

UNIWIN VERSION 9.7.0

ANALYSE DISCRIMINANTE QUALITATIVE

Révision : 02/09/2023

| | |
|-----------------------------------------------------|----|
| Définition..... | 1 |
| Entrée des données | 2 |
| Données manquantes | 3 |
| Exemple : Fichier CHIENS2 | 3 |
| L'option Rapports | 8 |
| L'option Graphiques | 10 |
| Exemple : Fichier ASSURANCE..... | 13 |
| Les variables internes créées par la procédure..... | 16 |

Définition

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.

Entrée des données

Cliquons sur l'icône ADQ dans le ruban Expliquer. La boîte de dialogue montrée ci-après s'affiche :

The dialog box is titled "Analyse discriminante qualitative". It contains a large empty list box on the left side. On the right side, there are four input fields, each preceded by a selection icon (a square with a right-pointing triangle). The labels for these fields are:

- Facteur de classement:
- Variables explicatives qualitatives:
- (Libellés des variables explicatives:)
- (Libellés des individus:)

At the bottom of the dialog box, there are five buttons: "Ok", "Annuler", "Sélection", "Supprimer", and "Aide".

Cette boîte de dialogue permet de définir le facteur de classement qualitatif, les variables explicatives qualitatives, la variable contenant les libellés de ces variables explicatives et la variable contenant les libellés des individus.

Données manquantes

Les données manquantes ne sont pas autorisées.

Exemple : Fichier CHIENS2

Nous utiliserons le fichier CHIENS2 pour illustrer cette procédure. Ce fichier contient 8 variables descriptives de 27 races de chiens :

| <i>Libellé long</i> | <i>Libellé court</i> |
|------------------------|----------------------|
| Beauceron | BEA |
| Basset | BAS |
| Berger allemand | BER |
| Boxer | BOX |
| Bull-dog | BUD |
| Bull-mastiff | BUM |
| Caniche | CAN |
| Chihuahua | CHI |
| Cocker | COC |
| Colley | COL |
| Dalmatien | DAL |
| Doberman | DOB |
| Dogue allemand | DOG |
| Epagneul breton | EPB |
| Epagneul français | EPF |
| Fox-Hound | FOH |
| Fox-terrier | FOT |
| Grand bleu de Gascogne | GRB |
| Labrador | LAB |
| Lévrier | LEV |
| Mastiff | MAS |
| Pékinois | PEK |
| Pointer | POI |
| Saint-Bernard | STB |
| Setter | SET |
| Teckel | TEC |
| Terre-Neuve | TER |

Les libellés courts sont dans la variable *librace*, les libellés longs dans la variable *race*.

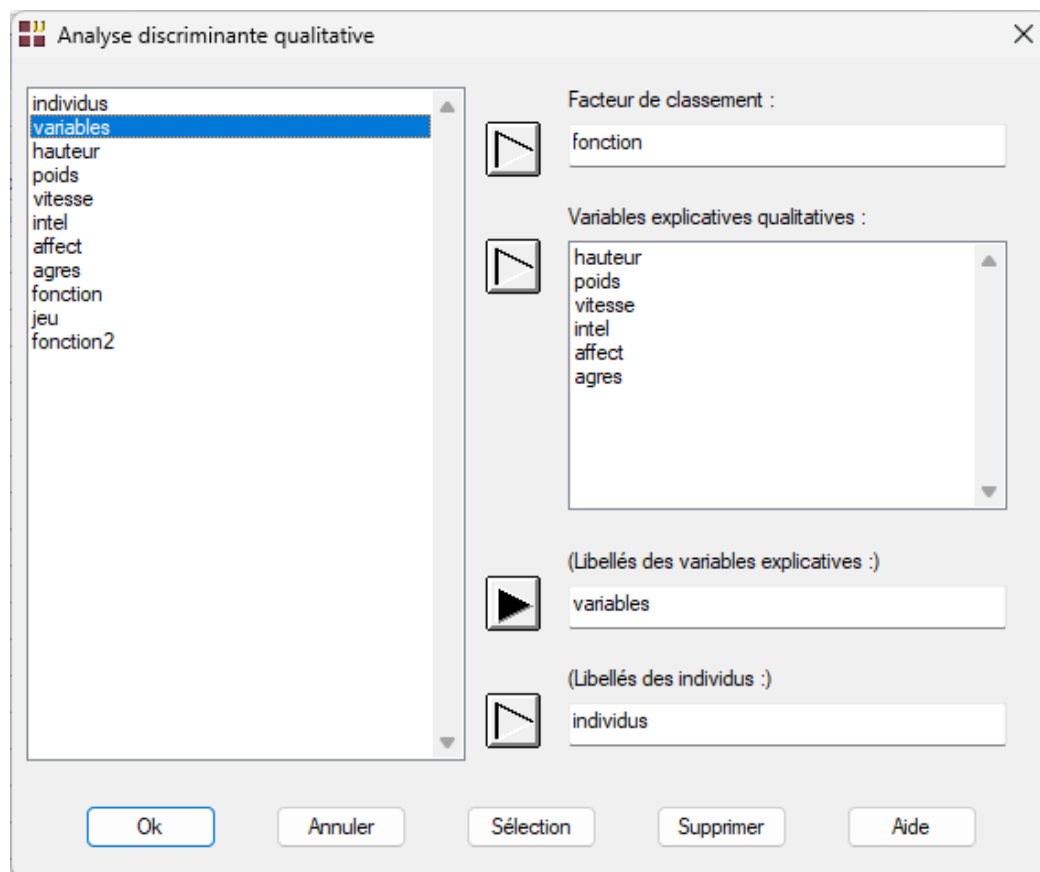
Les 8 variables descriptives sont les suivantes :

| Mesures | Variables | Modalités | Libellés des modalités |
|--------------|-----------|------------------------------|------------------------|
| Hauteur | hauteur | Ha1, Ha2, Ha3 | libhauteur |
| Poids | poids | Po1, Po2, Po3 | libpoids |
| Vitesse | vitesse | Vi1, Vi2, Vi3 | libvitesse |
| Intelligence | intel | In1, In2, In3 | libintel |
| Affectivité | affect | Af1, Af2 | libaffect |
| Agressivité | agres | Ag1, Ag2 | libagres |
| Fonction | fonction | Chasse Compagnie Utile | |
| Fonction2 | fonction2 | Autre Compagnie | |

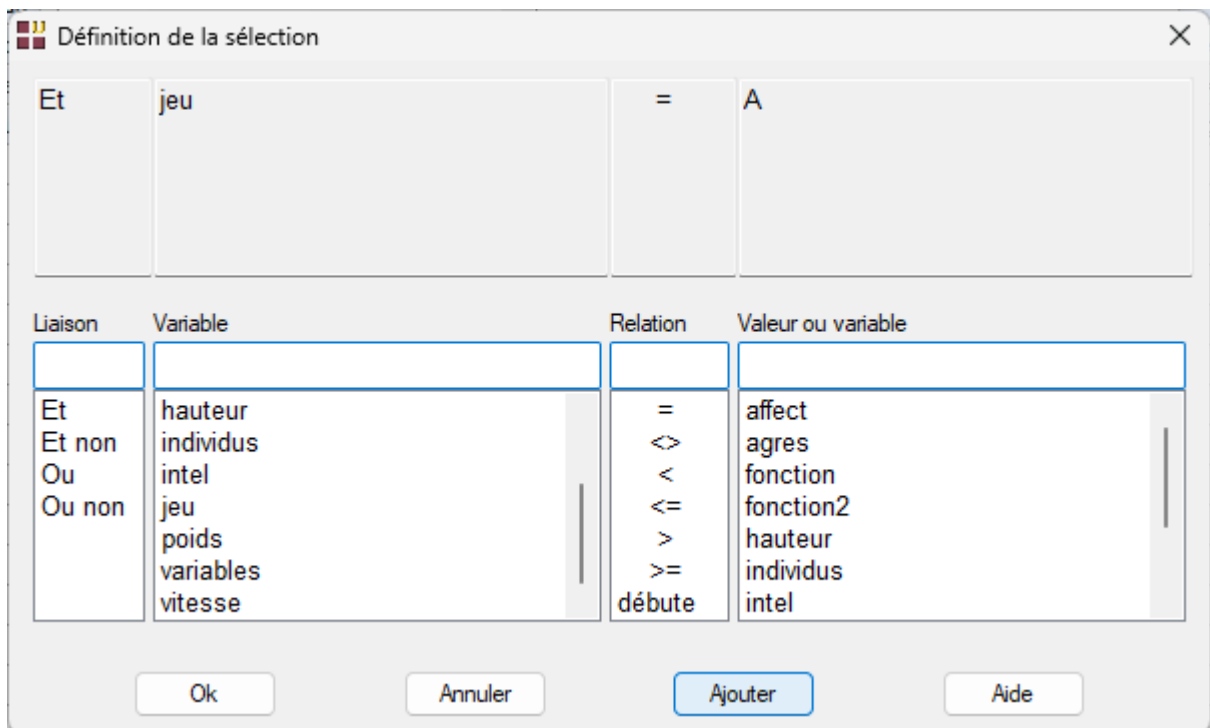
La variable 'jeu' permet de distinguer les populations d'apprentissage (A), de validation (V) et de prévision (P).

Les 6 premières variables descriptives seront utilisées comme variables explicatives, la 7ième variable (*fonction*) comme variable de classement définissant nos groupes de chiens.

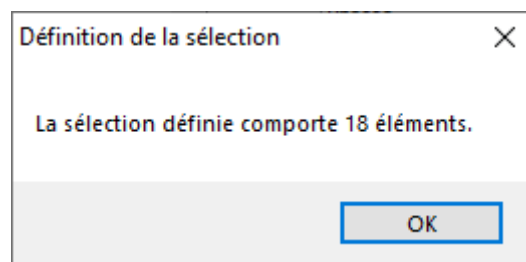
Cliquons sur l'icône ADQ dans le ruban Expliquer. La boîte de dialogue montrée ci-après s'affiche.



et sélectionnons la population d'apprentissage :



Il y a 18 chiens dans la population d'apprentissage :

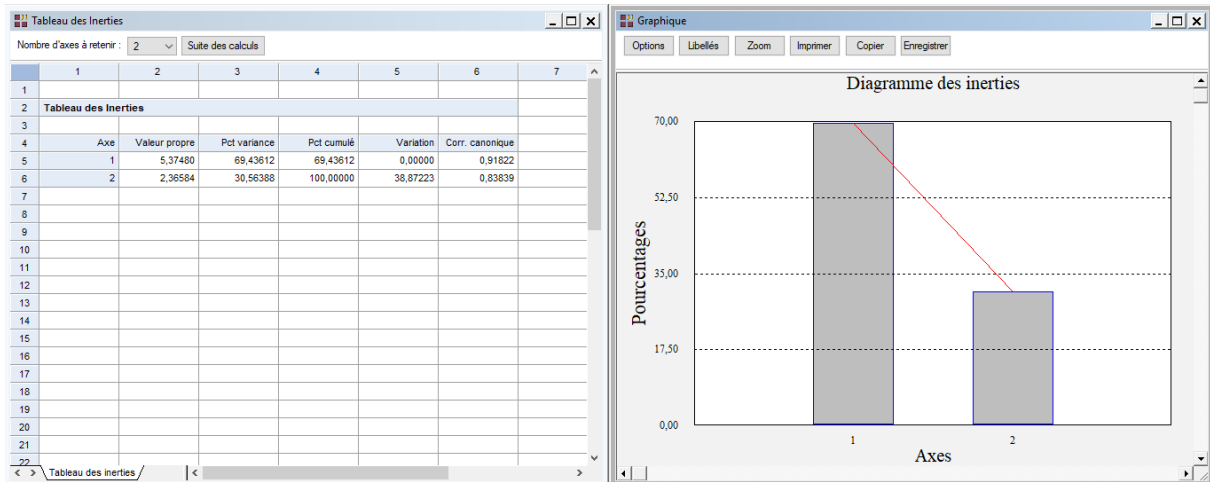


Les chiens non sélectionnés pour lesquels les valeurs du facteur de classement sont connues constituent la population de validation.

Les chiens non sélectionnés pour lesquels les valeurs du facteur de classement ne sont pas connues constituent la population de prévision.

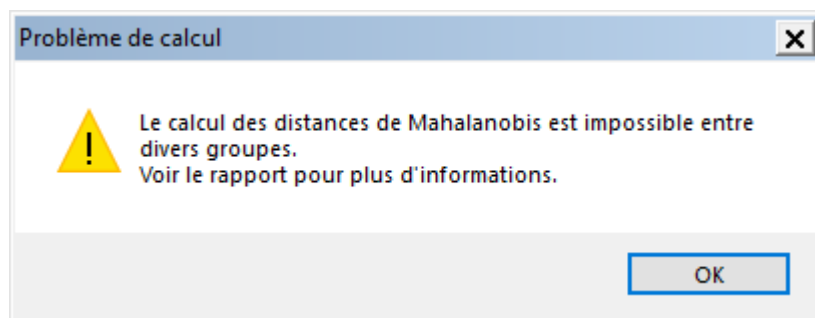
Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul de l'Analyse Discriminante Qualitative.

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

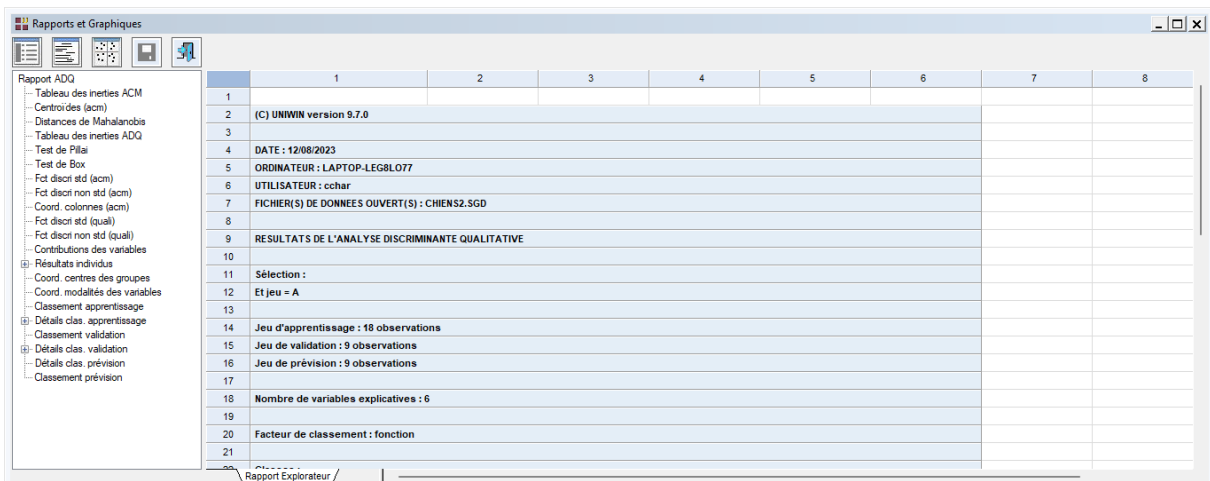



L'option 'Nombre d'axes à retenir' permet de préciser le nombre de composantes principales à extraire. Cliquons sur le bouton 'Suite des calculs'.


Le logiciel nous précise que les distances de Mahalanobis ne sont pas calculables pour tous les groupes. Plus de détails seront donnés dans le rapport.

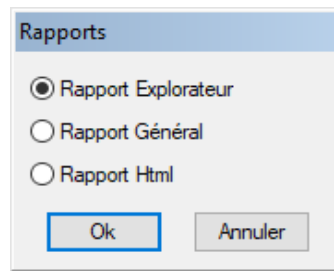



Après quelques instants, l'écran montré ci-dessous s'affiche :

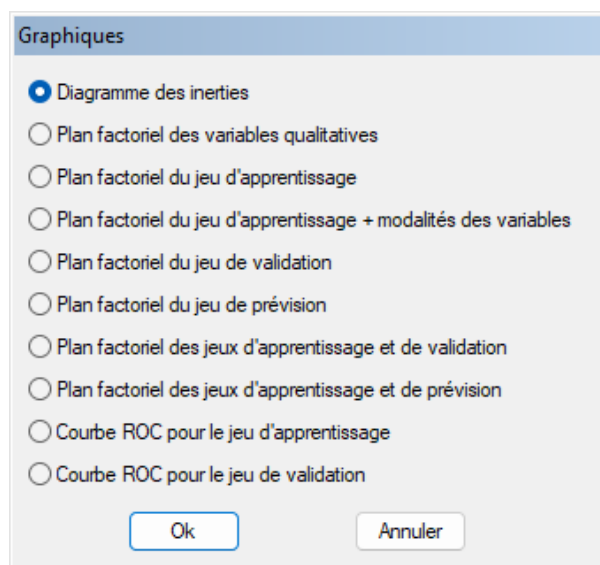



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

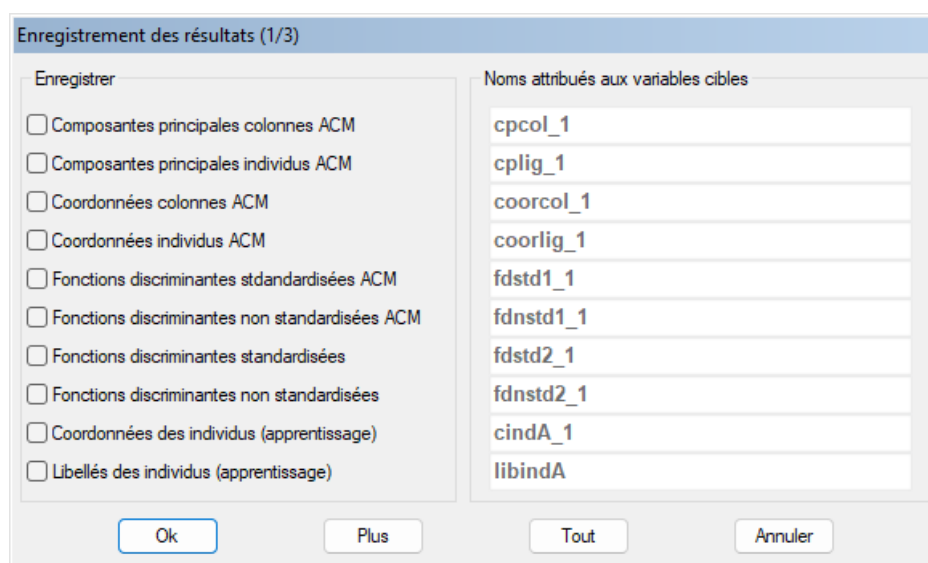
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



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



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

L'option Rapports

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

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

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

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----|-------------------------------------------------------------------------|-------|----------|------------|----------|----------|-----------|------------|
| 1 | | | | | | | | |
| 2 | RESULTATS POUR LES INDIVIDUS DU JEU D'APPRENTISSAGE POUR LE FACTEUR : 1 | | | | | | | |
| 3 | | | | | | | | |
| 4 | DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE | | | | | | | |
| 5 | COORD. = COORDONNÉES DES INDIVIDUS | | | | | | | |
| 6 | CONTRIB. = CONTRIBUTIONS A L'INERTIE | | | | | | | |
| 7 | COSINUS*2 = COSINUS CARRÉS | | | | | | | |
| 8 | COS*2 CUM. = SOMMES CUMULÉES DES COSINUS CARRÉS | | | | | | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |
| 11 | | GRUPE | INDIV/DU | DISTANCE*2 | COORD. | CONTRIB. | COSINUS*2 | COS*2 CUM. |
| 12 | Basset | 1 | 2 | 5,16027 | -1,36082 | 1,93662 | 0,35886 | 0,3588 |
| 13 | Epagneul Breton | 1 | 14 | 1,63173 | -1,16156 | 1,41099 | 0,82687 | 0,8268 |
| 14 | Epagneul Français | 1 | 15 | 9,18191 | -2,08161 | 4,53150 | 0,47192 | 0,4719 |
| 15 | Fox Hound | 1 | 16 | 6,74117 | 0,68287 | 0,48765 | 0,06917 | 0,0691 |
| 16 | Grand Bleu | 1 | 18 | 15,50031 | -1,27221 | 1,69262 | 0,10442 | 0,1044 |
| 17 | Boxer | 2 | 4 | 1,13274 | -0,90590 | 0,85823 | 0,72448 | 0,7244 |
| 18 | Bulldog | 2 | 5 | 7,02120 | -1,65597 | 2,86778 | 0,39056 | 0,3905 |
| 19 | Caniche | 2 | 7 | 3,25622 | -1,43330 | 2,14840 | 0,63090 | 0,6309 |
| 20 | Chihuahua | 2 | 8 | 6,93752 | -2,60805 | 7,10246 | 0,97895 | 0,9789 |
| 21 | Cocker | 2 | 9 | 6,11157 | -1,75046 | 3,20442 | 0,50137 | 0,5013 |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----|-----------------------------------------------------|--------|-----------|-------|-------|---|---|---|
| 1 | | | | | | | | |
| 2 | RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE | | | | | | | |
| 3 | | | | | | | | |
| 4 | EN LIGNES, LES GROUPES OBSERVES | | | | | | | |
| 5 | EN COLONNES, LES GROUPES PREVUS | | | | | | | |
| 6 | | | | | | | | |
| 7 | POURCENTAGE DE MAL CLASSES : 0,000 % | | | | | | | |
| 8 | POURCENTAGE DE BIEN CLASSES : 100,000 % | | | | | | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |
| 11 | | Chasse | Compagnie | Utile | Total | | | |
| 12 | Chasse | 5 | 0 | 0 | 5 | | | |
| 13 | Compagnie | 0 | 8 | 0 | 8 | | | |
| 14 | Utile | 0 | 0 | 5 | 5 | | | |
| 15 | Total | 5 | 8 | 5 | 18 | | | |
| 16 | | | | | | | | |
| 17 | | | | | | | | |
| 18 | | | | | | | | |
| 19 | | | | | | | | |
| 20 | | | | | | | | |
| 21 | | | | | | | | |

Rapports et Graphiques

TABLEAU DES INERTIES DE L'ACM

| | VALEUR PROPRE | PCT VARIANCE | PCT CUMULE | VARIATION |
|----|---------------|--------------|------------|-----------|
| 1 | 0,49618 | 29,77107 | 29,77107 | 0,00000 |
| 2 | 0,37708 | 22,62458 | 52,39564 | 7,14649 |
| 3 | 0,17626 | 10,57535 | 62,97099 | 12,04923 |
| 4 | 0,16433 | 9,85996 | 72,83095 | 0,71539 |
| 5 | 0,15550 | 9,33012 | 82,16107 | 0,52984 |
| 6 | 0,11283 | 6,77000 | 88,93107 | 2,56012 |
| 7 | 0,08642 | 5,18495 | 94,11601 | 1,58505 |
| 8 | 0,06701 | 4,02088 | 98,13690 | 1,16407 |
| 9 | 0,02116 | 1,26949 | 99,40639 | 2,75139 |
| 10 | 0,00989 | 0,59361 | 100,00000 | 0,67587 |

CENTROIDES DES GROUPES SUR LES COMPOSANTES PRINCIPALES DE L'ACM

| | CP ACM 1 | CP ACM 2 | CP ACM 3 | CP ACM 4 | CP ACM 5 | CP ACM 6 | CP ACM 7 | CP ACM 8 | CP ACM 9 | CP ACM 10 |
|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| Chasse | 0,21792 | -0,01235 | 0,14530 | -0,08772 | -0,42815 | -0,06289 | 0,10397 | 0,13350 | 0,00306 | 0,00587 |

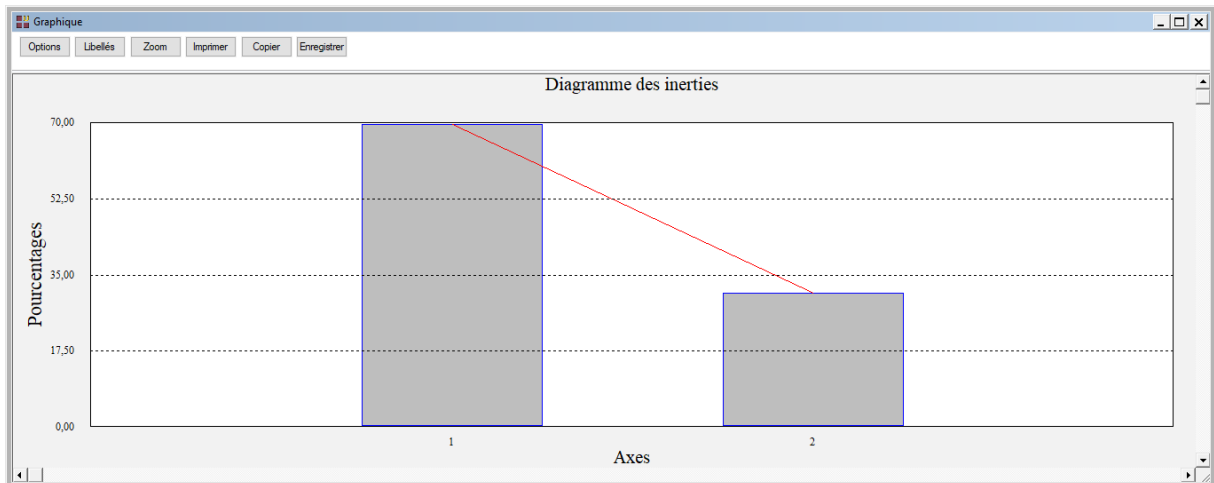
Ces rapports nous fournissent les renseignements suivants :

- Tableau des inerties de l'analyse des correspondances multiples (ACM)
- Centroides des groupes sur les composantes principales de l'ACM
- Distances de Mahalanobis entre les groupes, Fishers, niveaux de signification
- Tableau des inerties de l'analyse discriminante qualitative (ADQ)
- Test de Pillai
- Test de Box (égalité des matrices de covariances)
- Fonctions discriminantes standardisées de l'ACM
- Fonctions discriminantes non standardisées de l'ACM
- Coordonnées des variables dans l'ACM
- Fonctions discriminantes standardisées exprimées en fonction des variables qualitatives d'origine
- Fonctions discriminantes non standardisées exprimées en fonction des variables qualitatives d'origine
- Contributions des variables à l'inertie (carré du rapport de corrélation)
- Résultats pour les individus sur les différents facteurs (carré de la distance à l'origine, coordonnée, contribution, cosinus carré, cosinus carré cumulé)
- Coordonnées des centres des groupes
- Coordonnées des modalités des variables qualitatives
- Synthèse du classement (apprentissage)
- Détails du classement (apprentissage)
- VP, FN, FP, VN, sensibilité et spécificité pour la population d'apprentissage
- Synthèse du classement (validation)
- Détails du classement (validation)
- VP, FN, FP, VN, sensibilité et spécificité pour la population de validation
- Classement pour la population de prévision

L'option Graphiques

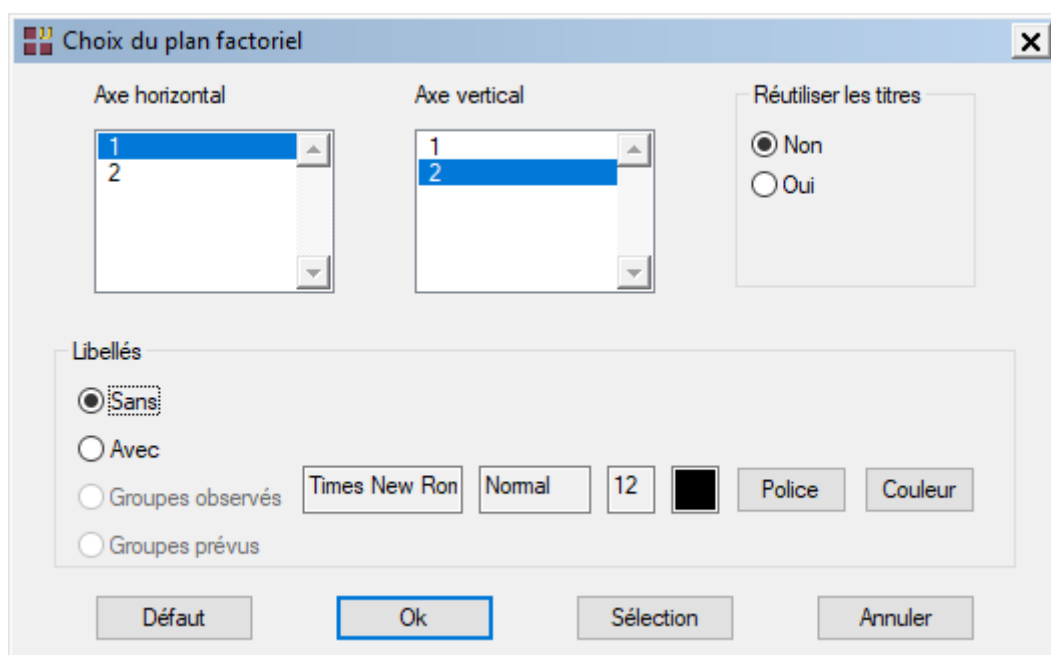
- Diagramme des inerties

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



- Plan factoriel des variables qualitatives

Cette option nous permet de représenter les variables qualitatives par les contributions de celles-ci à l'inertie (carré du rapport de corrélation). Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.



Boîte de dialogue : Choix du plan factoriel

Axe horizontal : 1

Axe vertical : 2

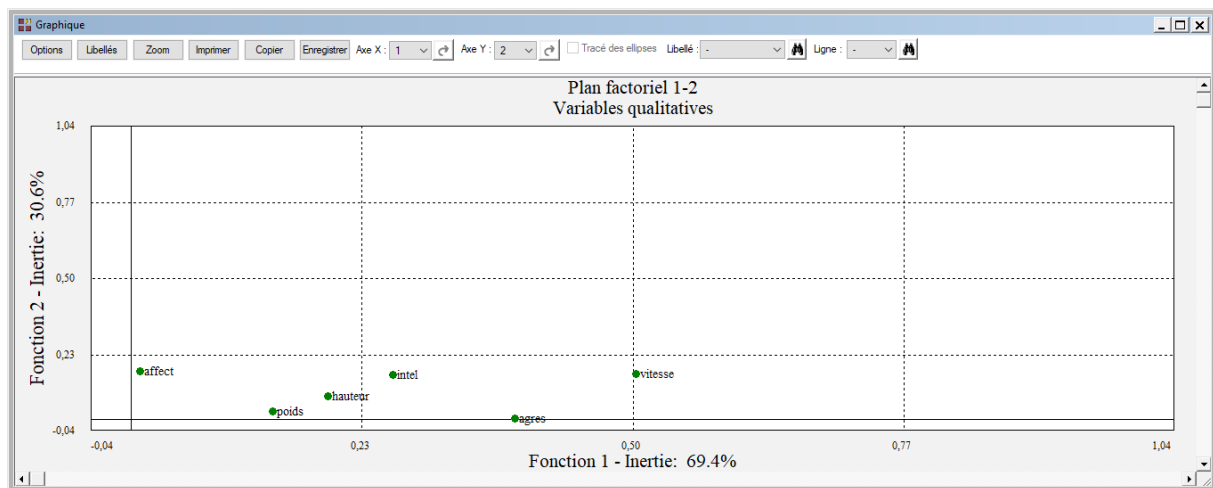
Réutiliser les titres : Non, Oui

Libellés : Sans, Avec

Groupes observés : Times New Ron, Normal, 12, [couleur], Police, Couleur

Groupes prévus : []

Boutons : Défaut, Ok, Sélection, Annuler



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

Cette option permet d'afficher des plans factoriels des individus des jeux d'apprentissage, de validation et de prévision ainsi que les centres des groupes. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.

Choix du plan factoriel

Axe horizontal: 1, 2

Axe vertical: 1, 2

Réutiliser les titres: Non, Oui

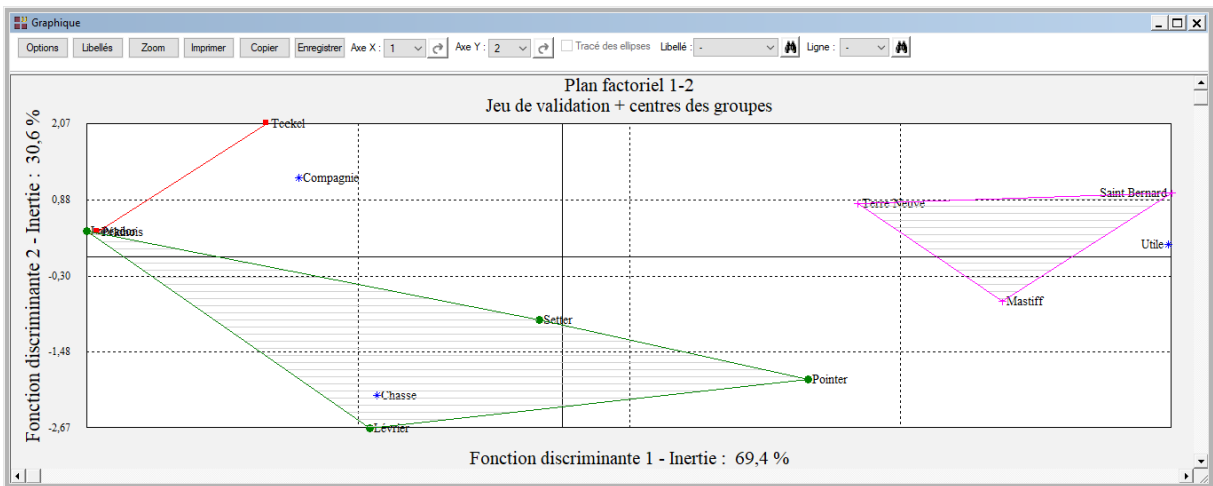
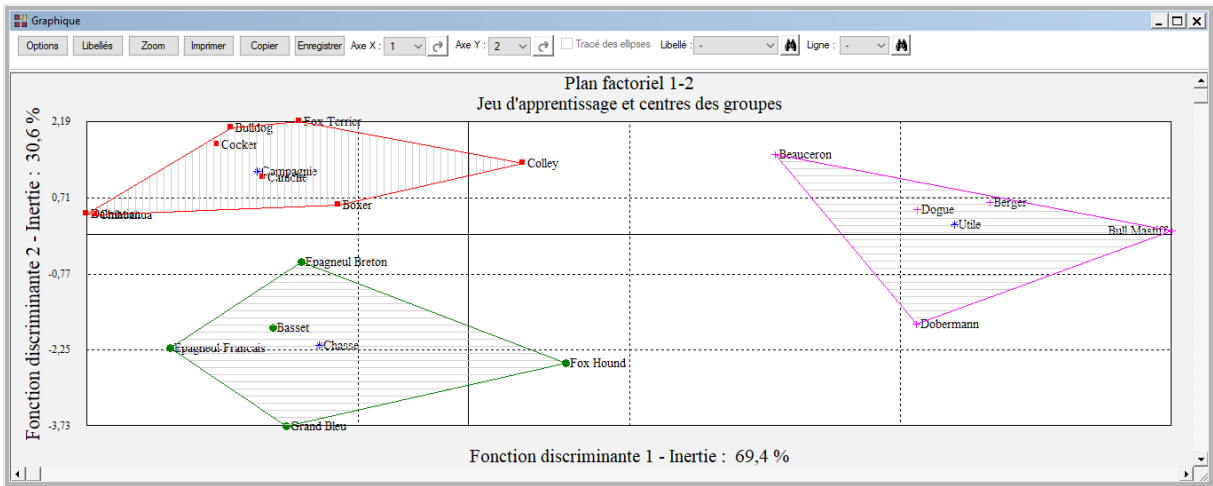
Libellés: Sans, Avec, Groupes observés, Groupes prévus

Groupes observés: Times New Ron, Normal, 12, [Color Swatch], Police, Couleur

Buttons: Défaut, Ok, Sélection, Annuler

Elle permet de préciser si l'on désire afficher ou non les libellés des individus, de préciser si ces libellés sont les codes des groupes observés ou les codes des groupes prévus, de choisir la couleur et la police pour ces libellés.

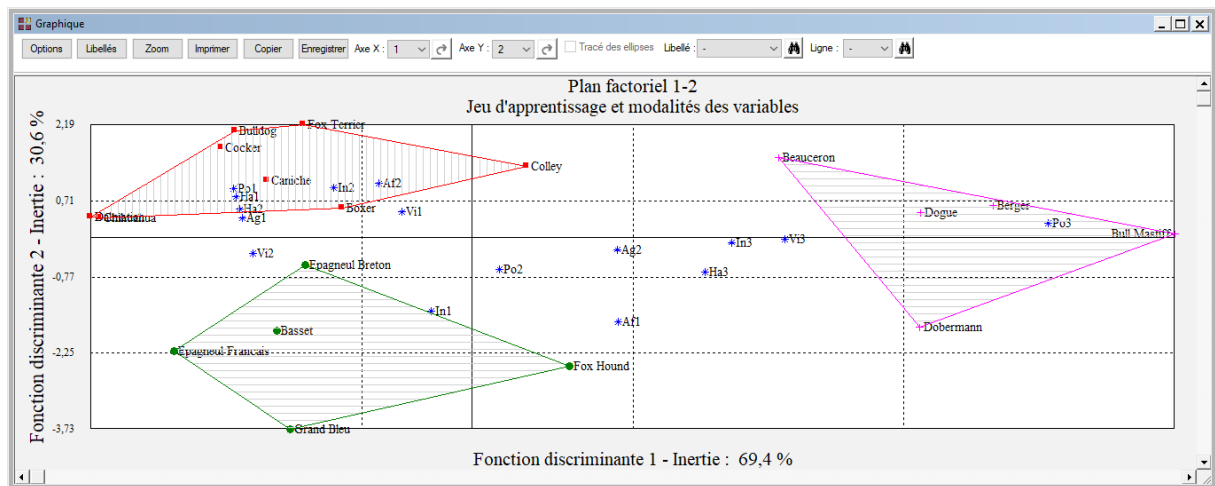
Il est également possible d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.



- Plan factoriel du jeu d'apprentissage et modalités des variables

Cette option permet d'afficher des plans factoriels des individus d'apprentissage et les modalités des variables qualitatives.

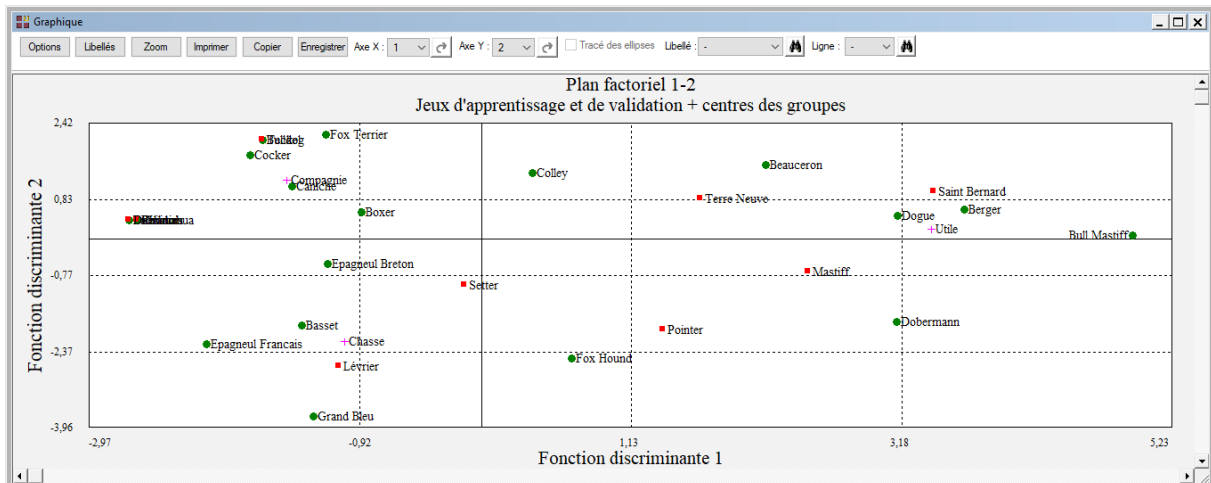
Comme pour l'option précédente, une boîte de dialogue permettant de choisir le plan factoriel et les libellés à afficher s'affiche.



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

Cette option permet d'afficher des plans factoriels dans lesquels les individus d'apprentissage et de validation ou de prévision sont visualisés simultanément.

Comme pour les options précédentes, une boîte de dialogue permettant de choisir le plan factoriel et les libellés à afficher s'affiche.



- Courbes ROC

Ces deux options ne sont pas actives dans cet exemple car le facteur de classement possède plus de 2 classes.

Exemple : Fichier ASSURANCE

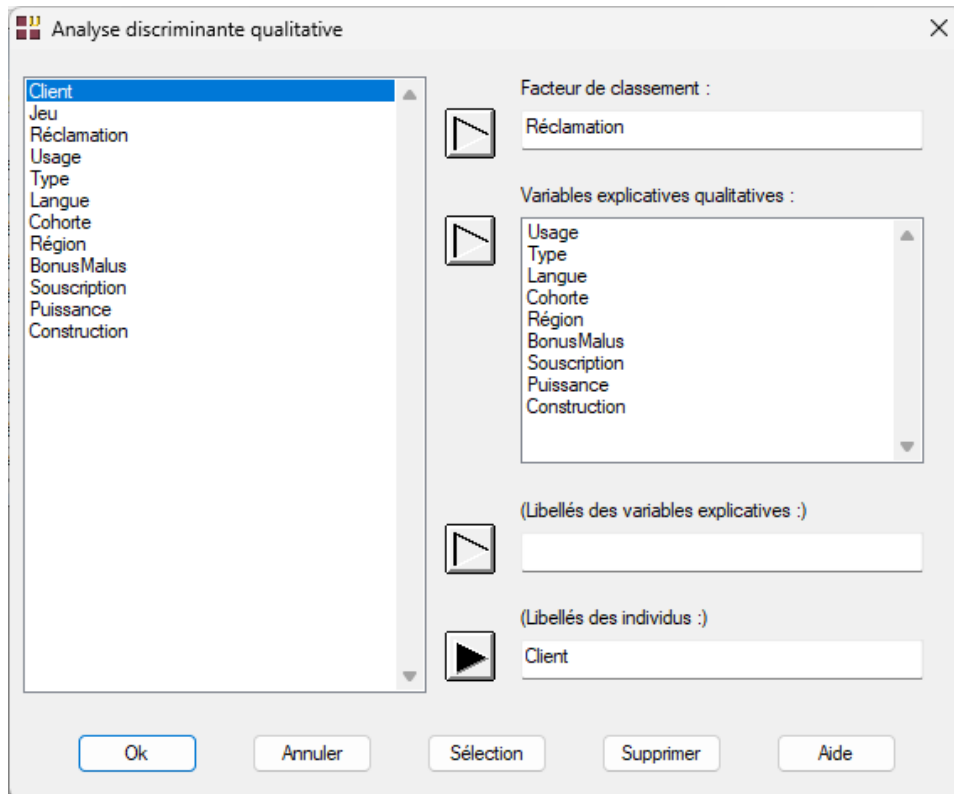
Le fichier ASSURANCE contient des informations collectées en 1992 concernant 1106 contrats d'assurance d'automobilistes belges.

- | | |
|-------------------------------------|-----------------------------------------------------------------|
| • Réclamation | Bonne, Mauvaise |
| • Usage du véhicule | Usage privé, Usage professionnel |
| • Type d'assurance | Usage entreprise, Usage femme Usage homme |
| • Langue parlée | Langue flamande, Langue française |
| • Cohorte de naissance | Cohorte 1890 à 1949, Cohorte 1950 à 1973 Cohorte inconnue |
| • Région géographique | Autres régions, Bruxelles |
| • Niveau de bonus-malus | Bonus moins, Bonus plus |
| • Année de souscription | Souscription 86 et au-delà Souscription avant 86 |
| • Puissance du véhicule | Puissance inférieure à 40 Puissance supérieure ou égale à 40 |
| • Année de construction du véhicule | Construction 33-89, Construction 90-91 |

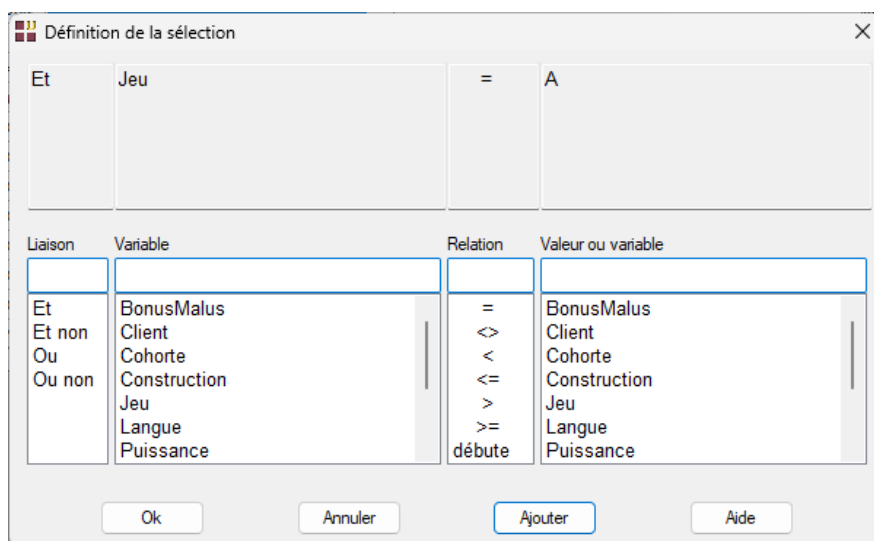
Référence : Saporta G., Niang N. (2006) Correspondence Analysis and Classification. In *Multiple Correspondence Analysis and Related Methods*, M. Greenacre and J. Blasius, Eds., pp 371-392. Chapman & Hall/CRC, Boca Raton, Florida, USA.

La variable 'Jeu' permet de distinguer les populations d'apprentissage (A) et de validation (V).

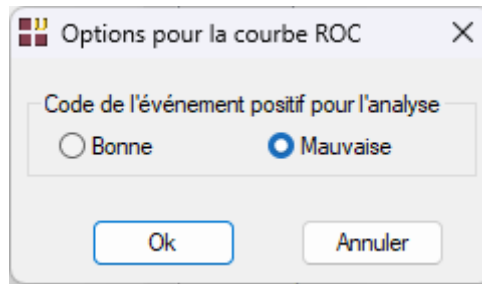
Cliquons sur l'icône ADQ dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après :



Sélectionnons la population d'apprentissage :



Puisque la variable à expliquer possède deux classes, la procédure demande de préciser le code de l'événement positif qui sera utilisé pour le tracé de la courbe ROC.



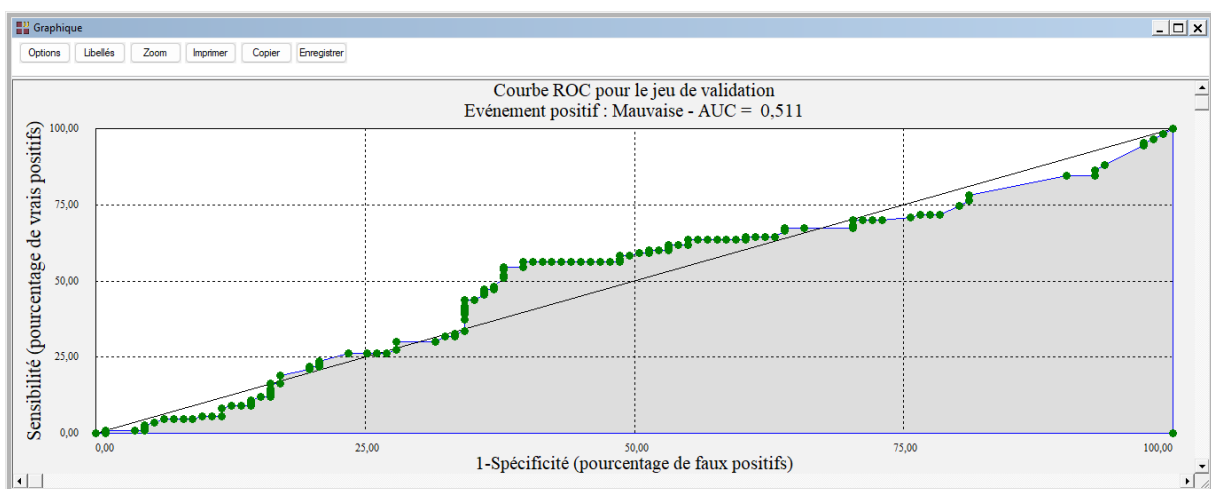
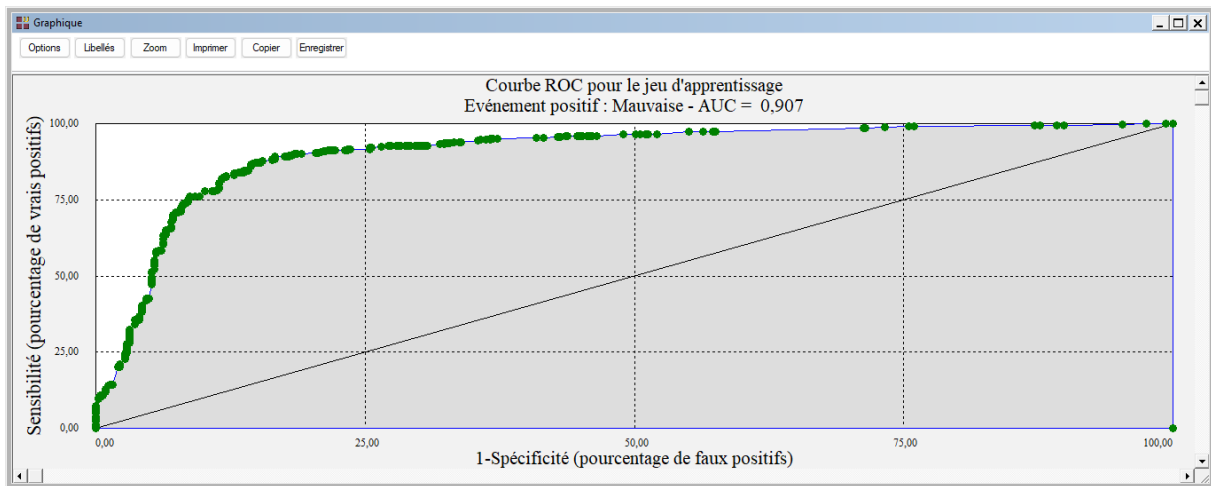
Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.

Visualisons les résultats des classements des populations d'apprentissage et de validation.

| RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE | | | | |
|-----------------------------------------------------|-------|----------|-------|--|
| EN LIGNES, LES GROUPES OBSERVES | | | | |
| EN COLONNES, LES GROUPES PREVUS | | | | |
| POURCENTAGE DE MAL CLASSES : 14,237 % | | | | |
| POURCENTAGE DE BIEN CLASSES : 85,763 % | | | | |
| | Bonne | Mauvaise | Total | |
| Bonne | 372 | 73 | 445 | |
| Mauvaise | 53 | 387 | 440 | |
| Total | 425 | 460 | 885 | |

| RESULTATS DU CLASSEMENT POUR LE JEU DE VALIDATION | | | | |
|---------------------------------------------------|-------|----------|-------|--|
| EN LIGNES, LES GROUPES OBSERVES | | | | |
| EN COLONNES, LES GROUPES PREVUS | | | | |
| POURCENTAGE DE MAL CLASSES : 15,385 % | | | | |
| POURCENTAGE DE BIEN CLASSES : 84,615 % | | | | |
| | Bonne | Mauvaise | Total | |
| Bonne | 91 | 20 | 111 | |
| Mauvaise | 14 | 96 | 110 | |
| Total | 105 | 116 | 221 | |

Le rapport affiche les sensibilités et spécificités qui permettent les tracés des courbes ROC des populations d'apprentissage et de validation.



Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure. Ces variables peuvent notamment être utilisées avec l'option 'Sélection'. A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

| <i>Variable</i> | <i>Contenu</i> |
|-----------------|--------------------------------------------------|
| cpcol | Composantes principales colonnes (ACM) |
| cplig | Composantes principales individus (ACM) |
| coorcol | Coordonnées des colonnes (ACM) |
| coorig | Coordonnées des individus (ACM) |
| fdstd1 | Fonctions discriminantes standardisées (ACM) |
| fdnstd1 | Fonctions discriminantes non standardisées (ACM) |
| fdstd2 | Fonctions discriminantes standardisées (ADQ) |
| fdnstd2 | Fonctions discriminantes non standardisées (ADQ) |
| cindA | Coordonnées des individus (apprentissage) |
| libindA | Libellés des individus (apprentissage) |

| | |
|--------------|-------------------------------------------------------------|
| clindA | Classes d'origine des individus (apprentissage) |
| distind | Distances carrées des individus à l'origine (apprentissage) |
| cosind | Cosinus carrés des individus (apprentissage) |
| conind | Contributions des individus (apprentissage) |
| cindV | Coordonnées des individus (validation) |
| libindV | Libellés des individus (validation) |
| clindV | Classes d'origine des individus (validation) |
| vpA | Vrais positifs (apprentissage) |
| fnA | Faux négatifs (apprentissage) |
| fpA | Faux positifs (apprentissage) |
| vnA | Vrais négatifs (apprentissage) |
| specificiteA | Spécificité (apprentissage) |
| sensibiliteA | Sensibilité (apprentissage) |
| vpV | Vrais positifs (validation) |
| fnV | Faux négatifs (validation) |
| fpV | Faux positifs (validation) |
| vnV | Vrais négatifs (validation) |
| specificiteV | Spécificité (validation) |
| sensibiliteV | Sensibilité (validation) |
| classeA | Classement apprentissage |
| classeV | Classement validation |
| classeP | Classement prévision |
| libindP | Libellés des individus (prévision) |