

UNIWIN VERSION 10.4.0

METHODE SIMCA

Révision : 15/09/2025

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichier IRIS3	3
L'option Rapports	6
L'option Graphiques	10
Exemple 2 : Fichier WINES2	15
Calculs de la matrice de confusion et des indicateurs	19
Les variables créées par la procédure.....	20
Références	20

Définition

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 utilisant uniquement les observations de cette classe. Les différents modèles obtenus pour chacune des classes peuvent 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'.

Entrée des données

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

Méthode SIMCA

type
lonsepal
larsepal
lonpetal
larpetal
codesp1
codesp2
numiris
mesures
nomesp

Facteur de classement qualitatif :

Variables explicatives quantitatives :

(Libellé du facteur de classement :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Nombres de composantes :

Paramètre alpha (Cooman) : 0,05

Centrage des données

Réduction des données

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de définir le facteur de classement qualitatif, les variables explicatives quantitatives, les libellés optionnels du facteur de classement, des variables explicatives et des observations.

Elle permet également de préciser si les données doivent être centrées et réduites, la valeur du paramètre alpha (niveau de signification pour les limites statistiques dans le graphique de Cooman) et les nombres de composantes pour chacun des modèles (obtenus par exemple par des analyses préalables ACP ou NIPALS des jeux d'apprentissage de chacun des modèles). Si ces nombres ne sont pas spécifiés, toutes les composantes sont extraites.

Données manquantes

Les valeurs manquantes du facteur de classement définissent l'échantillon de prévision. Les données manquantes pour les variables explicatives sont autorisées.

Exemple 1 : Fichier IRIS3

Pour ce premier exemple, nous utiliserons le fichier IRIS3 pour illustrer cette procédure.

Ce fichier contient les données relatives à 150 iris de 3 espèces : Iris Setosa, Iris Versicolor et Iris Virginica. Les mesures effectuées sont, en millimètres, la longueur du sépale (lonsepal), la longueur du pétale (lonpetal), la largeur du sépale (larsepal) et la largeur du pétale (larpetal).

Renseignons la boîte de dialogue comme montré ci-dessous.

Méthode SIMCA

type
lonsepal
larsepal
lonpetal
larpetal
codesp1
codesp2
numiris
mesures
nomesp

Facteur de classement qualitatif :
codesp2

Variables explicatives quantitatives :
lonsepal
larsepal
lonpetal
larpetal

(Libellé du facteur de classement :)

(Libellés des variables explicatives :)
mesures

(Libellés des observations :)
numiris

Nombres de composantes :
Paramètre alpha (Cooman) : 0.05

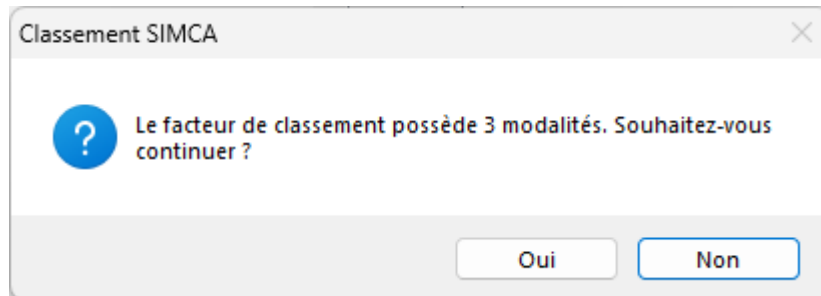
Centrage des données
 Réduction des données

Ok Annuler Sélection Supprimer Aide

Sélectionnons la variable *codesp2* comme facteur de classement et les variables *lonsepal*, *larsepal*, *lonpetal* et *lonsepal* comme variables explicatives.

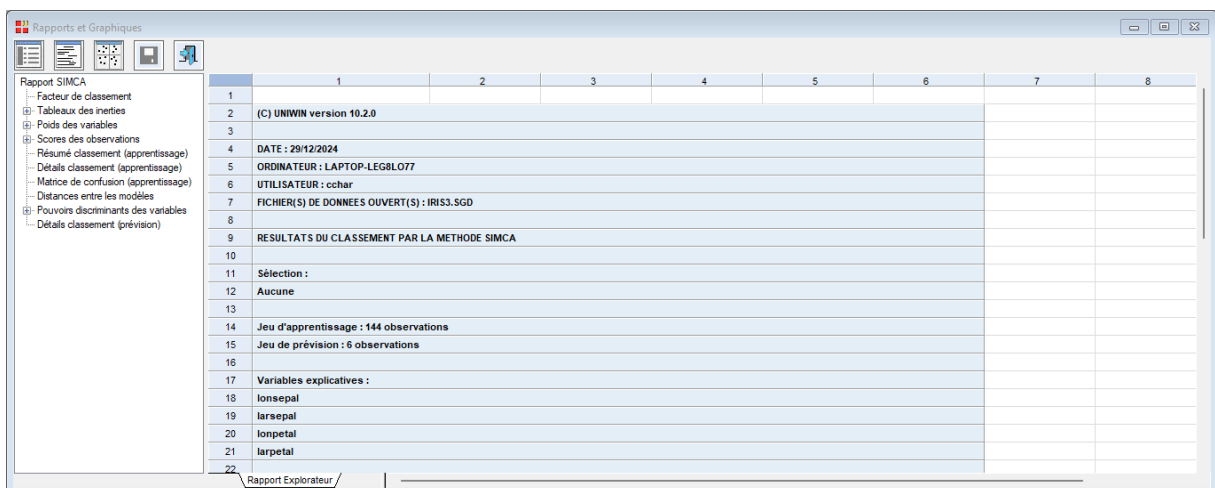
Conservons les autres paramètres aux valeurs par défaut.


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




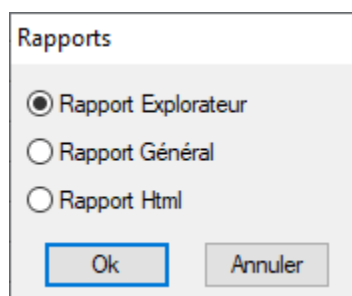
Après avoir visualisé le message informatif nous indiquant le nombre de modalités du facteur de classement, nous continuons l'analyse.


Après quelques instants, la fenêtre « Rapports et Graphiques » s'affiche :

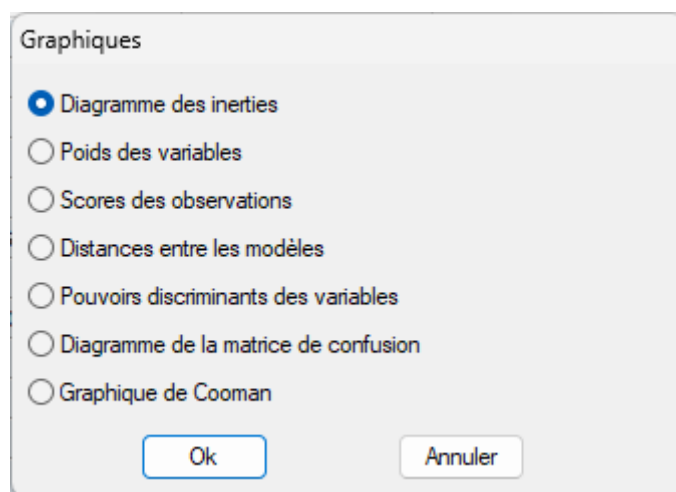


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

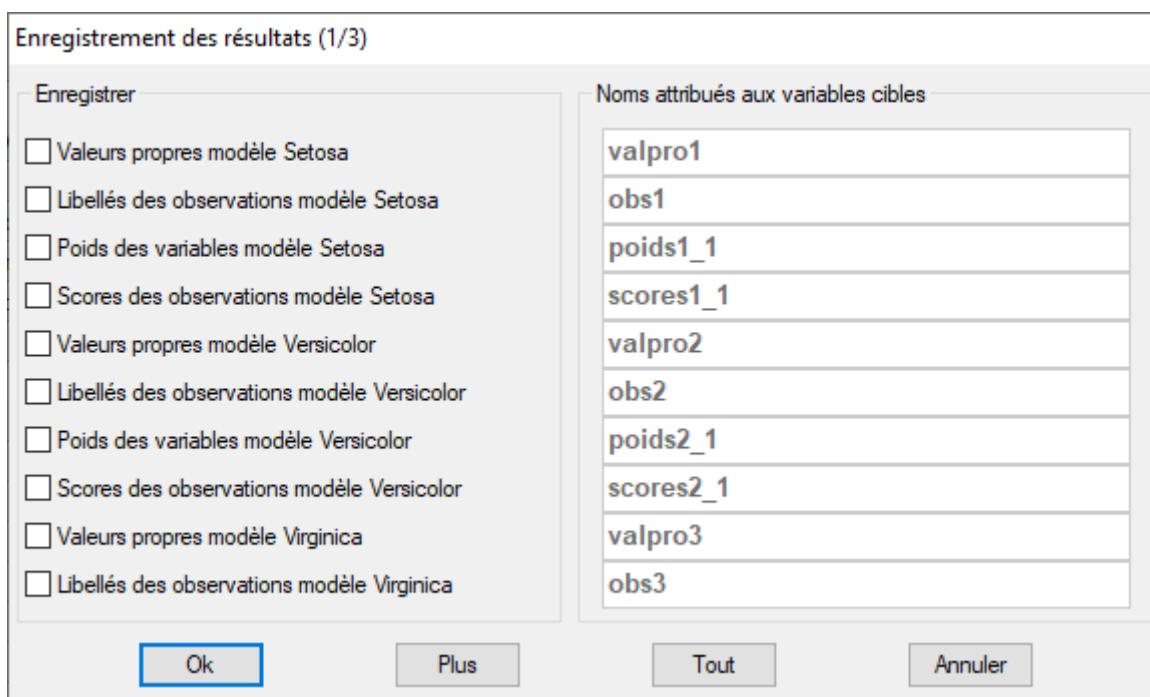
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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.



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.

Le premier tableau affiche les nombres d'observations pour les jeux d'apprentissage et de prévision et rappelle les paramètres de l'étude.

	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.2.0							
3								
4	DATE : 29/12/2024							
5	ORDINATEUR : LAPTOP-LEGBLO77							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : IRIS3.SGD							
8								
9	RESULTATS DU CLASSEMENT PAR LA METHODE SIMCA							
10								
11	Sélection :							
12	Aucune							
13								
14	Jeu d'apprentissage : 144 observations							
15	Jeu de prévision : 6 observations							
16								
17	Variables explicatives :							
18	lonsepal							
19	larsepal							
20	lonpetal							
21	larpetal							
22								

Le deuxième tableau affiche un tri à plat du facteur de classement.

	1	2	3	4	5	6	7	8
1								
2	FACTEUR DE CLASSEMENT							
3								
4	Le tableau affiche les effectifs, les fréquences (%), les effectifs cumulés et les fréquences cumulées (%).							
5								
6								
7		Effectifs	Fréquences	Effectifs cumulés	Fréquences cumulées			
8	Setosa	48	33,33	48	33,33			
9	Versicolor	48	33,33	96	66,67			
10	Virginica	48	33,33	144	100,00			
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Le troisième tableau affiche les inerties pour chacun des modèles construits.

	1	2	3	4	5	6	7	8
1								
2	TABLEAU DES INERTIES DE LA CLASSE Setosa							
3								
4								
5		Valeur propre	Pct variance	Pct cumulé	Variation			
6	Composante 1	2,03541	50,88514	50,88514	0,00000			
7	Composante 2	1,03772	25,94309	76,82823	24,94204			
8	Composante 3	0,67801	16,95037	93,77860	6,99272			
9								

Le quatrième tableau affiche les poids des variables pour chacun des modèles construits.

Rapports et Graphiques

Rapport SIMCA

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

	1	2	3	4	5	6	7	8
1								
2	POIDS DES VARIABLES POUR LA CLASSE Setosa							
3								
4								
5		Composante 1	Composante 2	Composante 3				
6	lonsepal	-0,61160	-0,31795	-0,08530				
7	larsepal	-0,57875	-0,44003	-0,00106				
8	lonpetal	-0,36282	0,64136	-0,66796				
9	larpetal	-0,39920	0,54216	0,73929				
10								
11								

Le cinquième tableau affiche les scores des observations pour chacun des modèles construits.

Rapports et Graphiques

Rapport SIMCA

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

	1	2	3	4	5	6	7	8
1								
2	SCORES DES OBSERVATIONS POUR LA CLASSE Setosa							
3								
4								
5		Composante 1	Composante 2	Composante 3				
6	1	0,08446	-0,65714	-0,07542				
7	2	1,18049	0,09382	-0,02827				
8	4	1,33184	0,62079	-0,34482				
9	5	0,10505	-0,68267	-0,05181				
10	6	-2,41258	0,75245	0,06170				
11	7	0,71907	0,40858	0,73414				
12	8	0,19463	-0,07911	-0,44119				
13	9	2,18759	0,65362	0,09345				
14	10	1,19050	-0,15225	-1,10632				
15	11	-0,94257	-0,77901	-0,53757				
16	12	0,32540	0,47338	-0,78335				
17	13	1,72427	-0,32303	-0,69223				
18	14	3,21609	-1,00102	0,59701				
19	15	-1,44438	-2,60213	0,53586				
20	16	-3,25627	-0,83636	0,76853				
21	17	-1,56538	-0,74517	1,62144				
22	18	-0,28804	-0,15124	0,61442				

Le sixième tableau affiche pour chacun des modèles élaborés en utilisant le jeu d'apprentissage les nombres de composantes extraites, les vrais positifs (VP), les faux positifs (FP), les vrais négatifs (VN), les faux négatifs (FN), les exactitudes, les spécificités et les sensibilités.

Rapports et Graphiques

Rapport SIMCA

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

	1	2	3	4	5	6	7	8	9
1									
2	RESUME DU CLASSEMENT SIMCA - JEU D'APPRENTISSAGE								
3									
4	Le tableau affiche pour chacun des modèles les nombres de composantes, les pourcentages de bien classés, les vrais positifs (VP), les faux positifs (FP), les vrais négatifs (VN), les faux négatifs (FN), les spécificités et les sensibilités.								
5									
6									
7	Exactitude = $(VP + VN) / (VP + VN + FP + FN)$								
8	Sensibilité = $VP / (VP + FN)$								
9	Spécificité = $VN / (VN + FP)$								
10									
11									
12		Nb composantes	VP	FP	VN	FN	Exactitude	Spécificité	Sensibilité
13	Setosa	3	45	0	96	3	0,979	1,000	0,938
14	Versicolor	3	46	2	94	2	0,972	0,979	0,958
15	Virginica	3	45	6	90	3	0,938	0,938	0,938

Le tableau suivant affiche les détails du classement pour le jeu d'apprentissage.

Observation - Classe observée	Setosa	Versicolor	Virginica
1 - Setosa	1	-1	-1
2 - Setosa	1	-1	-1
4 - Setosa	1	-1	-1
5 - Setosa	1	-1	-1
6 - Setosa	1	-1	-1
7 - Setosa	1	-1	-1
8 - Setosa	1	-1	-1
9 - Setosa	1	-1	-1
10 - Setosa	1	-1	-1
11 - Setosa	1	-1	-1
12 - Setosa	1	-1	-1
13 - Setosa	1	-1	-1
14 - Setosa	1	-1	-1
15 - Setosa	1	-1	-1
16 - Setosa	1	-1	-1
17 - Setosa	1	-1	-1
18 - Setosa	1	-1	-1
19 - Setosa	1	-1	-1
20 - Setosa	1	-1	-1
21 - Setosa	1	-1	-1
22 - Setosa	1	-1	-1

Chaque observation est affectée (1) ou non affectée (-1) à un modèle.

Une observation peut être affectée à plusieurs modèles, cela indique des recouvrements des classes observées.

Une observation peut n'être affectée à aucun des modèles, cela indique un possible point aberrant ou une observation définissant une éventuelle nouvelle classe.

Le tableau suivant affiche la matrice de confusion :

	Setosa	Versicolor	Virginica	Sans affectation	Précision	Rappel	Score F1
Setosa	45	0	0	3	1,000	0,938	0,968
Versicolor	0	46	6	2	0,958	0,958	0,958
Virginica	0	2	45	3	0,882	0,938	0,909

Les tableaux suivants affiche les distances entre les modèles et les pouvoirs discriminants des variables pour chacun des modèles :

Rapports et Graphiques

Rapport SIMCA

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

	1	2	3	4	5	6	7	8
1								
2	DISTANCES ENTRE LES MODELES							
3								
4	Ce tableau affiche les distances entre un modèle et les autres modèles.							
5								
6								
7		Setosa	Versicolor	Virginica				
8	Setosa	0,95743	2,49632	4,57974				
9	Versicolor	2,49632	0,95743	2,06086				
10	Virginica	4,57974	2,06086	0,95743				
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

Rapports et Graphiques

Rapport SIMCA

- Facteur de classement
- Tableaux des inerties
- Poids des variables
- Scores des observations
- Résumé classement (apprentissage)
- Détails classement (apprentissage)
- Matrice de confusion (apprentissage)
- Distances entre les modèles
- Pouvoirs discriminants des variables**
- Variable Ionsepal**
- Variable Ionsepal
- Variable Ionpetal
- Variable Iappetal
- Détails classement (prévision)

	1	2	3	4	5	6	7	8
1								
2	POUVOIR DISCRIMINANT DE LA VARIABLE Ionsepal							
3								
4								
5		Setosa	Versicolor	Virginica				
6	Setosa	0,95743	2,70613	4,55709				
7	Versicolor	2,70613	0,95743	2,08476				
8	Virginica	4,55709	2,08476	0,95743				
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

Voir l'option 'Graphiques' pour les détails sur les calculs effectués.

Le dernier tableau affiche les détails du classement pour les observations du jeu de prévision.

Rapports et Graphiques

Rapport SIMCA

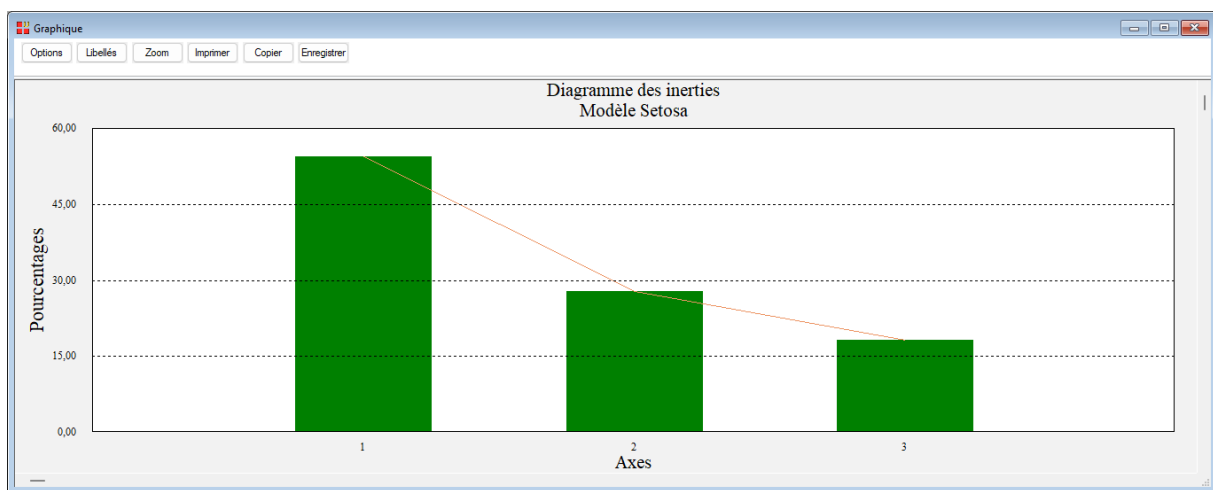
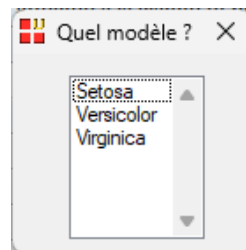
- Facteur de classement
- Tableaux des inerties
- Poids des variables
- Scores des observations
- Résumé classement (apprentissage)
- Détails classement (apprentissage)
- Matrice de confusion (apprentissage)
- Distances entre les modèles
- Pouvoirs discriminants des variables
- Variable Ionsepal
- Variable Ionsepal
- Variable Ionpetal
- Variable Iappetal
- Détails classement (prévision)**

	1	2	3	4	5	6	7	8
1								
2	DETAILS DU CLASSEMENT SIMCA - JEU DE PREVISION							
3								
4	Ce tableau indique pour chaque observation par 1 qu'elle est affectée à la classe et par -1 qu'elle n'est pas affectée à la classe.							
5								
6								
7		Setosa	Versicolor	Virginica				
8	3	1	-1	-1				
9	36	1	-1	-1				
10	62	-1	1	-1				
11	84	-1	-1	1				
12	104	-1	-1	1				
13	125	-1	-1	1				
14								

L'option Graphiques

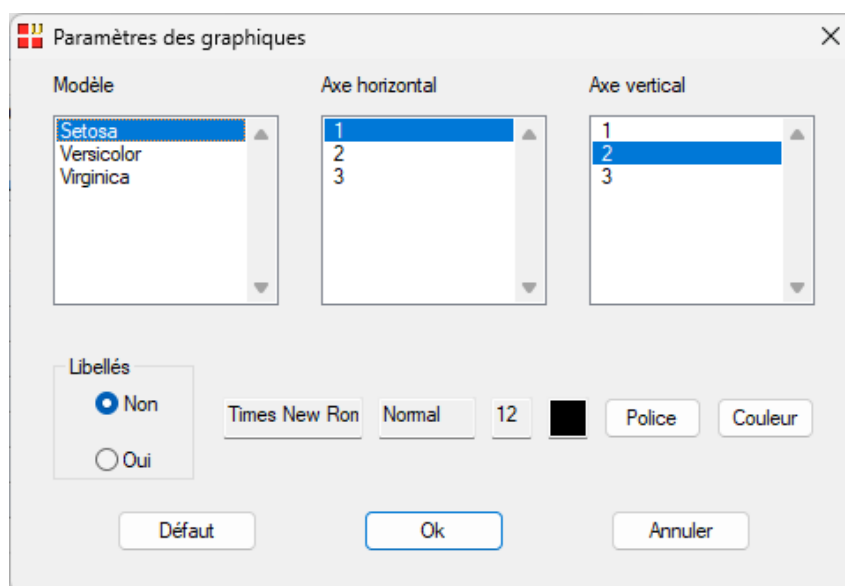
- Diagramme des inerties

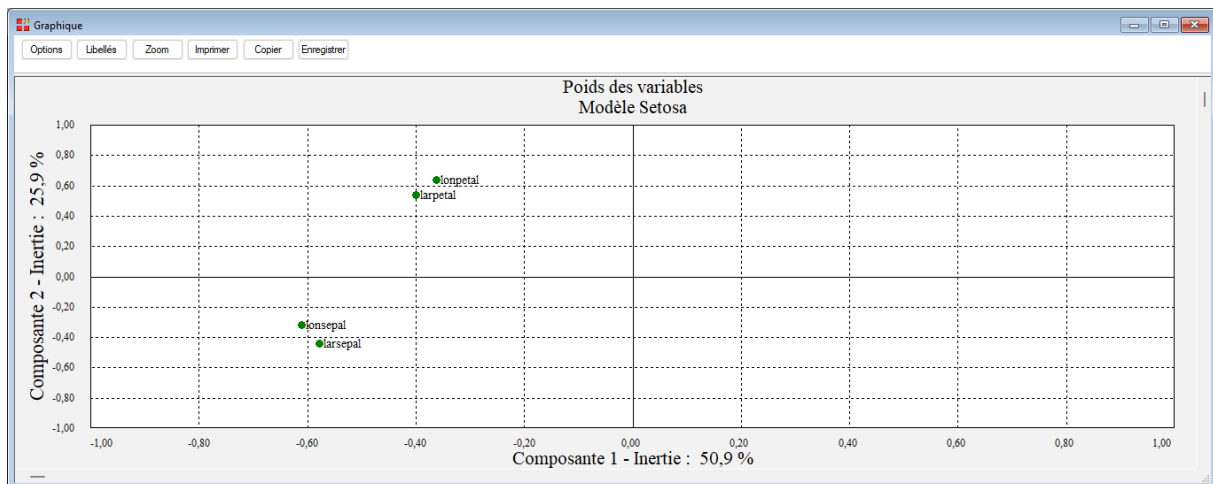
Ce graphique affiche le diagramme des inerties pour chacun des modèles. Il permet de choisir le modèle.



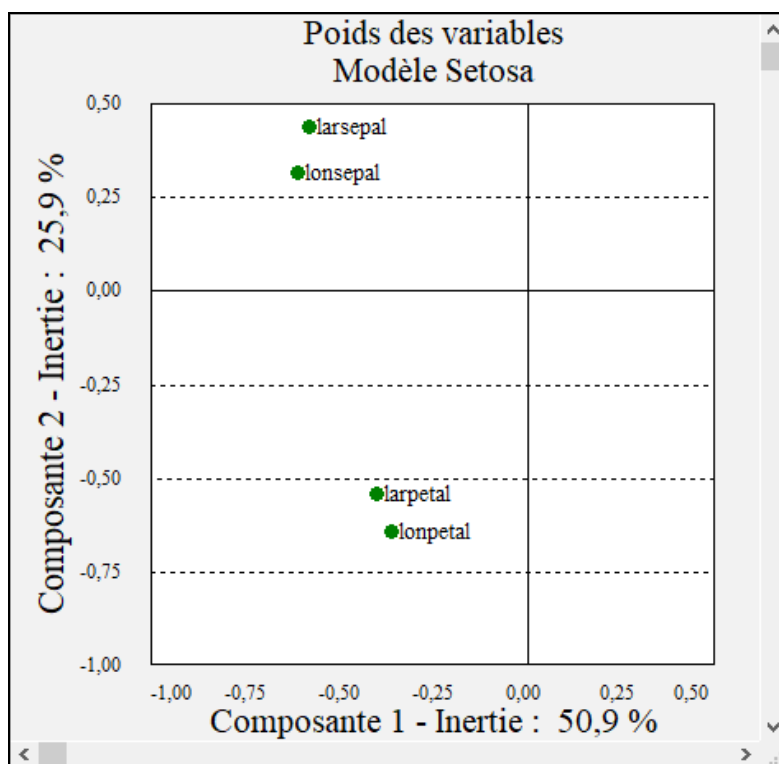
- Poids des variables

Ce graphique affiche les poids des variables pour chacun des modèles. Il permet de choisir le plan factoriel désiré et d'afficher ou non les libellés des variables.





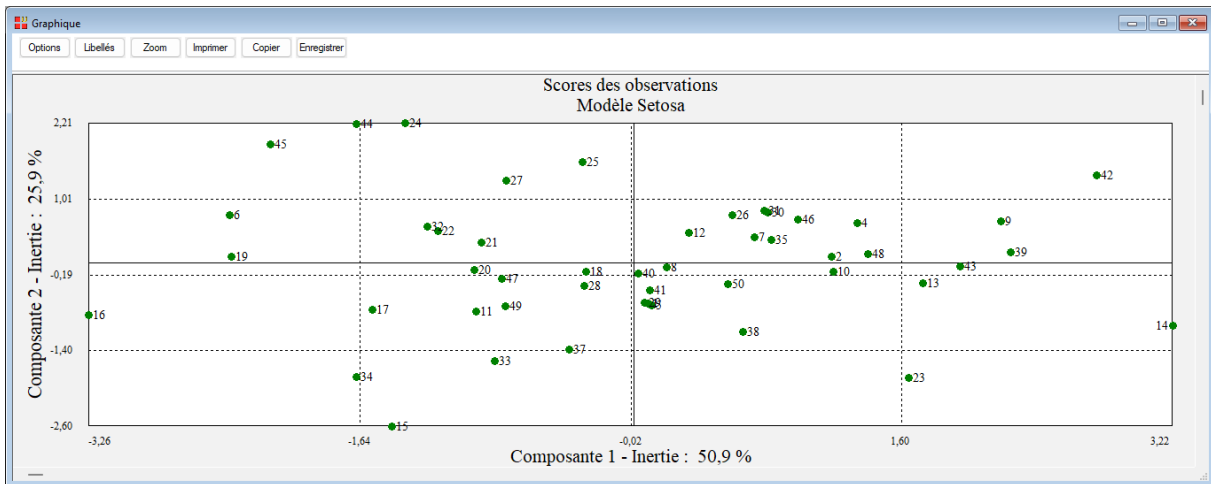
Par les options graphiques (échelles des axes), il est possible d'afficher un graphique orthonormé.



- Scores des observations

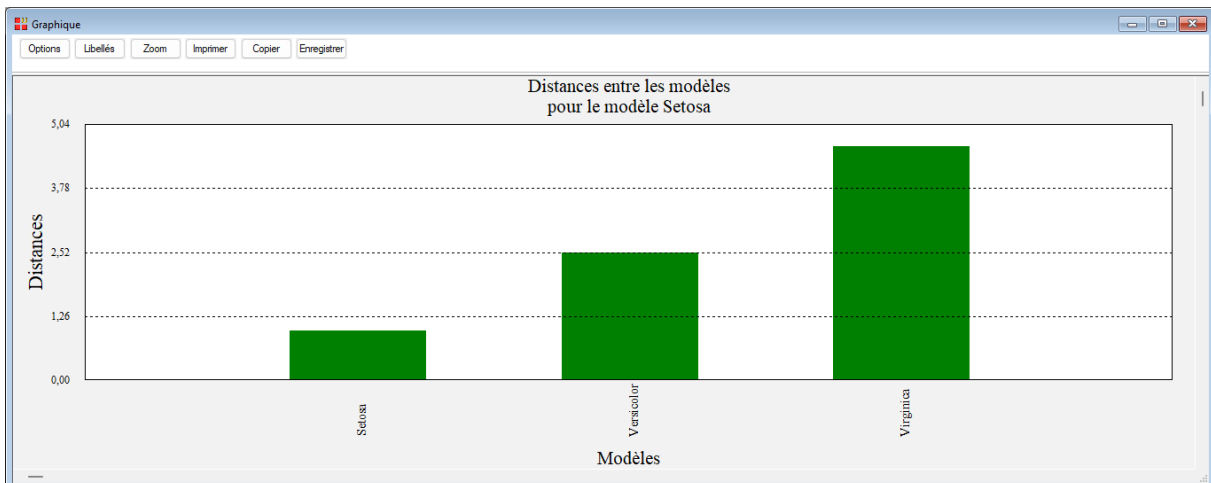
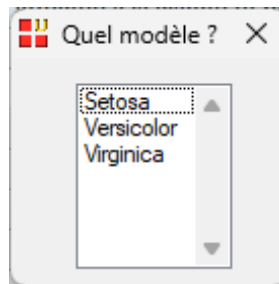
Ce graphique affiche les scores des observations pour chacun des modèles.

Il permet comme pour le poids des variables de choisir le plan factoriel désiré et d'afficher ou non les libellés des observations.



- Distances entre les modèles

Ce graphique affiche les similarités entre un modèle sélectionné et les autres modèles en utilisant les variances résiduelles.



Plus précisément, voici l'algorithme mis en œuvre.

Soient m_1 et m_2 deux modèles ayant les nombres respectifs de composantes A_1 et A_2 . Ces modèles ont été élaborés en utilisant les jeux d'apprentissage X_1 et X_2 ayant respectivement n_1 et n_2 observations.

Faisons les calculs suivants :

1. Etablissons le modèle m1 pour X2 et calculons les résidus E12
2. Calculons la variance des résidus $s_{12} = \text{somme}(E_{12}^2) / n_1$
3. Etablissons le modèle m2 pour X1 et calculons les résidus E21
4. Calculons la variance des résidus $s_{21} = \text{somme}(E_{21}^2) / n_2$
5. Calculons la variance des résidus pour m1 avec $s_1 = \text{somme}(E_1^2) / (n_1 - A_1 - 1)$
6. Calculons la variance des résidus pour m2 avec $s_2 = \text{somme}(E_2^2) / (n_2 - A_2 - 1)$

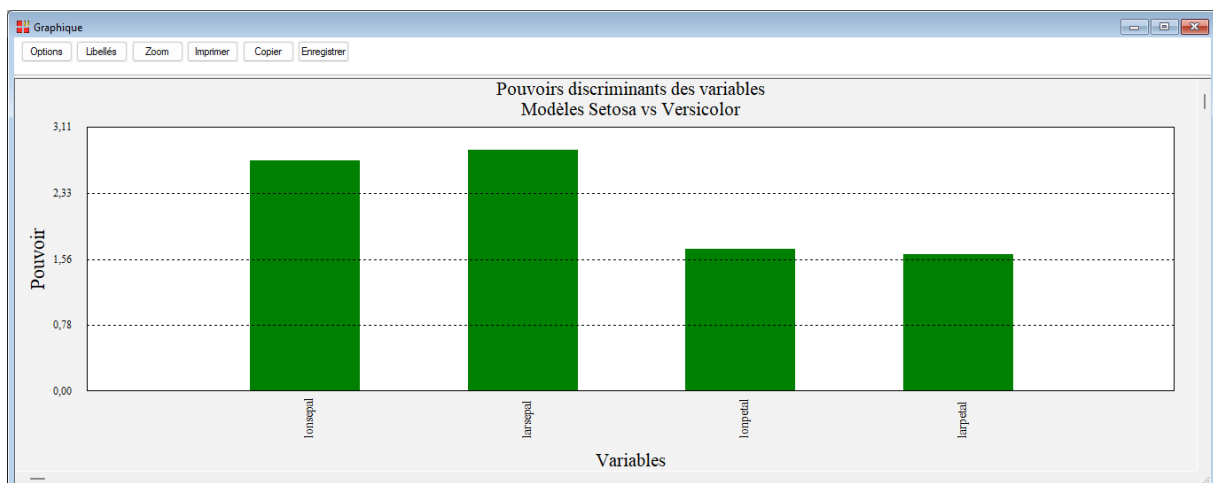
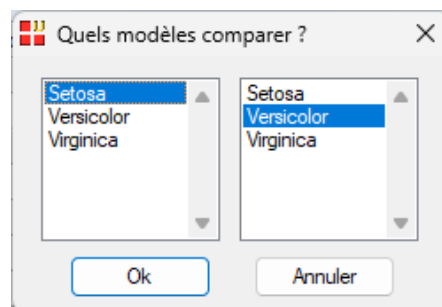
La distance au modèle est calculée par : $d = \text{racine}((s_{12} + s_{21}) / (s_1 + s_2))$

Si les deux modèles et les deux jeux d'apprentissage sont identiques, alors la distance est égale à $\text{racine}((n - A - 1) / n)$.

En général, si la distance entre les modèles est inférieure à 1, cela indique que les classes se recouvrent. Si la distance est supérieure à 3, les classes sont bien séparées.

- Pouvoirs discriminants des variables

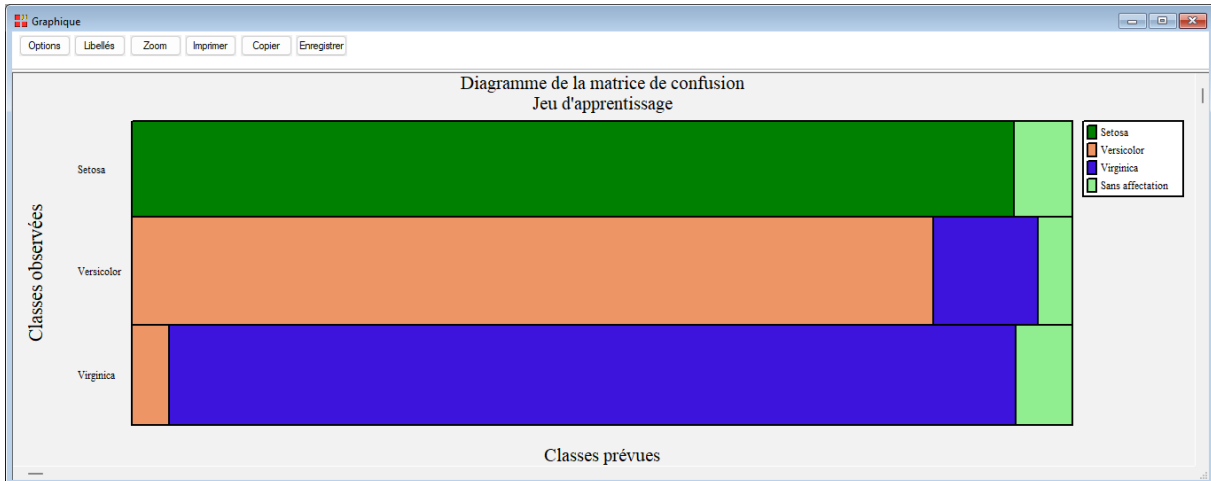
Ce graphique affiche pour deux modèles sélectionnés les pouvoirs discriminants des variables c'est-à-dire la capacité de ces variables à séparer les classes.



Le pouvoir discriminant est calculé comme la distance au modèle en utilisant la variance des résidus. Toutefois, dans ce cas, au lieu de sommer la variance sur toutes les variables, le calcul est fait de façon séparée pour chaque variable. Un pouvoir discriminant égal ou supérieur à 3 est considéré comme élevé.

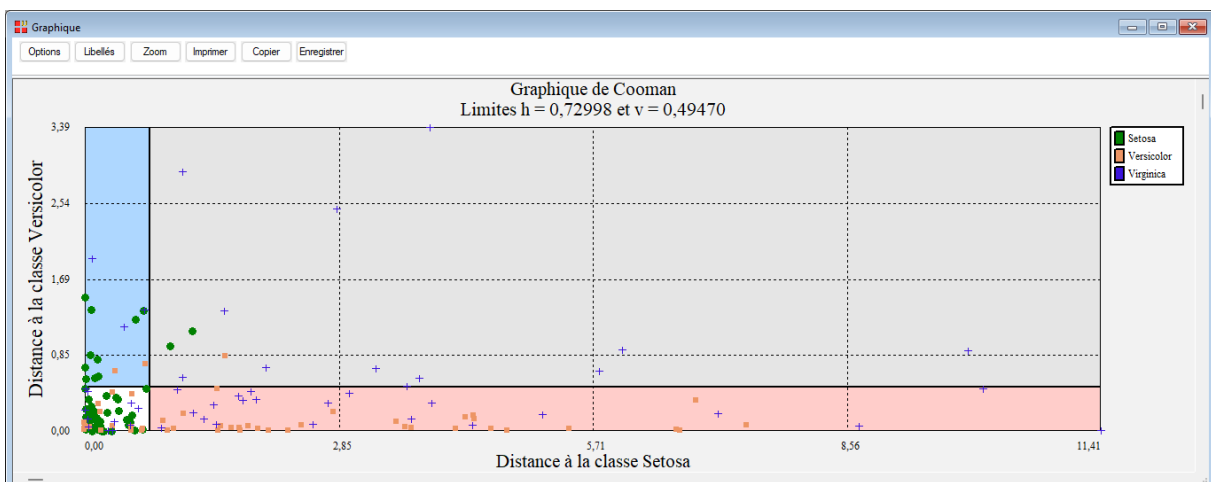
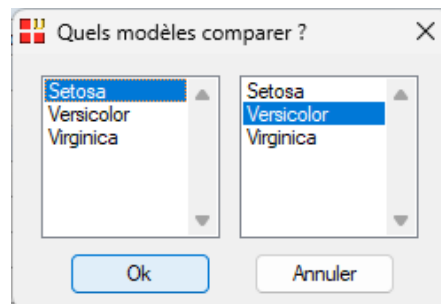
- Diagramme de la matrice de confusion

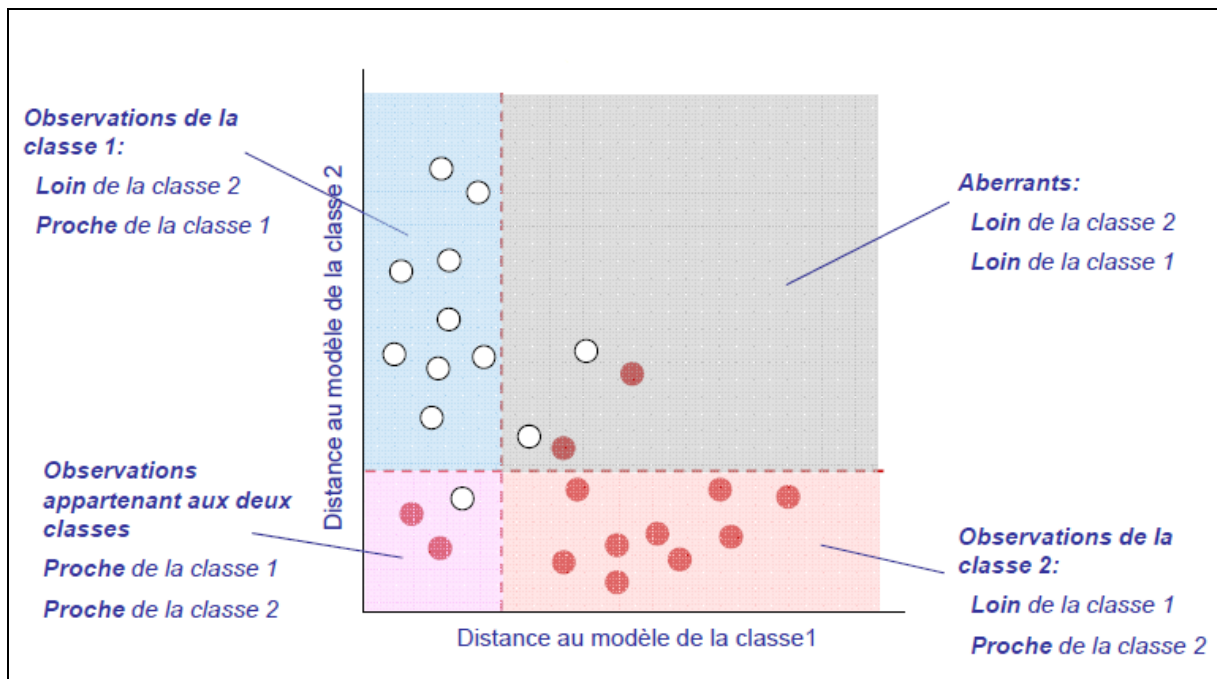
Le diagramme en mosaïque affiche le tableau de confusion en traçant des rectangles dont les aires sont proportionnelles aux effectifs des cellules. De plus, les largeurs des bâtons sont proportionnelles aux pourcentages de la distribution du facteur de classement.



- Graphique de Cooman

Cette option affiche pour les deux modèles sélectionnés le graphique de Cooman.





Origine : Julien Boccard – Université de Genève

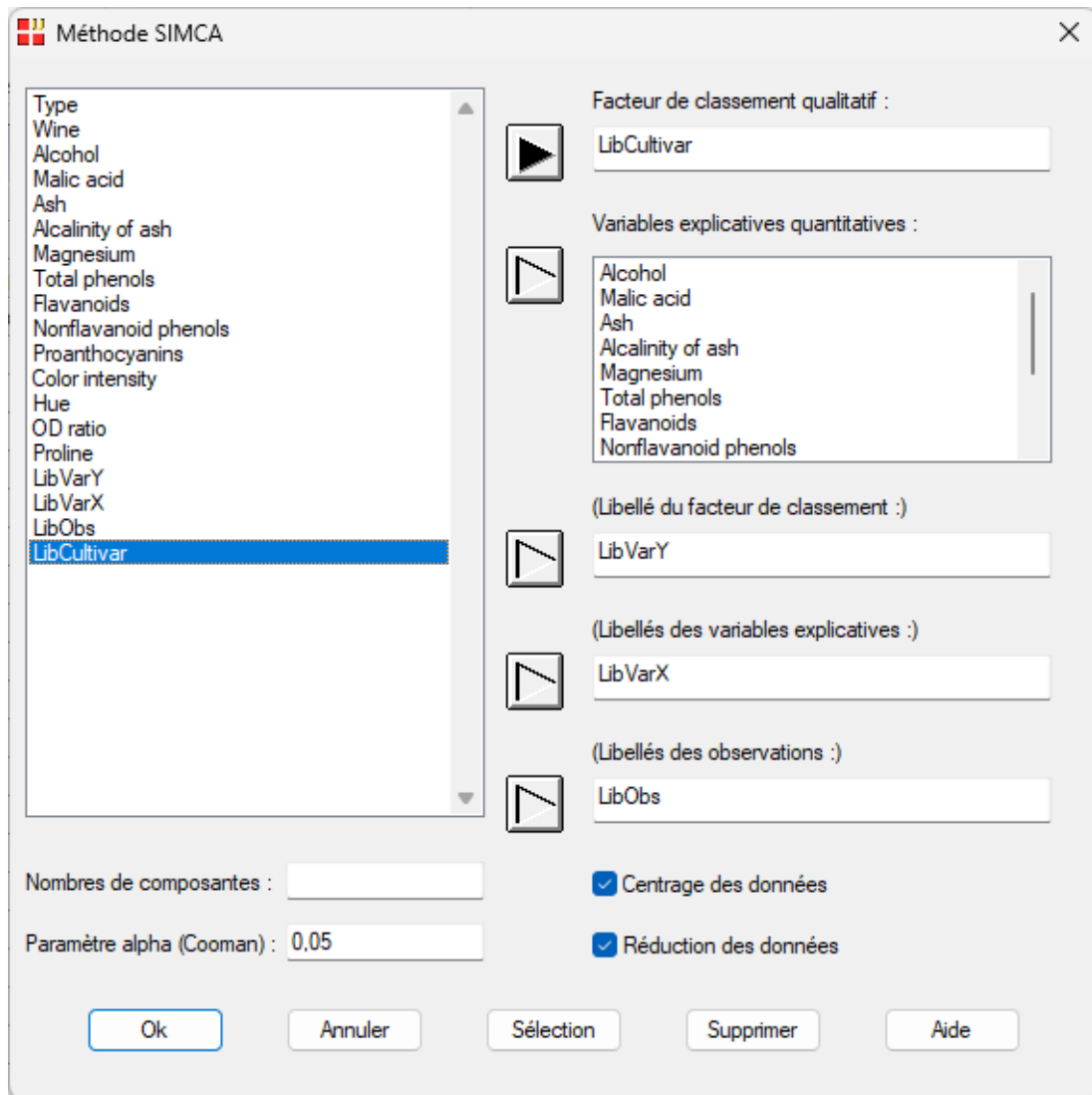
Les limites horizontale et verticale sont calculées en utilisant la loi du Khi-carré avec le niveau de signification alpha.

Exemple 2 : Fichier WINES2

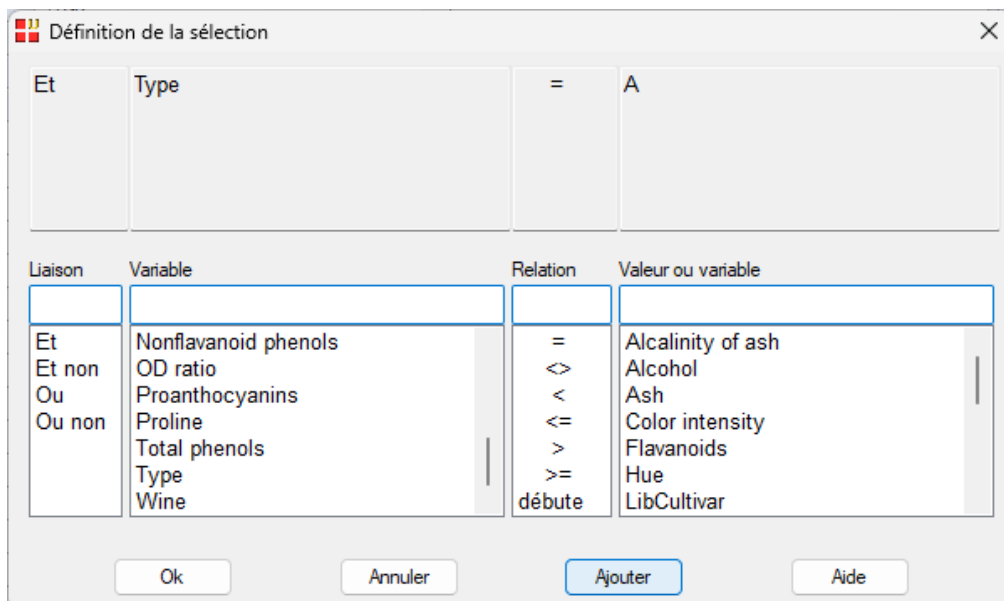
Les données sont le résultat de l'analyse chimique de 180 échantillons de vins issus de trois cultivars différents et provenant d'une même région en Italie (source : <https://archive.ics.uci.edu/ml/datasets/wine>). Elles sont constituées des treize caractéristiques chimiques et spectroscopiques suivantes :

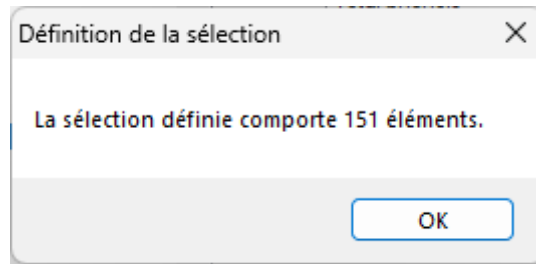
wine	Numéro du cultivar
alcohol	Alcool
malic acid	Acide malique
ash	Cendres
ash alkalinity	Alcalinité des cendres
magnesium	Magnésium
tot. phenols	Total des phénols
flavonoids	Flavonoïdes
non-flav. phenols	Phénols non-flavonoïdes
proanth	Proanthocyanidines
col. int.	Intensité de couleur
col. hue	Teinte de couleur
OD ratio	Rapport OD
proline	Proline

Cliquons sur l'icône SIMCA dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après (les variables explicatives de 'Alcohol' à 'Proline' sont sélectionnées).



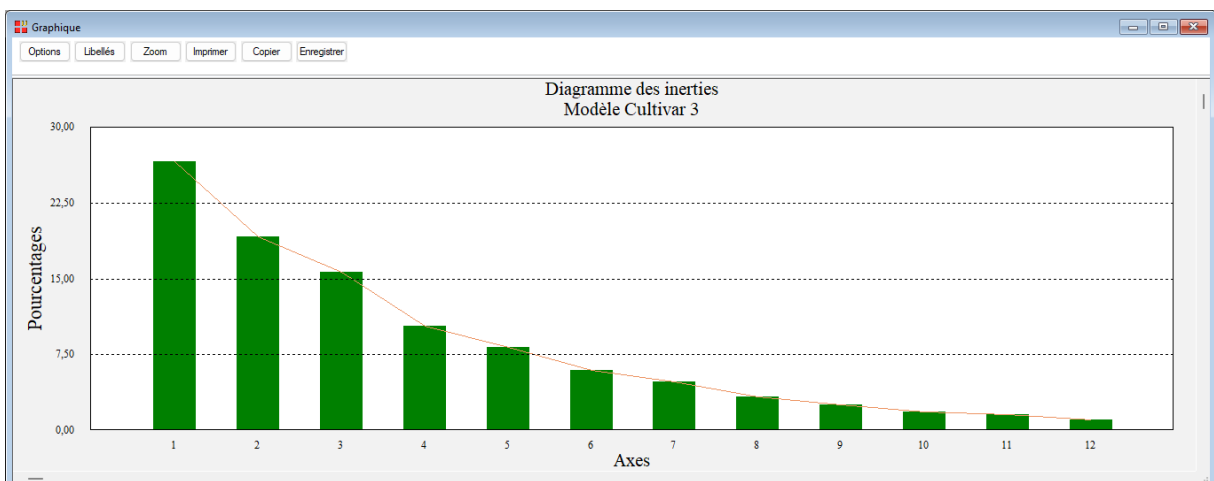
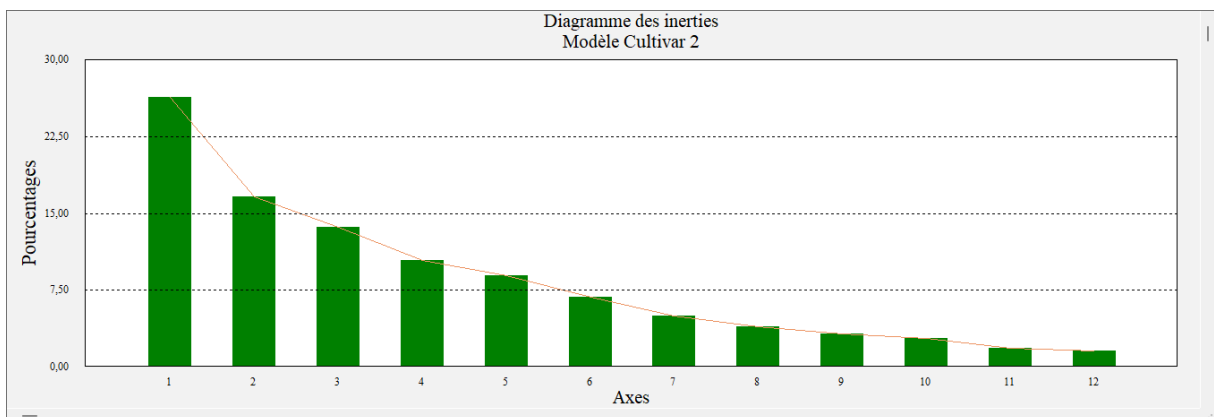
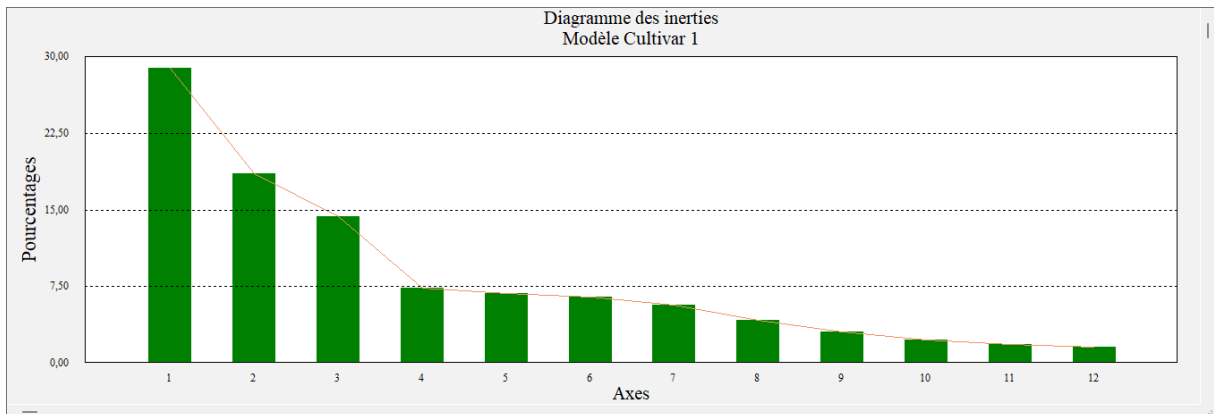
Cliquons sur le bouton 'Sélection' pour définir les données du jeu d'apprentissage et utilisons la colonne 'Type' du fichier des données pour cela.






Cliquons sur Ok puis confirmons le nombre de modalités. Après quelques instants la fenêtre 'Rapports et Graphiques' s'affiche.

Visualisons les diagrammes des inerties pour chacun des trois modèles.



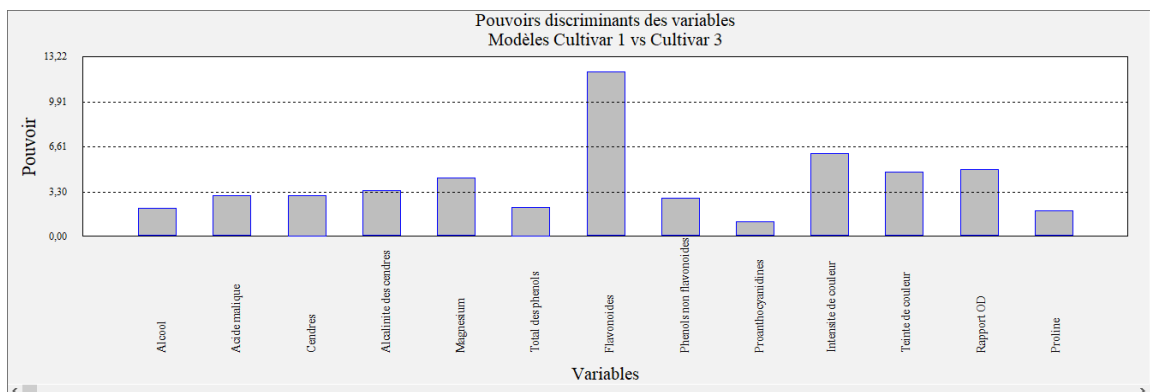
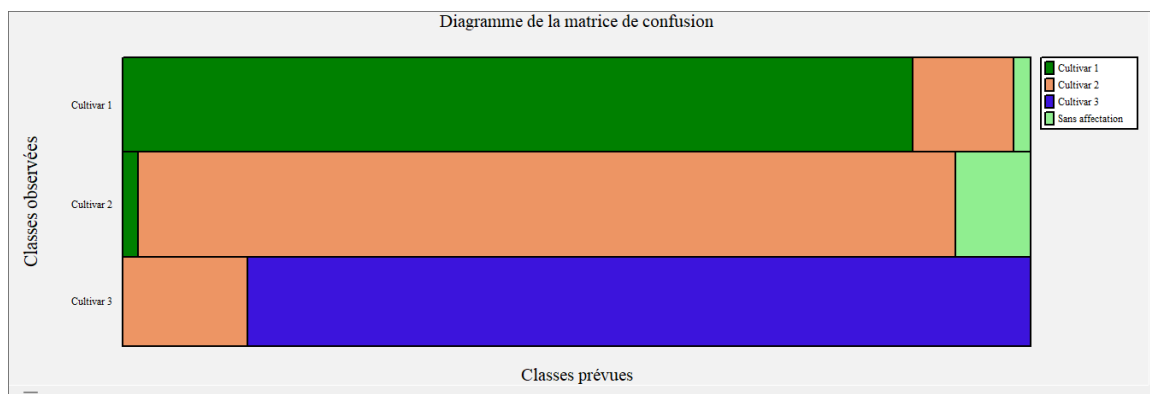
Ces résultats nous indiquent que les nombres de composantes à retenir pour chacun des modèles sont probablement respectivement de 3, 5 et 5.

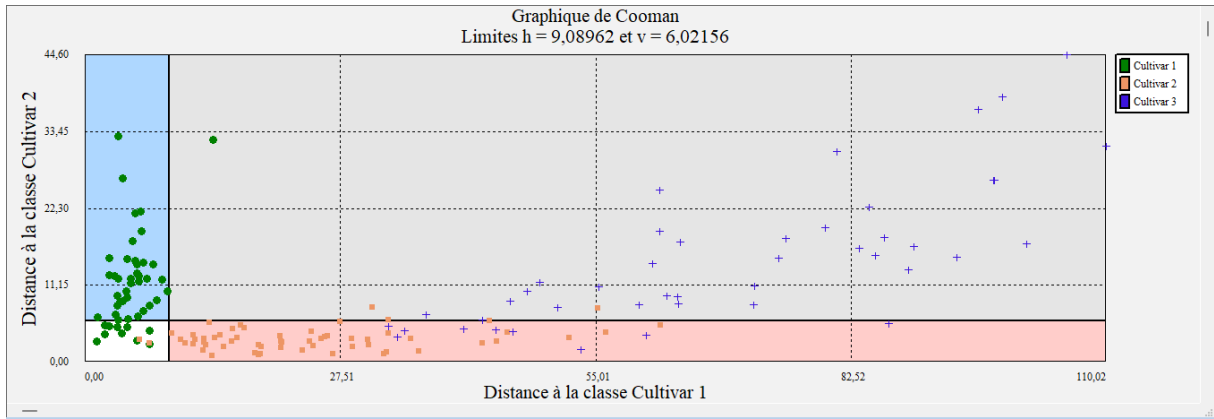
Exécutons donc à nouveau l'analyse en cliquant sur l'icône  de la barre d'outils et entrons ces nombres dans le champ 'Nombres de composantes'.

Cliquons sur Ok. Voici quelques-uns des résultats obtenus.

RESUME DU CLASSEMENT SIMCA - JEU D'APPRENTISSAGE									
Le tableau affiche pour chacun des modèles les nombres de composantes, les pourcentages de bien classés, les vrais positifs (VP), les faux positifs (FP), les vrais négatifs (VN), les faux négatifs (FN), les spécificités et les sensibilités.									
Exactitude = $(VP + VN) / (VP + VN + FP + FN)$									
Sensibilité = $VP / (VP + FN)$									
Spécificité = $VN / (VN + FP)$									
	Nb composantes	VP	FP	VN	FN	Exactitude	Spécificité	Sensibilité	
Cultivar 1	3	47	1	102	1	0,987	0,990	0,979	
Cultivar 2	5	54	13	79	5	0,881	0,859	0,915	
Cultivar 3	5	44	0	107	0	1,000	1,000	1,000	

MATRICE DE CONFUSION DU CLASSEMENT SIMCA - JEU D'APPRENTISSAGE							
La matrice de confusion affiche les affectations des données aux classes existantes ou à aucune classe.							
En lignes, les classes observées, en colonnes les classes prévues.							
Rappel = $VP / (VP + FN)$							
Précision = $VP / (VP + FP)$							
Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
	Cultivar 1	Cultivar 2	Cultivar 3	Sans affectation	Précision	Rappel	Score F1
Cultivar 1	47	6	0	1	0,979	0,979	0,979
Cultivar 2	1	54	0	5	0,806	0,915	0,857
Cultivar 3	0	7	44	0	1,000	1,000	1,000





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 créées par la procédure

Voici la liste des variables créées par la procédure.

<i>Variable</i>	<i>Contenu</i>
valpro	Valeurs propres pour le modèle
obs	Libellés des observations de la classe
poids	Poids des variables pour le modèle
scores	Scores des observations pour le modèle
obsapp	Libellés des observations (jeu d'apprentissage)
cpa	Classes prévues (jeu d'apprentissage)
coa	Classes observées apprentissage
resume	Résumé statistique
confusion	Matrice de confusion
distmod	Distances entre les modèles
discrivar	Pouvoirs discriminants de chaque variable
obsprev	Libellés des observations (jeu de prévision)
cpp	Classes prévues (jeu de prévision)

Références

[Documentation du package R 'mdatools' \(2024\)](#)

<https://cran.r-project.org/web/packages/mdatools/mdatools.pdf>

Exemple 2

<https://archive.ics.uci.edu/ml/datasets/wine>