

UNIWIN VERSION 10.4.0

ANALYSE FACTORIELLE DISCRIMINANTE

Révision : 15/09/2025

Définition.....	1
Entrée des données	2
Données manquantes	2
Exemple 1 : Fichier IRIS3.....	3
L'option Rapports	7
L'option Graphiques	8
Exemple 2 : Fichier INFARCT2	13
L'option Rapports	15
L'option Graphiques	16
Exemple 3 : Fichier BORDEAUX.....	19
Exemple 4 : Fichier TITANIC	21
Ellipses de tolérance	23
Calculs de la matrice de confusion et des indicateurs	24
Les variables internes créées par la procédure	25

Définition

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. Les données des variables explicatives sont automatiquement centrées et réduites.

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 :

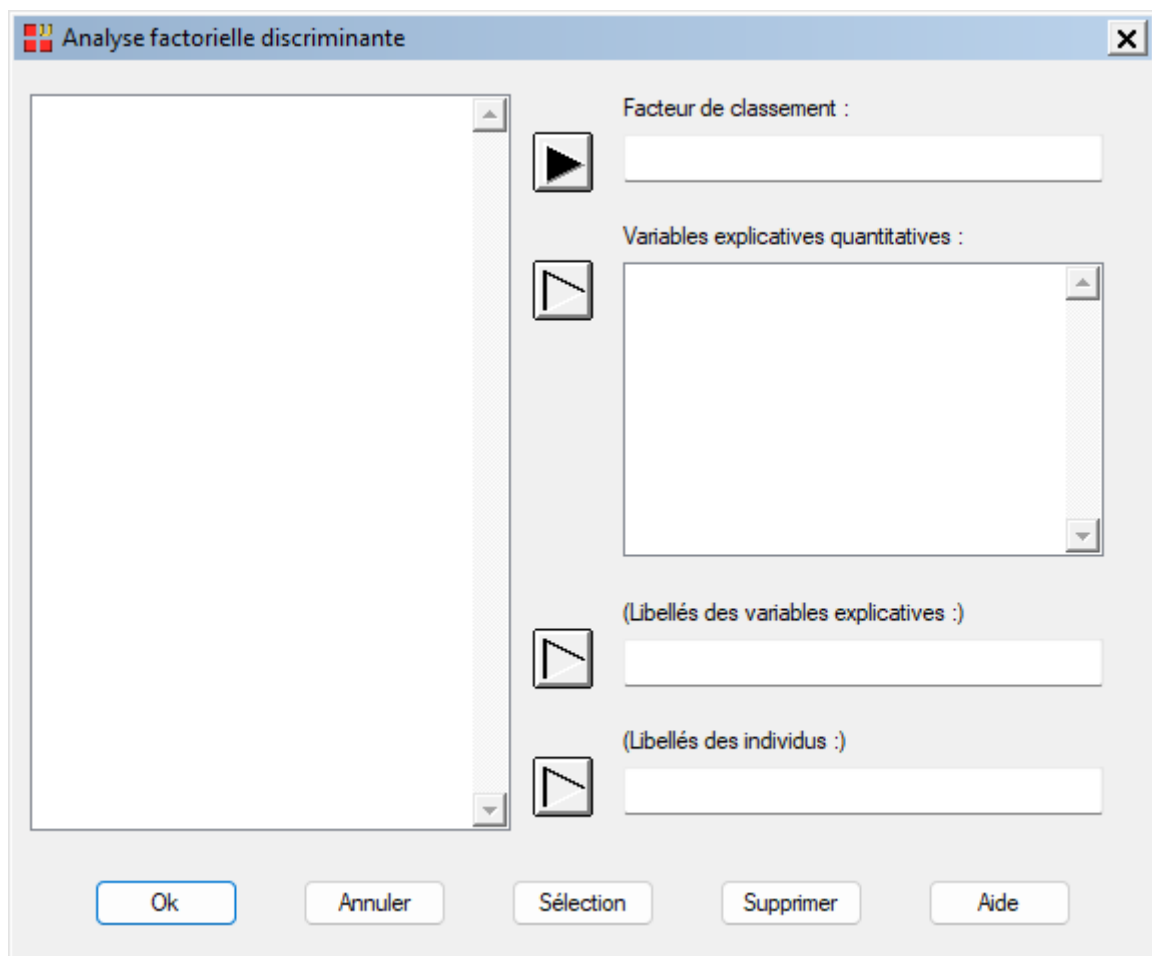
- une étude initiale sur la population d'apprentissage de taille n_1
- une étude de validation sur la population de validation de taille n_2
- une étude prospective sur une population de prévision de taille n_3

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 AFD dans le ruban Expliquer. La boîte de dialogue montrée ci-après apparaît.

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



Données manquantes

Les données manquantes ne sont pas autorisées pour le facteur de classement.

Elles sont autorisées pour les variables quantitatives.

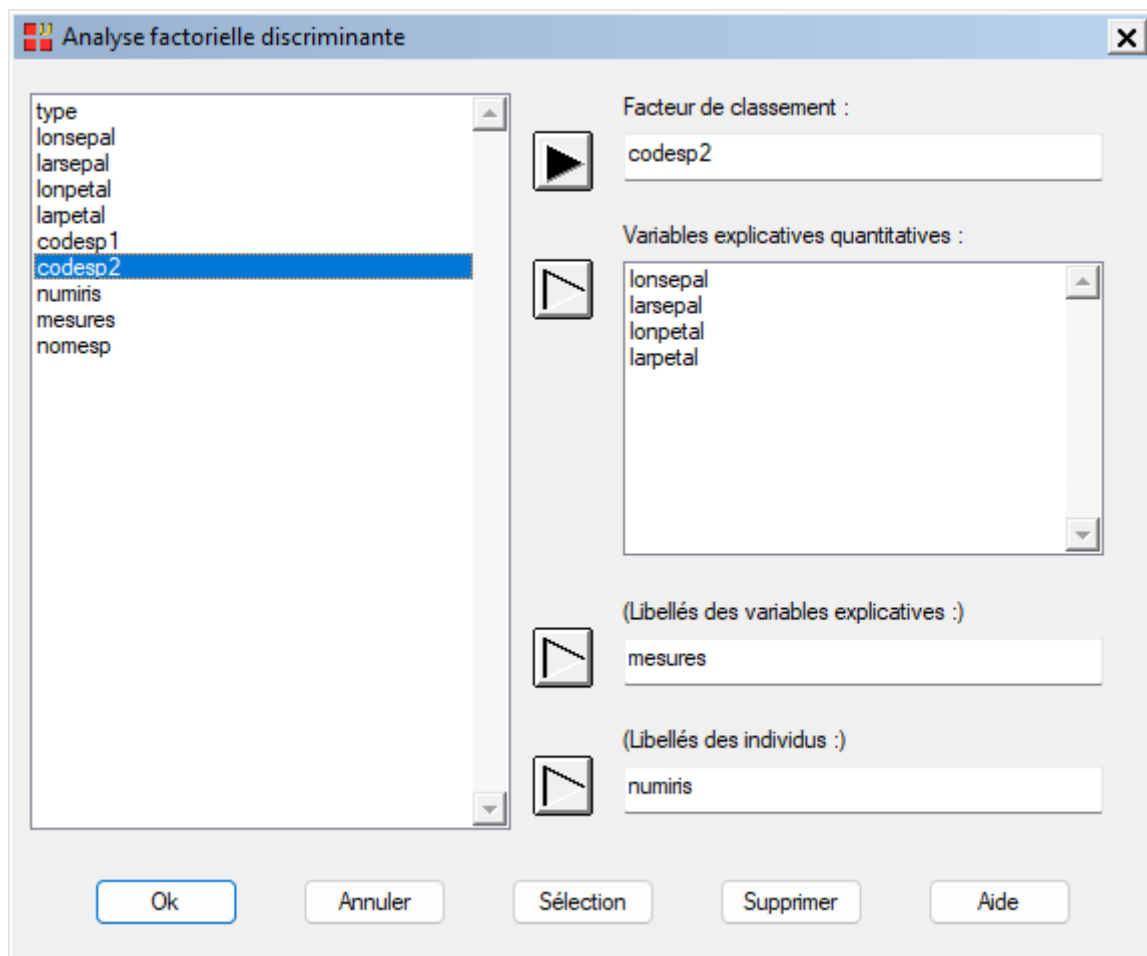
Exemple 1 : Fichier IRIS3

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

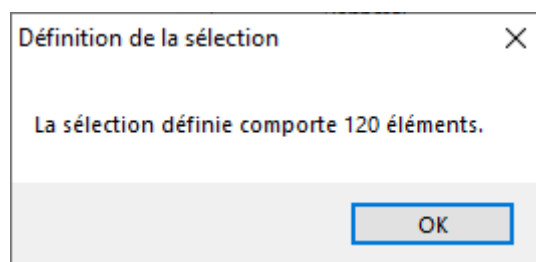
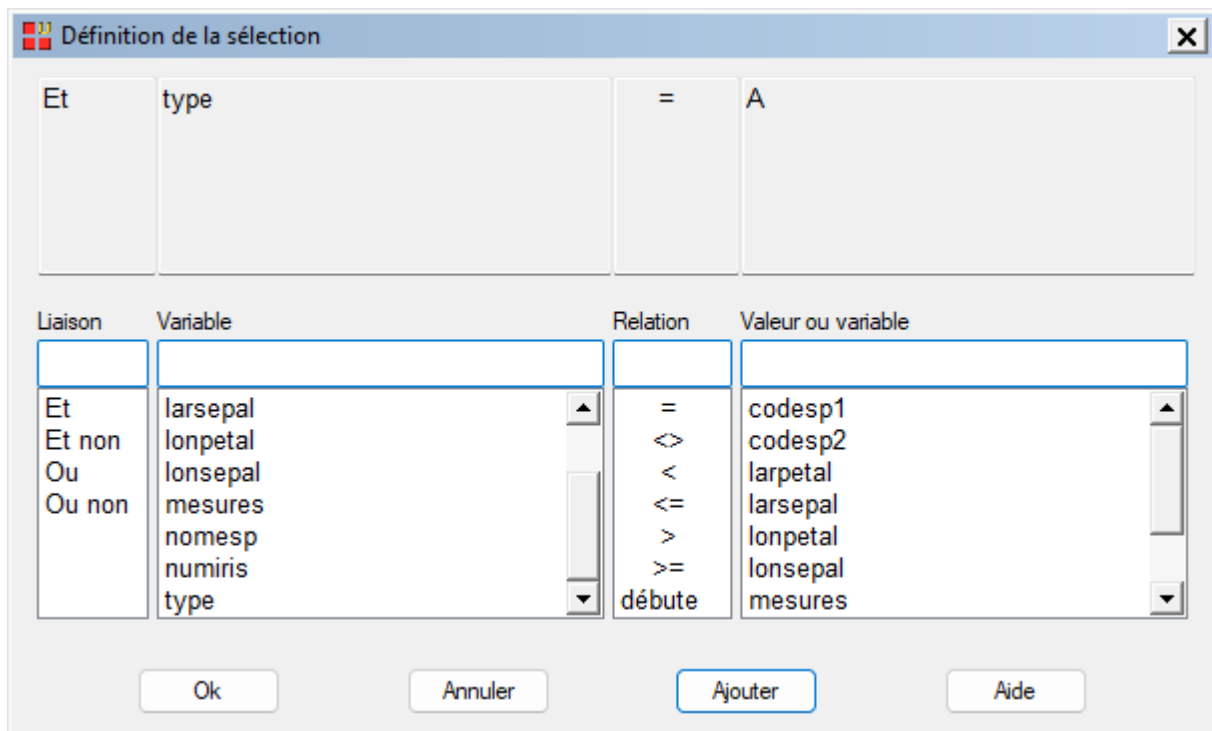
Cliquons sur l'icône AFD dans le menu Expliquer.

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, la variable *numiris* comme variable contenant les libellés des individus.

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



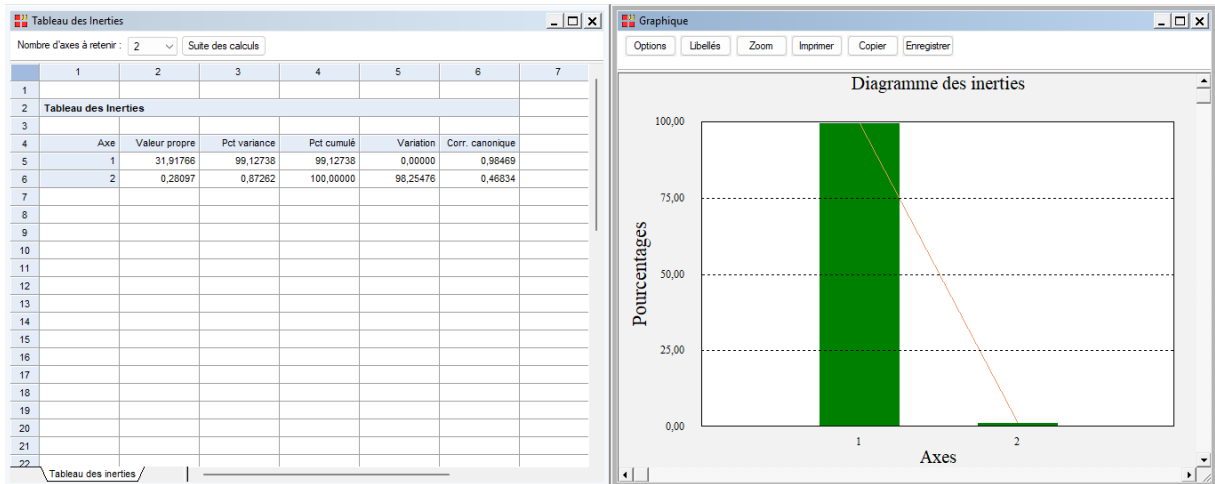
Cette sélection comporte 120 individus qui constituent la population d'apprentissage.

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 Factorielle Discriminante.

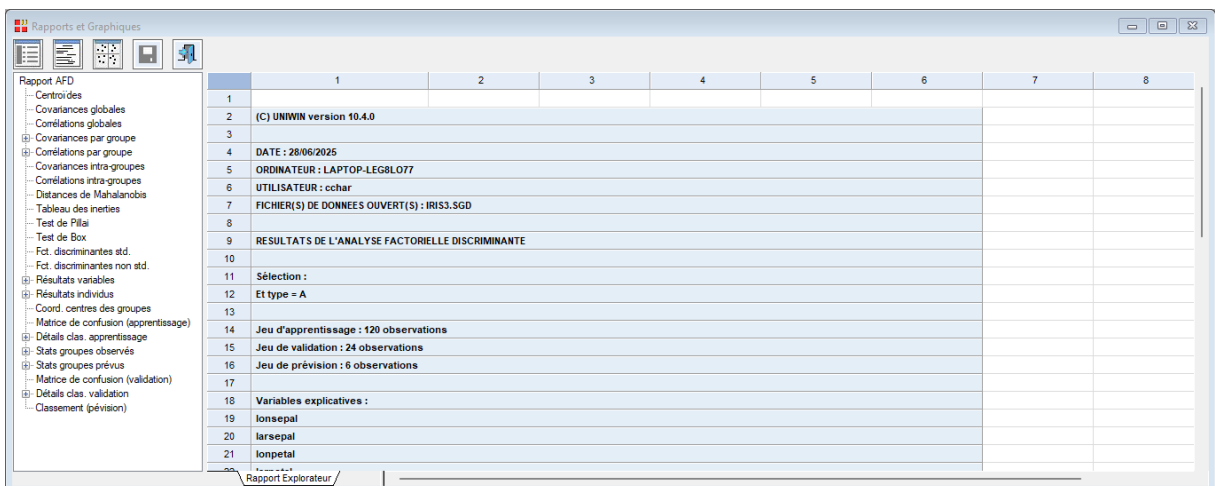
Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.





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

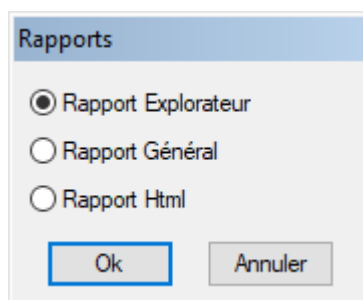
Cliquons sur le bouton 'Suite des calculs'.


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

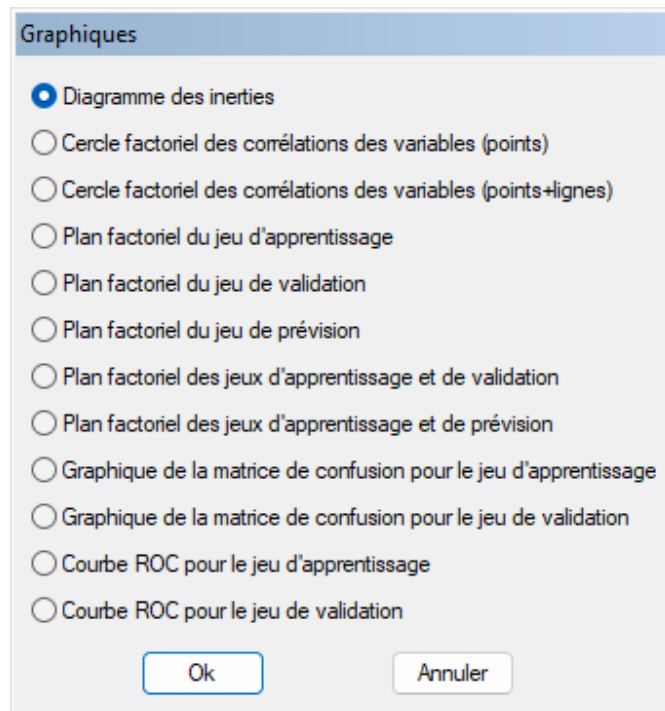



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

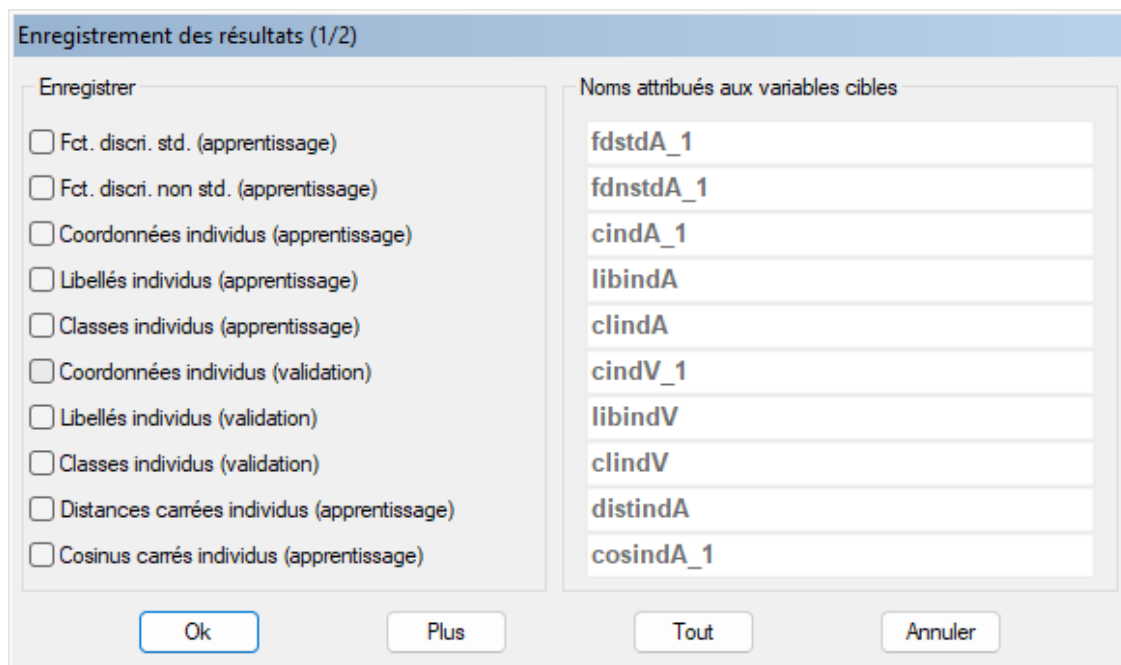
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



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

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

L'option Rapports

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

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

Rapports et Graphiques

Rapport AFD

- Centroides
- Covariances globales
- Corrélations globales
- Covariances par groupe
- Corrélations par groupe
- Covariances intra-groupes
- Corrélations intra-groupes
- Distances de Mahalanobis
- Tableau des inertes
- Test de Pillai
- Test de Box
- Fct. discriminantes std.
- Fct. discriminantes non std.
- Résultats variables
- Résultats individus
- Facteur 1
- Facteur 2
- Points multiples individus
- Coord. centres des groupes
- Matrice de confusion (apprentissage)
- Détails clas. apprentissage
- Stats groupes observés
- Stats groupes prévus
- Matrice de confusion (validation)
- Détails clas. validation
- Classement (prévision)

	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	COORDONNEE = COORDONNÉES DES INDIVIDUS								
6	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE								
7	COSINUS*2 = COSINUS CARRÉS								
8	COS*2 CUMULE = SOMMES CUMULÉES DES COSINUS CARRÉS								
9									
10									
11			GRUPE	INDIV/DU	DISTANCE*2	COORDONNEE	CONTRIBUTION	COSINUS*2	COS*2 CUMUL
12	1		1	1	57,32730	7,56828	1,48724	0,99916	0,9991
13	2		1	2	45,52882	6,89560	1,16403	0,98467	0,9846
14	4		1	3	40,65299	6,34950	1,04384	0,98890	0,9889
15	5		1	4	58,20534	7,61805	1,50807	0,99655	0,9965
16	6		1	5	53,47025	7,18429	1,34015	0,98528	0,9852
17	7		1	6	45,09110	6,70541	1,16744	0,99715	0,9971
18	8		1	7	50,68633	7,11765	1,31540	0,99989	0,9998
19	9		1	8	38,24358	6,10453	0,96759	0,97442	0,9744
20	11		1	9	62,60347	7,89480	1,61833	0,99560	0,9956
21	12		1	10	45,10741	6,71478	1,17071	0,99958	0,9995

Rapport Explorateur /

Rapports et Graphiques

	1	2	3	4	5	6	7	8	9	10	11	12	13
139	TABLEAU DES INERTIES												
140													
141	La corrélation canonique indique la force de la relation entre la fonction discriminante et le facteur de classement.												
142													
143	Le lambda de Wilks teste l'hypothèse d'égalité des vecteurs moyens des différentes classes.												
144													
145	Les fonctions discriminantes qui ont des probabilités inférieures à 0,05 sont statistiquement significatives au niveau de confiance de 95,0%.												
146													
147													
148		VALEUR PROPRE	PCT VARIANCE	PCT CUMULE	VARIATION	CORR. CANON.	LAMBDA WILKS	KHI-2	DEGRES LIBERTE	NIV. SIG.			
149													
150	1	31,91766	99,12738	0,00000	0,98469	0,02372	432,15810	8	0,00000				
151	2	0,28097	0,87282	100,00000	98,25476	0,46834	0,78066	28,60003	3	0,00000			
152													
153	TEST DE PILLAI												
154													
155	Il permet de tester l'hypothèse nulle H0 d'égalité des vecteurs moyens des différentes classes par rapport à l'hypothèse alternative Ha qu'au moins un des vecteurs moyens est différent d'un autre.												
156													
157	Si la probabilité du test est inférieure au niveau de signification usuel de 5%, H0 doit être rejeté.												
158													
159													

Rapport Général /

Rapports et Graphiques

CENTROIDES DES GROUPES ET GLOBAL

	lonsepal	larsepal	lonpetal	larpetal
Setosa	5,02326	3,44419	1,46512	0,25349
Versicolor	5,92368	2,76842	4,19737	1,30263
Virginica	6,54615	2,93590	5,51282	2,03590
Global	5,80333	3,06500	3,64583	1,16500

MATRICE DES COVARIANCES GLOBALES

	lonsepal	larsepal	lonpetal	larpetal
lonsepal	0,66399	-0,05988	1,24626	0,50862
larsepal	-0,05988	0,19511	-0,37915	-0,14239
lonpetal	1,24626	-0,37915	3,11415	1,30627
larpetal	0,50862	-0,14239	1,30627	0,59227

MATRICE DES CORRELATIONS GLOBALES

	lonsepal	larsepal	lonpetal	larpetal
lonsepal				
larsepal				
lonpetal				
larpetal				

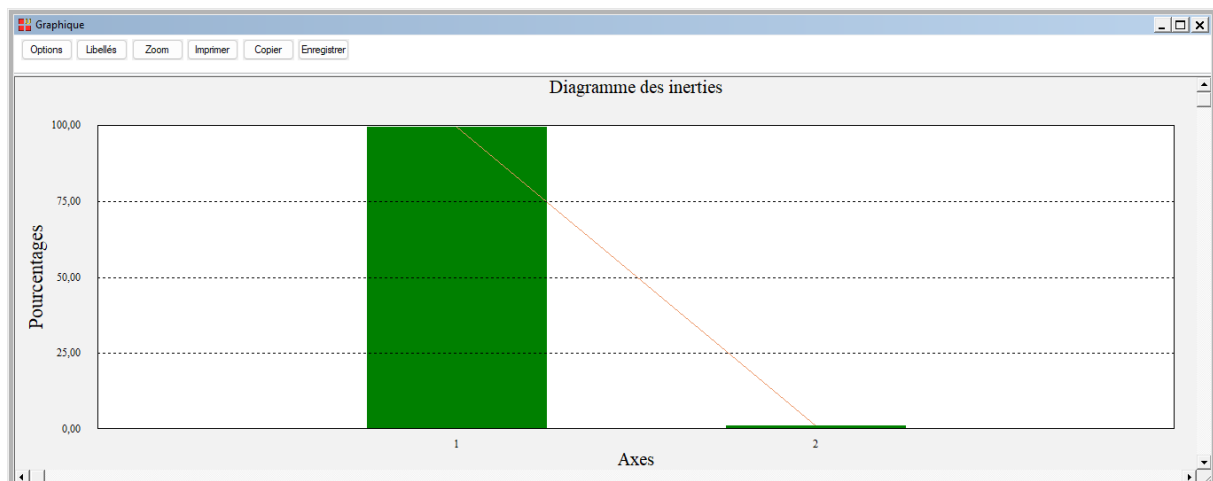
Ces rapports nous fournissent les renseignements suivants :

- 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
- Résultats pour les individus
- Coordonnées des centres des groupes
- Coefficients des fonctions de classement
- Détails du classement de la population d'apprentissage et matrice de confusion
- Statistiques pour les groupes observés et prévus (apprentissage)
- Détails du classement de la population de validation et matrice de confusion
- Classement de la population de prévision

L'option Graphiques

- Diagramme des inerties

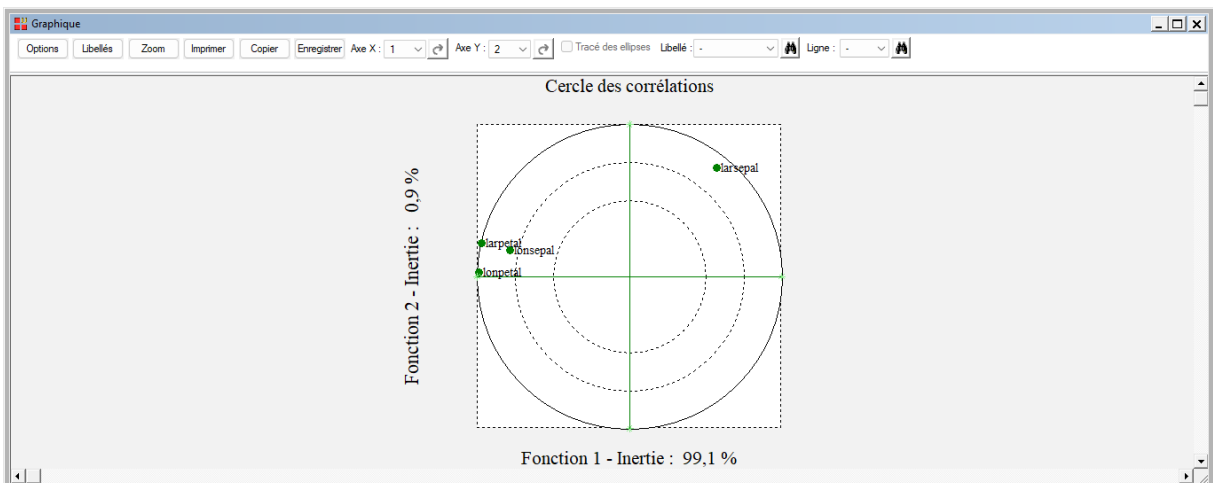
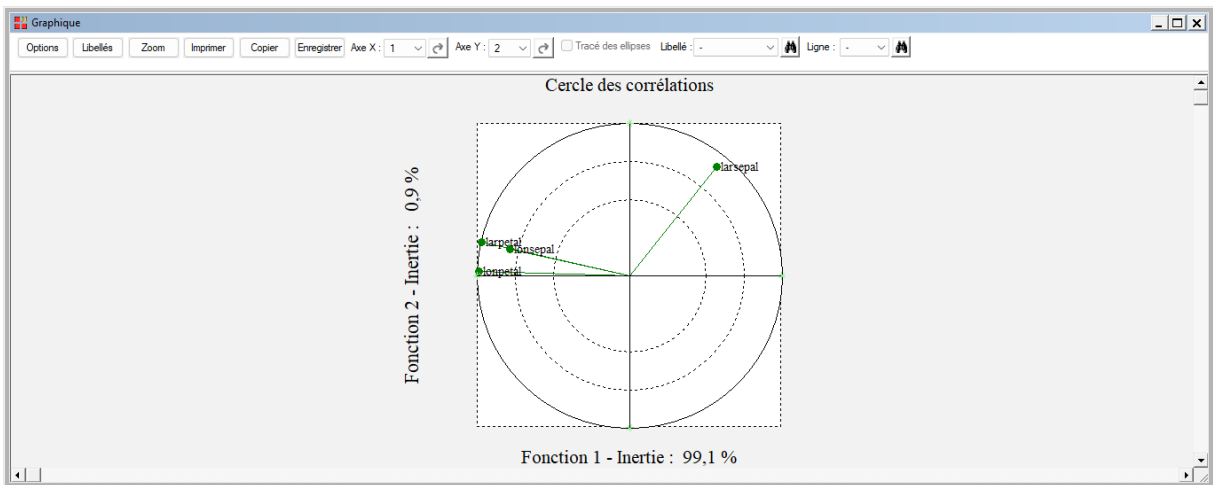
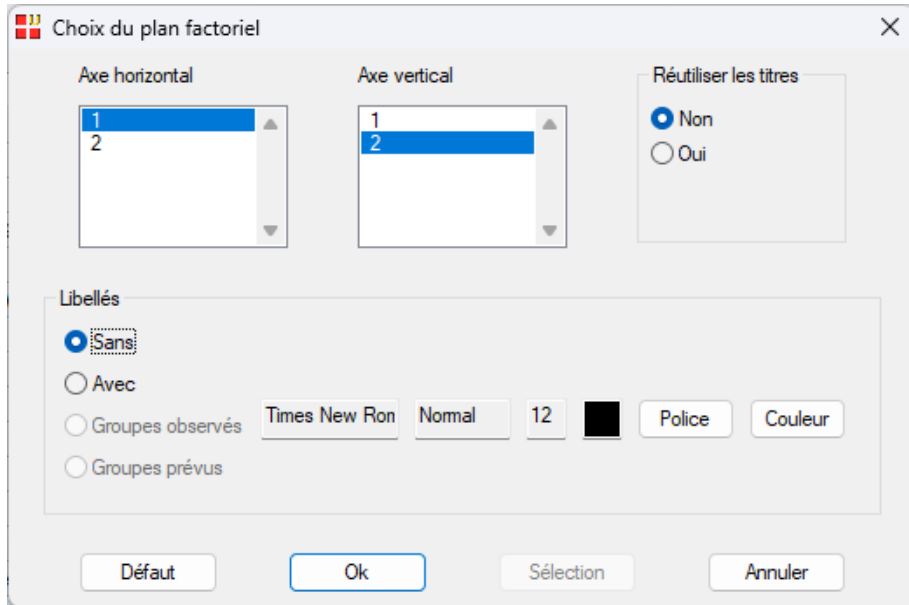
Ce graphique affiche les pourcentages d'inertie pour chacun des axes factoriels.



- Les options 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.

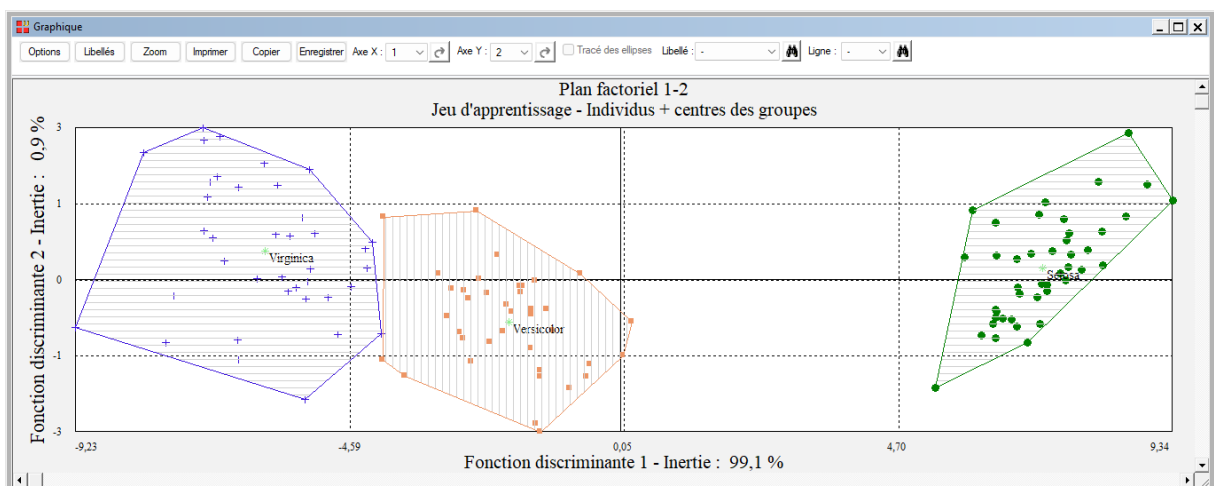
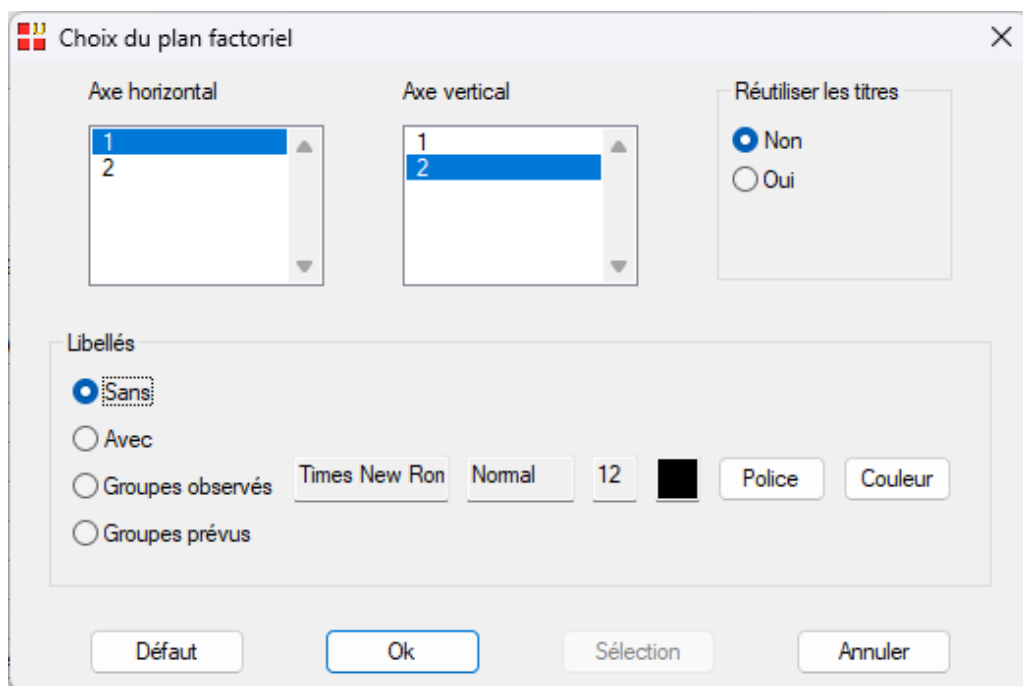


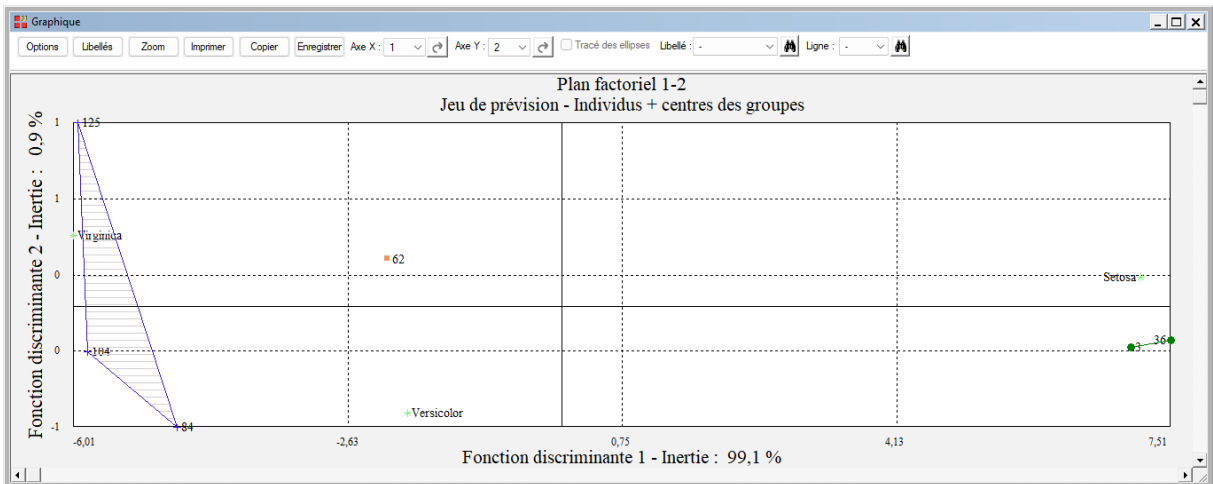
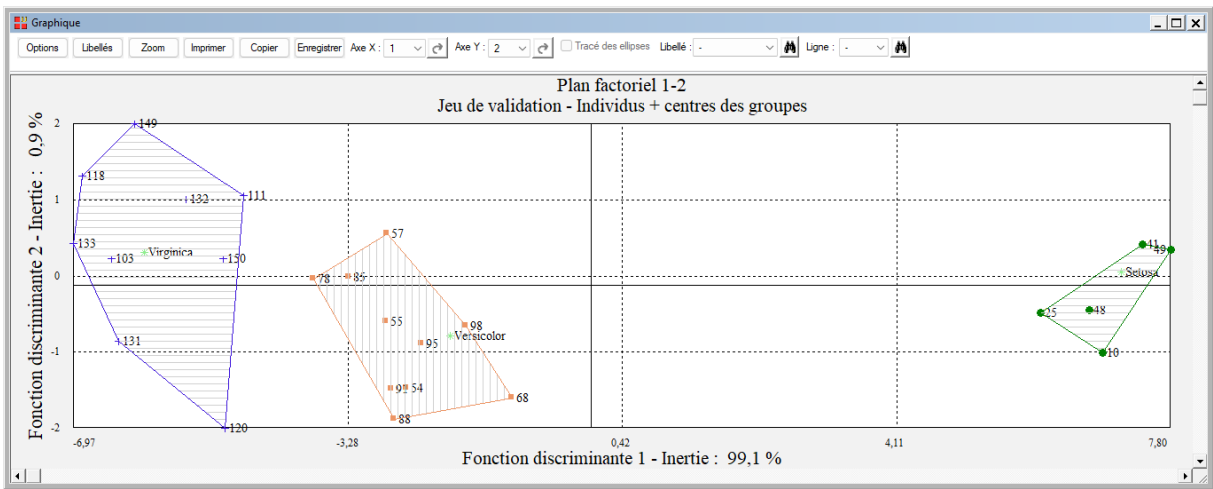
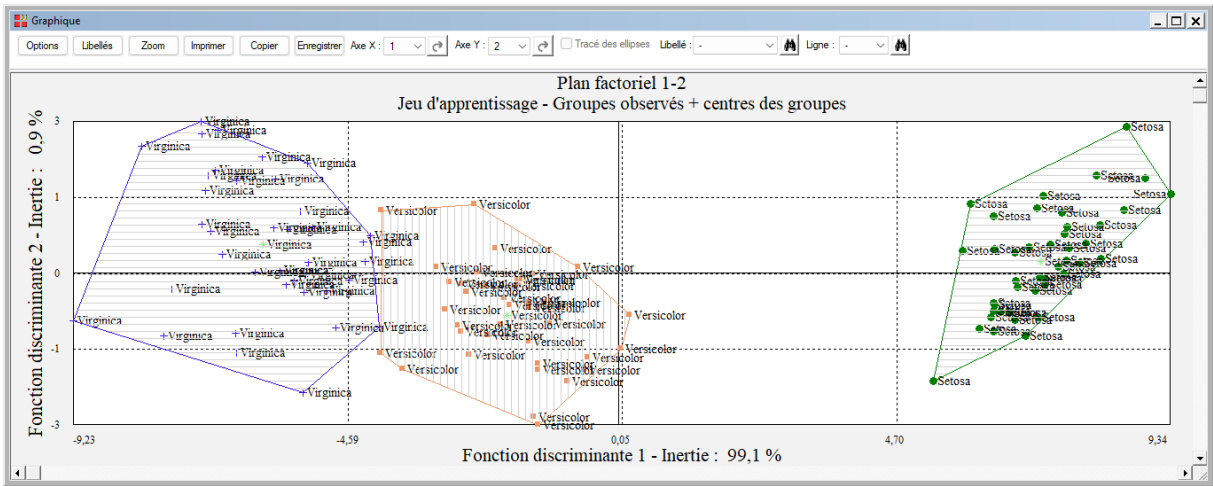
- Les options Plan factoriel des individus et centres des groupes

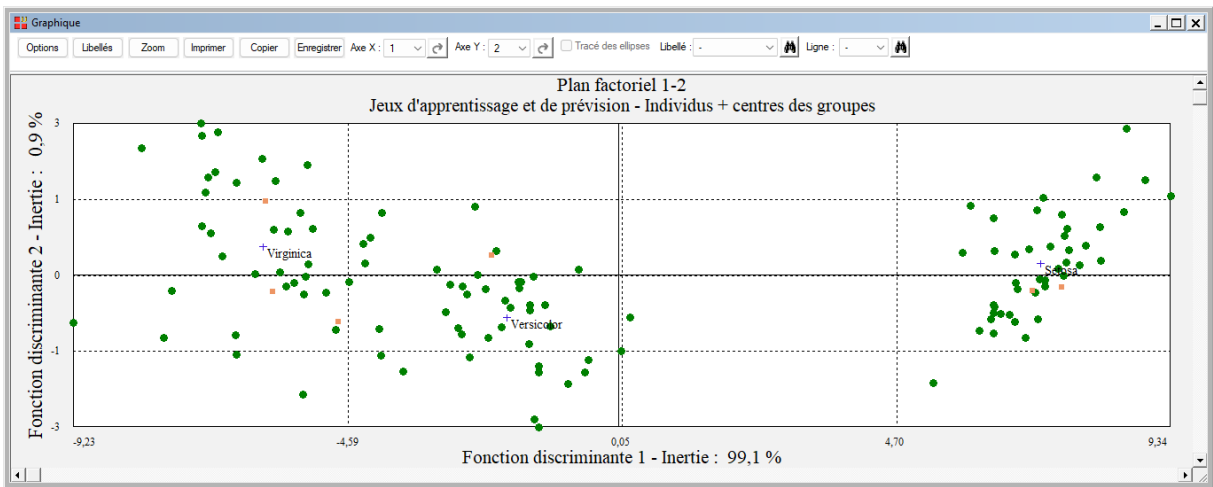
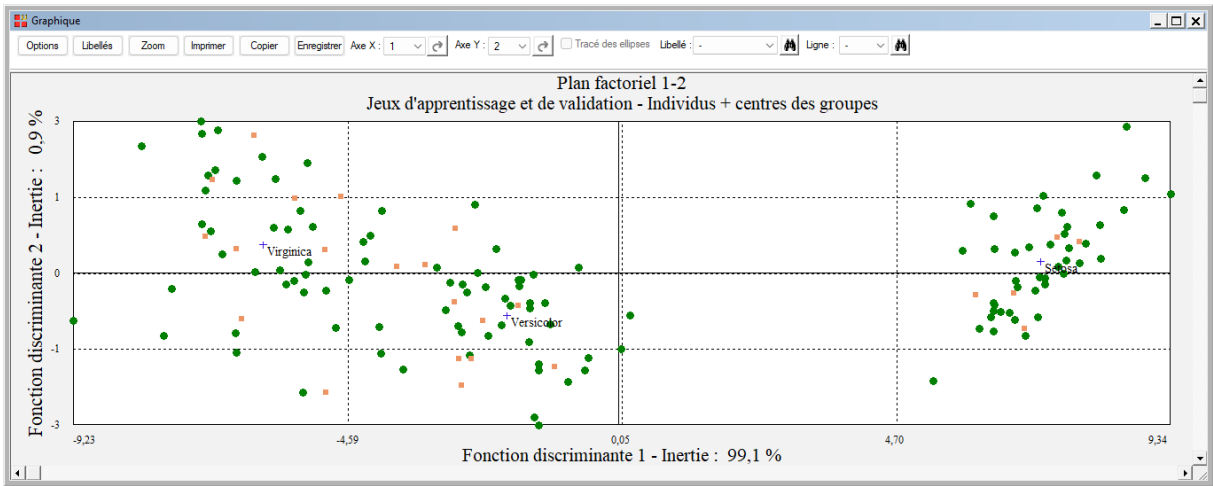
Ces options permettent d'afficher des plans factoriels des individus et des centres des groupes pour les populations d'apprentissage, de validation et de prévision. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.

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.

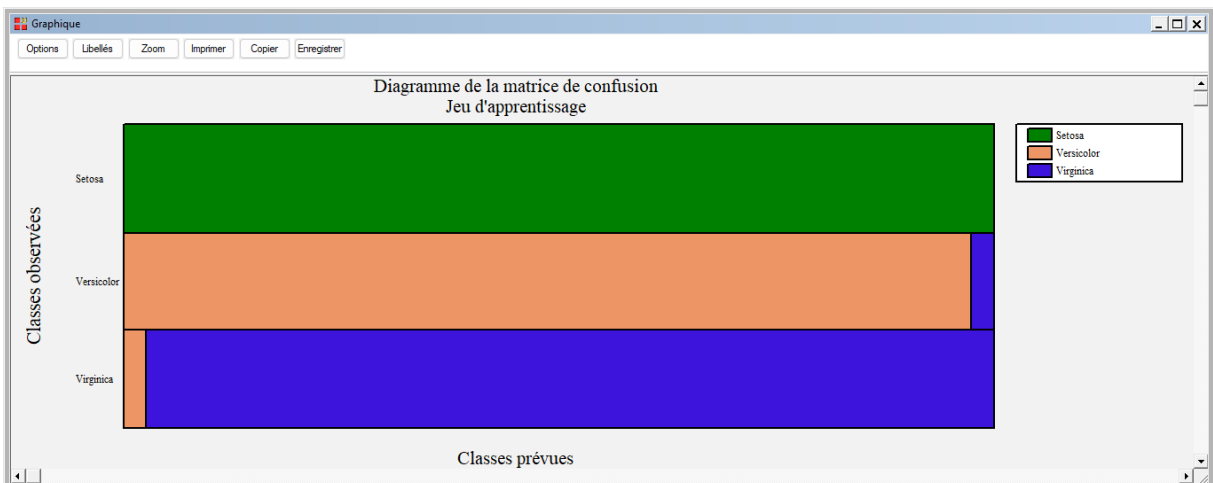


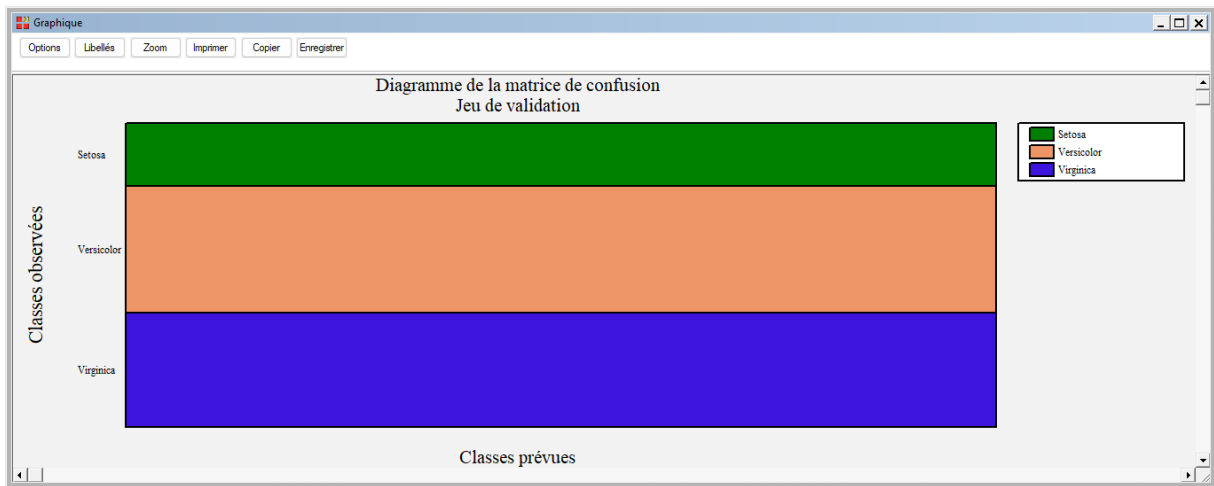




- Les options Graphiques des matrices de confusion

Ces graphiques affichent les matrices de confusion des jeux d'apprentissage et de validation sous la forme de diagrammes en mosaïque.





- Les options Courbes ROC

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

Exemple 2 : Fichier INFARCT2

Pour ce deuxième exemple, nous utiliserons le fichier INFARCT2. Ce fichier contient des informations concernant 101 victimes d'un infarctus du myocarde.

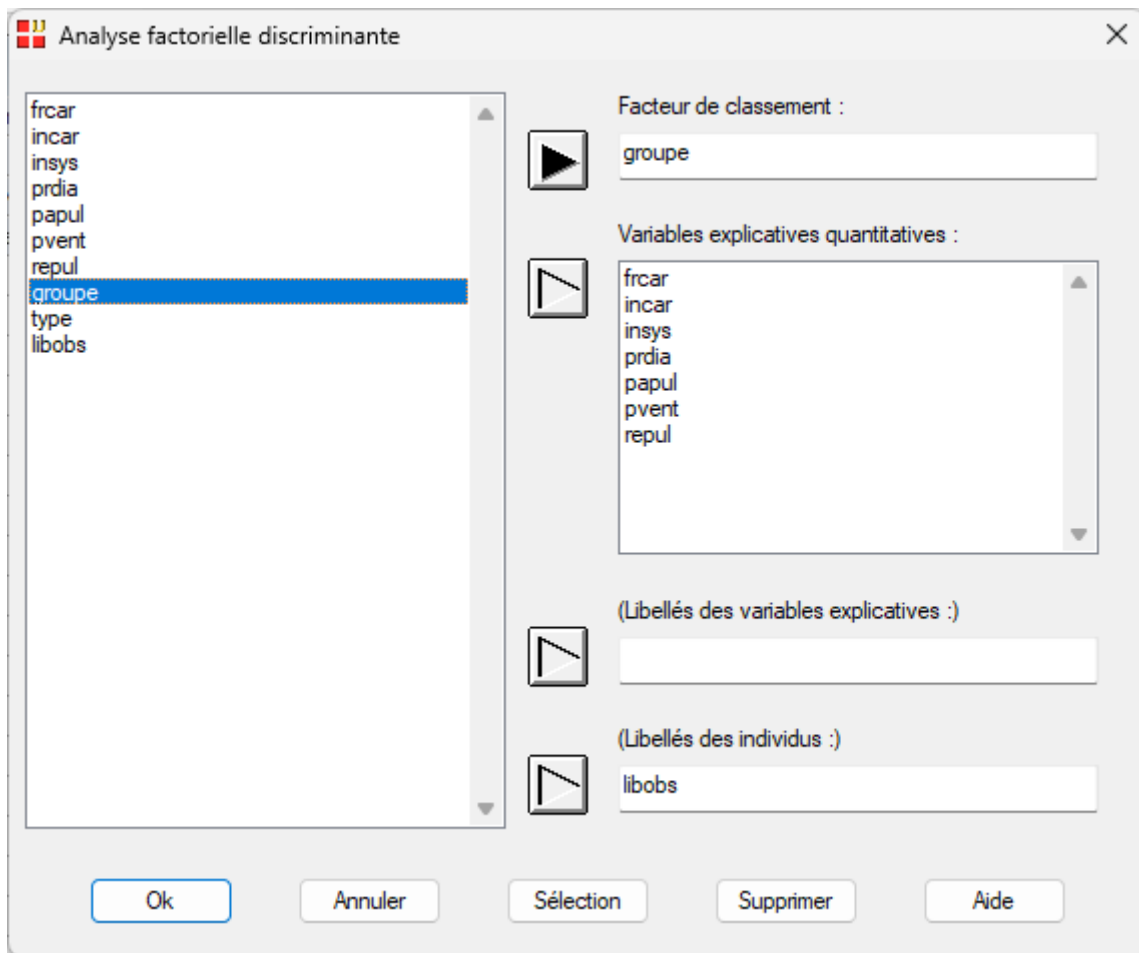
Cette population est constituée d'une population d'apprentissage de 81 individus et d'une population de validation de 20 individus. La population d'apprentissage comprend les individus 11 à 51 (groupe Décès) et les individus 52 à 91 (groupe Survie). La population de validation comprend les individus 1 à 10 (groupe Décès) et les individus 92 à 101 (groupe Survie).

Les variables mesurées sont :

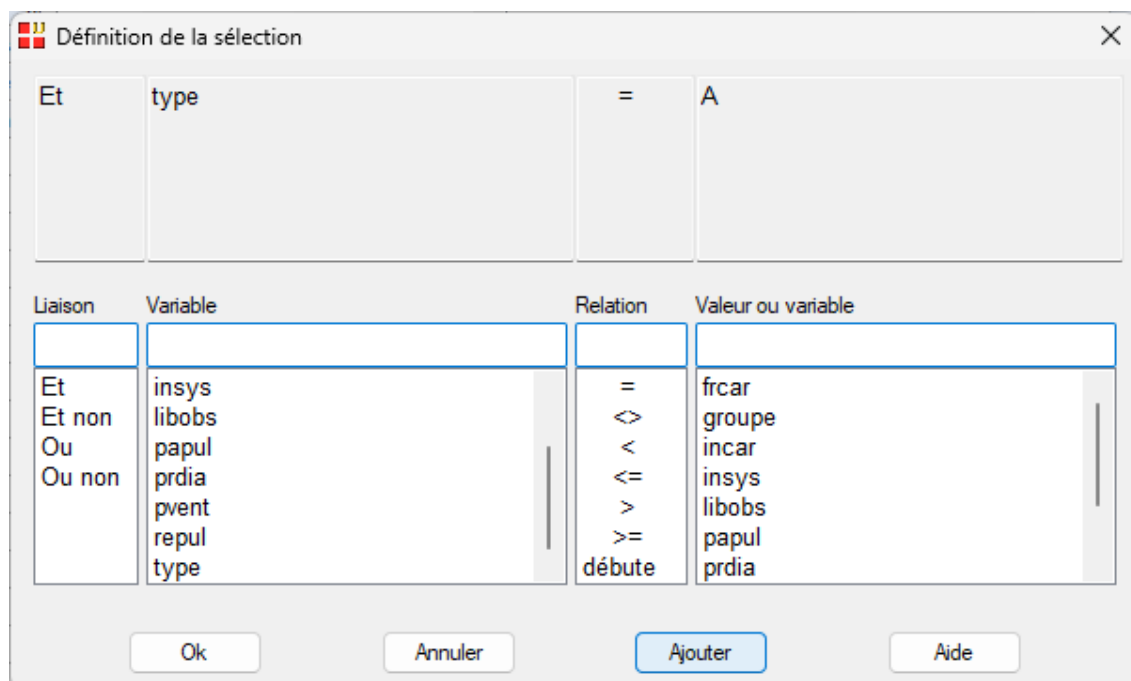
Variable	Mesure
<i>frcar</i>	fréquence cardiaque
<i>incar</i>	index cardiaque
<i>insys</i>	index systolique
<i>prdia</i>	pression diastolique
<i>papul</i>	pression artérielle pulmonaire
<i>pvent</i>	pression ventriculaire
<i>repul</i>	résistance pulmonaire

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.

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



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



Un message nous indique que 81 individus sont sélectionnés.

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.

	1	2	3	4	5	6	7	8	9	10
1										
2	Tableau des Inerties									
3										
4	Axe	Valeur propre	Pct variance	Pct cumulé	Variation	Corr. canonique				
5	1	1,34960	100	100	0	0,75789				
6										
7										
8										

Cliquons sur 'Suite des calculs'. L'écran 'Rapports et Graphiques' s'affiche.

	1	2	3	4	5	6	7	8
1								
2	(C) UNWIN version 10.4.0							
3								
4	DATE :	29/06/2025						
5	ORDINATEUR :	LAPTOP-LEGL077						
6	UTILISATEUR :	cchar						
7	FICHER(S) DE DONNEES OUVERT(S) :	INFARCT2.SGD						
8								
9	RESULTATS DE L'ANALYSE FACTORIELLE DISCRIMINANTE							
10								
11	Sélection :							
12	Et type =	A						
13								
14	Jeu d'apprentissage :	81 observations						
15	Jeu de validation :	20 observations						
16	Jeu de prévision :	0 observation						
17								
18	Variables explicatives :							
19	frcar							
20	incar							
21	insys							

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.

	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	COORDONNEE = COORDONNÉES DES INDIVIDUS							
6	CONTRIBUTION = CONTRIBUTIONS A L'INERTIE							
7	COSINUS^2 = COSINUS CARRÉS							
8	COS^2 CUMULE = SOMMES CUMULÉES DES COSINUS CARRÉS							
9								
10								
11		GRUPE	INDIV/DU	DISTANCE^2	COORDONNEE	CONTRIBUTION	COSINUS^2	COS^2 CUMUL
12	Obs011	1	1	0,12121	-0,34815	0,06531	1	
13	Obs012	1	2	2,88426	1,69831	1,55397	1	
14	Obs013	1	3	0,00002	-0,00455	0,00001	1	
15	Obs014	1	4	4,83268	2,19833	2,60372	1	
16	Obs015	1	5	1,90155	1,37897	1,02451	1	
17	Obs016	1	6	2,85242	1,68891	1,53681	1	
18	Obs017	1	7	1,29329	1,13723	0,69679	1	
19	Obs018	1	8	0,22116	0,47028	0,11915	1	
20	Obs019	1	9	3,80729	1,95123	2,05127	1	
21	Obs020	1	10	1,19438	1,09288	0,64350	1	

Ce rapport fournit les renseignements suivants :

- 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
- Résultats pour les individus
- Coordonnées des centres des groupes
- Matrice de confusion pour la population d'apprentissage
- Détails du classement de la population d'apprentissage
- Sensibilités et spécificités pour la population d'apprentissage
- Statistiques pour les groupes observés et prévus
- Matrice de confusion pour la population de validation (si existe))
- Détails du classement de la population de validation (si existe)
- Sensibilités et spécificités pour la population de validation (si existe)
- Classement de la population de prévision (si existe)

L'option Graphiques

- Diagramme des inerties

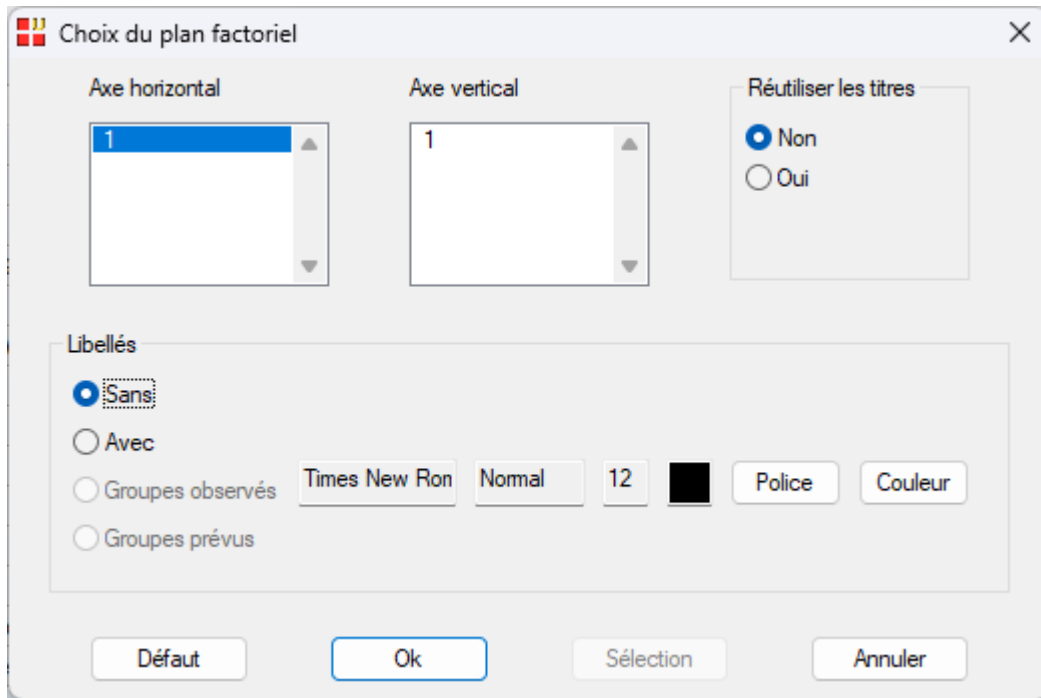
Ce diagramme n'est pas disponible dans cet exemple car il n'y a qu'un seul axe factoriel.

- 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.

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.

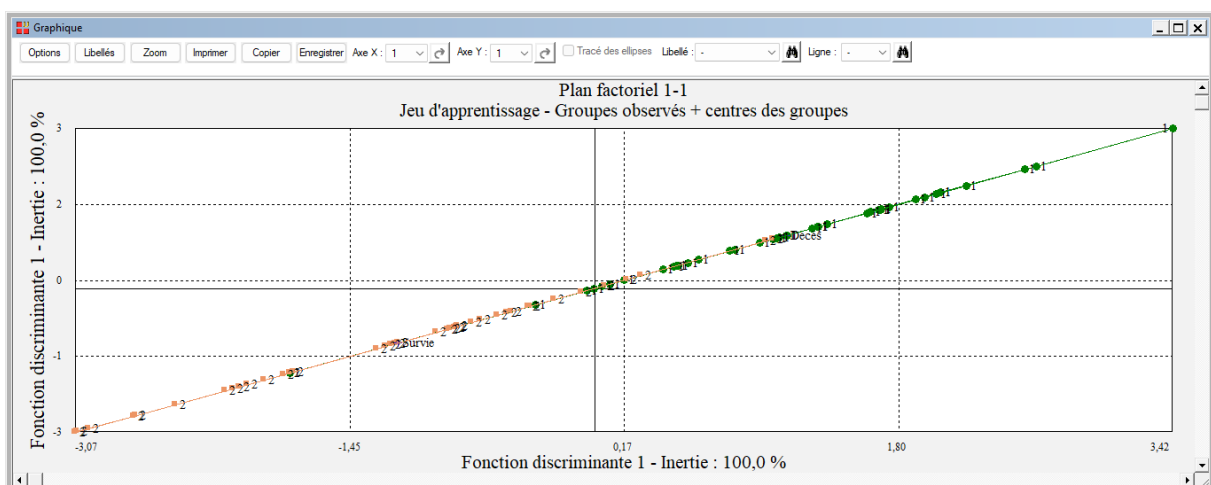


A noter que dans notre exemple, il n'y a qu'un seul axe factoriel et donc ce graphique n'est pas disponible.

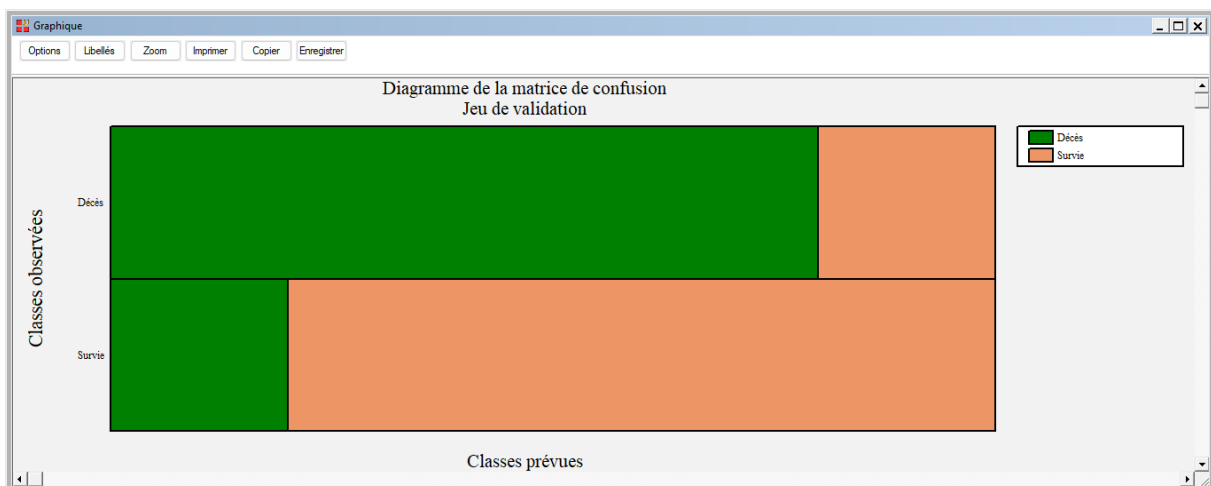
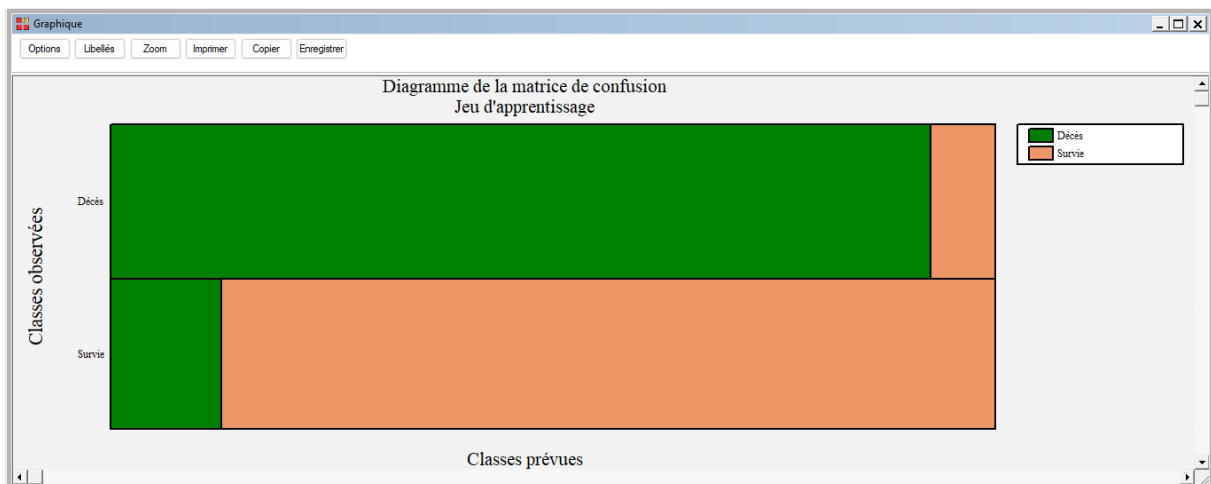
- Plan factoriel des individus et centres des groupes

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.

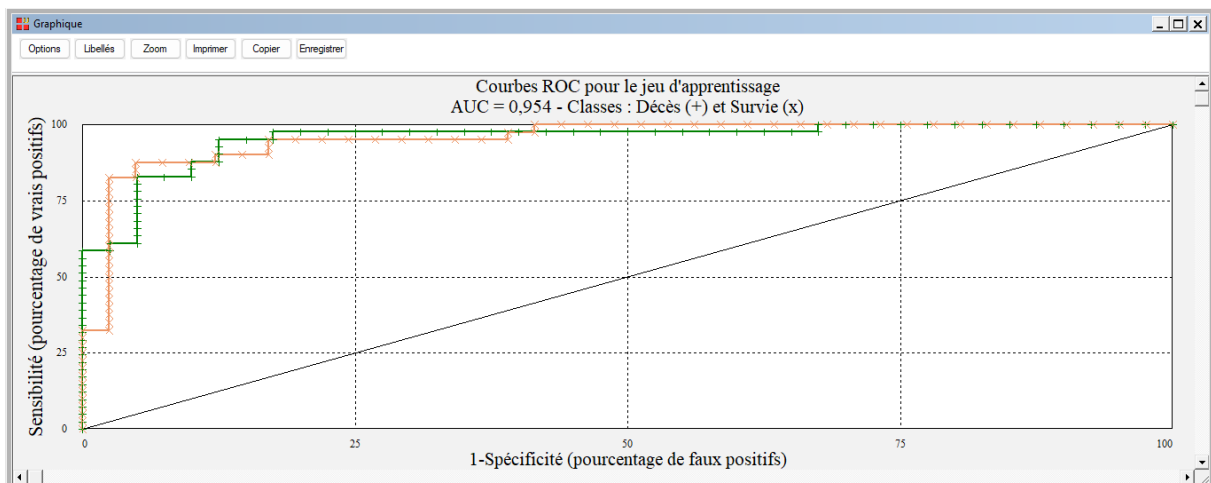


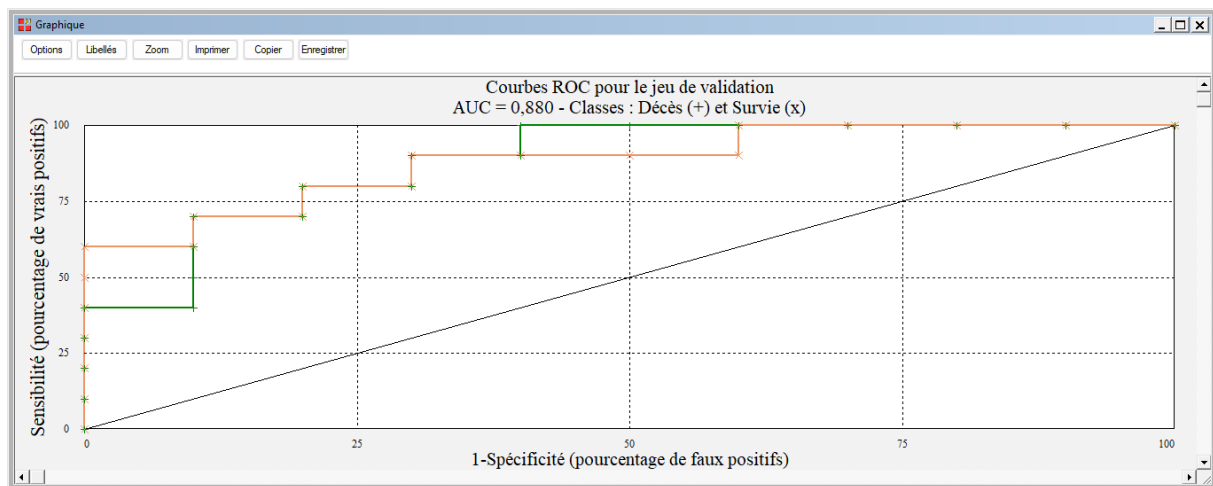
- Graphiques des matrices de confusion



- Courbes ROC

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





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* contient les libellés des individus.

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

Visualisons les résultats du classement et demandons également le cercle factoriel et le plan factoriel pour les axes 1 et 2.

Analyse factorielle discriminante

Année
Temp
Insol
Chaleur
Pluie
Qualité

Facteur de classement :
Qualité

Variables explicatives quantitatives :
Temp
Insol
Chaleur
Pluie

(Libellés des variables explicatives :)

(Libellés des individus :)
Année

Ok Annuler Sélection Supprimer Aide

Rapports et Graphiques

Rapport AFD

- 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. discriminantes std.
- Fct. discriminantes non std.
- Résultats variables
- Résultats individus
- Coord. centres des groupes
- Matrice de confusion (apprentissage)
- Détails clas. apprentissage
- Stats groupes observés
- Stats groupes prévus

	1	2	3	4	5	6	7	8	9
1									
2	MATRICE DE CONFUSION POUR LE JEU D'APPRENTISSAGE								
3									
4	En lignes, les groupes observés								
5	En colonnes, les groupes prévus								
6									
7	Pourcentage de mal classés : 20,588 %								
8	Pourcentage de bien classés (exactitude) : 79,412 %								
9									
10	Précision = VP / (VP + FP)								
11	Rappel = VP / (VP + FN)								
12	Score F1 = 2 x (Précision x Rappel) / (Précision + Rappel)								
13									
14									
15		Taille	Bon	Moyen	Médiocre	Précision	Rappel	Score F1	
16	Bon	11	9	2	0	0,81818	0,81818	0,81818	
17	Moyen	11	2	8	1	0,66667	0,72727	0,69565	
18	Médiocre	12	0	2	10	0,90909	0,83333	0,86957	
19									
20									
21									

Rapport Explorateur

Rapports et Graphiques

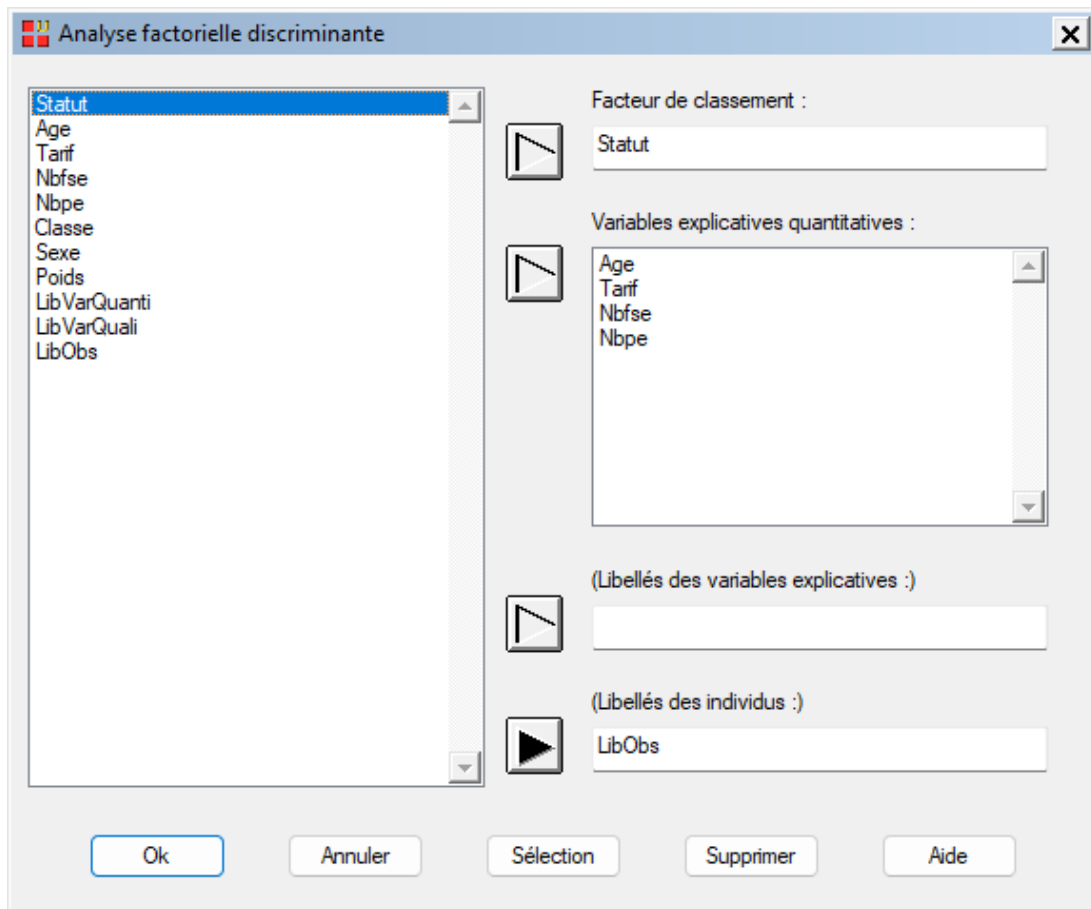
Rapport AFD

- 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. discriminantes std.
- Fct. discriminantes non std.
- Résultats variables
- Résultats individus
- Coord. centres des groupes
- Matrice de confusion (apprentissage)
- Détails clas. apprentissage
- Stats groupes observés
- Stats groupes prévus

	1	2	3	4	5	6	7	8
1								
2	CLASSEMENTS ET DISTANCES AUX CENTROIDES DES GROUPES							
3								
4	GROUPE OBSERVE : Bon							
5								
6	JEU D'APPRENTISSAGE							
7								
8	L'INDIVIDU EST AFFECTE AU GROUPE DONT IL EST LE PLUS PROCHE DU CENTROIDE							
9								
10								
11	INDIVIDU - AFFECTATION							
12	1928 - Bon		4,00053	5,34872	10,21839			
13	1929 - Bon		0,05702	5,34034	18,64652			
14	1934 - Bon		0,10178	4,33567	18,16603			
15	1937 - Bon		0,97337	1,97671	13,75642			
16	1943 - Bon		0,62290	5,12625	20,62039			
17	1945 - Bon		3,44810	11,66118	32,85423			
18	1947 - Bon		3,92572	15,26169	30,65577			
19	1949 - Bon		4,89371	18,79791	39,42416			
20	1952 - Bon		0,58081	4,08430	18,53601			
21	1953 - Moyen *		3,15229	0,45168	5,92207			

Rapport Explorateur

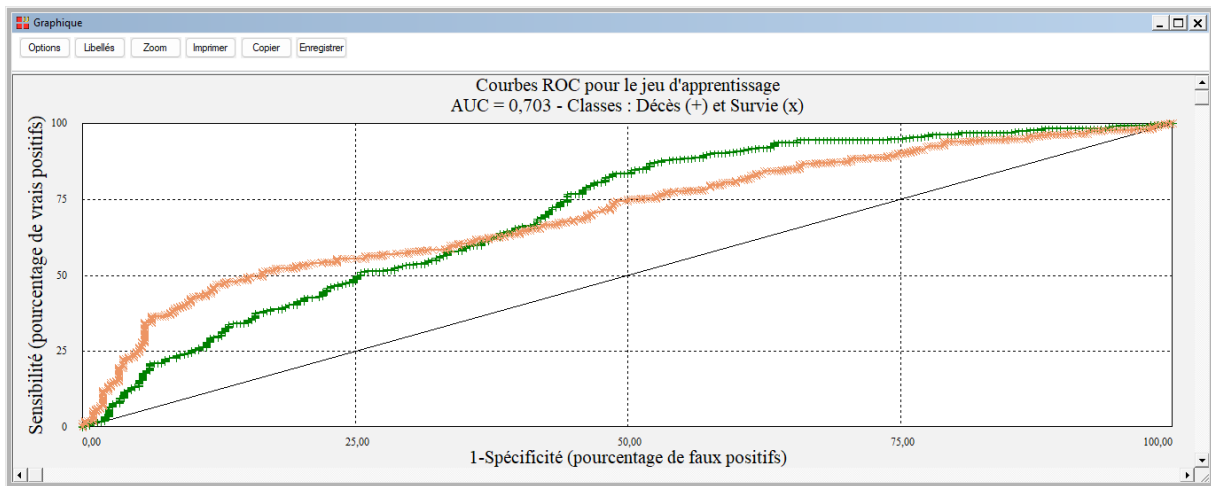
Cliquons sur l'icône AFD dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-dessous.



Après exécution de la procédure, visualisons la matrice de confusion des données d'apprentissage et la courbe ROC associée.

	1	2	3	4	5	6	7	8
1								
2	MATRICE DE CONFUSION POUR LE JEU D'APPRENTISSAGE							
3								
4	En lignes, les groupes observés							
5	En colonnes, les groupes prévus							
6								
7	Pourcentage de mal classés : 29,132 %							
8	Pourcentage de bien classés (exactitude) : 70,868 %							
9								
10	Précision = VP / (VP + FP)							
11	Rappel = VP / (VP + FN)							
12	Score F1 = 2 x (Précision x Rappel) / (Précision + Rappel)							
13								
14								
15		Taille	Décès	Survie	Précision	Rappel	Score F1	
16	Décès	424	367	57	0,70849	0,86557	0,77919	
17	Survie	290	151	139	0,70918	0,47931	0,57202	
18								
19								
20								
21								

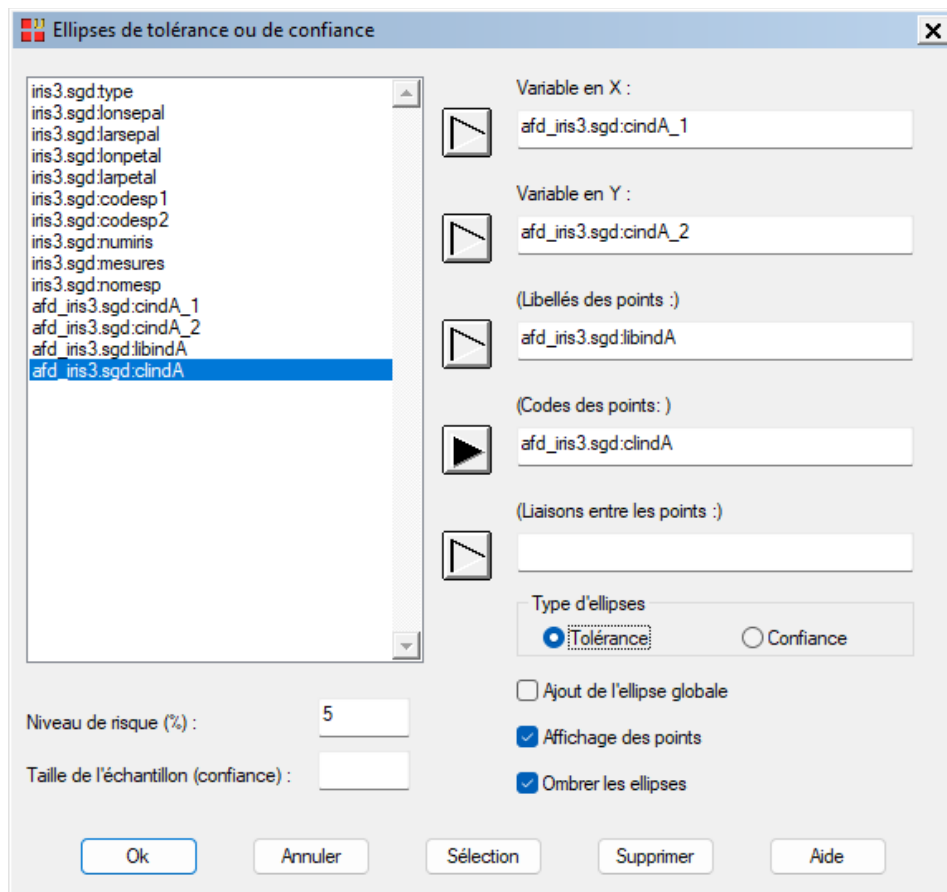
Environ 71 % des passagers sont bien classés par cette analyse et l'aire sous la courbe ROC est proche de 0,7.



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.

Ellipses de tolérance

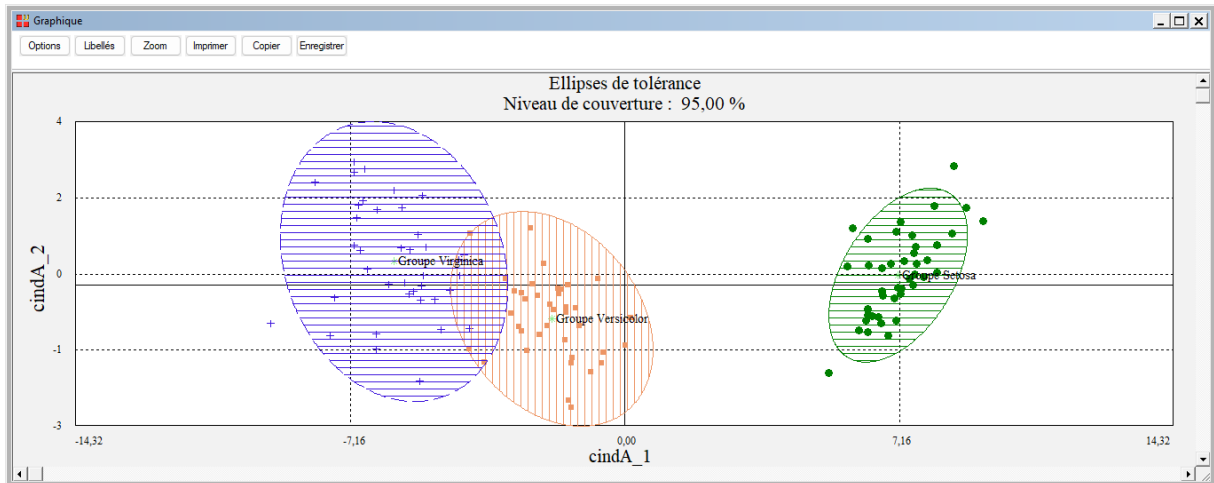
Reprenons l'exemple IRIS3. Il est possible d'afficher les ellipses de tolérance autour des centres de gravité des classes en enregistrant dans la procédure d'analyse factorielle discriminante les variables 'Coordonnées individus', 'Libellés individus' et 'Classes individus'.



Appelons AFD_Iris3.sgd le fichier dans lequel ces trois variables sont enregistrées.

Par le menu Décrire, exécutons la procédure Ellipses et renseignons la boîte de dialogue comme montré ci-après.

Demandons l'affichage du graphique des ellipses avec ombrage.



Calculs de la matrice de confusion et des indicateurs

Dans le cas de deux classes A et B, nous avons le tableau suivant :

	Prévu A	Prévu B	Total	% correct
Observé A	VP	FN	VP + FN	$\frac{100 * VP}{(VP + FN)}$
Observé B	FP	VN	FP + VN	$\frac{100 * VN}{(VN + FP)}$
Total	VP + FP	FN + VN	VP + FP + VN + FN	
% correct	$\frac{100 * VP}{(VP + FP)}$	$\frac{100 * VN}{(FN + VN)}$		$\frac{100 * (VP + VN)}{(VP + VN + FP + FN)}$
				% total correctement prévu

Dans le cas multi-classes (plus de 2 classes), chaque classe est étudiée par rapport une classe virtuelle réunissant l'ensemble des autres classes.

Définition des indicateurs :

- la sensibilité $VP / (VP+FN)$
- la spécificité $VN / (VN+FP)$
- l'exactitude $(VP+VN) / (VP+VN+FP+FN)$
- la précision $VP / (VP+FP)$
- le rappel $VP / (VP+FN)$
- le score F1 $2 \times (\text{précision} \times \text{rappel}) / (\text{précision} + \text{rappel})$

La sensibilité (ou rappel) indique la capacité du modèle à prévoir les vrais positifs.

La spécificité (ou taux de vrais négatifs) permet de mesurer la capacité du modèle à prévoir les vrais négatifs.

L'exactitude mesure le pourcentage de prévisions correctes par rapport à toutes les prévisions positives et négatives. Elle varie entre 0 et 1 et est sensible aux données déséquilibrées. Plus elle est proche de 1, meilleure est la prévision globale.

Le rappel (ou sensibilité ou taux de vrais positifs) varie entre 0 et 1 et n'est pas sensible aux données déséquilibrées. Un rappel égal à 1 indique une prévision parfaite des positifs.

La précision mesure le pourcentage de prévisions positives correctes. Elle varie entre 0 et 1 et n'est pas sensible aux données déséquilibrées. Une précision égale à 1 indique que tous les positifs sont prédits positifs.

Le score F1 combine la précision et le rappel en utilisant les moyennes harmoniques. Il varie entre 0 et 1. Maximiser ce score revient à maximiser la précision et le rappel. Il n'est pas sensible aux données déséquilibrées.

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)
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)
ccentA	Coordonnées des centroïdes (apprentissage)
seuilA	Seuils (apprentissage)
specificiteA	Spécificités (apprentissage)
sensibiliteA	Sensibilités (apprentissage)
aireA	Aires sous les courbes ROC (apprentissage)
classeA	Classement (apprentissage)
cindV	Coordonnées des individus (validation)
libindV	Libellés des individus (validation)
clindV	Classes des individus (validation)
seuilV	Seuils (validation)
specificiteV	Spécificités (validation)
sensibiliteV	Sensibilités (validation)
aireV	Aires sous les courbes ROC (validation)
classeV	Classement (validation)
classeP	Classement (prévision)
libindP	Libellés des individus (prévision)