

## UNIWIN VERSION 9.7.0

# ANALYSE DISCRIMINANTE BAYESIENNE

Révision : 02/09/2023

Définition.....	1
Entrée des données .....	2
Données manquantes .....	3
Exemple 1 : Fichier IRIS3.....	3
L'option Rapports .....	8
L'option Graphiques .....	9
Exemple 2 : Fichier INFARCT2 .....	13
L'option Rapports .....	16
L'option Graphiques .....	18
Exemple 3 : Fichier BORDEAUX.....	20
Exemple 4 : Fichier TITANIC .....	23
Les variables internes créées par la procédure .....	25

### Définition

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.

## Entrée des données

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

Boîte de dialogue intitulée "Analyse discriminante bayésienne".

Facteur de classement: [Champ de saisie]

Variables explicatives quantitatives: [Liste déroulante vide]

(Libellés des variables explicatives:) [Champ de saisie]

(Libellés des individus:) [Champ de saisie]

(Probabilités initiales:) [Champ de saisie]

Centrage et réduction  
 Oui  Non

Boutons: Ok, Annuler, Sélection, Supprimer, Aide

Cette boîte de dialogue permet de préciser la variable qualitative définissant facteur de classement, les variables explicatives quantitatives, la variable contenant les libellés des variables explicatives et la variable contenant les libellés des individus.

Elle permet également de définir la variable contenant les probabilités a priori si on ne désire pas qu'UNIWIN les calcule à partir des fréquences des groupes dans les données.

Enfin, l'option de centrage et réduction des données est proposée. Cette option est utile si le classement de certains individus n'est pas possible suite à un dépassement de capacité lors des calculs. Il se peut que cela soit lié aux grandeurs des données et les centrer-réduire peut résoudre ce problème de calcul.

### **Données manquantes**

Les données manquantes ne sont pas autorisées pour le facteur de classement. Elles sont autorisées pour les variables quantitatives.

### **Exemple 1 : Fichier IRIS3**

Nous utiliserons le fichier IRIS3 pour illustrer cette procédure. Ce fichier contient pour 150 iris de trois espèces différentes les mesures des quatre caractéristiques suivantes exprimées en millimètres : longueur du sépale, largeur du sépale, longueur du pétale et largeur du pétale

Les trois espèces sont : Iris Setosa, Iris Versicolor et Iris Virginica

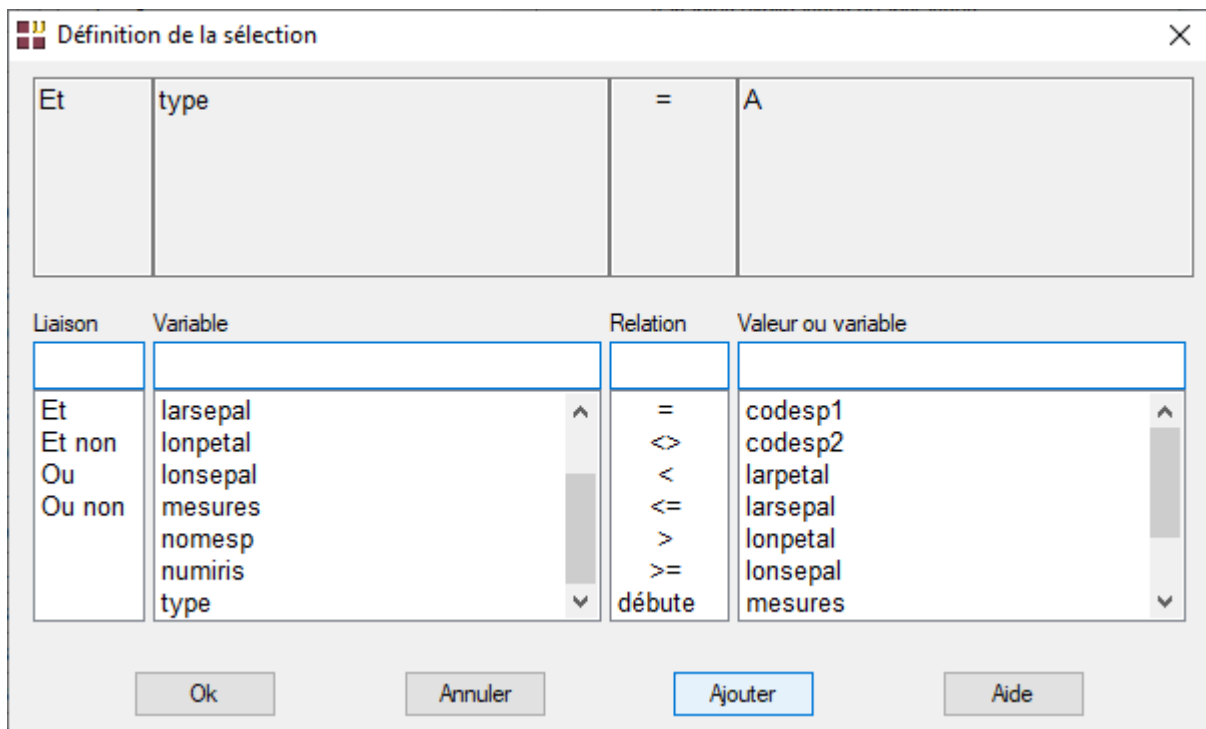
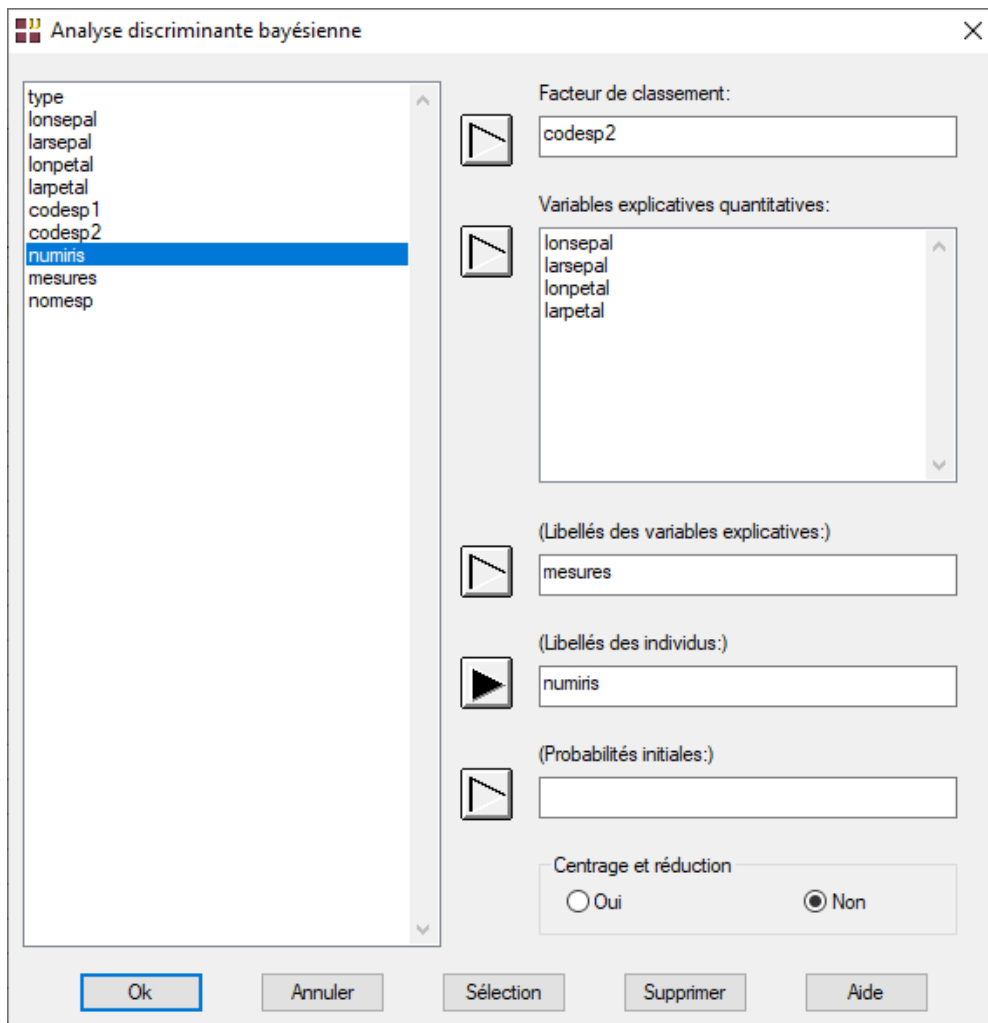
Cliquons sur l'icône ADB dans le ruban Expliquer.

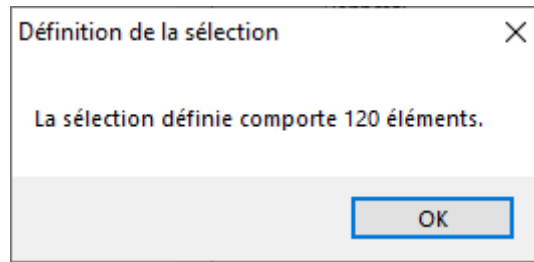
La boîte de dialogue montrée ci-après s'affiche.

Choisissons les variables *lonsepal* à *larpetal* comme variables quantitatives, la variable *codesp2* comme facteur de classement, la variable *mesures* comme variable contenant les libellés des variables quantitatives et la variable *numiris* comme variable contenant les libellés des individus.

Réalisons une analyse non centrée-réduite et laissons à UNIWIN le soin de calculer les probabilités initiales.

Cliquons sur le bouton Sélection pour définir la population d'apprentissage.





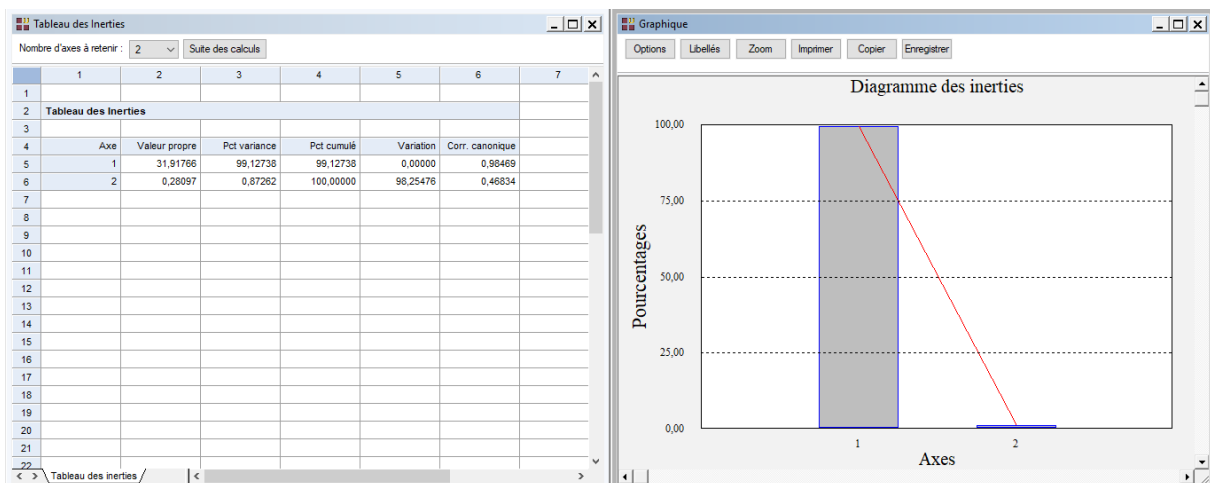
Cette sélection comporte 120 individus qui constituent la population d'apprentissage.

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

Les individus 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 Bayésienne.

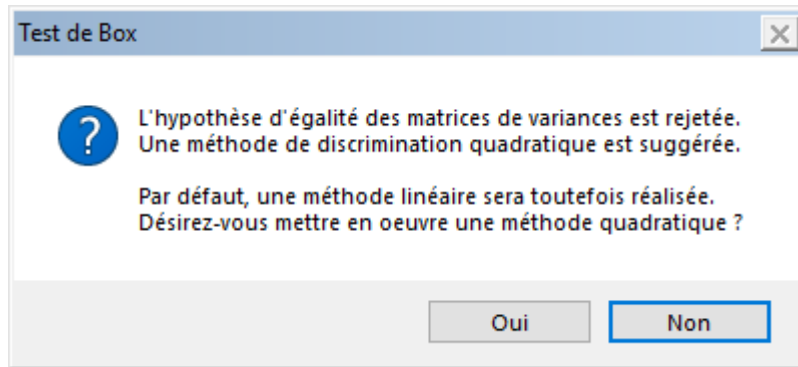
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.



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

Cliquons sur le bouton 'Suite des calculs'.

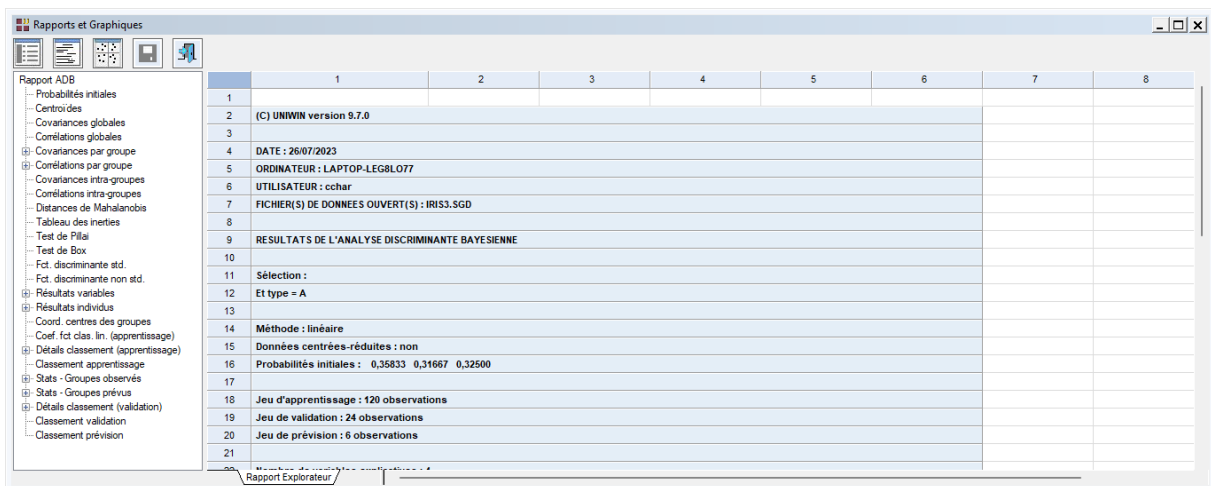
Après quelques instants, UNIWIN affiche une boîte de dialogue vous indiquant si l'hypothèse d'égalité des matrices de variances est vérifiée ou non.





Dans notre exemple, l'hypothèse doit être rejetée et donc une analyse discriminante quadratique est suggérée.

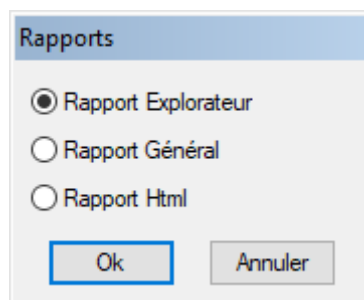
Choisissons cependant de mettre en oeuvre une analyse linéaire.


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

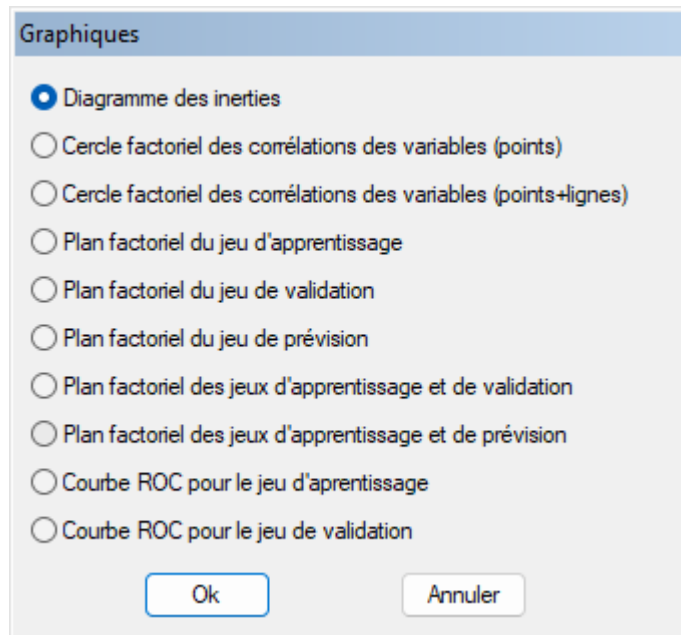



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

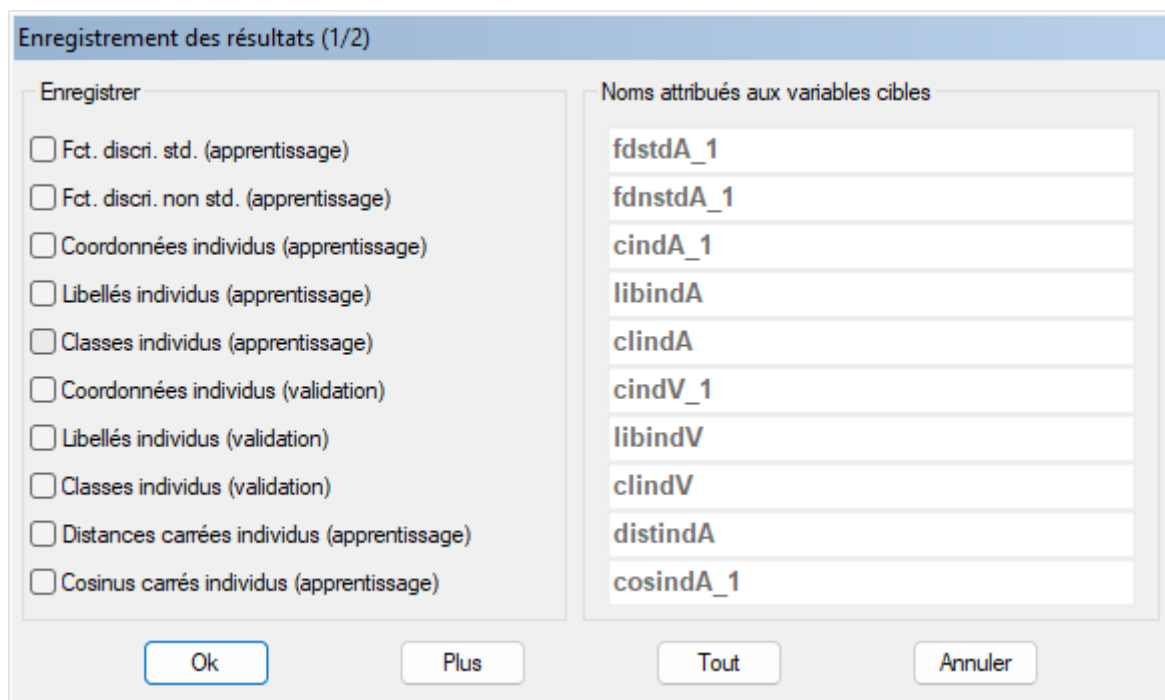
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



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

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

## L'option Rapports

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

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

**Rapports et Graphiques**

Rapport ADB

- Probabilités initiales
- Centroides
- Covariances globales
- Corrélations globales
- Covariances par groupe
- Corrélations par groupe
- Covariances intra-groupes
- Corrélations intra-groupes
- Distances de Mahalanobis
- Tableau des inerties
- Test de Pillai
- Test de Box
- Fct. discriminante std.
- Fct. discriminante non std.
- Résultats variables
- Résultats individus
  - Facteur 1**
  - Facteur 2
- Points multiples individus
- Coord. centres des groupes
- Coeff. fct clas. lin. (app)
- Détails classement (app)
- Classement apprentissage
- Stats - Groupes observés
- Stats - Groupes prévus
- Détails classement (vald)
- Classement validation
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>RESULTATS INDIVIDUS POUR LE FACTEUR : 1</b>							
3								
4	<b>DISTANCE*2 = CARRÉS DES DISTANCES A L'ORIGINE OU AU BARYCENTRE</b>							
5	<b>COORD. = COORDONNÉES DES INDIVIDUS</b>							
6	<b>CONTRIB. = CONTRIBUTIONS A L'INERTIE</b>							
7	<b>COSINUS*2 = COSINUS CARRÉS</b>							
8	<b>COS*2 CUM. = SOMMES CUMULÉES DES COSINUS CARRÉS</b>							
9								
10								
11								
		GRUPE	INDIV/DU	DISTANCE*2	COORD.	CONTRIB.	COSINUS*2	COS*2 CUM.
12	1	1	1	57,32730	7,56828	1,48724	0,99916	0,99916
13	2	1	2	45,52882	6,69560	1,16403	0,98467	0,98467
14	4	1	3	40,65299	6,34050	1,04384	0,98890	0,98890
15	5	1	4	58,20534	7,61605	1,50607	0,99555	0,99555
16	6	1	5	53,47025	7,18429	1,34015	0,98528	0,98528
17	7	1	6	45,09110	6,70541	1,16744	0,99715	0,99715
18	8	1	7	50,66633	7,11765	1,31540	0,99989	0,99989
19	9	1	8	38,24358	6,10453	0,96759	0,97442	0,97442
20	10	1	9	62,60347	7,89480	1,61833	0,99560	0,99560
21	11	1	10	45,10741	6,71478	1,17071	0,99958	0,99958

**Rapports et Graphiques**

	1	2	3	4	5	6	7	8	9	10	11	12	13
494													
495	<b>ANALYSE LINEAIRE - COEFFICIENTS DES FONCTIONS DE CLASSEMENT</b>												
496													
497													
498		Setosa	Versicolor	Virginica									
499	Constante	-85,34195	-72,20247	-103,06840									
500	lonsep	22,72256	14,66467	10,78294									
501	larsep	23,46884	8,91552	5,92113									
502	lonpetal	-14,98624	5,89627	13,90215									
503	larpetal	-17,29620	4,45679	19,29284									
504													
505	<b>DISCRIMINATION LINEAIRE</b>												
506													
507	<b>GRUPE OBSERVE : Setosa</b>												
508													
509	<b>JEU D'APPRENTISSAGE</b>												
510													
511	<b>L'INDIVIDU EST AFFECTE AU GROUPE DE PLUS FORTE PROBABILITE</b>												
512													
513													
514	INDIV/DU-GROUPE	P(Setosa)	P(Versicolor)	P(Virginica)									

**Rapports et Graphiques**

i148 - Virginica    0,00000    0,00198    0,99802

**ANALYSE LINEAIRE - RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE**

EN LIGNES, LES GROUPES OBSERVES  
EN COLONNES, LES GROUPES PREVUS

POURCENTAGE DE MAL CLASSES : 1,667 %  
POURCENTAGE DE BIEN CLASSES : 98,333 %

	Setosa	Versicolor	Virginica	Total
Setosa	43	0	0	43
Versicolor	0	37	1	38
Virginica	0	1	38	39
Total	43	38	39	120

**STATISTIQUES GROUPE OBSERVE : Setosa**

	lonsep	larsep	lonpetal	larpetal
Effectif	43,00000	43,00000	43,00000	43,00000
Moyenne	5,02326	3,44419	1,46512	0,25349



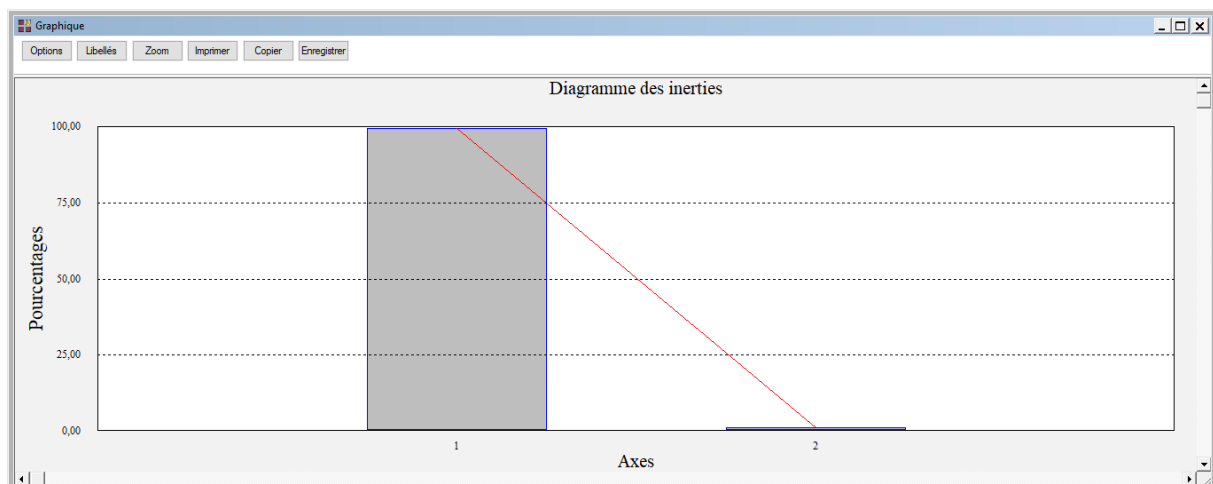
Ces rapports nous fournissent les renseignements suivants :

- Probabilités initiales
- Centroïdes des groupes et global
- Matrices des covariances et corrélations globales et des groupes
- Matrice des covariances et des corrélations intra-groupes
- Distances de Mahalanobis entre les groupes, Fishers, niveaux de signification
- Tableau des inerties (avec corrélation canonique, lambda de Wilks, Khi-2, degrés de liberté et niveau de signification)
- Test de Pillai
- Test de Box
- Fonctions discriminantes standardisées et non standardisées
- Résultats pour les variables et pour les individus
- Coordonnées des centres des groupes
- Coefficients des fonctions de classement (analyses linéaire et/ou quadratique)
- Résumé du classement d'apprentissage (analyses linéaire et/ou quadratique)
- Tableaux détaillés des classements par groupe (apprentissage)
- Statistiques pour les groupes observés et prévus (apprentissage)
- Résumé et détails du classement de la population de validation
- Classement de la population de prévision

## L'option Graphiques

- Diagramme des inerties

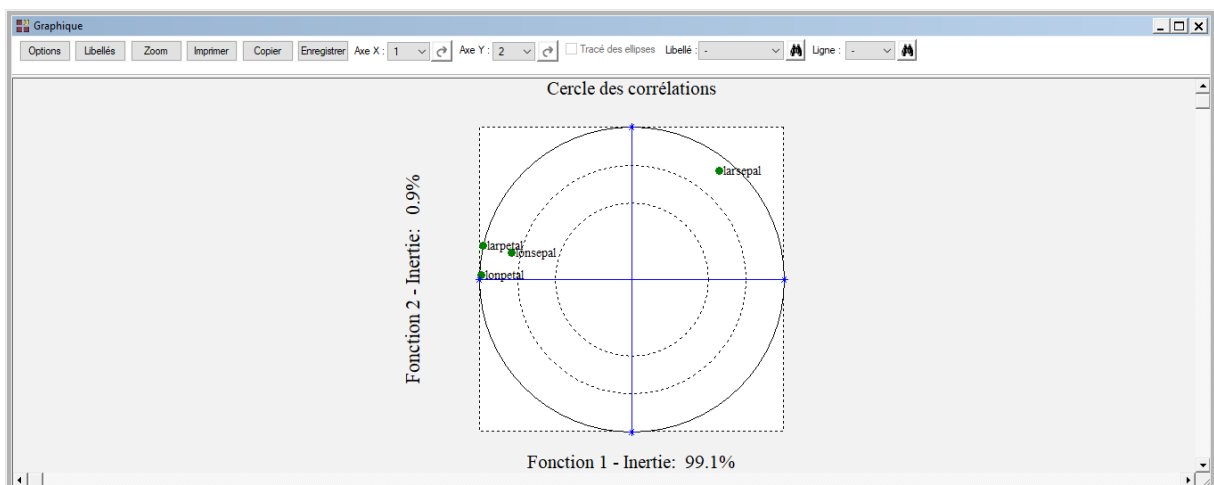
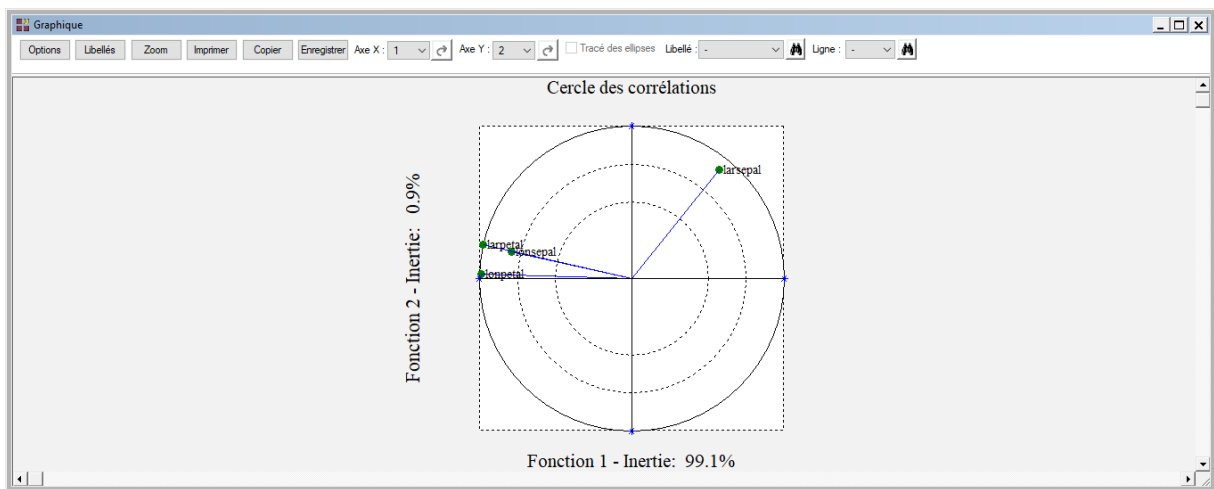
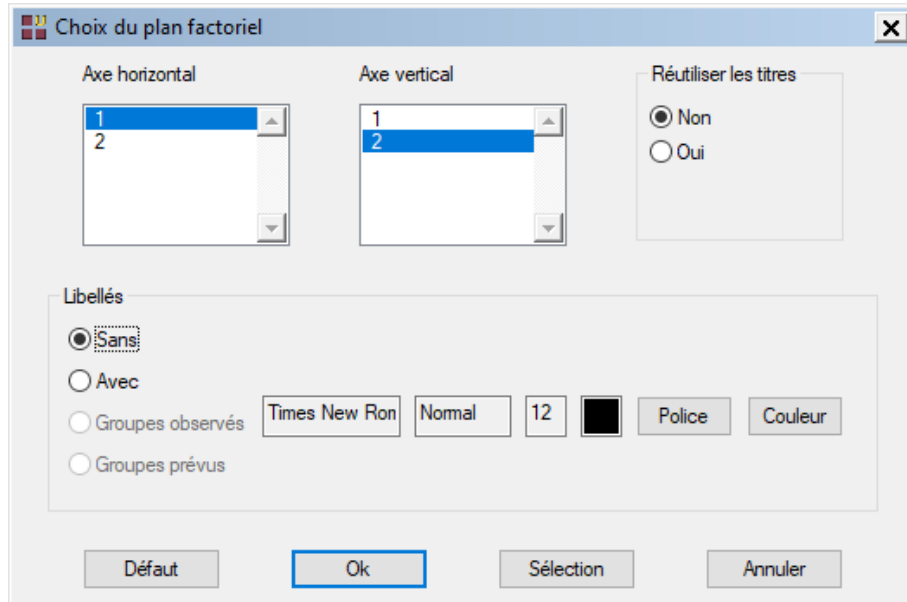
Ce graphique affiche les pourcentages d'inertie pour chacun des axes factoriels.



- Cercle factoriel des corrélations des variables

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

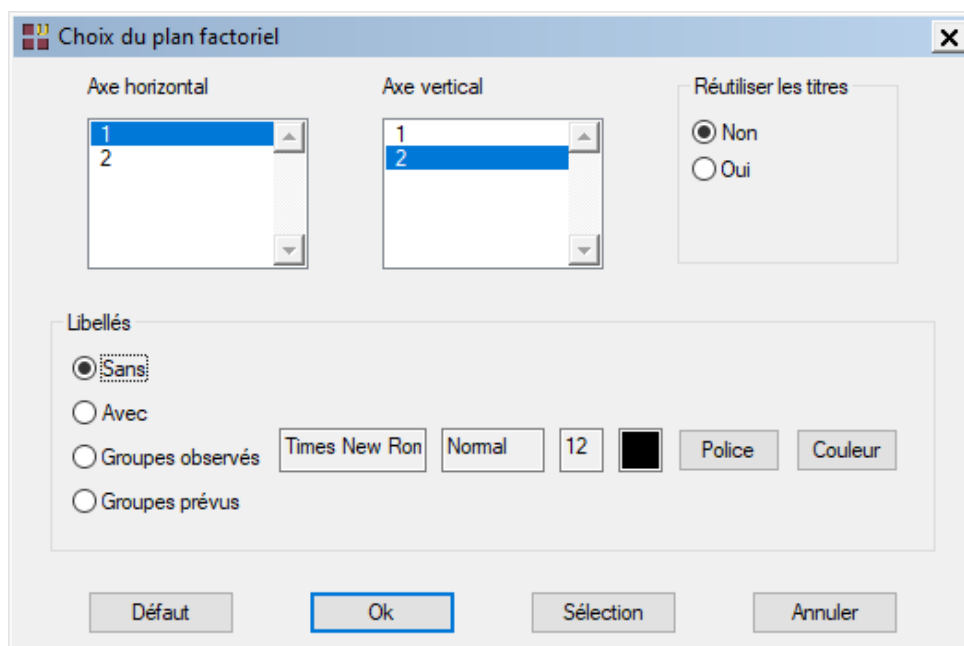
Choisissons les variables avec lignes puis sans lignes. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.



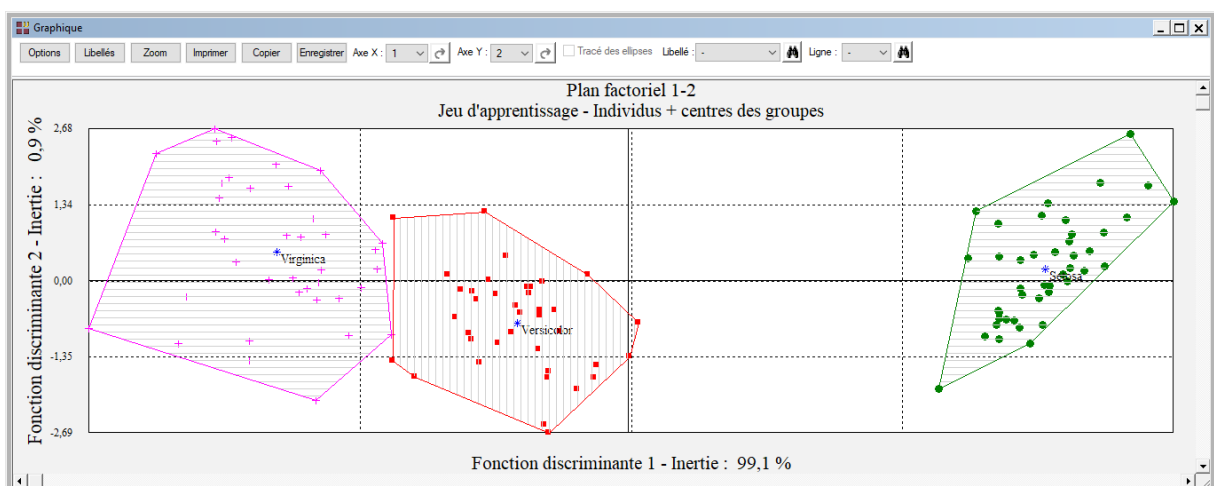
- Plan factoriel des individus et centres des groupes

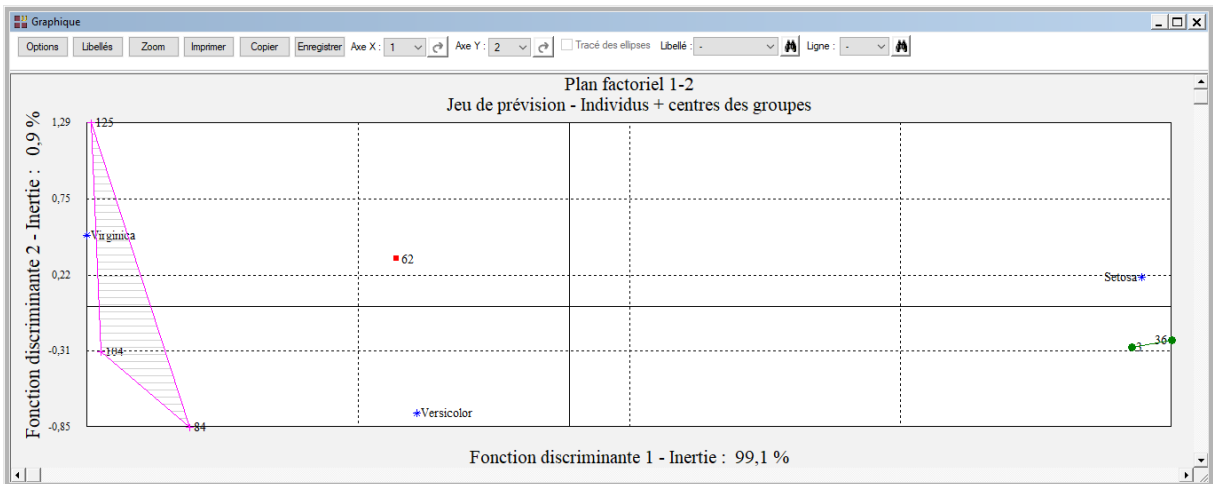
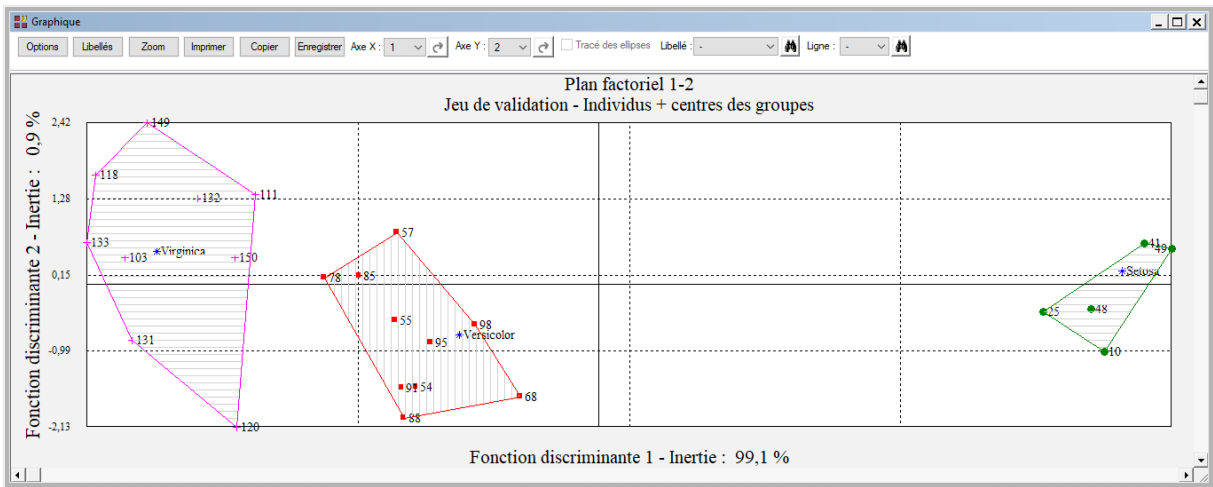
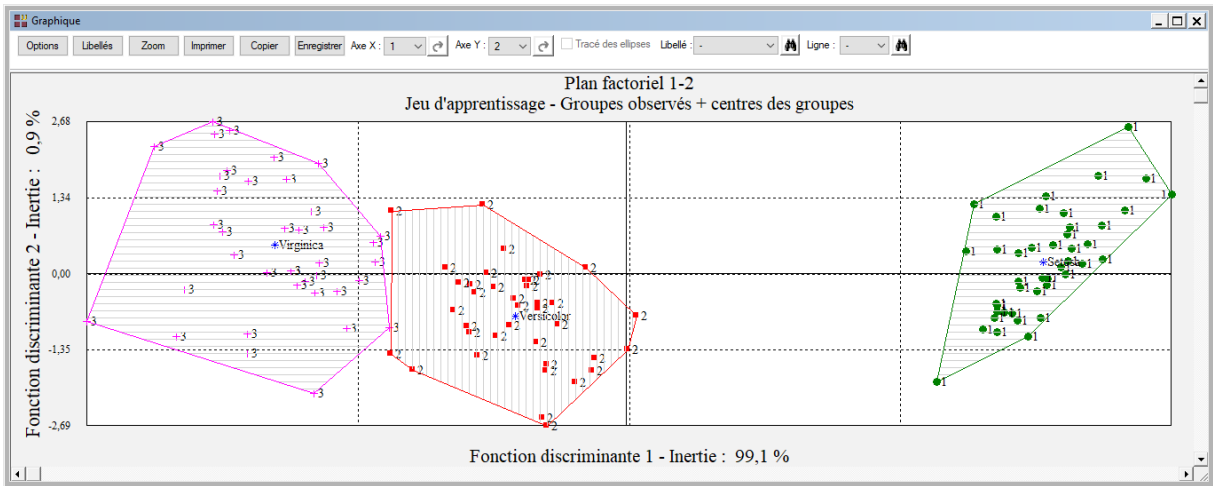
Ces options permettent d'afficher des plans factoriels des individus et des centres des groupes pour les populations d'apprentissage, de validation et de prévision. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.

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.



Des exemples de plans factoriels sont montrés ci-après.





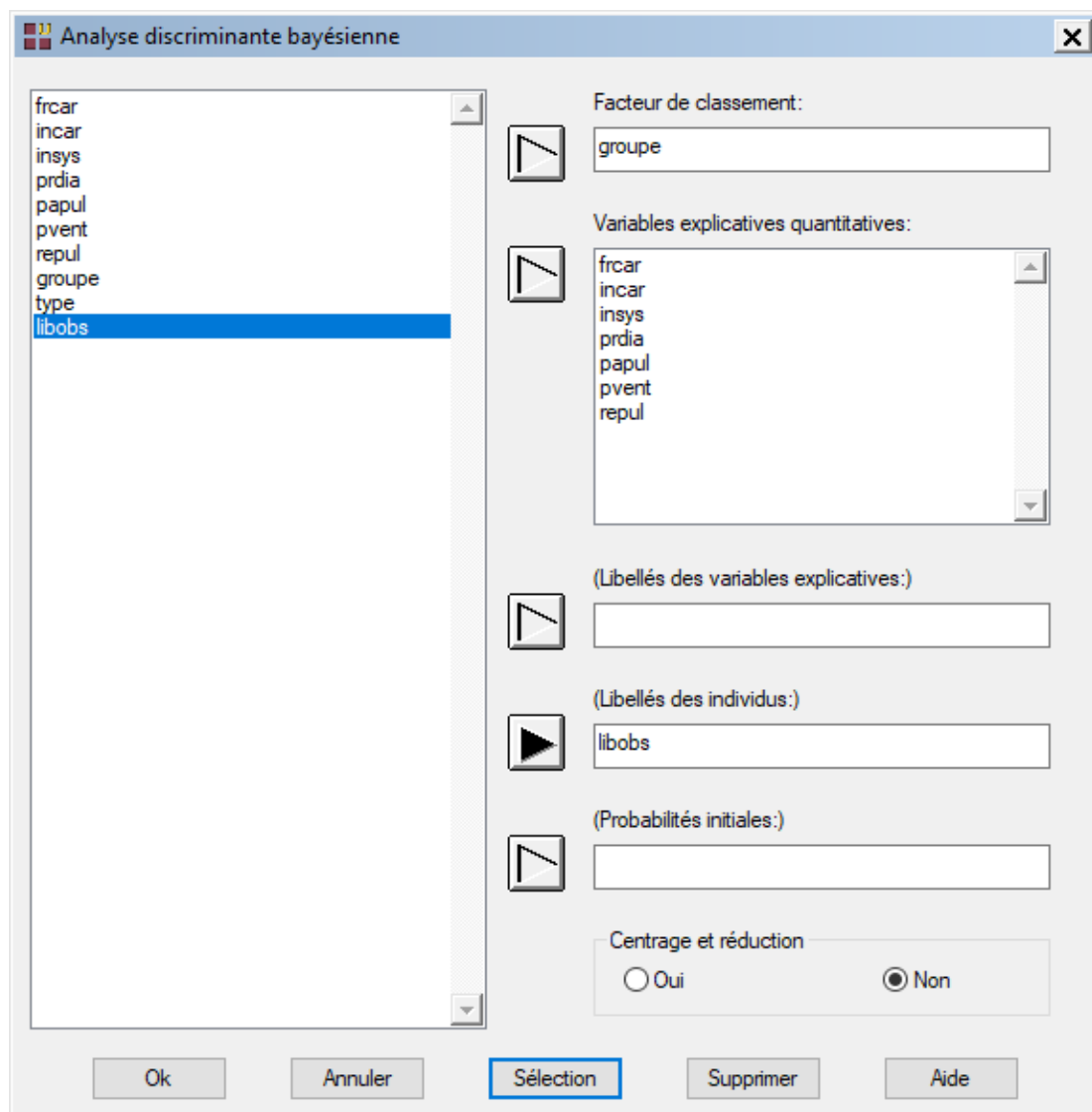


Les variables mesurées sont :

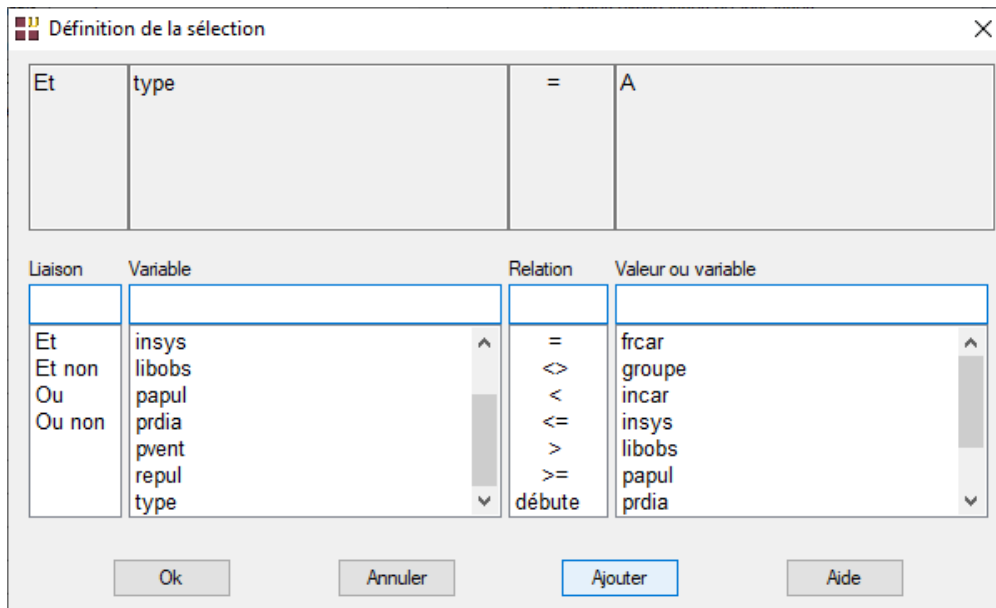
Variable	Mesure
<i>frcar</i>	fréquence cardiaque
<i>incar</i>	index cardiaque
<i>insys</i>	index systolique
<i>prdia</i>	pression diastolique
<i>papul</i>	pression artérielle pulmonaire
<i>pvent</i>	pression ventriculaire
<i>reput</i>	résistance pulmonaire

La variable *groupe* indique le groupe d'appartenance de chaque individu (Décès ou Survie). La variable *type* précise la population d'appartenance de chaque individu (A si apprentissage, V si validation). La variable *libobs* contient les libellés des individus des populations d'apprentissage et de validation.

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

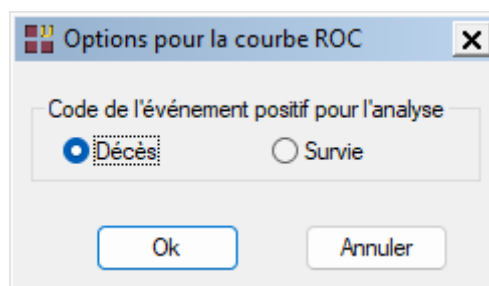


Cliquons sur le bouton Sélection pour définir la population d'apprentissage.



Un message nous indique que 81 individus sont sélectionnés.

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.



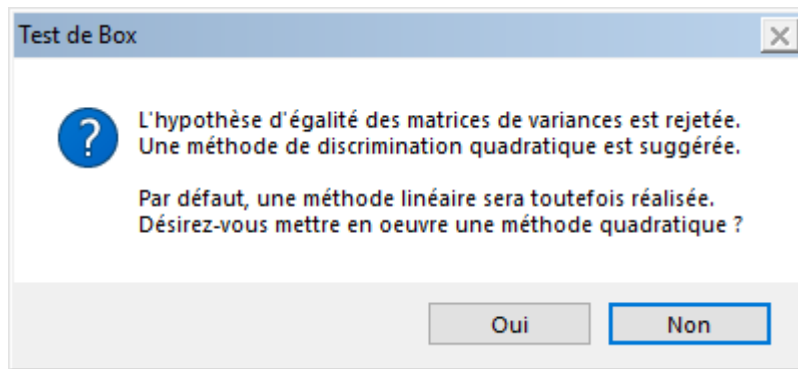
Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.

Après quelques instants, un tableau précisant l'inertie expliquée par l'unique vecteur propre issu de l'analyse s'affiche.

Tableau des Inerties										
Nombre d'axes à retenir : 1 Suite des calculs										
	1	2	3	4	5	6	7	8	9	10
1										
2	Tableau des Inerties									
3										
4	Composante	Valeur propre	Pct de variance	Pct cumulé	Variation	Corrél. canon.				
5	1	1,34960	100	100	0	0,75789				
6										

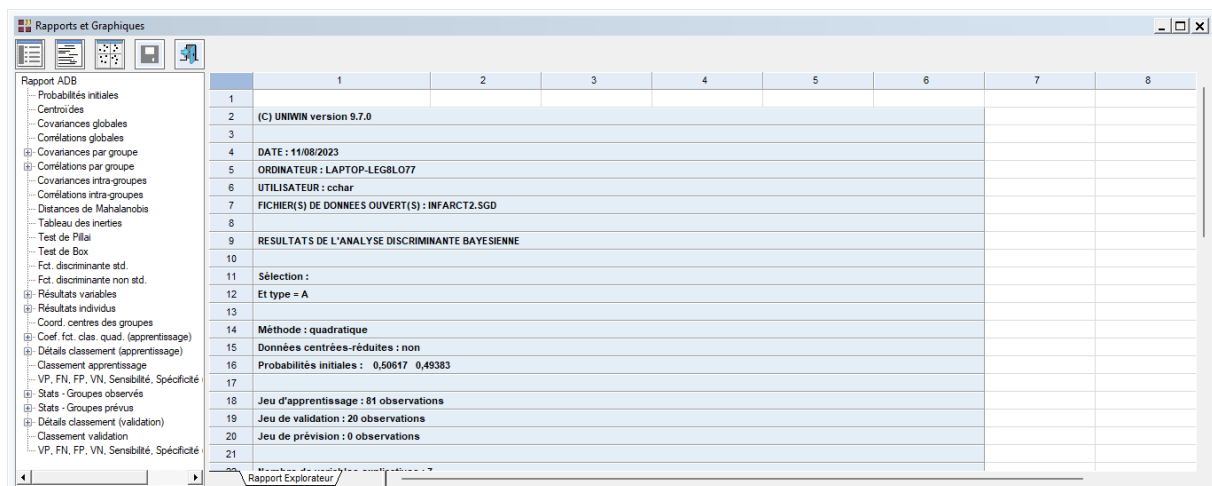
Aucun graphique n'est proposé, car il n'y a qu'une unique valeur propre.

Après quelques instants, UNIWIN affiche une boîte de dialogue vous indiquant si l'hypothèse d'égalité des matrices de variances est vérifiée ou non.



Dans notre exemple, l'hypothèse doit être rejetée et donc une analyse discriminante quadratique est suggérée. Choisissons donc de mettre en œuvre une analyse quadratique.

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



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données' de rappeler la boîte de dialogue d'entrée des données, par l'icône 'Rapports' d'afficher la boîte de dialogue des options pour les rapports et par et par l'icône 'Graphiques' d'afficher la boîte de dialogue des options pour les graphiques.

## 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.

Ces rapports nous fournissent les renseignements suivants :

- Probabilités initiales
- Centroides des groupes et global
- Matrices des covariances et corrélations globales et des groupes
- Matrice des covariances et des corrélations intra-groupes



- Distances de Mahalanobis entre les groupes, Fishers, niveaux de signification
- Tableau des inerties (avec corrélation canonique, lambda de Wilks, Khi-2, degrés de liberté et niveau de signification)
- Test de Pillai
- Test de Box
- Fonctions discriminantes standardisées et non standardisées
- Résultats pour les variables et pour les individus
- Coordonnées des centres des groupes
- Coefficients des fonctions de classement (analyses linéaire et/ou quadratique)
- Résumé du classement pour la population d'apprentissage (analyses linéaire et/ou quadratique)
- Tableaux détaillés des classements par groupe (apprentissage)
- VP, FN, FP, VN, sensibilité et spécificité pour la population d'apprentissage
- Statistiques pour les groupes observés et prévus (apprentissage)
- Résumé et détails du classement de la population de validation
- VP, FN, FP, VN, sensibilité et spécificité pour la population de validation
- Classement de la population de prévision (si existe)

	1	2	3	4	5	6	7	8
1								
2	RESULTATS INDIVIDUS 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								
		GRUPE	INDIVDU	DISTANCE^2	COORD.	CONTRIB.	COSINUS^2	COS^2 CUM
12	I1	1	1	0,12121	-0,34815	0,06531	1	
13	I2	1	2	2,68426	1,68931	1,55397	1	
14	I3	1	3	0,00002	-0,00455	0,00001	1	
15	I4	1	4	4,83268	2,19833	2,60372	1	
16	I5	1	5	1,90155	1,37897	1,02451	1	
17	I6	1	6	2,85242	1,68891	1,53881	1	
18	I7	1	7	1,29329	1,13723	0,69679	1	
19	I8	1	8	0,22116	0,47028	0,11915	1	
20	I9	1	9	3,80729	1,95123	2,05127	1	
21	I10	1	10	1,19438	1,09288	0,64350	1	

	1	2	3	4	5	6	7	8
1								
2	ANALYSE QUADRATIQUE - RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE							
3								
4	EN LIGNES, GROUPES OBSERVES							
5	EN COLONNES, GROUPES PREVUS							
6								
7	POURCENTAGE DE MAL CLASSES : 11,111 %							
8	POURCENTAGE DE BIEN CLASSES : 88,889 %							
9								
10								
		Décès	Survie	Total				
12	Décès	36	5	41				
13	Survie	4	36	40				
14	Total	40	41	81				
15								
16								
17								
18								
19								
20								
21								

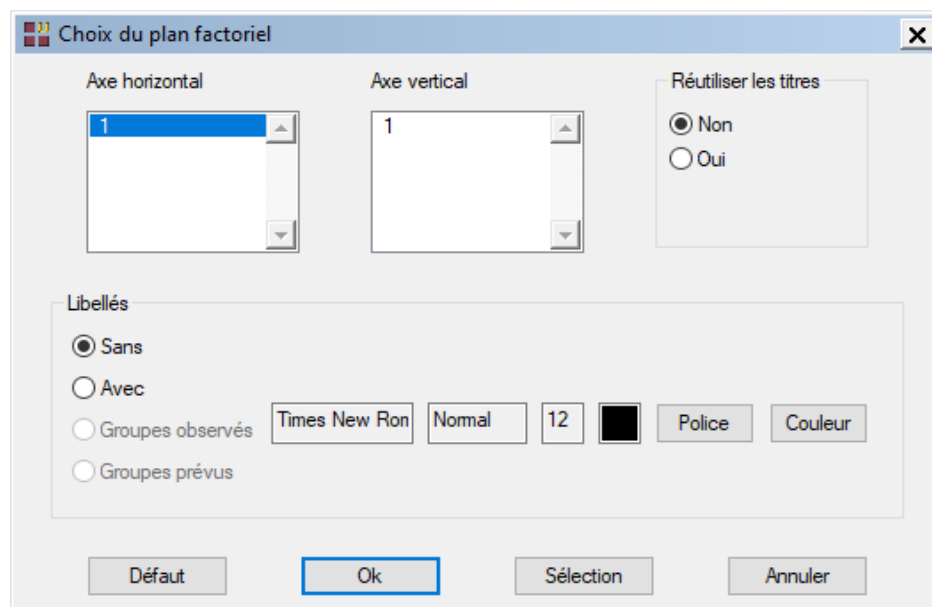
## L'option Graphiques

- Diagramme des inerties

Ce diagramme n'est pas disponible dans cet exemple car il n'y a qu'une seule composante.

- Cercle factoriel des corrélations des variables

Ces options permettent d'afficher le cercle de corrélations des variables et de choisir si on désire tracer les lignes reliant les points à l'origine du cercle. L'option sans ces lignes est utile lorsqu'il y a un grand nombre de variables représentées. Choisissons les variables avec lignes puis sans lignes. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.

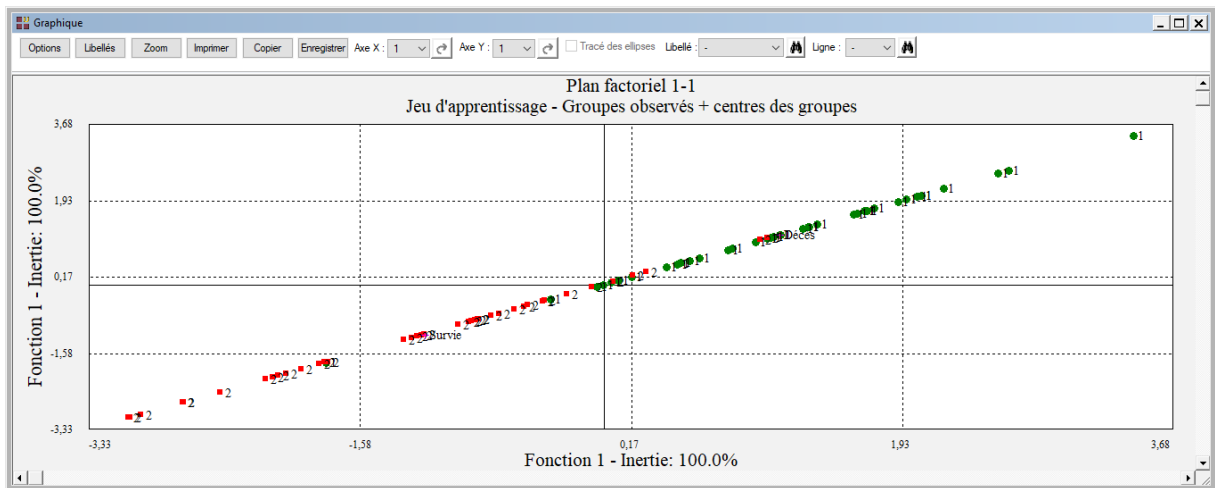
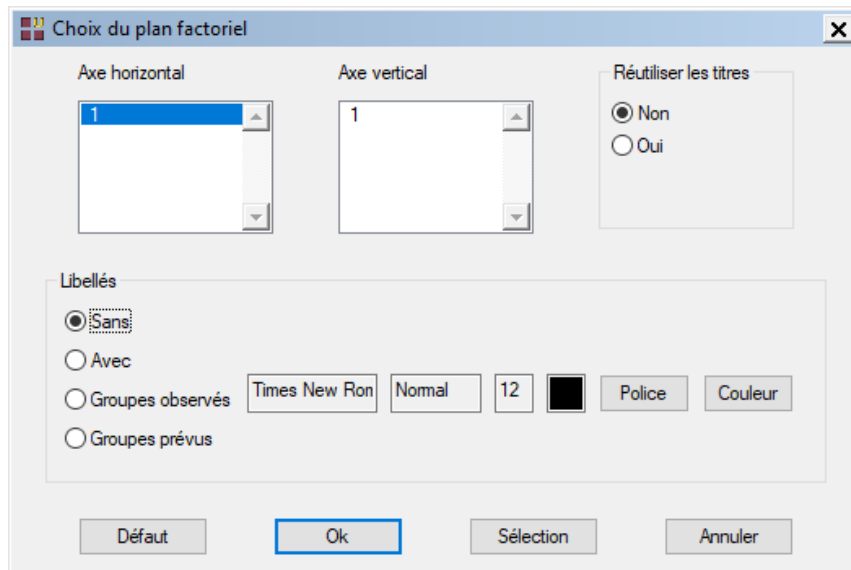


A noter que dans notre exemple, il n'y a qu'un axe factoriel et donc le tracé n'est pas proposé.

- Plan factoriel des individus et centres des groupes

Ces options permettent d'afficher des plans factoriels des individus et des centres des groupes pour les populations d'apprentissage, de validation et de prévision.

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



Refaisons maintenant cette même analyse en choisissant une méthode linéaire.

Pour cela, il faut refaire l'analyse actuelle en choisissant 'Non' lorsque la boîte de dialogue 'Test de Box' s'affiche.

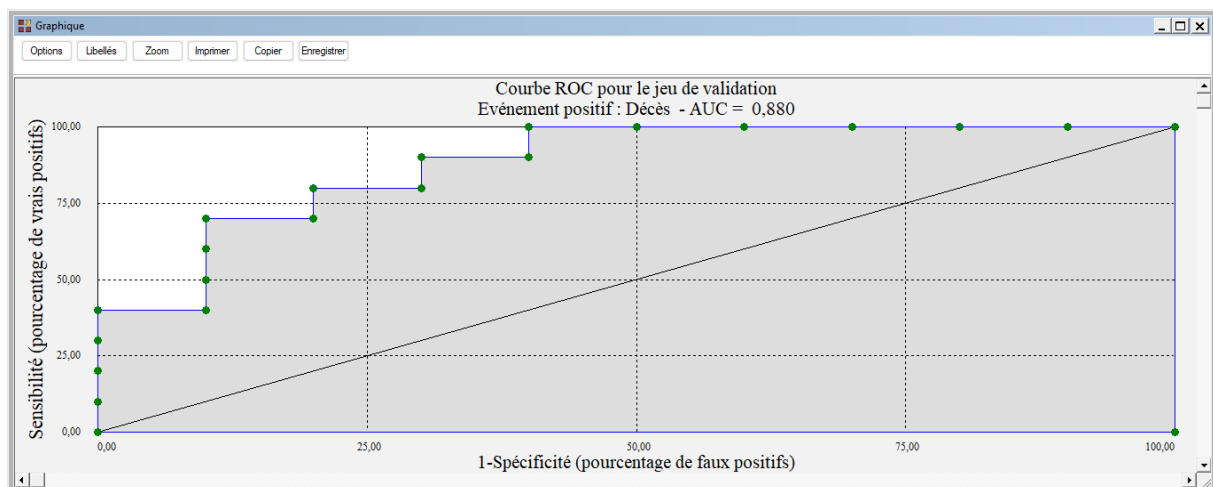
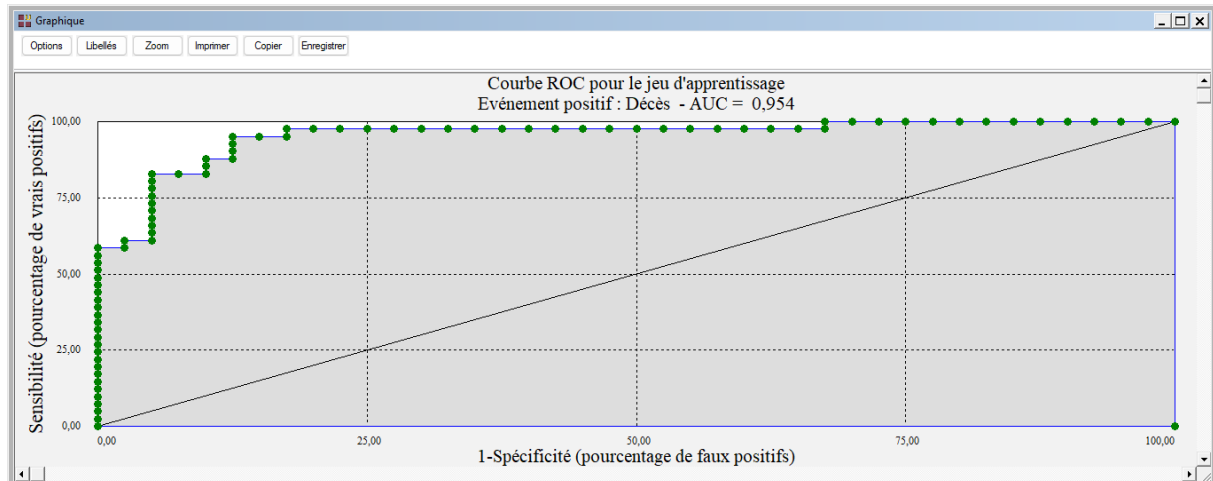
	1	2	3	4	5	6	7	8
1								
2	<b>ANALYSE LINEAIRE - RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE</b>							
3								
4	<b>EN LIGNES, LES GROUPES OBSERVES</b>							
5	<b>EN COLONNES, LES GROUPES PREVUS</b>							
6								
7	<b>POURCENTAGE DE MAL CLASSES : 9,877 %</b>							
8	<b>POURCENTAGE DE BIEN CLASSES : 90,123 %</b>							
9								
10								
11			Décès	Survie	Total			
12	Décès	38	3	41				
13	Survie	5	35	40				
14	Total	43	38	81				
15								
16								
17								
18								
19								
20								
21								

Le pourcentage d'erreur de classement (9,88 %) est meilleur que pour l'analyse quadratique (11,11 %) mais l'hypothèse d'égalité des matrices des variances, non observée, y est supposée.

- Courbe ROC

Le tracé de la courbe ROC et le calcul de l'aire sous la courbe (AUC) sont possibles car la variable à expliquer possède deux modalités.

Visualisons les courbes pour le jeu d'apprentissage et le jeu de validation.



### Exemple 3 : Fichier BORDEAUX

Pour ce troisième exemple, nous utiliserons le fichier BORDEAUX.

Ce fichier contient des informations sur la qualité de vins de Bordeaux en relation avec les conditions météorologiques.

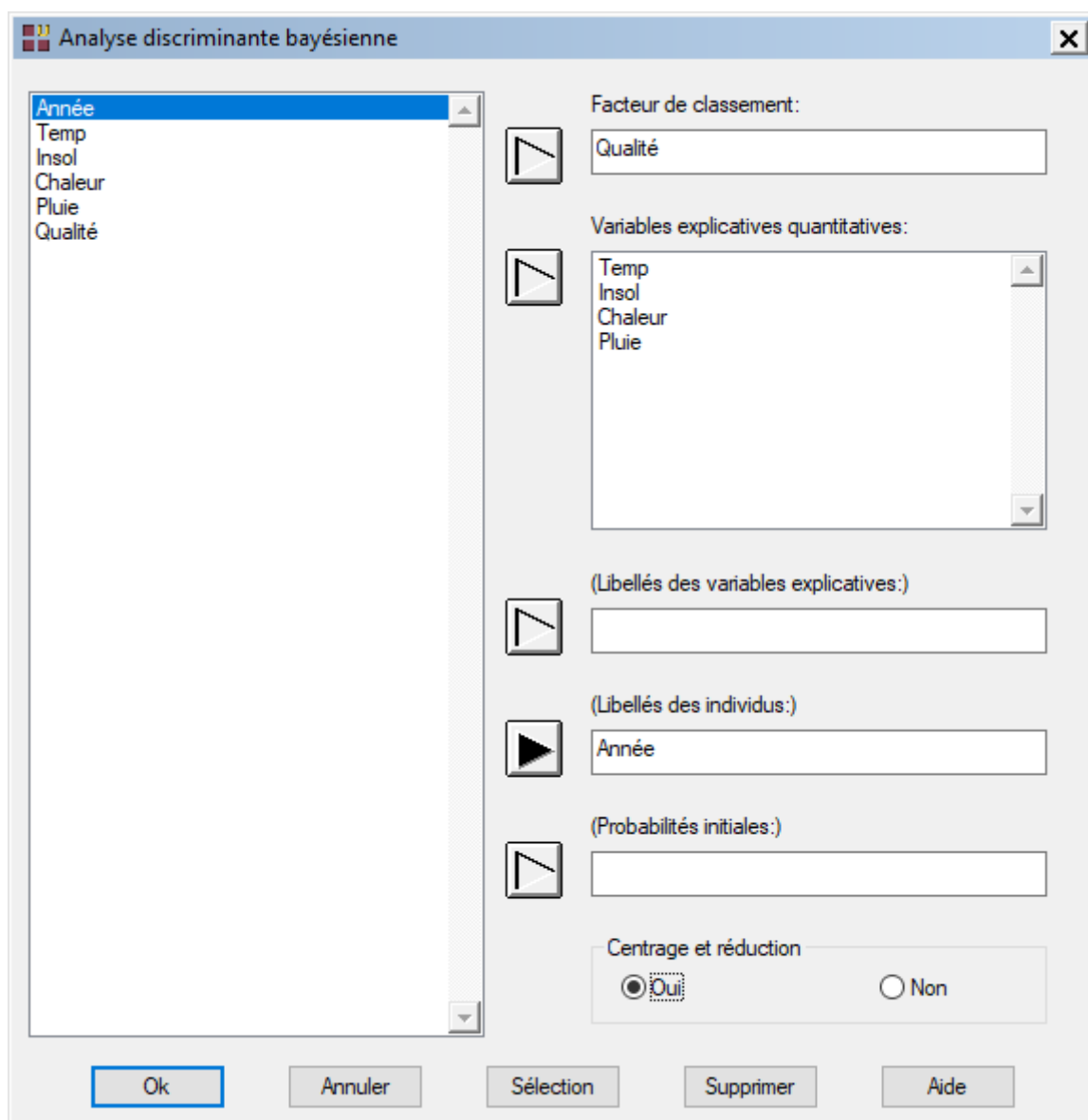
La variable qualitative *Qualité*, facteur de classement, prend trois modalités : Bon, Moyen et Médiocre.

Les variables explicatives sont :

- *Temp* somme des températures moyennes journalières (° C)
- *Insol* durée d'insolation (heures)
- *Chaleur* nombre de jours de grande chaleur
- *Pluie* hauteur des pluies (millimètres)

La variable *Année* fournit les libellés des individus.

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



Nous sélectionnons une analyse centrée et réduite.

La fenêtre 'Rapports et Graphiques' s'affiche pour la méthode linéaire.

Visualisons les résultats du classement.

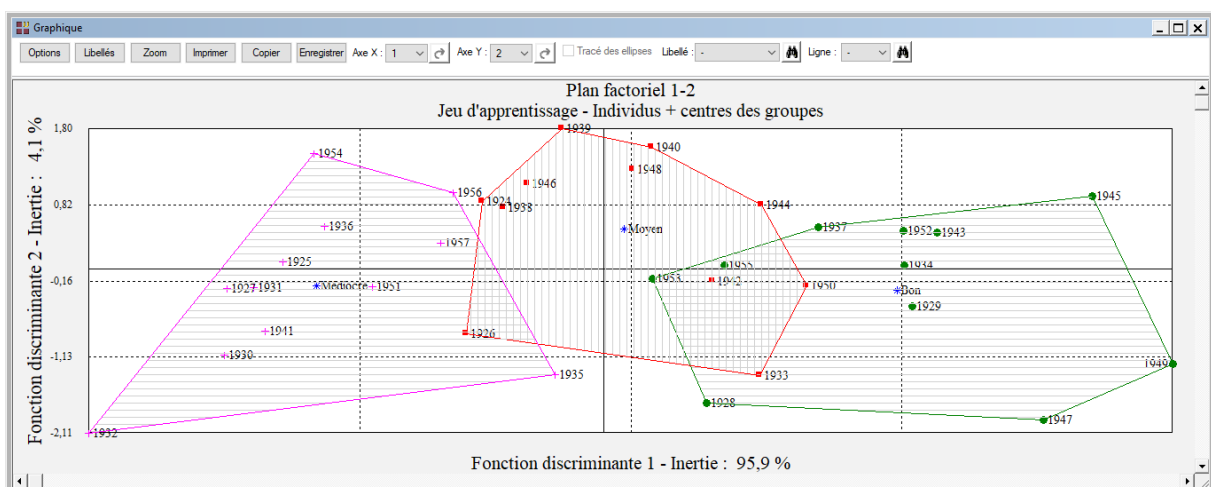
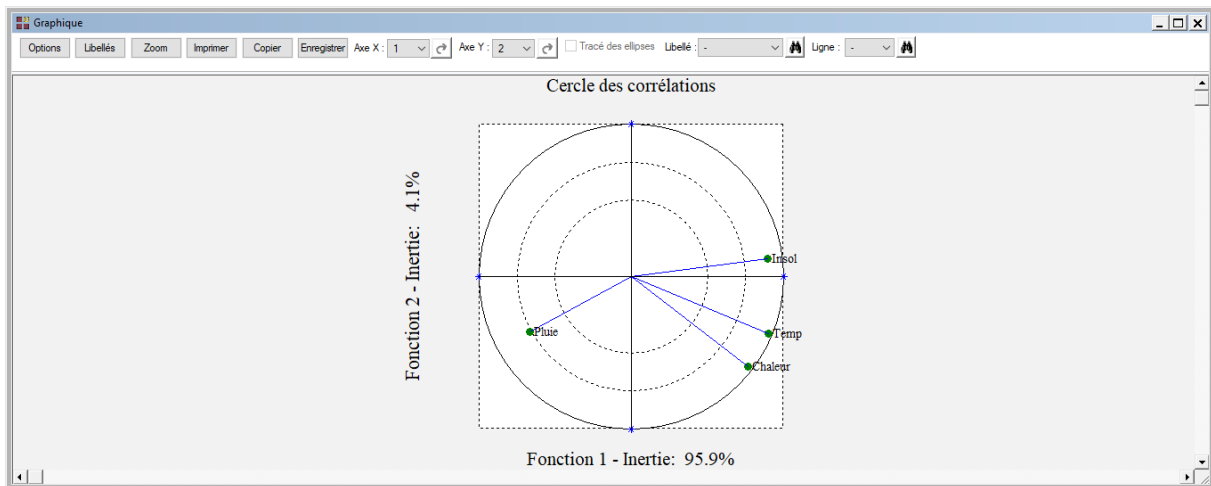
Les vins mal classés sont les vins des années :

1953, 1955      bons classés moyens  
 1933, 1950      moyens classés bons  
 1926              moyen classé médiocre  
 1935, 1956      médiocres classés moyens

	1	2	3	4	5	6	7	8
1								
2	ANALYSE LINEAIRE - 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 : 20,588 %							
8	POURCENTAGE DE BIEN CLASSES : 79,412 %							
9								
10								
11								
12	Bon		Bon	Moyen	Médiocre	Total		
13	Bon	9	2	0	11			
14	Moyen	2	8	1	11			
15	Médiocre	0	2	10	12			
16	Total	11	12	11	34			

	1	2	3	4	5	6	7	8
1								
2	DISCRIMINATION LINEAIRE							
3								
4	GROUPE OBSERVE : Bon							
5								
6	JEU D'APPRENTISSAGE							
7								
8	L'INDIVIDU EST AFFECTE AU GROUPE DE PLUS FORTE PROBABILITE							
9								
10								
11	INDIVIDU-GROUPE	P(Bon)	P(Moyen)	P(Médiocre)				
12	1928 - Bon	0,64171	0,32703	0,03128				
13	1929 - Bon	0,93341	0,06650	0,00009				
14	1934 - Bon	0,89244	0,10745	0,00012				
15	1937 - Bon	0,62214	0,37672	0,00114				
16	1943 - Bon	0,90475	0,09520	0,00004				
17	1945 - Bon	0,98380	0,01620	0,00000				
18	1947 - Bon	0,99656	0,00344	0,00000				
19	1949 - Bon	0,99904	0,00096	0,00000				
20	1952 - Bon	0,85207	0,14781	0,00012				
21	1953 - Moyen *	0,19487	0,75191	0,05322				
22	1955 - Moyen *	0,38292	0,60530	0,01178				

Demandons également le cercle factoriel et le plan factoriel pour les axes 1 et 2.



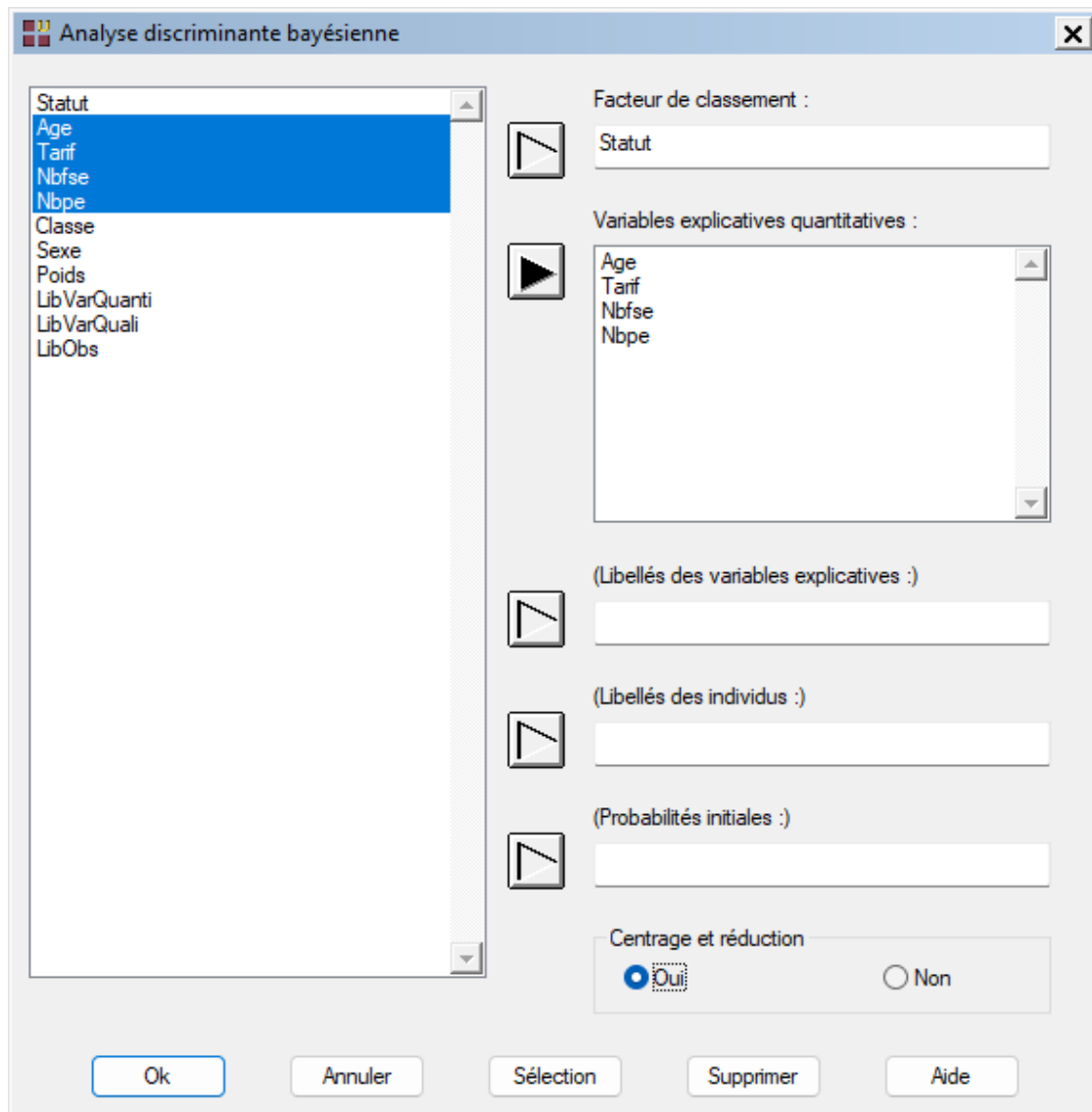
### Exemple 4 : Fichier TITANIC

Pour ce quatrième exemple, nous utiliserons le fichier TITANIC.

Ce fichier contient des informations concernant 714 passagers :

Statut	Survie ou Décès
Classe	Classe du passager (1 <sup>ère</sup> , 2 <sup>ème</sup> ou 3 <sup>ème</sup> )
Sexe	Homme ou Femme
Age	Age du passager
Nbfse	Nombre de frères, sœurs ou époux, épouses à bord
Nbpe	Nombre de parents ou enfants à bord
Tarif	Tarif passager (en £)

Cliquons sur l'icône ADB dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-dessous.



Après exécution de la procédure, visualisons le tableau de classement des données d'apprentissage et la courbe ROC associée.

Rapports et Graphiques

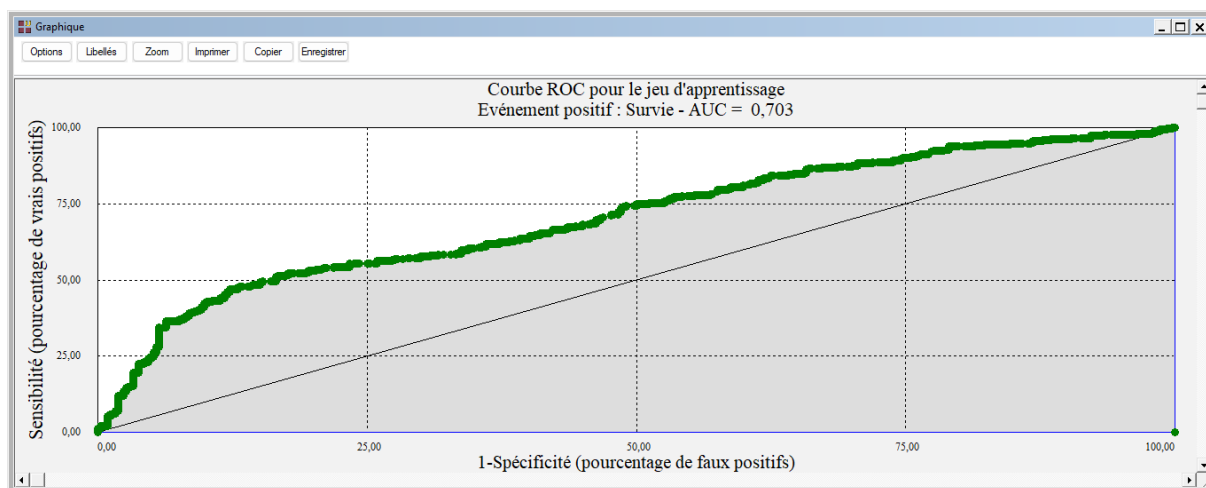
Rapport ADB

- Probabilités initiales
- Centroides
- Covariances globales
- Corrélations globales
- Covariances par groupe
- Corrélations par groupe
- Covariances intra-groupes
- Corrélations intra-groupes
- Distances de Mahalanobis
- Tableau des inerties
- Test de Pits
- Test de Box
- Fct. discriminante std.
- Fct. discriminante non std.
- Résultats variables
- Résultats individuels
- Coord. centres des groupes
- Coef. fct. clas. quad. (apprentissage)
- Détails classement (apprentissage)
- Classement apprentissage**
- VP, FN, FP, VN, Sensibilité, Spécificité
- Stats - Groupes observés
- Stats - Groupes prévus

	1	2	3	4	5	6	7	8
1								
2	<b>ANALYSE QUADRATIQUE - RESULTATS DU CLASSEMENT POUR LE JEU D'APPRENTISSAGE</b>							
3								
4	<b>EN LIGNES, GROUPES OBSERVES</b>							
5	<b>EN COLONNES, GROUPES PREVUS</b>							
6								
7	POURCENTAGE DE MAL CLASSES : 30,952 %							
8	POURCENTAGE DE BIEN CLASSES : 69,048 %							
9								
10								
11			Décès	Survie	Total			
12	Décès		402	22	424			
13	Survie		199	91	290			
14	Total		601	113	714			
15								
16								
17								
18								
19								
20								
21								

Rapport Explorateur /





Environ 69 % des passagers sont bien classés par cette analyse et l'aire sous la courbe ROC est proche de 0,7.

Note : Pour comparer les performances de plusieurs méthodes d'analyse, cet exemple est traité dans les six analyses AFD, ADB, KNN, BAYES, ANN et ARBRE.

### 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>
fdstdA	Fonctions discriminantes standardisées (apprentissage)
fdnstdA	Fonctions discriminantes non standardisées (apprentissage)
cindA	Coordonnées des individus (apprentissage)
libindA	Libellés des individus (apprentissage)
clindA	Classes des individus (apprentissage)
cindV	Coordonnées des individus (validation)
libindV	Libellés des individus (validation)
clindV	Classes des individu (validation)
distindA	Distances carrées à l'origine pour les individus (apprentissage)
cosindA	Cosinus carrés pour les individus (apprentissage)
conindA	Contributions pour les individus (apprentissage)
cvarA	Coordonnées des variables (apprentissage)
disvarA	Distances carrées à l'origine des variables (apprentissage)
cosvarA	Cosinus carrés des variables (apprentissage)
convarA	Contributions des variables (apprentissage)

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)

si modèle linéaire :

coefcl	Fonctions de classement linéaire (apprentissage)
classA	Classement linéaire (apprentissage)
classV	Classement linéaire (validation)
classP	Classement linéaire (prévision)
libindP	Libellés des individus (prévision)

si modèle quadratique :

ccq(i)	Constantes classement quadratique groupe i (apprentissage)
coefcq(i)	Fonctions de classement quadratique groupe i (apprentissage)
classqA	Classement quadratique (apprentissage)
classqV	Classement quadratique (validation)
classqP	Classement quadratique (prévision)
libindP	Libellés des individus (prévision)