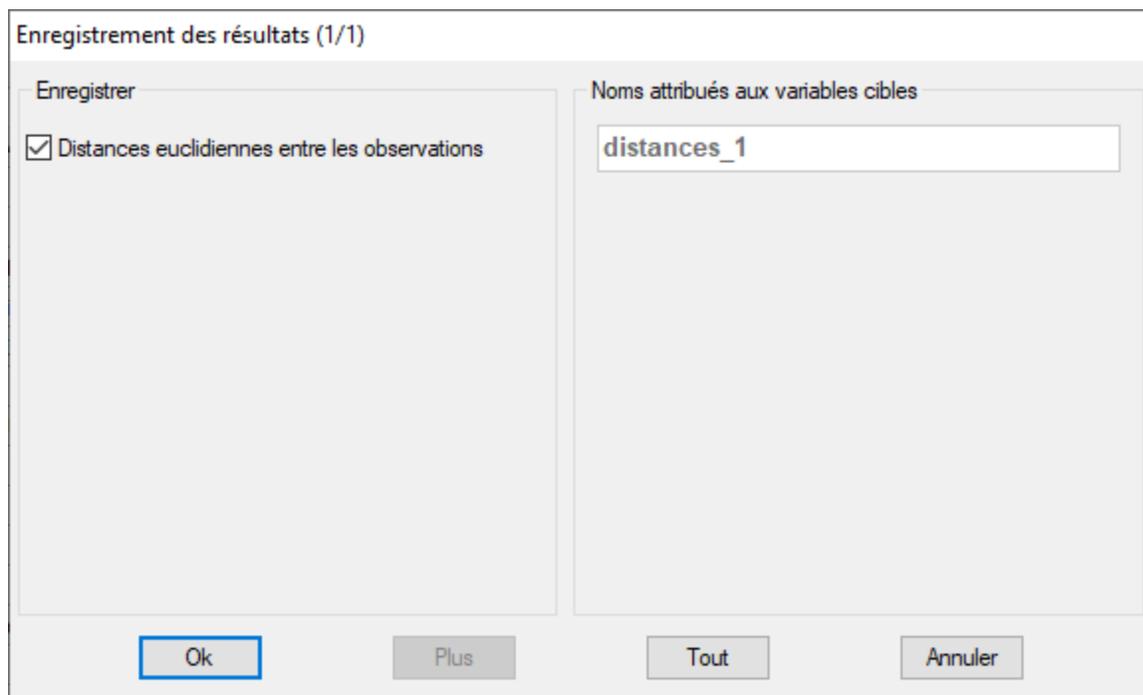


The screenshot shows a software window titled "Rapports et Graphiques". On the left, there is a sidebar with a tree view containing "Rapport DISTANCES" and "Distances euclidiennes". The main area displays a report with the following content:

1								
2	(C) UNIWIN version 9.6.0							
3								
4	DATE : 28/05/2023							
5	ORDINATEUR : LAPTOP-LEG8L077							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : DISTANCE.SGD							
8								
9	RESULTATS DU CALCUL DU TABLEAU DES DISTANCES EUCLIDIENNES ENTRE LES OBSERVATIONS							
10								
11	Sélection :							
12	Aucune							
13								
14	Nombre d'observations : 10							
15								
16	Variables utilisées :							
17	lonsepai							
18	larsepai							
19	lonpetal							
20	larpetal							
21								

At the bottom of the window, the text "Rapport Explorateur /" is visible.

Cliquons sur l'icône  pour enregistrer les variables créées :



The dialog box is titled "Enregistrement des résultats (1/1)". It contains two main sections:

- Enregistrer**: A checkbox labeled "Distances euclidiennes entre les observations" is checked.
- Noms attribués aux variables cibles**: A text input field contains the value "distances\_1".

At the bottom of the dialog, there are four buttons: "Ok" (highlighted with a blue border), "Plus", "Tout", and "Annuler".

Ouvrons ce fichier :

Données										
	Dist_1	Dist_2	Dist_3	Dist_4	Dist_5	Dist_6	Dist_7	Dist_8	Dist_9	Dist_10
	Distances 1	Distances 2	Distances 3	Distances 4	Distances 5	Distances 6	Distances 7	Distances 8	Distances 9	Distances 10
	Type = N									
	Longueur = 10									
1	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0
2	0,53852	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0
3	0,50990	0,30000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0
4	0,64807	0,33166	0,24495	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0
5	0,14142	0,60828	0,50990	0,64807	0,00000	0,00000	0,00000	0,00000	0,00000	0
6	0,61644	1,09087	1,08628	1,16619	0,61644	0,00000	0,00000	0,00000	0,00000	0
7	0,51962	0,50990	0,26458	0,33166	0,45826	0,99499	0,00000	0,00000	0,00000	0
8	0,17321	0,42426	0,41231	0,50000	0,22361	0,70000	0,42426	0,00000	0,00000	0
9	0,92195	0,50990	0,43589	0,30000	0,92195	1,45945	0,54772	0,78740	0,00000	0
10	0,46904	0,17321	0,31623	0,31623	0,52915	1,00995	0,47958	0,33166	0,55678	0
11										

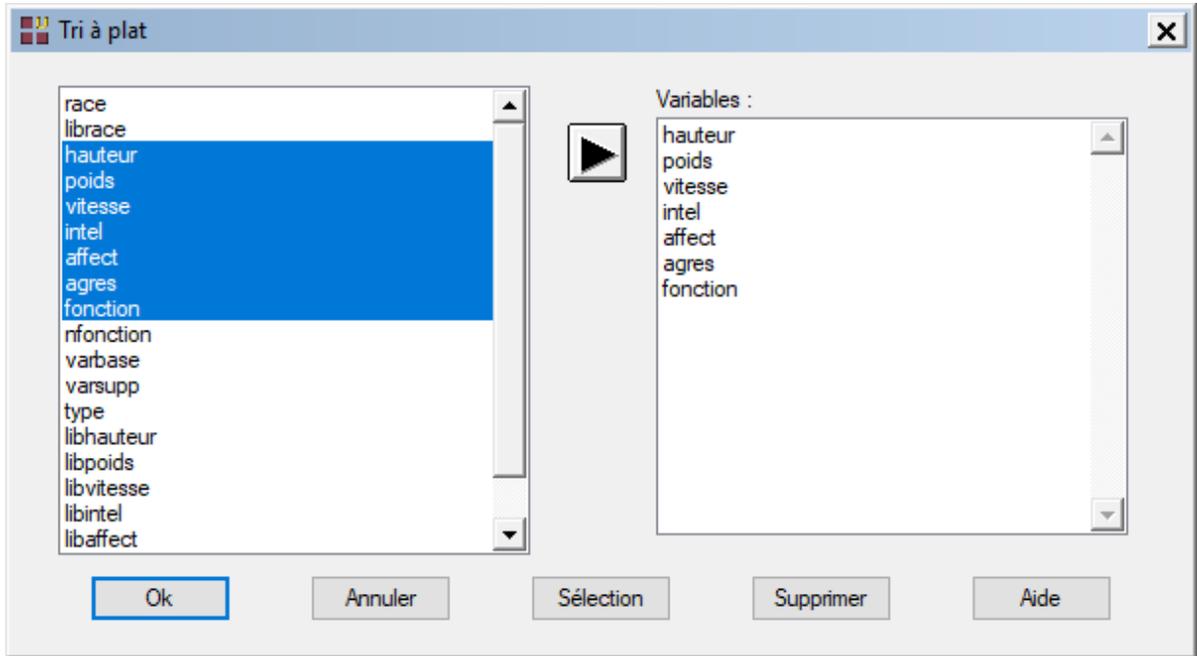
## 7.3 Tri à plat et tri croisé

La procédure ‘Tri à plat’ est conçue pour résumer une unique colonne de données qualitatives. Elle calcule les nombres de fois où chacune des valeurs uniques de cette colonne apparaît.

La procédure ‘Tri croisé’ est conçue pour résumer deux ensembles de colonnes de données qualitatives. Elle construit un tableau à deux entrées affichant les comptages des occurrences de toutes les paires de valeurs uniques dans les deux ensembles de colonnes. Ce tableau est, par exemple, directement utilisable dans une Analyse Factorielle des Correspondances simples (AFC).

### Tri à plat

Prenons un exemple en utilisant le fichier CHIENS. Renseignons la boîte de dialogue comme montré ci-dessous :



Sept tris à plat seront effectués.

Cliquons sur Ok. La fenêtre suivante s'affiche :

	Effectifs	Fréquences	Effectifs cumulés	Fréquences cumulées
A11	13	48,15	13	48,15
A12	14	51,85	27	100,00



L'explorateur à gauche dans cette fenêtre permet d'afficher chaque tri à plat. L'icône affiche la boîte de dialogue permettant de sélectionner les résultats à enregistrer :

**Enregistrement des résultats (1/2)**

Enregistrer

- Comptages affect
- Modalités affect
- Comptages agres
- Modalités agres
- Comptages fonction
- Modalités fonction
- Comptages hauteur
- Modalités hauteur
- Comptages intel
- Modalités intel

Noms attribués aux variables cibles

compt_affect
modal_affect
compt_agres
modal_agres
compt_fonction
modal_fonction
compt_hauteur
modal_hauteur
compt_intel
modal_intel

Ok Plus Tout Annuler

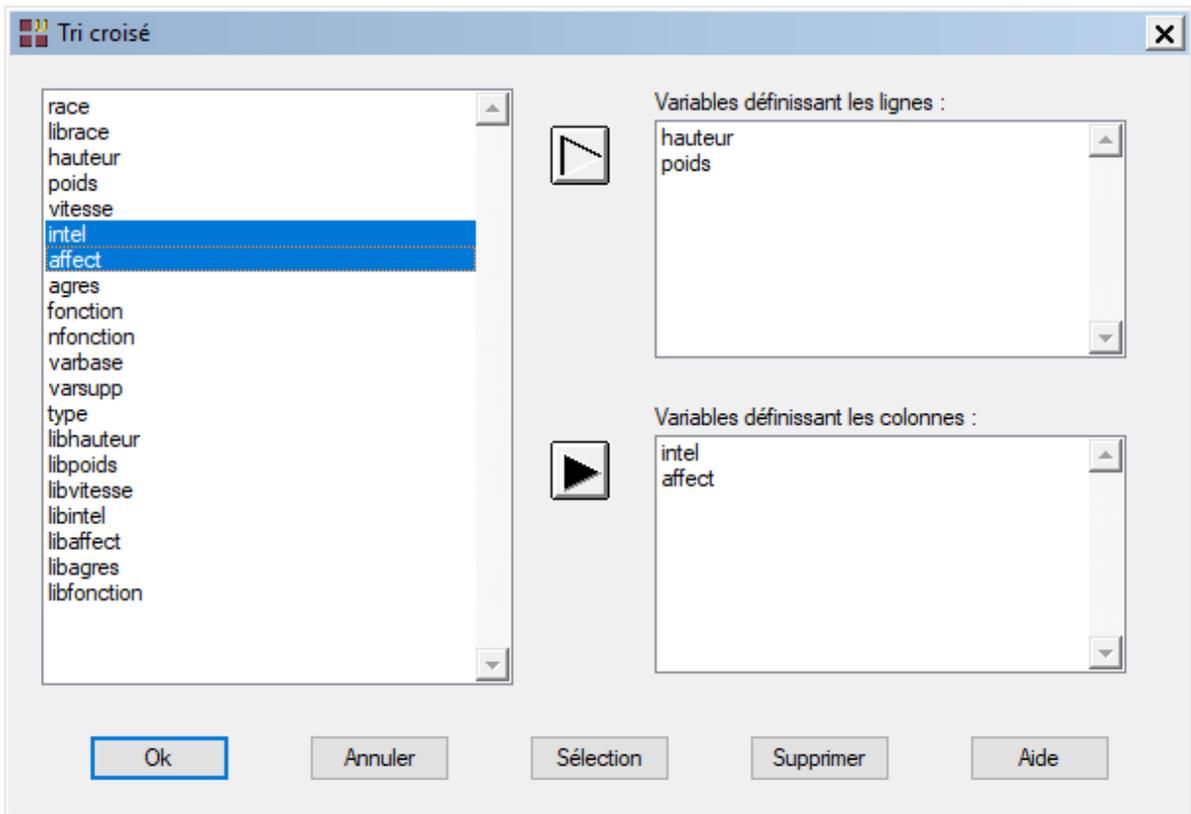
Cliquer dans les cases à cocher pour indiquer les résultats à enregistrer.

Ouvrons le fichier des résultats :

Données						
	compt_affect	modal_affect	compt_agres	modal_agres	compt_fonction	modal_fonction
	Comptages affect	Modalités affect	Comptages agres	Modalités agres	Comptages fonction	Modalités fonction
	Type = N	Type = C	Type = N	Type = C	Type = N	Type = C
	Longueur = 2	Longueur = 2	Longueur = 2	Longueur = 2	Longueur = 3	Longueur = 3
1	13	Af1	14	Ag1	9	Chasse
2	14	Af2	13	Ag2	10	Compagnie
3					8	Utile

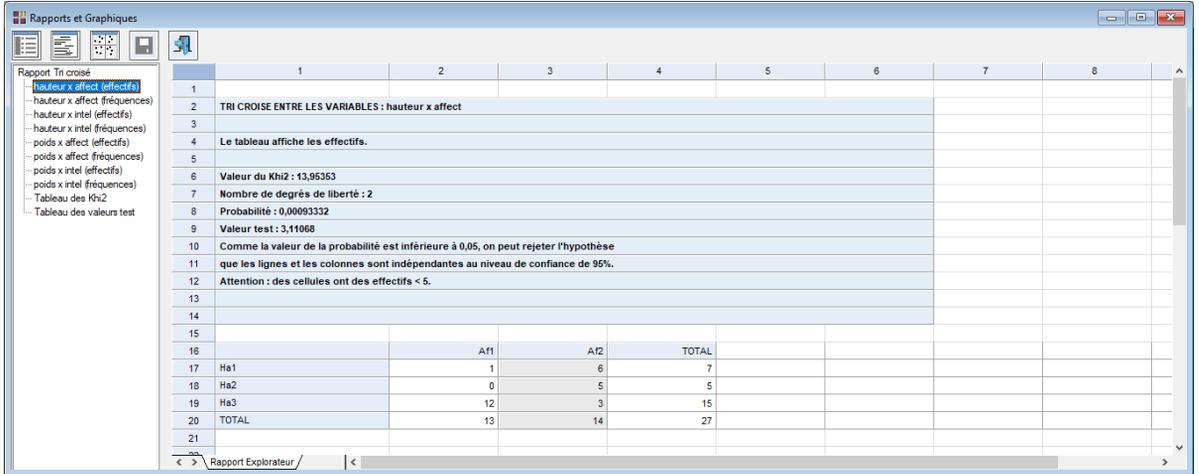
## Tri croisé

Prenons un exemple en utilisant le fichier CHIENS. Renseignons la boîte de dialogue comme montré ci-dessous :

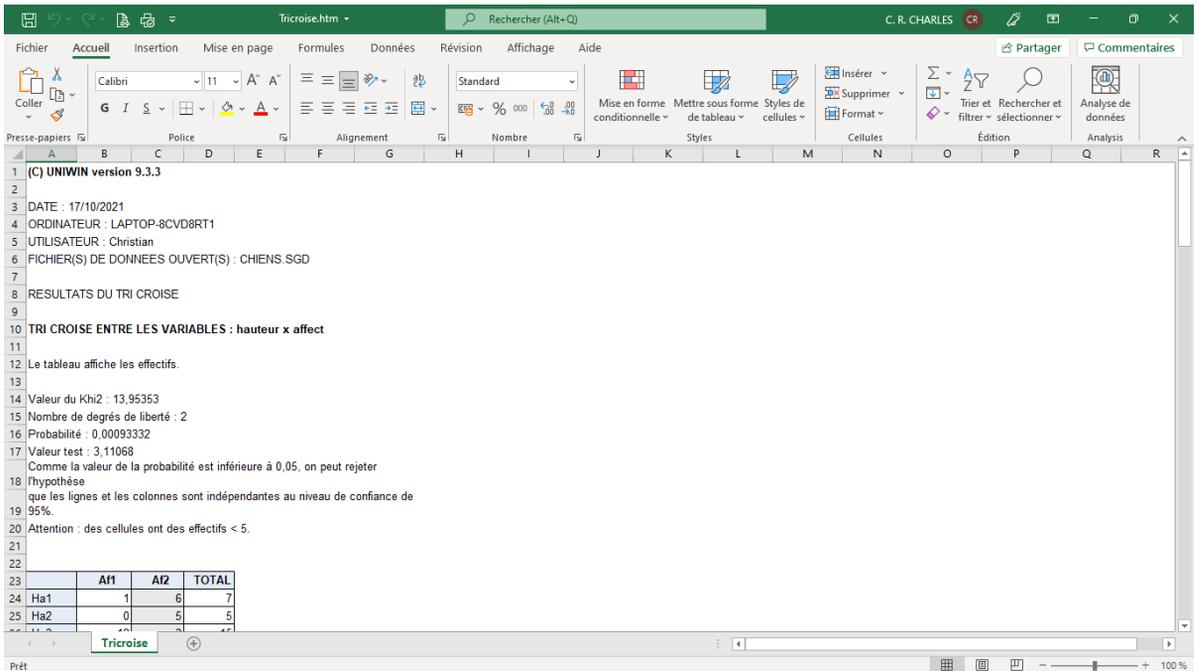


Quatre tris croisés seront effectués.

Cliquons sur Ok. La fenêtre suivante s'affiche :



Le rapport peut être affiché sous trois formats : Explorateur, Général et HTML. Dans ce dernier cas, un fichier au format HTML contenant l'ensemble des résultats est créé. Ce fichier peut être ouvert dans d'autres applications comme Microsoft Excel :



Les données des tableaux sont placées dans des cellules Excel et peuvent donc être utilisées pour effectuer des calculs ou tracer des graphiques sous Microsoft Excel.

L'explorateur à gauche dans cette fenêtre permet d'afficher chaque tri croisé.

L'icône  affiche la boîte de dialogue permettant de sélectionner les résultats à enregistrer :

Enregistrement des résultats (1/1)

Enregistrer

- Comptages
- Modalités lignes
- Modalités colonnes
- Tableau des khi2
- Tableau des valeurs test
- Variables lignes
- Variables colonnes

Noms attribués aux variables cibles

comptages\_1

libmodlig

libmodcol

khi2\_1

vt\_1

varlig

varcol

Ok Plus Tout Annuler

Données							
	comptages_1	comptages_2	comptages_3	comptages_4	comptages_5	libmodlig	libmodcol
	Comptages	Comptages	Comptages	Comptages	Comptages	Modalités lignes	Modalités colonnes
	Type = N	Type = C	Type = C				
	Longueur = 6	Longueur = 5					
1	1	6	3	3	1	Ha1	Af1
2	0	5	0	4	1	Ha2	Af2
3	12	3	5	6	4	Ha3	In1
4	1	7	3	4	1	Po1	In2
5	7	7	3	7	4	Po2	In3
6	5	0	2	2	1	Po3	
7							

Le tableau 'comptages' enregistré est constitué des juxtapositions des quatre tris croisés.

## 7.4 Analyse en composantes principales



ACP

### Fonction

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 une représentation simultanée (Biplot). La possibilité d'analyse d'individus et de variables supplémentaires (quantitatives ou qualitatives) est offerte. 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.

### Tableaux

Moyennes, écarts-types et coefficients de variation
Corrélations
Tableau des inerties
Tests de factorisation
Individus de base
Variables de base
Individus supplémentaires
Variables supplémentaires
Corrélations variables supplémentaires
Modalités variables qualitatives

## Graphiques

Diagramme des inerties
Cercle factoriel - Variables de base (points)
Cercle factoriel - Variables de base (points + lignes)
Cercle factoriel - Variables de base et supplémentaires (points)
Cercle factoriel - Variables de base et supplémentaires (points + lignes)
Plan factoriel - Individus de base
Plan factoriel - Individus de base et supplémentaires
Plan factoriel - Variables qualitatives
Biplot
Biplot (corrélations)
Contributions
Cosinus carrés
Cosinus carrés cumulés
Distances carrées à l'origine
T2 de Hotelling - Individus de base
T2 de Hotelling - Individus de base et supplémentaires

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse en composantes principales.pdf](#)

## 7.5 Analyse en composantes principales symbolique



### Fonction

Les méthodes classiques d'analyse factorielle ne sont applicables qu'à des objets caractérisés par des variables monovaluées (la valeur prise par une variable pour un objet est une valeur unique). L'ACPS est une extension de l'ACP à des objets caractérisés par des variables multivaluées décrivant de la variation ou de l'imprécision (la valeur prise par une variable pour un objet est un intervalle de valeurs).

La procédure ACPS proposée (méthode des centres) accepte deux types de données en entrée :

1. Un tableau de données individus x variables qui sera transformé en un tableau de données (concepts) contenant les intervalles des objets symboliques par utilisation de variables qualitatives.
2. Un tableau de données (concepts) contenant les intervalles des objets symboliques.

### Tableaux

Tableau des intervalles des concepts
Tableau des centres des concepts
Tableau des inerties du nuage des centres des concepts
Composantes principales (centres)
Composantes principales (approche symbolique)
Coordonnées des variables (centres)
Coordonnées des variables (approche symbolique)
Cosinus carrés et cosinus carrés cumulés des variables (centres)
Contribution des variables à l'inertie (centres)
Distances carrées des concepts à l'origine (centres)
Cosinus carrés et cosinus carrés cumulés des concepts (centres)
Contributions des concepts à l'inertie (centres)
Contributions des concepts à l'inertie totale (centres)

## Graphiques

Diagramme des inerties
Cercle des corrélations (approche classique)
Plan factoriel des variables (approche symbolique)
Plan factoriel des concepts (approche
Plan factoriel des concepts (approche symbolique)

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse en composantes principales symbolique.pdf](#)

## 7.6 Analyse NIPALS



NIPALS

### Fonction

L'analyse NIPALS (Nonlinear Iterative Partial Least Squares) est une méthode d'analyse en composantes principales autorisant la présence de données manquantes dans les données. L'algorithme estime les données manquantes.

### Tableaux

Valeurs propres
R-carrés (R <sup>2</sup> X)
Poids des variables
Corrélations des variables
Scores des observations
Contributions
Cosinus carrés
Cosinus carrés cumulés
Distances carrées à l'origine
Distances normalisées au modèle (DModX)
Données reconstituées
Ecart

### Graphiques

Diagramme des inerties
Graphique des poids des variables
Cercles des corrélations
Graphique des scores des observations
Graphique des contributions
Graphique des cosinus carrés cumulés
Graphique des distances carrées à l'origine
Graphique des distances normalisées au modèle (DModX)

Localisation dans les menus

Décrire - Analyses factorielles

Documentation

[Analyse NIPALS.pdf](#)

## 7.7 Analyse factorielle des correspondances



AFC

### Fonction

La méthode d'Analyse Factorielle des Correspondances Simples ou Binaires (AFC) permet de représenter graphiquement un tableau de contingence créé par le ou les croisements (tris croisés) de deux ou plusieurs variables qualitatives.

Voir la rubrique '[Tri croisé](#)' pour plus d'informations sur la façon de construire ce tableau.

La méthode vise à rassembler sur un ou plusieurs graphiques (plan factoriel) la plus grande partie possible de l'information contenue dans le tableau en s'attachant non pas aux valeurs absolues mais aux correspondances entre les caractéristiques, c'est-à-dire aux valeurs relatives.

La possibilité d'analyse de points lignes ou colonnes supplémentaires est offerte.

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 l'ensemble des graphiques des plans factoriels lignes, colonnes, simultanés, barycentriques et de ceux relatifs aux lignes et colonnes supplémentaires.

De nombreux outils d'aide à l'interprétation sont fournis, notamment contributions et cosinus carrés.

### Tableaux

Tableau des inerties
Probabilités conditionnelles lignes
Probabilités conditionnelles colonnes
Axes factoriels lignes
Axes factoriels colonnes
Résultats lignes
Résultats colonnes

Colonnes supplémentaires
--------------------------

Lignes supplémentaires
------------------------

## Graphiques

Diagramme des inerties
------------------------

Plan de base - Représentation des lignes
--

Plan de base - Représentation des colonnes
--

Plan de base - Représentation simultanée
--

Plan de base - Représentation barycentrique en lignes
---

Plan de base - Représentation barycentrique en colonnes
---

Plan supplémentaire - Représentation des lignes
---

Plan supplémentaire - Représentation des colonnes
---

Plan supplémentaire - Représentation simultanée
---

Contributions - lignes
------------------------

Contributions - colonnes
--------------------------

Cosinus carrés - lignes
-------------------------

Cosinus carrés - colonnes
---------------------------

Cosinus carrés cumulés - lignes
---------------------------------

Cosinus carrés cumulés - colonnes
-----------------------------------

Distances carrées à l'origine - lignes
--

Distances carrées à l'origine - colonnes
--

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse factorielle des correspondances.pdf](#)

## 7.8 Analyse des correspondances multiples



ACM

### Fonction

L'Analyse des Correspondances Multiples (ACM) permet d'analyser un tableau individus x variables lorsque les variables sont qualitatives. Cette méthode effectue une analyse des correspondances sur le tableau disjonctif complet obtenu en remplaçant dans le tableau d'origine chaque variable qualitative par l'ensemble des variables indicatrices des différentes modalités de cette variable. La possibilité d'analyse d'individus ou de variables supplémentaires est offerte. La transformation du tableau des données initiales en tableau disjonctif complet est automatiquement effectuée.

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 l'ensemble des graphiques des plans factoriels individus, variables, simultanés, barycentriques et de ceux relatifs aux individus et variables supplémentaires. De nombreux outils d'aide à l'interprétation sont fournis, notamment contributions, cosinus carrés, distances à l'origine, valeurs test.

### Tableaux

Tableau de contingence (Burt)
Tableau de contingence des colonnes supplémentaires
Tableau de contingence des individus supplémentaires
Tableau des inerties et tableau des inerties corrigées
Résultats colonnes
Résultats individus
Contributions totales des colonnes
Valeurs tests des colonnes supplémentaires
Coordonnées des colonnes supplémentaires
Coordonnées des individus supplémentaires

Représentation des individus de base et supplémentaires
Points multiples des individus de base et supplémentaires
Distances à l'origine des individus supplémentaires
Représentation des colonnes de base et supplémentaires
Distances à l'origine des colonnes supplémentaires

## Graphiques

Diagramme des inerties
Individus de base
Colonnes de base
Individus et colonnes de base
Représentation barycentrique des individus de base
Représentation barycentrique des colonnes de base
Individus supplémentaires
Colonnes de base et supplémentaires
Individus et colonnes supplémentaires
Contributions des individus
Contributions des colonnes
Cosinus carrés des individus
Cosinus carrés des colonnes
Cosinus carrés cumulés des individus
Cosinus carrés cumulés des colonnes
Distances carrées à l'origine des individus
Distances carrées à l'origine des colonnes
Valeurs test des colonnes

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse des correspondances multiples.pdf](#)

## 7.9 Analyse sur tableau de distances ou de dissimilarités



ATD

### Fonction

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 rubrique '[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.

### Tableaux

Distances carrées observées
Matrice initiale à diagonaliser
Matrice modifiée à diagonaliser
Tableau initial des inerties
Tableau modifié des inerties
Vecteurs propres

Facteurs colonnes
-------------------

Distances carrées calculées
-----------------------------

Différences avec les distances initiales
--

## Graphiques

Diagramme des inerties
------------------------

Plan factoriel des individus
------------------------------

Distances observées et calculées
----------------------------------

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse sur tableau de distances ou de dissimilarités.pdf](#)

## 7.10 Analyse factorielle de données mixtes



AFDM

### Fonction

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

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

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

Un rapport général de synthèse est proposé ainsi que les graphiques des cercles factoriels, des représentations simultanées des variables quantitatives et qualitatives, des plans factoriels des individus et des modalités.

De nombreux outils d'aide à l'interprétation sont fournis, notamment pondérations des indicatrices des variables qualitatives, liaisons entre les variables initiales, contributions des variables à l'inertie et pour les plans factoriels les contributions et cosinus carrés.

### Tableaux

Moyennes, écarts-types et coefficients de variation des variables quantitatives
Pondérations des indicatrices des variables qualitatives
Liaisons entre les variables initiales
Tableau des inerties
Contributions des variables à l'inertie
Facteurs propres
Résultats individus
Résultats variables quantitatives

## Résultats variables qualitatives

### Graphiques

Diagramme des inerties

Cercle factoriel - Variables quantitatives (points)

Cercle factoriel - Variables quantitatives

Représentation simultanée des variables

Plan factoriel des individus

Plan factoriel des individus et des modalités

Contributions

Cosinus carrés

Cosinus carrés cumulés

Distances carrées à l'origine

### Localisation dans les menus

Décrire - Analyses factorielles

### Documentation

[Analyse factorielle de données mixtes.pdf](#)

## 7.11 Analyse en composantes indépendantes



ICA

### Fonction

L'Analyse en Composantes Indépendantes (Independent Component Analysis - ICA) est une méthode d'analyse des données qui relève des statistiques, des réseaux de neurones et du traitement du signal. Elle est notoirement et historiquement connue en tant que méthode de séparation aveugle de signaux mélangés dans des proportions inconnues et est aujourd'hui appliquée à de nombreux problèmes, notamment en chimimétrie pour l'analyse de spectres (Wikipédia).

Cette procédure est basée sur les packages R 'ica' et 'JADE'.

Les méthodes 'JADE', 'fastICA', 'Infomax' et 'SOBI' peuvent être mises en œuvre.

### Tableaux

Données initiales
Données centrées
Matrice de rotation initiale
Matrice de mélange
Matrice de démixage
Composantes principales
Matrice de blanchiment
Matrice de rotation
Composantes indépendantes
Variances des composantes
Résidus
Coefficients de Durbin-Watson

## Graphiques

Graphique des données initiales
Graphique d'une composante principale
Graphique d'une composante indépendante
Histogramme des données initiales
Histogramme des données estimées

## Localisation dans les menus

Décrire - Analyses factorielles

## Documentation

[Analyse en composantes indépendantes.pdf](#)

## 7.12 Classification ascendante hiérarchique



### Fonction

La méthode de Classification Ascendante Hiérarchique (CAH) permet de construire une typologie (ou partition) d'un ensemble d'individus en classes telles que les individus appartenant à une même classe sont proches alors que les individus appartenant à des classes différentes sont éloignés.

La méthode utilisée est celle de Ward, méthode dite des voisins réciproques. C'est une méthode ascendante partant d'un nombre de classes égal au nombre des individus et regroupant à chaque itération un (ou plusieurs) individus ou classes en fonction de la diminution de l'inertie inter-classes.

L'algorithme s'applique à des données quantitatives individus x variables (avec standardisation ou non), à un tableau de contingence ou à un tableau de variables qualitatives. Dans le cas d'un tableau de contingence, la classification peut se faire suivant les lignes ou les colonnes.

Un tableau résumé de la classification et le tracé de l'arbre de classification (avec zoom et troncature) et de nuages codifiés sont proposés ainsi que des outils d'analyse de la partition et des nœuds.

Cette méthode peut notamment s'associer à une [Analyse en Composantes Principales](#) (ACP) et à une [Analyse des Correspondances Multiples](#) (ACM).

### Tableaux

Agrégations obtenues et restantes
Définition des nœuds de la hiérarchie
Troncature - Table d'appartenance
Statistiques pour les classes
Contributions des variables quantitatives aux classes
Contributions des classes aux variables quantitatives

## Graphiques

Diagramme des distances

Diagramme des indices de niveau

Nuages codifiés

Dendrogramme vertical (distances)

Dendrogramme horizontal (distances)

Dendrogramme vertical (indices)

Dendrogramme horizontal (indices)

## Localisation dans les menus

Décrire - Classifications

## Documentation

[Classification ascendante hiérarchique.pdf](#)

## 7.13 Classification par cartes auto-organisatrices



SOM

### Fonction

Les cartes auto-adaptatives, cartes auto-organisatrices ou cartes topologiques forment une classe de réseau de neurones artificiels fondée sur des méthodes d'apprentissage non-supervisées. Elles sont souvent désignées par le terme anglais « self organizing maps » (SOM), ou encore cartes de Kohonen du nom du statisticien ayant développé le concept en 1984. La littérature utilise aussi les dénominations : « réseau de Kohonen », « réseau auto-adaptatif » ou « réseau auto-organisé ». Elles sont utilisées pour cartographier un espace réel, c'est-à-dire pour étudier la répartition de données dans un espace à grande dimension. En pratique, cette cartographie peut servir à réaliser des tâches de discrétisation, quantification vectorielle ou classification (source Wikipédia).

Cette procédure est basée sur le package R 'kohonen'.

### Tableaux

Effectifs dans les nœuds
Affectations des nœuds
Affectations des observations
Profils des nœuds

### Graphiques

Graphique de la progression de l'apprentissage
Graphique des effectifs dans les nœuds
Graphique des observations dans les nœuds
Graphique des distances entre les nœuds
Graphique de la qualité des nœuds
Graphique des profils des nœuds
Carte thermique individuelle
Ensemble des cartes thermiques

## Classification ascendante hiérarchique

Localisation dans les menus

Décrire - Classifications

Documentation

[Classification par cartes auto-organisatrices.pdf](#)

## 7.14 Ellipses de tolérance ou de confiance



ELL

### Fonction

Les Ellipses de Tolérance ou de Confiance (ELL) permettent de définir des régions autour de nuages de points décrivant les différentes catégories de variables supplémentaires dans un plan d'une analyse factorielle.

Plus généralement, elles permettent de définir les régions de confiance autour des centres de gravité de groupes définis par différentes catégories.

Ellipses de tolérance ou de confiance sont proposées. Le choix du niveau de risque alpha est possible et pour les ellipses de confiance, les calculs peuvent se faire sous l'hypothèse d'un échantillon de taille finie ou infinie.

La variable de liaison entre les points permet de définir comment les points seront reliés pour créer des courbes.

Il est possible ou non d'afficher les symboles des points lorsque le tracé est sous la forme de courbes.

Des graphiques permettent de visualiser les différentes ellipses, les centres de gravité, les codes des catégories et les numéros des points.

Un rapport général de synthèse est proposé.

Ces ellipses peuvent notamment s'utiliser après une [Analyse en Composantes Principales](#) (ACP).

### Tableaux

Statistiques descriptives des groupes
---------------------------------------

## Graphiques

Ellipses de tolérance ou de confiance

### Localisation dans les menus

Décrire – Ellipses

### Documentation

[Ellipses de confiance ou de tolérance.pdf](#)

## 7.15 Analyse factorielle multiple



AFM

### Fonction

L'Analyse Factorielle Multiple (AFM) est spécialement conçue pour étudier une population d'individus caractérisée par un certain nombre de groupes de variables.

Ces groupes de variables peuvent être constitués de variables mesurées à différents instants, mais aussi de sous-tableaux issus d'un unique tableau : ces sous-tableaux correspondent alors à des regroupements de variables selon des critères.

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

Un rapport général de synthèse est proposé ainsi que les graphiques des cercles factoriels, des plans factoriels du compromis, de l'interstructure, de l'intrastructure, des trajectoires des individus.

De nombreux outils d'aide à l'interprétation sont fournis, notamment coefficients de liaison entre les groupes, comparaison des projections du compromis et des groupes, contributions, cosinus carrés, distances à l'origine pour les individus et les groupes.

### Tableaux

Moyennes et écarts-types des variables
Inerties des groupes
Inerties du compromis
Décomposition de l'inertie
Coefficients Lg
Coefficients RV
Rapport inertie inter / inertie totale
Résultats individus compromis
Résultats variables compromis

Résultats individus / groupes
-------------------------------

Résultats groupes de variables
--------------------------------

## Graphiques

Diagramme des inerties
------------------------

Compromis - Corrélations des variables (points)
---

Compromis - Corrélations des variables (points + lignes)
--

Compromis - Plan des individus
--------------------------------

Interstructure - Groupes de variables
---------------------------------------

Intrastructure - Plan des individus
-------------------------------------

## Localisation dans les menus

Décrire - Tableaux multiples

## Documentation

[Analyse factorielle multiple.pdf](#)

## 7.16 Méthode STATIS



STATIS

### Fonction

La méthode Statis (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 individus, les variables pouvant être éventuellement différentes selon les tableaux.

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

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 (individus), de l'interstructure (groupes des variables) et de l'intrastructure (individus et trajectoires des individus).

De nombreux outils d'aide à l'interprétation sont fournis, notamment normes et traces des groupes, coefficients RV, distances entre les groupes, décompositions de ces distances par individu, pondérations des groupes, résultats pour les groupes et le compromis (distances à l'origine, contributions, cosinus carrés) ainsi que pour les individus pour les divers groupes.

### Tableaux

Moyennes et écarts-types des variables
Inerties des groupes
Normes et traces des groupes
Coefficients RV

Distances entre les groupes
Décomposition des distances
Inerties de l'interstructure
Pondérations des groupes
Produits scalaires et distances compromis-groupes
Décomposition de l'inertie selon les groupes
Résultats groupes
Inerties du compromis
Résultats compromis-lignes
Décomposition de l'inertie selon les lignes
Corrélations variables - axes du compromis
Résultats individus - groupes

### Graphiques

Diagramme des inerties du compromis
Compromis - Plan des individus
Compromis - Corrélations des variables
Interstructure - Groupes de variables
Intrastructure - Plan des individus
Intrastructure - Trajectoires des individus

### Localisation dans les menus

Décrire - Tableaux multiples

### Documentation

[Méthode STATIS.pdf](#)

## 7.17 Méthode STATIS duale



STATIS

### Fonction

La méthode Statist 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.

### Tableaux

Moyennes et écarts-types des variables
--

Matrices des corrélations
---------------------------

Inerties des groupes
Normes des groupes
Produits scalaires entre matrices des corrélations
Produits scalaires normés entre matrices des corrélations
Distances entre matrices des corrélations
Décomposition de la distance entre les matrices des corrélations
Coefficients de pondération des matrices des corrélations
Matrice des corrélations du compromis
Produits scalaires et distances entre matrice compromis et matrices des corrélations
Décomposition de l'inertie selon les matrices des corrélations
Inerties de l'interstructure
Résultats groupes
Inerties du compromis
Résultats compromis
Résultats variables -groupes

### Graphiques

Diagramme des inerties du compromis
Compromis - Plan des variables
Interstructure - Groupes de variables
Intrastructure - Plan des variables
Intrastructure - Trajectoires des variables

### Localisation dans les menus

Décrire - Tableaux multiples

### Documentation

[Méthode STATIS duale.pdf](#)

## 7.18 Analyse procrustéenne généralisée



### Fonction

L'analyse procrustéenne généralisée (GPA) est une technique mathématique permettant de comparer des formes. Elle est utilisée pour déformer un objet afin de le rendre autant que faire se peut semblable à une référence, ne laissant apparaître entre l'objet et la référence que les différences que les transformations (rotation, translation et mise à l'échelle) n'ont pu gommer. La déformation supprime les différences qui ne sont pas dues à la forme intrinsèque de l'objet. Celles qui subsistent sont donc considérées objectives, et permettent d'évaluer le degré de ressemblance entre l'objet et la référence.

Cette technique a été nommée ainsi à partir de Procruste, un bandit de la mythologie grecque qui forçait ses victimes à s'allonger sur un lit et modifiait violemment leur taille pour que celle-ci corresponde à la taille du lit.

L'analyse procrustéenne généralisée est utilisée dans tous les champs où l'analyse de formes peut être utile : analyse sensorielle, biologie, archéologie, médecine.

Cette procédure est basée sur le package R 'shapes'.

### Tableaux

Distances initiales entre les configurations
Distances finales entre les configurations
Configurations finales (observations)
Configurations finales (configurations)
Consensus
Distances au consensus
Inerties ACP (observations)
Inerties ACP (configurations)
Corrélations des variables
Scores du consensus
Scores des observations x configurations

Scores des configurations

Graphiques

Distances des configurations au consensus

Corrélations des variables x configurations (points)

Corrélations des variables x configurations (points + lignes)

Plan factoriel du consensus des observations

Plan factoriel des observations x configurations

Plan factoriel des configurations

Localisation dans les menus

Décrire - Tableaux multiples

Documentation

[Analyse procrustéenne généralisée.pdf](#)

## 7.19 Analyse APRIORI



### Fonction

L'algorithme APRIORI est un algorithme d'exploration de données conçu en 1994 par Rakesh Agrawal et Ramakrishnan Srikant dans le domaine de l'apprentissage des règles d'association.

Il sert à reconnaître des propriétés qui reviennent fréquemment dans un ensemble de données et à détecter les relations ou les associations entre des modalités spécifiques de variables catégorielles dans des fichiers de données volumineux.

Cette technique exploratoire puissante possède toute une gamme d'applications dans de nombreux domaines : analyse des tickets de caisse, campagne marketing, analyse de log Web, détection de fraudeurs, analyse de défaillances de process.

Cette procédure est basée sur le package R 'arules'.

### Tableaux

Libellés des modalités
Motifs fréquents
Motifs fréquents fermés
Motifs fréquents maximaux
Détails des règles
Règles pour la colonne cible
Occurrences et fréquences des modalités de la colonne cible

### Graphiques

Graphique du lift vs conséquent x antécédent
Graphique de la confiance vs conséquent x antécédent
Graphique du support vs conséquent x antécédent
Graphique du lift par rapport au support
Graphique du lift par rapport à la confiance

Graphique du lift par rapport au support pour la colonne cible

Graphique du lift par rapport à la confiance pour la colonne cible

Localisation dans les menus

Décrire - Associations

Documentation

[Analyse APRIORI.pdf](#)

## Le ruban Expliquer

Le ruban 'Expliquer' permet de mettre en œuvre une analyse discriminante pas à pas, une analyse factorielle discriminante, une analyse discriminante bayésienne, une analyse discriminante qualitative, une méthode de scoring, la méthode de classement des k plus proches voisins, la méthode de classement bayésienne naïve, la méthode de classement SIMCA, la régression sur composantes principales, la régression PLS, l'analyse par réseaux de neurones artificiels et les arbres de décision et de régression.



## 8.1 Analyse discriminante pas à pas



DISP

### Fonction

L'Analyse Discriminante Pas à Pas (DISP) permet de sélectionner à partir d'un ensemble de variables quantitatives et d'une variable qualitative découpant la population en plusieurs groupes (2 ou plus), le sous-ensemble des variables quantitatives les plus explicatives des groupes qui seront alors utilisées pour définir des fonctions discriminantes robustes.

La méthode proposée permet de préciser les valeurs des Fishers pour l'entrée et la sortie d'une variable, les critères d'arrêt de l'algorithme (variation minimale du lambda de Wilks si plus de 2 groupes ou de la distance de Mahalanobis si 2 groupes, nombre maximal de variables à sélectionner, nombre maximal d'itérations).

La méthode itérative peut être "manuelle" (affiche toutes les étapes) ou "automatique" (affiche uniquement le tableau final).

L'algorithme est formé d'une succession d'étapes ascendantes (entrée d'une variable) et descendantes (sortie d'une variable).

Un rapport général de synthèse est proposé : il permet de visualiser toutes les étapes de l'algorithme.

Cette méthode peut être complétée par une [Analyse Discriminante Bayésienne](#) (ADB) ou par une [Analyse Factorielle Discriminante](#) (AFD).

## Tableaux

Tableaux des itérations ascendantes et descendantes

Tableau de synthèse trié par valeurs décroissantes des Fishers

## Localisation dans les menus

Expliquer - Analyses discriminantes

## Documentation

[Analyse discriminante pas à pas.pdf](#)

## 8.2 Analyse factorielle discriminante



AFD

### Fonction

L'Analyse Factorielle Discriminante (AFD) est une méthode géométrique permettant de construire à partir d'un ensemble de variables quantitatives et d'une variable qualitative découpant la population en plusieurs groupes (2 ou plus) des fonctions discriminantes qui les séparent au mieux dans l'échantillon d'apprentissage.

En fonction des données et des paramètres définis par l'utilisateur, l'analyse AFD réalise automatiquement les études de la population d'apprentissage et des éventuelles populations de validation et de prévision. De façon plus précise, la méthode peut se décomposer en trois étapes. Supposons une population de  $n$  individus. Découpons cette population en trois sous-populations de tailles  $n_1$ ,  $n_2$  et  $n_3$  avec  $n_1 + n_2 + n_3 = n$ . Les trois étapes sont :

- une étude initiale sur la population d'apprentissage de taille  $n_1$
- une étude de validation sur la population de validation de taille  $n_2$
- une étude prospective sur une population de prévision de taille  $n_3$

Des tableaux résumés et détaillés des classements sont calculés. Le tracé de plans factoriels et un rapport général de synthèse sont proposés.

### Tableaux

Centroides des groupes et global
Matrice des covariances globales
Matrice des corrélations globales
Covariances par groupe
Corrélations par groupe
Covariances intra-groupes
Corrélations intra-groupes
Distances de Mahalanobis, valeurs des Fishers et tests de signification
Tableau des inerties
Critères statistiques (Hotelling, Pillai, Roy)

Test de Box
Fonctions discriminantes standardisées
Fonctions discriminantes non standardisées
Résultats variables
Résultats individus
Coordonnées des centres des groupes
Classement apprentissage
Détails du classement apprentissage
VP, FN, FP, VN, sensibilité, spécificité apprentissage
Statistiques pour les groupes observés
Statistiques pour les groupes prévus
Classement validation
Détails du classement validation
VP, FN, FP, VN, sensibilité, spécificité validation
Classement prévision

## Graphiques

Diagramme des inerties
Cercle factoriel des corrélations des variables (points)
Cercle factoriel des corrélations des variables (points + lignes)
Plan factoriel du jeu d'apprentissage
Plan factoriel du jeu de validation
Plan factoriel du jeu de prévision
Plan factoriel des jeux d'apprentissage et de validation
Plan factoriel des jeux d'apprentissage et de prévision
Courbes ROC pour les jeux d'apprentissage et de validation

## Localisation dans les menus

Expliquer - Analyses discriminantes

## Documentation

[Analyse factorielle discriminante.pdf](#)

## 8.3 Analyse discriminante bayésienne



ADB

### Fonction

L'Analyse Discriminante Bayésienne (ADB) permet de construire à partir d'un ensemble de variables quantitatives et d'une variable qualitative découpant la population en plusieurs groupes (2 ou plus), des fonctions discriminantes qui définissent une règle de décision optimale à partir de laquelle on peut affecter des individus de validation et de prévision aux différents groupes.

Cette technique suppose que l'on connaisse a priori les probabilités d'appartenance aux différents groupes et que les données suivent une loi multi-normale.

La méthode proposée permet de traiter les cas linéaire (égalité des matrices de variances) et quadratique (non-égalité des matrices de variances).

L'entrée des probabilités a priori est proposée. Par défaut, le système utilise les probabilités issues des fréquences des groupes dans les données entrées.

En fonction des données et des paramètres définis par l'utilisateur, l'analyse ADB réalise automatiquement les études de la population d'apprentissage et des éventuelles populations de validation et de prévision.

De façon plus précise, la méthode peut se décomposer en trois étapes. Supposons une population de  $n$  individus. Découpons cette population en trois sous-populations de tailles  $n_1$ ,  $n_2$  et  $n_3$  avec  $n_1 + n_2 + n_3 = n$ . Les trois étapes sont :

- une étude initiale sur la population d'apprentissage de taille  $n_1$
- une étude de validation sur la population de validation de taille  $n_2$
- une étude prospective sur une population de prévision de taille  $n_3$

Des tableaux résumés et détaillés des classements sont calculés.

Le tracé de plans factoriels et un rapport général de synthèse sont proposés.

## Tableaux

Probabilités initiales
Centroïdes des groupes et global
Matrice des covariances globales
Matrice des corrélations globales
Covariances par groupe
Corrélations par groupe
Covariances intra-groupes
Corrélations intra-groupes
Distances de Mahalanobis, valeurs des Fishers et tests de signification
Tableau des inerties
Critères statistiques (Hotelling, Pillay, Roy)
Test de Box
Fonctions discriminantes standardisées
Fonctions discriminantes non standardisées
Résultats variables
Résultats individus
Coordonnées des centres des groupes
Coefficients des fonctions de classement (linéaires ou quadratiques)
Classement apprentissage
Détails du classement apprentissage
VP, FN, FP, VN, sensibilité, spécificité apprentissage
Statistiques des groupes observés
Statistiques des groupes prévus
Classement validation
Détails du classement validation
VP, FN, FP, VN, sensibilité, spécificité validation
Classement prévision

## Graphiques

Diagramme des inerties
Cercle factoriel des corrélations des variables (points)
Cercle factoriel des corrélations des variables (points + lignes)
Plan factoriel du jeu d'apprentissage
Plan factoriel du jeu de validation
Plan factoriel du jeu de prévision
Plan factoriel des jeux d'apprentissage et de validation
Plan factoriel des jeux d'apprentissage et de prévision
Courbe ROC pour le jeu d'apprentissage
Courbe ROC pour le jeu de validation

## Localisation dans les menus

Expliquer - Analyses discriminantes

## Documentation

[Analyse discriminante bayésienne.pdf](#)

## 8.4 Analyse discriminante qualitative



ADQ

### Fonction

L'Analyse Discriminante Qualitative (ADQ) est une généralisation de l'Analyse Factorielle Discriminante (AFD) dans le cas où les variables explicatives sont qualitatives et non plus quantitatives.

La première étape de l'analyse consiste à mettre en œuvre une Analyse des Correspondances Multiples (ACM) des variables qualitatives.

La deuxième étape remplace les variables qualitatives d'origine par les coordonnées sur les axes factoriels issus de l'ACM et effectue sur ces données une Analyse Factorielle Discriminante (AFD).

Les fonctions discriminantes sont ensuite exprimées en fonction des indicatrices des modalités des variables qualitatives d'origine.

En fonction des données et des paramètres définis par l'utilisateur, l'analyse ADQ réalise automatiquement les études de la population d'apprentissage et des éventuelles populations de validation et de prévision.

De façon plus précise, la méthode peut se décomposer en trois étapes. Supposons une population de  $n$  individus. Découpons cette population en trois sous-populations de tailles  $n_1$ ,  $n_2$  et  $n_3$  avec  $n_1 + n_2 + n_3 = n$ . Les trois étapes sont :

- une étude initiale sur la population d'apprentissage de taille  $n_1$
- une étude de validation sur la population de validation de taille  $n_2$
- une étude prospective sur une population de prévision de taille  $n_3$

Des tableaux résumés et détaillés des classements sont calculés. Le tracé de plans factoriels et un rapport général de synthèse sont proposés.

## Tableaux

Tableau des inerties de l'ACM
Centroïdes des groupes sur les composantes de l'ACM
Distances de Mahalanobis, valeurs des Fishers et tests de signification
Tableau des inerties de l'ADQ
Critères statistiques (Hotelling, Pillai, Roy)
Test de Box
Fonctions discriminantes standardisées (ACM)
Fonctions discriminantes non standardisées (ACM)
Coordonnées des variables (ACM)
Fonctions discriminantes standardisées exprimées en fonction des variables qualitatives
Fonctions discriminantes non standardisées exprimées en fonction des variables qualitatives
Contributions des variables à l'inertie
Résultats individus
Coordonnées des centres des groupes
Coordonnées des modalités des variables
Classement apprentissage
Détails du classement apprentissage
VP, FN, FP, VN, sensibilité, spécificité apprentissage
Classement validation
Détails du classement validation
VP, FN, FP, VN, sensibilité, spécificité validation
Détails du classement prévision
Classement prévision

## Graphiques

Diagramme des inerties
Plan factoriel des variables qualitatives
Plan factoriel du jeu d'apprentissage
Plan factoriel du jeu d'apprentissage + modalités des variables
Plan factoriel du jeu de validation
Plan factoriel du jeu de prévision

Plans factoriels des jeux d'apprentissage et de validation
--

Plans factoriels des jeux d'apprentissage et de prévision
---

Courbes ROC pour les jeux d'apprentissage et de validation
--

Localisation dans les menus

Expliquer - Analyses discriminantes

Documentation

[Analyse discriminante qualitative.pdf](#)

## 8.5 Scoring



Scoring

### Fonction

Le scoring est une technique permettant de classer des individus caractérisés par des variables descriptives qualitatives dans deux catégories préexistantes en leur affectant un score d'appartenance.

La première étape de cette technique consiste à transformer les variables descriptives qualitatives d'origine en coordonnées sur les axes factoriels issus d'une Analyse des Correspondances Multiples (ACM) des variables qualitatives. La deuxième étape est une Analyse Factorielle Discriminante (AFD) de ces données, avec une possible sélection des axes factoriels significatifs de l'ACM.

La fonction discriminante obtenue est ensuite exprimée en fonction des indicatrices des modalités des variables qualitatives d'origine et les coefficients de la fonction de score sont évalués.

L'analyse se poursuit en calculant les scores des individus, la sensibilité, la spécificité, le niveau de qualité ROC et l'indice de Gini pour la courbe de Lift.

Un rapport général de synthèse est proposé ainsi que des graphiques des coefficients de la fonction de score, des courbes de répartition (avec ou sans zone d'indécision), des courbes de densité, de la courbe ROC et de la courbe de Lift.

Des tableaux résumés et détaillés des classements de l'Analyse Factorielle Discriminante (AFD) et du Scoring sont calculés.

Une option Simulation permet de visualiser l'évolution du score d'un individu en fonction des modifications de ses modalités. L'étude d'une population de prévision est également proposée.

## Tableaux

Variables utilisées
Fonction discriminante (ACM)
Tableau des inerties AFD
Fonction discriminante (facteurs significatifs de l'ACM)
Fonction discriminante (modalités des variables)
Classement issu de l'AFD
Coefficients de la fonction de score
Scores des individus
Valeurs ROC
Valeurs Lift
Classement issu du scoring
Scores des individus (prévision)

## Graphiques

Histogramme des coefficients de la fonction de score
Courbes de répartition des scores des individus (3 zones)
Courbes de répartition des scores des individus (2 zones)
Courbes de répartition des scores des individus
Courbes de densité des scores des individus
Courbe ROC
Courbe de Lift

## Localisation dans les menus

Expliquer – Scoring

## Documentation

[Scoring.pdf](#)

## 8.6 Méthode KNN



### Fonction

La méthode des K plus proches voisins (KNN) a pour objectif de classer des observations dont les classes sont inconnues (échantillon de prévision) en fonction de leurs distances euclidiennes (calculées en utilisant les variables explicatives quantitatives précisées) à des observations dont les classes sont connues (échantillon d'apprentissage).

Une plage de valeurs de K est précisée.

La première étape de cette méthode consiste à classer les observations de l'échantillon d'apprentissage par validation croisée. Le taux d'erreur de classement est alors calculé pour chaque valeur de K et le K optimal est déterminé.

La seconde étape consiste à classer les données de l'échantillon de prévision en utilisant ce K optimal. Chaque nouvelle observation est affectée à la classe la plus fréquente de ses K voisins.

Cette procédure est basée sur le package R 'class'.

### Tableaux

Classement apprentissage (pour chaque valeur de K)
--

Synthèse apprentissage (pour chaque valeur de K)
--

VP, FN, FP, VN, sensibilité, spécificité (pour le K optimal)
--

Classement prévision (pour chaque valeur de K)
--

### Graphiques

Graphique de l'évolution du taux d'erreur
---

Graphique des données d'apprentissage
---------------------------------------

Graphique des données de prévision
------------------------------------

Courbe ROC (pour le K optimal)
--------------------------------

Localisation dans les menus

Expliquer – Classement

Documentation

[Méthode KNN.pdf](#)

## 8.7 Méthode SIMCA



### Fonction

La méthode SIMCA (Soft Independent Modeling of Class Analogy) est une technique de classement proposée par Svante Wold dans les années 1970. Cette méthode supervisée est basée sur l'analyse en composantes principales (ACP). Pour chaque classe, une ACP est réalisée en utilisant uniquement les observations de cette classe. Les différents modèles obtenus pour chacune des classes peuvent ainsi avoir des nombres de composantes différents. Ces modèles permettent de prévoir l'appartenance ou non d'une observation du jeu d'apprentissage ou de prévision à une classe prédéfinie mais également de déterminer si une observation appartient à plusieurs classes (recouvrement) ou à aucune classe (possible point aberrant ou nouvelle classe). Cette procédure est basée sur le package R 'mdatools'.

### Tableaux

Facteur de classement
Tableaux des inerties
Poids des variables
Scores des observations
Résumé du classement (apprentissage)
Détails du classement (apprentissage)
Matrice de confusion (apprentissage)
Distances entre les modèles
Pouvoirs discriminants des variables
Détails du classement (prévision)

### Graphiques

Diagramme des inerties
Poids des variables
Scores des observations
Distances entre les modèles
Pouvoirs discriminants des variables
Graphique de Cooman

Localisation dans les menus

Expliquer - Classement

Documentation

[Méthode SIMCA.pdf](#)

## 8.8 Méthode de classement bayésienne naïve



### Fonction

La méthode de classement bayésienne naïve est basée sur le théorème de Bayes avec l'hypothèse, dite naïve, d'indépendance conditionnelle entre toutes les paires de descripteurs par rapport aux valeurs de la variable à prédire. Elle met en œuvre un classifieur bayésien naïf appartenant à la famille des classifieurs linéaires. Un terme plus approprié pour le modèle probabiliste sous-jacent pourrait être modèle à caractéristiques statistiquement indépendantes. En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique pour une classe est indépendante de l'existence d'autres caractéristiques. Un fruit peut être considéré comme une pomme s'il est rouge, arrondi, et fait une dizaine de centimètres. Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que le fruit est une pomme en considérant indépendamment ses caractéristiques de couleur, de forme et de taille.

Malgré le modèle de conception naïf et les hypothèses de base extrêmement simplistes, les classifieurs bayésiens naïfs ont fait preuve d'une efficacité plus que suffisante dans beaucoup de situations réelles complexes. L'avantage du classifieur bayésien naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires au classement.

Cette procédure est basée sur le package R 'naivebayes'.

### Tableaux

Facteur de classement
Informations issues de R
Classement apprentissage
Synthèse du classement apprentissage
VP, FN, FP, VN, sensibilité, spécificité apprentissage
Classement validation
Synthèse du classement validation
VP, FN, FP, VN, sensibilité, spécificité validation
Classement prévision

## Graphiques

Graphiques des lois marginales

Graphiques des lois conditionnelles

Graphiques des données d'apprentissage

Graphiques des données de validation

Graphiques des données de prévision

Courbe ROC pour les données d'apprentissage

Courbe ROC pour les données de validation

## Localisation dans les menus

Expliquer – Classement

Documentation

[Méthode bayésienne naïve.pdf](#)

## 8.9 Régression sur composantes principales



RCP

### Fonction

La méthode de Régression sur Composantes Principales (RCP) est une technique de régression utile lorsque de fortes colinéarités entre les variables explicatives sont présentes et que l'on ne désire pas utiliser les algorithmes de régression pas à pas pour éliminer les variables corrélées entre elles ou les régressions Ridge ou PLS.

Cette technique utilise à la fois l'Analyse en Composantes Principales (ACP) et la Régression Multiple pour élaborer un modèle dont les coefficients sont stables.

Après l'affichage du tableau et de l'histogramme des inerties, un rapport général de synthèse est proposé contenant notamment les résultats de l'Analyse en Composantes Principales (ACP), les descriptions des différents modèles de régression et le tableau de l'analyse de la variance.

Les graphiques des cercles factoriels, des plans factoriels, des régressions, des composants et des résidus sont également disponibles.

### Tableaux

Moyennes et écart-types des variables explicatives et à expliquer
Matrice des corrélations des variables explicatives
Tableau des inerties
Vecteurs propres
Composantes principales
Corrélations des variables avec les composantes principales normées
Coefficients de régression sur les composantes principales
Tableau de l'analyse de la variance
Coefficients des régressions sur les variables explicatives centrées-réduites
Coefficients des régressions sur les variables explicatives
R <sup>2</sup> et R <sup>2</sup> ajustés

Résultats des modèles
Prévisions

## Graphiques

Diagramme des inerties
Graphique des R-carrés
Graphique des R-carrés ajustés
Cercle factoriel (points)
Cercle factoriel (points et droites)
Plan factoriel
Régression - observés / estimés
Régression - composants + résidus
Régression - résidus

## Localisation dans les menus

Expliquer - Régressions

## Documentation

[Régression sur composantes principales.pdf](#)

## 8.10 Régression PLS



### Fonction

La méthode Régression PLS (partial least squares ou moindres carrés partiels) est conçue pour ajuster un modèle statistique reliant un ensemble de variables explicatives  $X$  à une variable à expliquer  $Y$  (PLS1) ou à plusieurs variables à expliquer  $Y$  (PLS2). La procédure est utile lorsqu'il y a de nombreux  $X$  et que le but principal est de prévoir simultanément les variables  $Y$ . Elle est recommandée dans le cas où un grand nombre de variables  $X$  est utilisé, lorsqu'il y a de fortes colinéarités entre ces variables  $X$ , lorsque le nombre de variables  $X$  est supérieur au nombre d'observations et lorsqu'il y a des données manquantes. La méthode PLS est notamment largement utilisée par les chimistes et les chimiométriciens pour l'étalonnage en spectrométrie.

Un rapport général de synthèse est proposé contenant notamment les PRESS,  $R^2$ ,  $Q^2$ ,  $R^2X$ ,  $R^2Y$ , les poids  $w$  et  $w^*$ , les scores  $t$  et  $u$ , les poids des variables  $X$  et  $Y$ , les corrélations des variables avec les composantes, les valeurs observées, ajustées et résidus du modèle, les distances au modèle en  $X$  et  $Y$ , les  $T^2$  de Hotelling et les VIP.

Les graphiques des  $R^2X$ ,  $R^2Y$ ,  $Q^2$ , des coefficients standardisés, des cercles des corrélations, des plans factoriels, des poids des variables, des  $T^2$ , des distances au modèle en  $X$  et  $Y$ , des VIP, des valeurs observées vs estimées et des résidus sont également disponibles.

La procédure implémentée est basée sur le package R 'plsdepot'.

### Tableaux

PRESS, $R^2$ , $Q^2$ , $Q^2$ cumulés
$R^2X$ , $R^2Y$
Corrélations $X$ vs $t$ , $Y$ vs $t$ , $X$ vs $u$ , $Y$ vs $u$ (PLS2)
Coefficients de régression standardisés
Coefficients de régression non standardisés
Poids $w$
Poids $w^*$
Scores $X$ ( $t$ )
Scores $Y$ ( $u$ )

Poids des variables X et Y
Corrélations variables - composantes
Valeurs observées, estimées et résidus pour chaque Y (apprentissage, validation, prévision)
Distances au modèle en X
Distances au modèle en Y
T2 de Hotelling (PLS1)
VIP (PLS2)

### Graphiques

Diagramme des R2X cumulés
Diagramme des R2Y cumulés (PLS2)
Diagramme des Q2 cumulés
Diagramme des coefficients standardisés
Cercle des corrélations (points)
Cercle des corrélations (points + lignes)
Plan factoriel des observations (t vs t)
Plan factoriel des observations (u vs t)
Plan factoriel des observations (u vs u)
Diagramme des poids des variables X et Y
Graphique des T2 de Hotelling (PLS1)
Graphique des distances au modèle en X
Graphique des distances au modèle en Y
Graphique des VIP (PLS2)
Graphique observés vs estimés pour chaque Y (apprentissage, validation)
Graphique des résidus pour chaque Y (apprentissage, validation)

### Localisation dans les menus

Expliquer – Régressions

### Documentation

[Régression PLS.pdf](#)

## 8.11 Réseaux de neurones artificiels



ANN

### Fonction

Inspirés des neurones du cerveau, les réseaux de neurones artificiels n'ont rien de biologique. Ce sont des fonctions mathématiques à plusieurs paramètres. L'analogie date des premiers automates proposés en 1943 par Warren McCulloch et Walter Pitts. Comme dans les neurones du cerveau où des connexions se créent, disparaissent ou se renforcent en fonction de différents stimuli et produisent une action, les réseaux de neurones artificiels ajustent des paramètres (appelés poids synaptiques en référence au fonctionnement biologique du cerveau) en fonction de données d'entrée afin de fournir la meilleure réponse possible.

En apprentissage machine, un neurone fait une combinaison linéaire des entrées qu'il reçoit, à laquelle il ajoute une valeur appelée biais. Une fonction non linéaire, dite d'activation (logistique ou tangente hyperbolique) est alors appliquée à la valeur de sortie. Cette valeur est ensuite transmise à la couche de neurones suivante si le seuil est franchi.

Chaque neurone effectue ainsi un calcul très rudimentaire et c'est la succession des couches de neurones qui permet d'obtenir des réseaux complexes. Durant cette phase dite d'apprentissage, le réseau va ajuster automatiquement les paramètres de chaque neurone, c'est-à-dire les valeurs des poids et du biais afin de minimiser l'erreur moyenne calculée sur l'ensemble des observations entre la sortie attendue et celle observée. L'hypothèse est qu'après cette phase d'apprentissage, le réseau sera capable de traiter de manière satisfaisante de nouvelles observations, dont la sortie est inconnue, en fonction de ce qu'il a appris.

Dans un réseau de neurones à deux couches, la première couche est constituée d'un ensemble de neurones connectés en parallèle et fournissant un ensemble de sorties, elles-mêmes combinées pour devenir les entrées d'un nouvel ensemble de neurones formant une seconde couche.

Cette procédure est basée sur le package R 'neuralnet'.

## Tableaux

Résumé de l'analyse
Poids ajustés
Poids généralisés
Détails apprentissage
Matrice de confusion apprentissage (classement)
VP, FN, FP, VN, sensibilité, spécificité apprentissage
Détails validation et prévision
Matrice de confusion validation (classement)
VP, FN, FP, VN, sensibilité, spécificité validation

## Graphiques

Graphique du réseau
Graphique des poids généralisés
Courbe ROC (classement – apprentissage)
Courbe ROC (classement - validation)
Graphique valeurs prévues vs valeurs observées (régression - apprentissage)
Graphique résidus vs valeurs observées (régression - apprentissage)
Graphique résidus vs valeurs prévues (régression - apprentissage)
Graphique valeurs prévues vs valeurs observées (régression - validation)
Graphique résidus vs valeurs observées (régression - validation)
Graphique résidus vs valeurs prévues (régression - validation)

## Documentation

[Réseaux de neurones artificiels.pdf](#)

## 8.12 Arbres de décision et de régression



ARBRE

### Fonction

Les arbres de décision et de régression sont des méthodes permettant d'obtenir des modèles explicatifs et prédictifs. Ils sont faciles à comprendre du fait de l'affichage des résultats sous la forme d'arbres et de la génération d'un ensemble de règles en langage naturel. Les arbres de décision (classement) permettent d'expliquer et de prévoir l'appartenance d'observations à une classe d'une variable qualitative en se basant sur un ensemble de variables explicatives quantitatives et qualitatives. Les arbres de régression permettent d'expliquer et de prévoir la valeur prise par une variable quantitative à expliquer en fonction de variables explicatives quantitatives et qualitatives.

La procédure propose l'étude des jeux d'apprentissage, de validation et de prévision. Un rapport général de synthèse est proposé ainsi que les graphiques des coefficients de complexité, de l'importance des variables, des arbres complet et élagué, de la courbe ROC (décision), des valeurs estimées par rapport aux valeurs observées (régression) et des résidus par rapport aux valeurs estimées (régression).

Cette procédure est basée sur les packages R 'rpart' et 'rpart.plot'.

### Tableaux

Coefficients de complexité
Résultats pour l'arbre complet
Règles pour l'arbre complet
Importances des variables
Résultats pour l'arbre élagué
Règles pour l'arbre élagué
Détails du classement pour les jeux d'apprentissage et de validation (décision)
Synthèse du classement pour les jeux d'apprentissage et de validation (décision)

VP, FN, FP, VN, sensibilité, spécificité pour les jeux d'apprentissage et de validation (décision)
Détails du classement pour le jeu de prévision (décision)
Valeurs observées, estimées et résidus pour les jeux d'apprentissage et de validation (régression)
Valeurs estimées pour le jeu de prévision (régression)

## Graphiques

Graphique des coefficients de complexité
Graphique de l'importance des variables
Graphique de l'arbre complet de décision ou de régression
Graphique de l'arbre élagué de décision ou de régression
Courbe ROC pour les jeux d'apprentissage et de validation (décision)
Graphique des valeurs estimées et observées pour l'apprentissage et la validation (régression)
Graphique des résidus pour l'apprentissage et la validation (régression)

## Bibilographie

Voici quelques références de livres et revues en langues française et anglaise. Ces livres vous donneront toutes les informations théoriques nécessaires sur les méthodes statistiques disponibles dans UNIWIN.

### 9.1 Bibliographie française

- M. BARDOS « Analyse discriminante, application au risque et scoring financier »  
Dunod (2001)
- J.P. BENZECRI « La place de l'a priori »  
Encyclopedia Universalis Tome 17, 11-23
- J.P. BENZECRI « L'analyse des données ». Tome 1 : « La taxinomie »  
Tome 2 : « L'analyse des correspondances »"  
3è édition Dunod (1979)
- J.P. BENZECRI « Histoire et préhistoire de l'analyse des données »  
Dunod (1983)
- J.P. BENZECRI « La pratique de l'analyse des données »  
F. BENZECRI Tome 1 : « Analyse des correspondances, exposé élémentaire »  
Dunod (1983)
- J.P. BENZECRI « La pratique de l'analyse des données »  
Tome 2 : « Abrégé théorique, études de cas modèles »  
Dunod (1980)

- J.P. BENZECRI « La pratique de l'analyse des données ». Tome 5  
Economie Dunod (1987)
- C. BIALES « L'analyse statistique des données »  
Chotard et associés (1988)
- J.M. BOUROCHE « Analyse des données en marketing »  
Masson (1977)
- J.M. BOUROCHE « Analyse des données multidimensionnelles »  
P. BERTIER PUF, 2<sup>e</sup> édition (1977)
- J.M. BOUROCHE « L'analyse des données ». Pour la Science n°5,23-35  
G. SAPORTA (1978)
- J.M. BOUROCHE « L'analyse des données ». Collection Que Sais-Je.  
G. SAPORTA n° 1854. PUF (2002)
- F. CAILLEZ « Introduction à l'analyse des données »  
J.P. PAGES Smash (1976)
- G. CELEUX « Classification automatique des données »  
E. DIDAY Dunod (1989)
- J.L. CHANDON « Analyse typologique »  
S. PINSON Masson (1980)
- P. CIBOIS « L'analyse factorielle ». Collection Que Sais Je  
PUF (1983)
- J.P. CRAUSER « Guide pratique de l'analyse des données »  
Editions d'organisation (1989)
- F. DAZY « L'analyse des données évolutives »  
J.F. LE BARZIC GERI - Editions Technip (1996)
- D. DESBOIS « Une introduction au positionnement multidimensionnel »  
Revue Modulad n° 32 (2005)

- E. DIDAY « Optimisation en classification automatique »  
2 tomes. INRIA (1979)
- E. DIDAY « Eléments d'analyse des données »  
Dunod (1983)
- B. ESCOPIER « Traitement simultané de variables quantitatives  
et qualitatives en analyse factorielle »  
Les Cahiers de l'Analyse des Données - 1979  
Volume IV 4 -2, pages 137-146
- B. ESCOPIER « Analyses factorielles simples et multiples »  
J. PAGES Dunod (1998)
- J.P. FENELON « Qu'est-ce que l'analyse des données ? »  
Lefonen (1982)
- G. GOVAERT « Analyse des données »  
Hermès-Lavoisier (2003)
- J.L. GUIGOU « Méthodologies multidimensionnelles : Analyse des  
données et choix à critères multiples »  
Dunod 2è édition (1977)
- M. HUGUES « Segmentation et typologie »  
Bordas (1970)
- M. JAMBU « Exploration informatique et statistique des données »  
Dunod (1989)
- M.O. LEBEAUX « Classification automatique pour l'analyse des  
données ». Tome 1 : « Méthodes et algorithmes ».  
M. JAMBU Dunod (1978)
- J. JUAN « Classification hiérarchique par l'algorithme des voisins  
réciproques »  
Les Cahiers de l'Analyse des Données – 1982  
Volume VII - n°2, pages 219-225

- J. DE LAGARDE « Initiation à l'analyse des données »  
Dunod (1983)
- C. LAVIT « Analyse conjointe de tableaux quantitatifs »  
Masson (1988)
- L. LEBART « L'analyse des données »  
E. DIDAY La Recherche n° 74, 15-25 (1977)
- L. LEBART « Traitement des données statistiques »  
A. MORINEAU Dunod (1979)  
J.P. FENELON
- L. LEBART « Techniques de la description statistique »  
A. MORINEAU Dunod (1977)  
N. TABARD
- L. LEBART « Statistique exploratoire multidimensionnelle »  
A. MORINEAU Dunod (2000)  
M. PIRON
- L. LEBART « Analyse statistique de données textuelles »  
A. SALEM Dunod (1988)
- I.C. LERMAN « Les bases de la classification automatique »  
Gauthier-Villars (1970)
- I.C. LERMAN « Classification et analyse ordinale des données »  
Dunod (1981)
- P. MICHAUD « Optimisation en analyse ordinale des données »  
Masson (1979)
- M. MASSON « Méthodologies générales du traitement statistique  
de l'information de masse »  
Cedic-Nathan (1980)

- J. MOSCAROLA « Enquêtes et analyse des données »  
Vuibert (1990)
- J. P NAKACHE « Exercices commentés de mathématiques pour l'analyse statistique  
des données ».  
Dunod (1981)
- J. P NAKACHE « Statistique explicative appliquée »  
J. CONFAIS Technip (2003)
- J. P NAKACHE « Approche pragmatique de la classification »  
J. CONFAIS Technip (2004)
- J. PAGES « Analyse factorielle multiple appliquée aux variables qualitatives et  
aux données mixtes »  
Revue de Statistique Appliquée, volume L, n°4 (2002)
- J. PAGES « Analyse factorielle de données mixtes »  
Revue de Statistique Appliquée, volume LII, n°4 (2004)
- C. ROBERT « Analyse descriptive multivariée »  
Flammarion (1989)
- J.M ROMEDER « Méthodes et programmes d'analyse discriminante »  
Dunod (1973)
- M. ROUX « Algorithmes de classification »  
Masson (1986)
- G. SAPORTA « Probabilités, analyse des données et statistique »  
Technip (1990)
- M. TENENHAUS « Méthodes Statistiques en Gestion »  
Dunod Entreprise (1994)
- M. TENENHAUS « Régression PLS : Théorie et Pratique »  
Technip (1998)

- M. TENENHAUS « L'approche PLS »  
Revue de Statistique Appliquée, volume 47, n°2 (1999)
- R. TOMASSONE « Discrimination et classement »  
Masson (1988)
- D. TRIVI Mémoire d'Ingénieur  
CNAM (1989)
- S. TUFFERY « Data Mining et statistique décisionnelle »  
Technip (2005)
- S. TUFFERY « Etude de cas en statistique décisionnelle »  
Technip (2009)
- M. VOLLE « L'analyse des données »  
Economie et Statistique n° 9 (1978)
- M. VOLLE « Analyse des données »  
Economica, 4<sup>e</sup> édition (1997)

## 9.2 Bibliographie anglo-saxonne

- ANDERBERG "Cluster analysis for applications"  
Academic Press (1973)
- T.W. ANDERSON "An introduction to multivariate statistical analysis".  
Wiley (1984)
- V. BARNETT "Interpreting multivariate data"  
Wiley (1981)
- L. BREIMAN "Classification and regression trees"  
Wadsworth (1984)
- CLEVELAND "Dynamic graphics for statistics"  
Wadsworth (1984)

- W. COOLEY "Multivariate data analysis"  
Wiley (1971)
- CORMARK "A review of classification"  
Journal of the Royal Statistical Association (1971)
- J.C. DEVILLE "Correspondence analysis with an extension towards nominal  
time series"  
Journal of Econometrics 22, 169-189 (1983)
- E. DIDAY "Data analysis and informatics"  
North Holland (1979, 1984, 1987, 1988)
- E. DIDAY "Data analysis, learning symbolic and numeric knowledge"  
Nova Science (1989)
- DRAPER SMITH Applied Regression Analysis
- W. GAUL "Classification as a tool of research"  
North-Holland (1986)
- A. GIFI "Nonlinear multivariate analysis"  
John Wiley & Sons (1990)
- GNANADESIKAN "Methods for statistical data analysis of multivariate  
observations"  
Wiley (1977)
- B. GREEN "Analyzing multivariate data"  
Holt Rinehart Winston (1978)
- GREENACRE "Theory and application of correspondence analysis".  
Academic Press (1984)
- J.A. HARTIGAN "Clustering algorithms"  
Wiley (1975)

- C. HAYASHI "Recent developments in clustering and data analysis".  
Academic Press (1988)
- J. JANSSEN "New trends in data analysis".  
North-Holland (1983)
- J. JANSSEN "Data analysis in real life environment"  
North-Holland (1985)
- J. JANSSEN "Data analysis". Plenum (1988)  
JOHNSON "Hierarchical clustering schemes"  
Psychometrika (1967)
- I.T. JOLLIFFE "Principal components analysis"  
Springer (1986)
- KOSTER "Mathematical aspects of multiple correspondence analysis  
for ordinal data"  
DSWO Press (1989)
- KROONENBERG "Three-mode principal components analysis"  
DSWO Press (1983)
- J.B. KRUSKAL "Multidimensional scaling". Sage (1978)
- E. MALINVAUD "Data analysis in official socio-economic statistics".  
J.R.S.S. serie A, vol 146, 335-361 (1983)
- K.V. MARDIA "Multivariate analysis"  
Academic Press (1979)
- L. LEBART "Multivariate descriptive statistical analysis  
A. MORINEAU correspondence analysis and related technics for  
K. WARWICK large matrices".  
Wiley (1984)
- S. SCHIFFMAN "Introduction to multidimensional scaling"  
Academic Press (1981)

- SEBER "Multivariate observations"  
Wiley (1984)
- K. TAKEUCHI "The foundations of multivariate analysis"  
Wiley Eastern (1982)
- P. VAN DER HEIJDEN "Correspondence analysis of longitudinal categorical data".  
DSWO Press (1987)
- RIJCKEVORSEL "The application of fuzzy coding and horseshoes in multiple correspondence analysis".  
DSWO Press (1987)
- RIJCKEVORSEL "Component and correspondence analysis"  
Wiley (1998)
- TORGERSON Multidimensional scaling. Theory and method.  
Psychometrika 17, 401-19.
- J. TUKEY "Exploratory data analysis"  
Addison-Wesley (1977)