

## UNIWIN VERSION 10.4.0

# METHODE BAYESIENNE NAIVE

Révision : 15/09/2025

Définition.....	1
Entrée des données .....	2
Définition des lois conditionnelles.....	3
Données manquantes .....	3
Exemple 1 : Fichier BAYES .....	3
L'option Rapports .....	6
L'option Graphiques .....	11
Exemple 2 : Fichier IRIS3.....	15
Exemple 3 : Fichier TITANIC.....	19
Calculs de la matrice de confusion et des indicateurs .....	21
Les variables créées par la procédure.....	22
Références .....	23

### Définition

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

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

Les données brutes sont utilisées car les résultats de cette méthode ne sont pas affectés par le prétraitement des données.

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

## Entrée des données

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

Méthode bayésienne naïve

Facteur de classement qualitatif :

Variables explicatives quantitatives et qualitatives :

(Libellé du facteur de classement :)

(Libellés des variables explicatives :)

(Libellés des observations :)

(Probabilités a priori :)

Paramètre de lissage de Laplace : 0

Estimation non paramétrique par noyau  
(variables numériques continues)

Loi de Poisson  
(variables numériques discrètes positives)

Ok Annuler Sélection Supprimer Aide

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

Elle permet également de préciser les probabilités a priori de chacun des niveaux du facteur de classement (dans l'ordre alphabétique de ces niveaux), la valeur du paramètre de lissage de Laplace, si les lois des variables numériques continues doivent être estimées de façon non paramétrique par la méthode du noyau et si les lois des variables numériques discrètes à valeurs non négatives doivent être estimées par la loi de Poisson.

Le lissage de Laplace permet d'éviter d'obtenir des probabilités nulles ou égales à un dans le cas de petits échantillons. Dans ce cas, une valeur usuelle de ce paramètre est 1.

### **Définition des lois conditionnelles**

Si « Estimation non paramétrique par noyau » et « Loi de Poisson » ne sont pas cochés, alors la loi normale est utilisée pour toutes les variables numériques.

Si « Estimation non paramétrique par noyau » est coché et « Loi de Poisson » n'est pas coché, alors l'estimation non paramétrique est utilisée pour toutes les variables numériques.

Si « Estimation non paramétrique par noyau » n'est pas coché et « Loi de Poisson » est coché, alors la loi normale est utilisée pour toutes les variables numériques continues et la loi de Poisson est utilisée pour toutes les variables numériques discrètes à valeurs non négatives.

Si « Estimation non paramétrique par noyau » et « Loi de Poisson » sont cochés, alors l'estimation non paramétrique est utilisée pour toutes les variables numériques continues et la loi de Poisson est utilisée pour toutes les variables numériques discrètes à valeurs non négatives.

### **Données manquantes**

Dans cette procédure, les valeurs manquantes du facteur de classement permettent de définir l'échantillon de prévision. Les lignes ayant des valeurs manquantes pour les variables explicatives sont automatiquement éliminées de l'analyse.

### **Exemple 1 : Fichier BAYES**

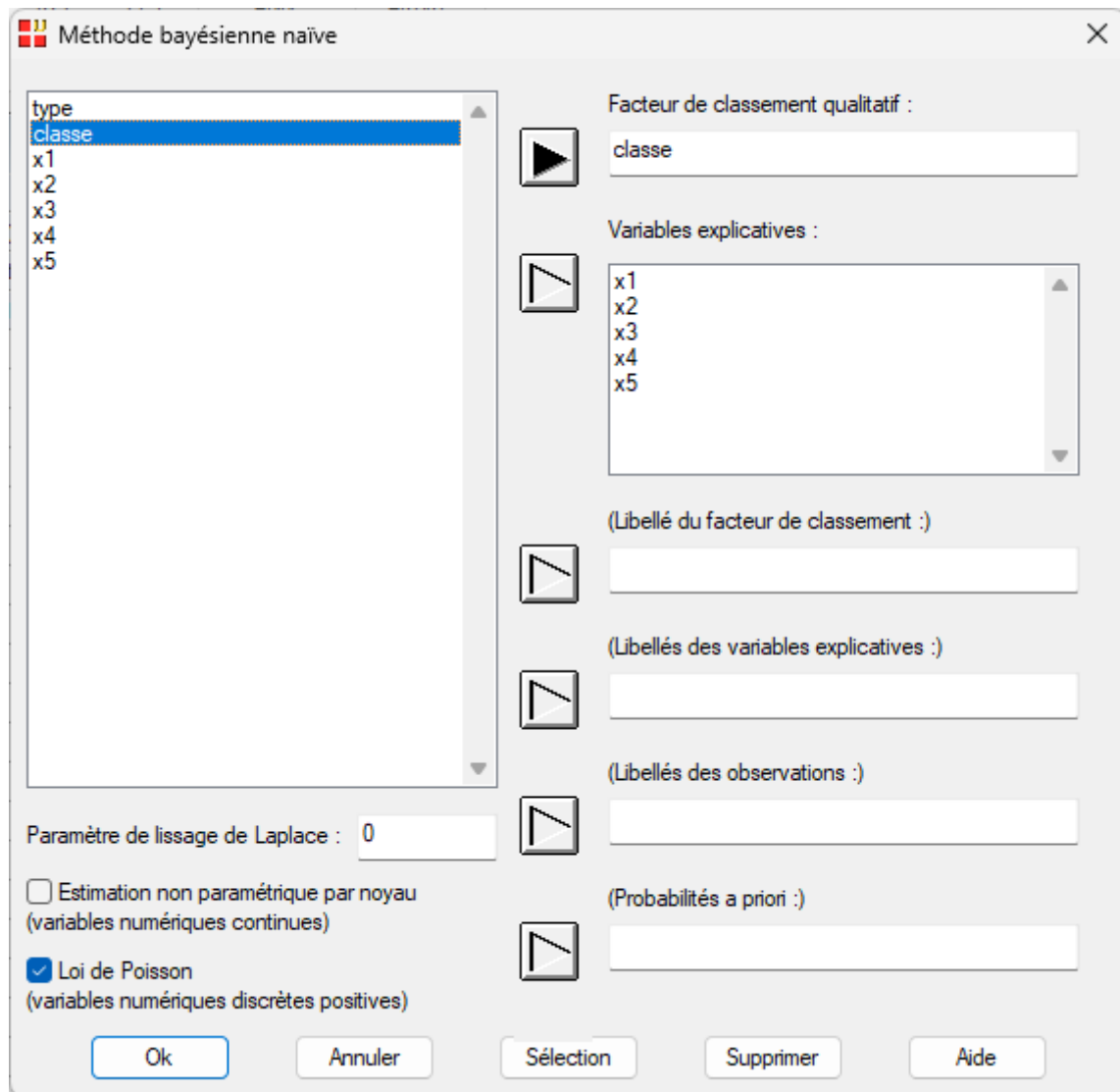
Pour ce premier exemple, nous utiliserons le fichier BAYES pour illustrer cette procédure. Ce fichier contient 100 observations pour lesquelles les variables suivantes sont collectées :

- type : A si l'observation est utilisée pour l'apprentissage, P si l'observation est utilisée pour la prévision
- classe : facteur de classement à deux modalités : classA, classB

- x1 : variable qualitative à deux modalités A et B (Bernoulli)
- x2 : variable qualitative à trois modalités a, b et c (catégorielle)
- x3 : variable prenant les valeurs 0 ou 1 (logique)
- x4 : variable numérique continue (normale)
- x5 : variable numérique discrète (comptage)

Ce fichier contient 5 observations pour lesquelles les classes d'appartenance sont inconnues. Elles définissent l'échantillon de prévision.

Renseignons la boîte de dialogue comme montré ci-après.

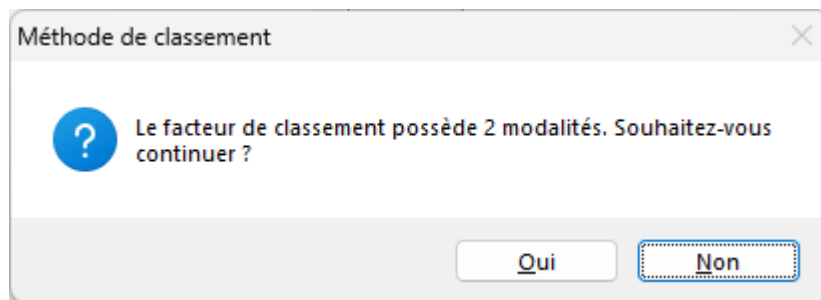


Sélectionnons la variable *classe* comme facteur de classement, les variables *x1* à *x5* comme variables explicatives.

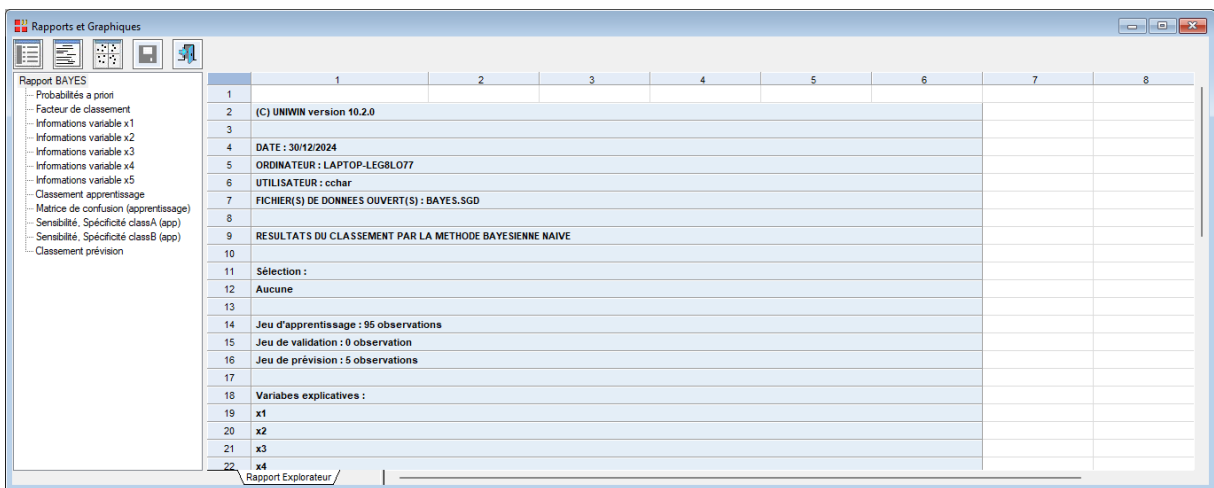
Cochons '*Loi de Poisson*' pour le traitement des variables discrètes à valeurs positives.


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

Un message informatif nous indique le nombre de modalités du facteur de classement. Cliques sur Oui.

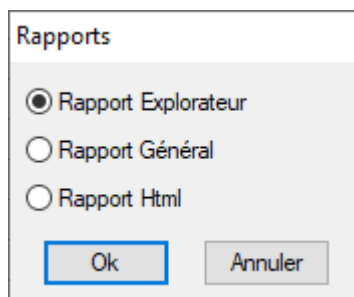



Après quelques instants, la fenêtre « Rapports et Graphiques » s'affiche :

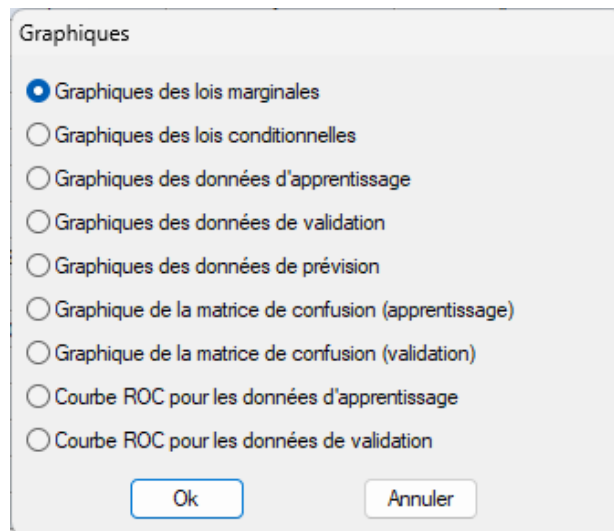


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

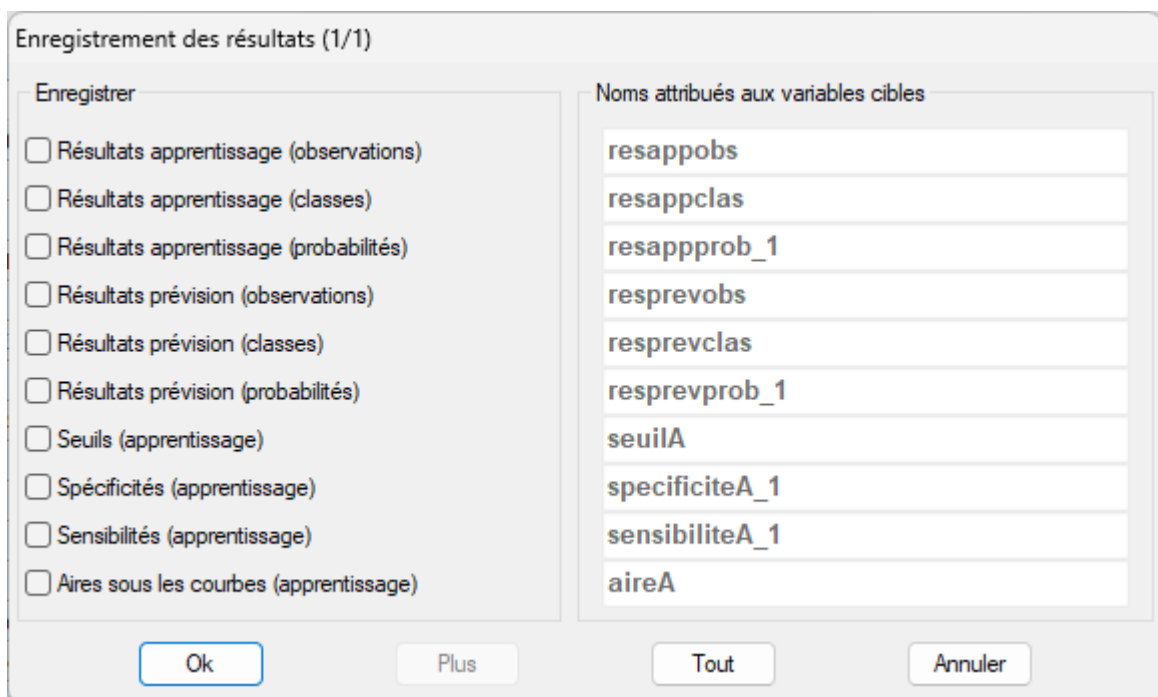
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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, de validation et de prévision. Il précise également les lois conditionnelles utilisées pour chacune des variables explicatives.

	1	2	3	4	5	6	7	8
14	Nombre d'observations : 95							
15	Nombre de variables explicatives : 5							
16								
17	Jeu d'apprentissage : 95 observations							
18	Jeu de validation : 0 observations							
19	Jeu de prévision : 5 observations							
20								
21	Facteur de classement : classe							
22								
23	Classes :							
24	classA							
25	classB							
26								
27	Code de l'évènement d'intérêt : classB							
28	Evènement d'intérêt si : mesures <= seuil							
29								
30	Lissage de Laplace : 0,000							
31	Méthode du noyau : non							
32	Loi de Poisson : oui							
33								
34	Lois conditionnelles des variables explicatives :							
35								

Le deuxième tableau affiche les probabilités a priori des classes.

	1	2	3	4	5	6	7	8
1								
2	PROBABILITES A PRIORI DES CLASSES							
3								
4								
5	Probabilité							
6	classA	0,48421						
7	classB	0,51579						
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Le troisième tableau affiche un tri à plat des données 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	classA	46	48,42	46	48,42			
9	classB	49	51,58	95	100,00			
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Les tableaux suivants affichent des informations concernant chacune des variables explicatives.

### X1 : Loi de Bernoulli

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE x1</b>							
3								
4	<b>Loi de Bernoulli</b>							
5								
6								
7		Classe classA	Classe classB					
8	Modalité A	0.5	0.55102					
9	Modalité B	0.5	0.44898					
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

### X2 : loi catégorielle

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE x2</b>							
3								
4	<b>Loi catégorielle</b>							
5								
6								
7		Classe classA	Classe classB					
8	Modalité a	0.41304	0.40816					
9	Modalité b	0.28261	0.38776					
10	Modalité c	0.30435	0.20408					
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

### X3 : Loi de Bernoulli

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE x3</b>							
3								
4	<b>Loi de Bernoulli</b>							
5								
6								
7		Classe classA	Classe classB					
8	Modalité FALSE	0.56522	0.28571					
9	Modalité TRUE	0.43478	0.71429					
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

## X4 : Loi normale

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE x4</b>							
3								
4	<b>Loi normale</b>							
5								
6								
7		Classe classA	Classe classB					
8	Moyenne	0.01676	0.02925					
9	Ecart-type	1.07402	0.94079					
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

## X5 : Loi de Poisson

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE x5</b>							
3								
4	<b>Loi de Poisson</b>							
5								
6								
7		Classe classA	Classe classB					
8	lambda	9.39130	9.81633					
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Les deux tableaux suivants affichent les résultats détaillés du classement de la population d'apprentissage et une synthèse de ce classement (matrice de confusion).

	1	2	3	4	5	6
1	<b>RESULTATS DU CLASSEMENT DE LA POPULATION D'APPRENTISSAGE</b>					
2						
3						
4	<b>Individu - Classe observée - Classe prévue - Probabilités a posteriori</b>					
5						
6						
7				classA	classB	
8	Individu : o1 - Observé : classA -> Prévu : classA			0.60798	0.39202	
9	Individu : o2 - Observé : classB -> Prévu : classB			0.39452	0.60548	
10	Individu : o3 - Observé : classA -> Prévu : classA			0.62860	0.37140	
11	Individu : o4 - Observé : classA -> Prévu : classB *			0.41998	0.58002	
12	Individu : o5 - Observé : classB -> Prévu : classA *			0.55237	0.44763	
13	Individu : o6 - Observé : classA -> Prévu : classA			0.70694	0.29306	
14	Individu : o7 - Observé : classA -> Prévu : classA			0.72215	0.27785	
15	Individu : o8 - Observé : classA -> Prévu : classB *			0.20031	0.79969	
16	Individu : o9 - Observé : classB -> Prévu : classA *			0.52362	0.47638	
17	Individu : o10 - Observé : classB -> Prévu : classA *			0.65487	0.34513	
18	Individu : o11 - Observé : classA -> Prévu : classA			0.73054	0.26946	
19	Individu : o12 - Observé : classA -> Prévu : classB *			0.44607	0.55393	
20	Individu : o13 - Observé : classA -> Prévu : classA			0.55105	0.44895	
21	Individu : o14 - Observé : classA -> Prévu : classA			0.65567	0.34433	
22	Individu : o15 - Observé : classA -> Prévu : classB *			0.39537	0.60463	

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable x1
- Informations variable x2
- Informations variable x3
- Informations variable x4
- Informations variable x5
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Sensibilité, Spécificité classA (app)
- Sensibilité, Spécificité classB (app)
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>MATRICE DE CONFUSION DE LA POPULATION D'APPRENTISSAGE</b>							
3								
4	<b>En lignes, les classes observées</b>							
5	<b>En colonnes, les classes prévues</b>							
6								
7	<b>Pourcentage de mal classés : 30,526 %</b>							
8	<b>Pourcentage de bien classés : 69,474 %</b>							
9								
10								
11		Taille	classA	classB	Précision	Rappel	Score F1	
12	classA	46	30	16	0.69767	0.65217	0.67416	
13	classB	49	13	36	0.69231	0.73469	0.71287	
14								
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

Environ 70% des observations sont bien classées.

Voir la documentation sur la courbe ROC dans le « Manuel de l'utilisateur » pour des détails sur l'exactitude, la précision, le rappel et le score F1.

Les tableaux suivants affichent les informations concernant les sensibilités et spécificités utilisées pour le tracé des courbes ROC.

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable x1
- Informations variable x2
- Informations variable x3
- Informations variable x4
- Informations variable x5
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Sensibilité, Spécificité classA (app)
- Sensibilité, Spécificité classB (app)
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE</b>							
3								
4	<b>Classe : classA</b>							
5								
6	<b>Sensibilité en %</b>							
7	<b>Spécificité en %</b>							
8								
9	<b>Aire sous la courbe (AUC) = 0.708</b>							
10								
11								
12		Seuil	Sensibilité	Spécificité				
13	1	-Infini	0.00000	100.00000				
14	2	0.18558	0.00000	100.00000				
15	3	0.22186	2.17391	100.00000				
16	4	0.23130	4.34783	100.00000				
17	5	0.25018	6.52174	100.00000				
18	6	0.25354	6.52174	97.95918				
19	7	0.25603	6.52174	95.91837				
20	8	0.25644	6.52174	93.87755				
21	9	0.25945	8.69565	93.87755				
22	10	0.26946	8.69565	91.83673				

Rapport Explorateur /

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable x1
- Informations variable x2
- Informations variable x3
- Informations variable x4
- Informations variable x5
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Sensibilité, Spécificité classA (app)
- Sensibilité, Spécificité classB (app)
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE</b>							
3								
4	<b>Classe : classB</b>							
5								
6	<b>Sensibilité en %</b>							
7	<b>Spécificité en %</b>							
8								
9	<b>Aire sous la courbe (AUC) = 0.708</b>							
10								
11								
12		Seuil	Sensibilité	Spécificité				
13	1	-Infini	100.00000	0.00000				
14	2	0.18558	100.00000	0.00000				
15	3	0.22186	100.00000	2.17391				
16	4	0.23130	100.00000	4.34783				
17	5	0.25018	100.00000	6.52174				
18	6	0.25354	97.95918	6.52174				
19	7	0.25603	95.91837	6.52174				
20	8	0.25644	93.87755	6.52174				
21	9	0.25945	93.87755	8.69565				
22	10	0.26946	91.83673	8.69565				

Rapport Explorateur /

Enfin le dernier tableau affiche les résultats du classement de la population de prévision.

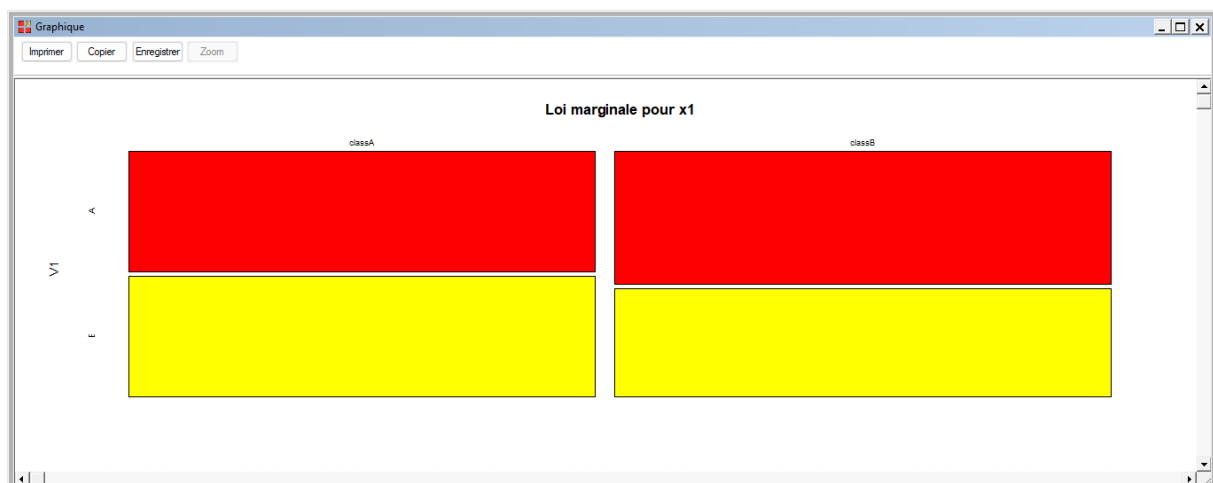
	1	2	3	4	5	6	7
1							
2	<b>RESULTATS DU CLASSEMENT DE LA POPULATION DE PREVISION</b>						
3							
4	<b>Individu - Classe prévue - Probabilités a posteriori</b>						
5							
6							
7		classA	classB				
8	Individu : o96 -> Prévu : classA	0,67082	0,32918				
9	Individu : o97 -> Prévu : classB	0,27928	0,72072				
10	Individu : o98 -> Prévu : classA	0,62148	0,37852				
11	Individu : o99 -> Prévu : classA	0,58069	0,41931				
12	Individu : o100 -> Prévu : classA	0,70748	0,29252				
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

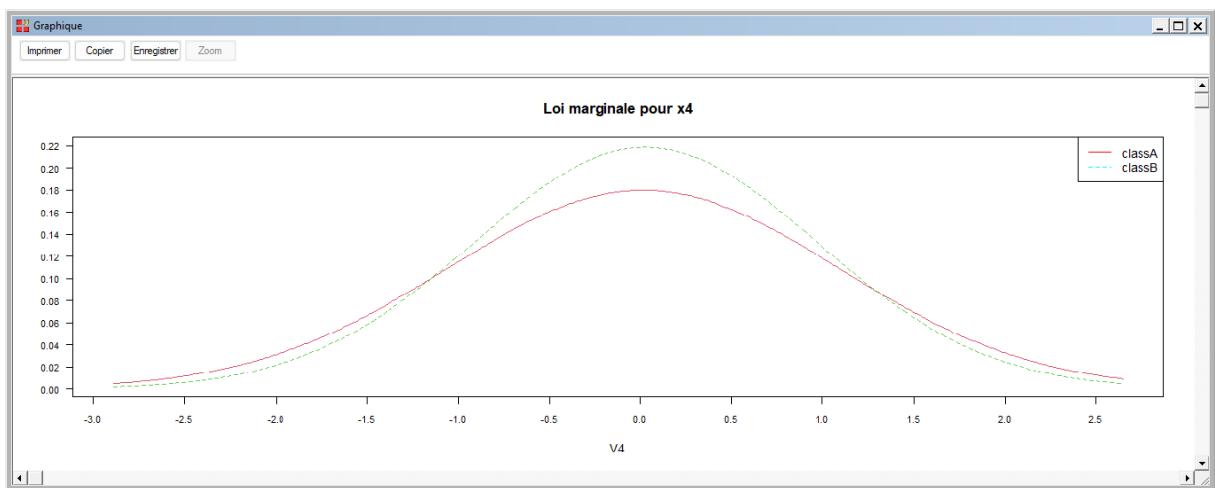
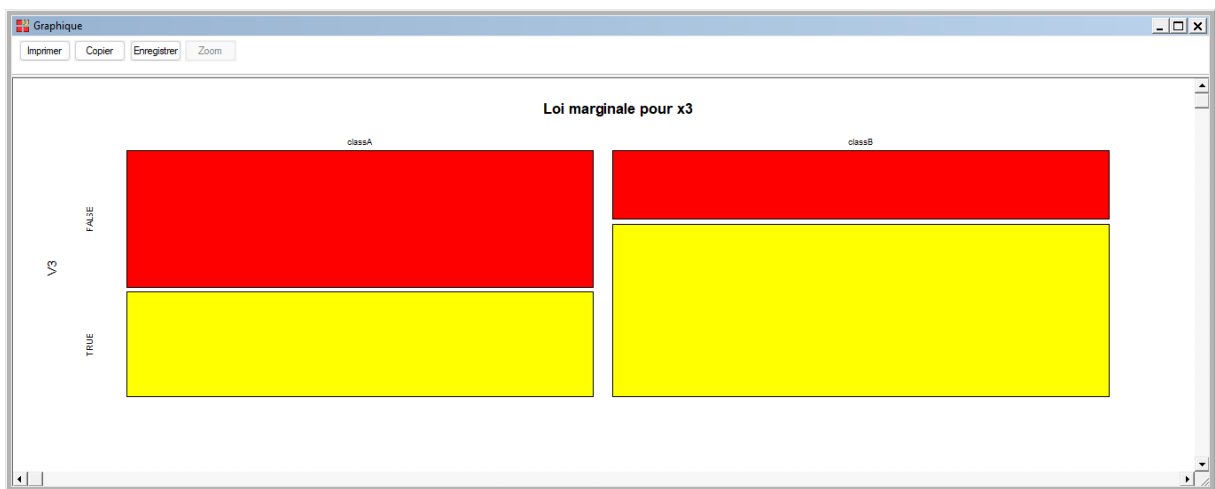
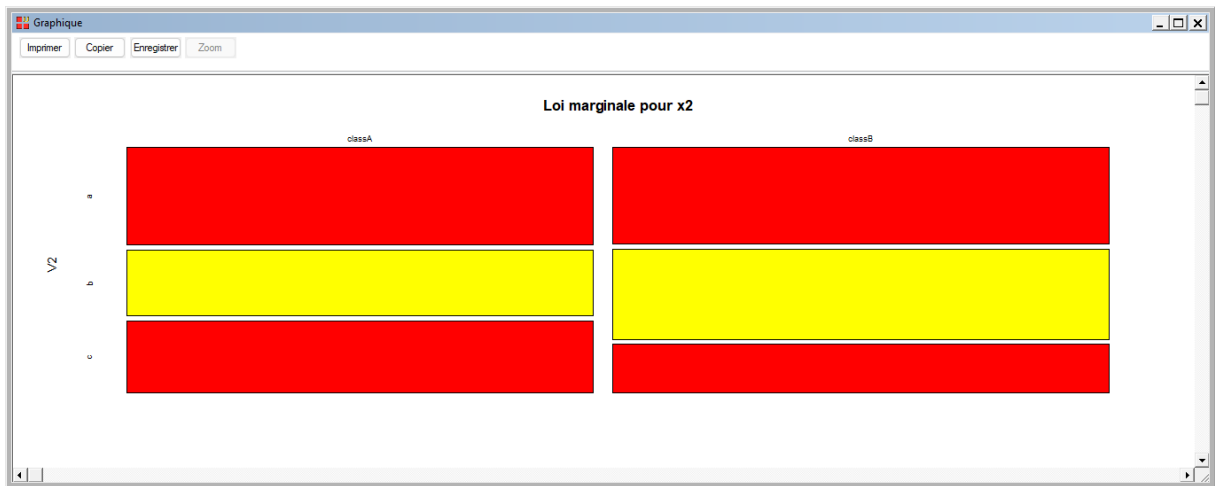
## L'option Graphiques

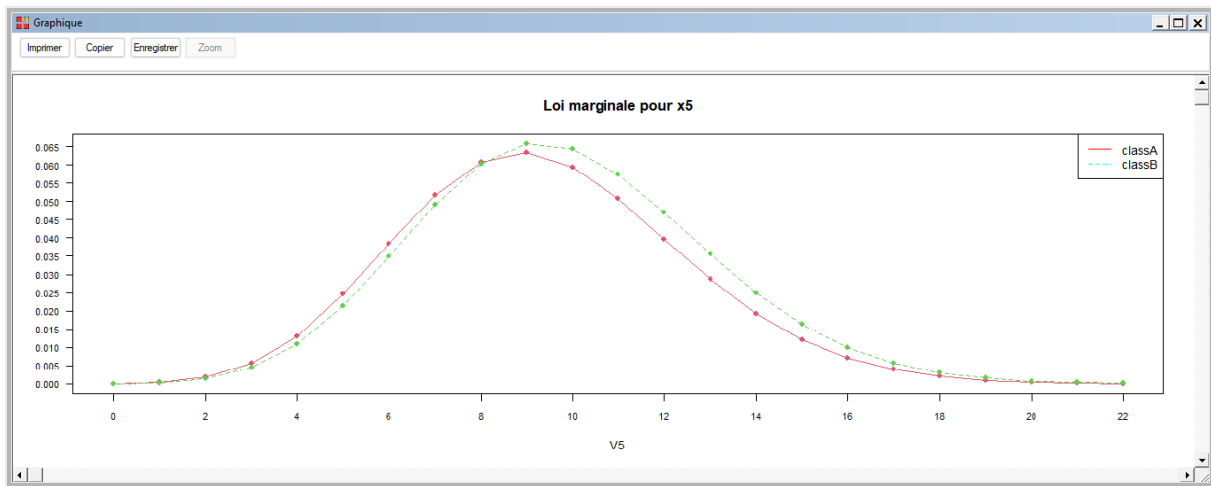
- Graphiques des lois marginales

Ce graphique affiche la loi marginale  $P(X_i / \text{classe}) \times P(\text{classe})$  pour chaque variable explicative :

- diagramme en mosaïque pour les variables Bernoulli
- diagramme en mosaïque pour les variables catégorielles
- fonction de densité pour les variables modélisées par une loi normale
- fonction de masse pour les variables modélisées par une loi de Poisson

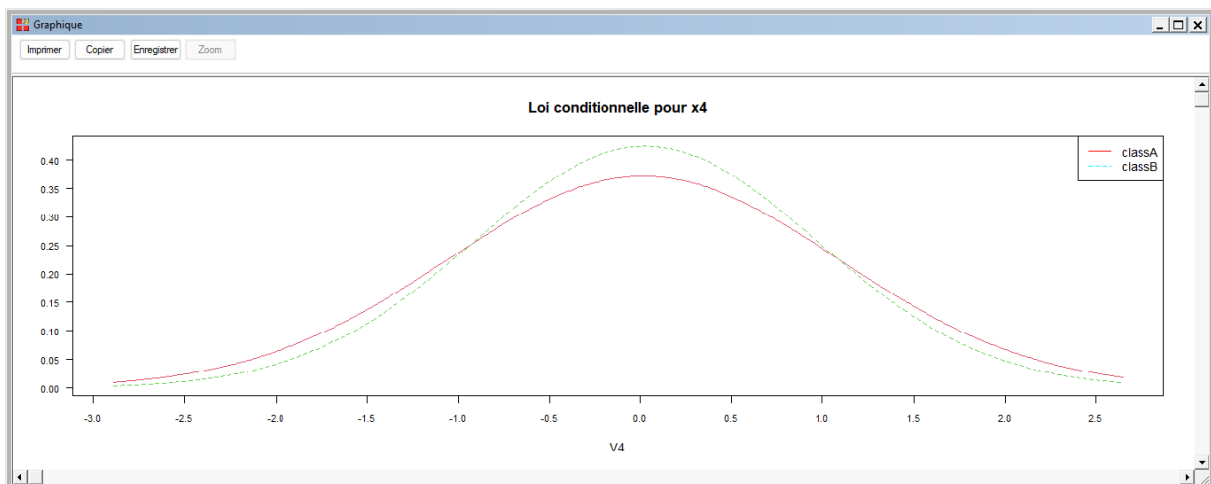




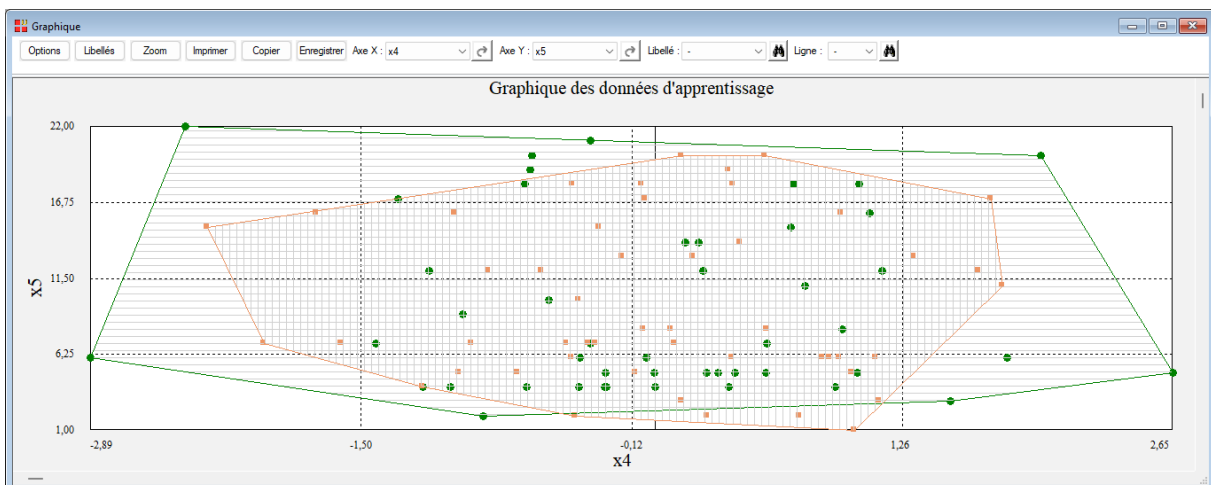


- Graphique des lois conditionnelles

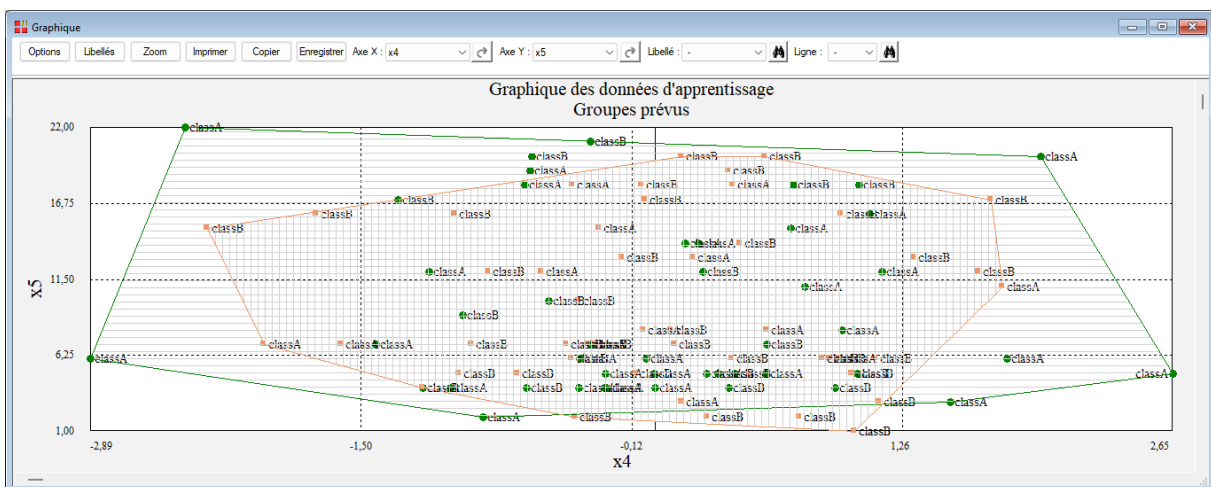
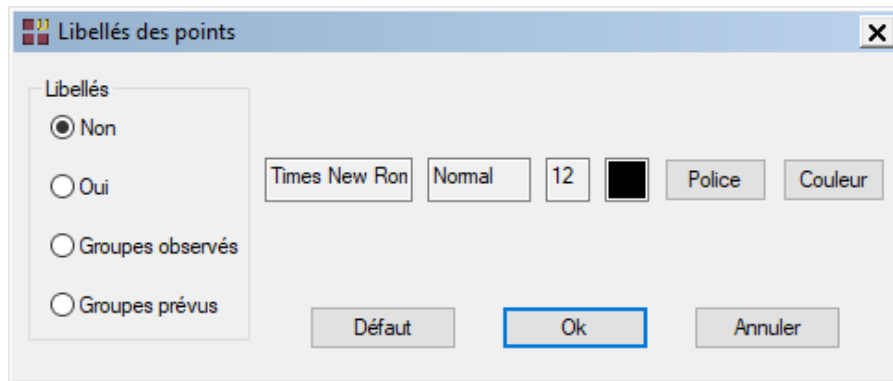
Ce graphique affiche la loi conditionnelle  $P(X_i / \text{classe})$  pour chaque variable explicative.



- Graphique des données d'apprentissage



La barre d'outils permet de préciser les variables numériques affichées sur les axes X et Y et le bouton 'Libellés' permet de préciser les libellés à afficher :

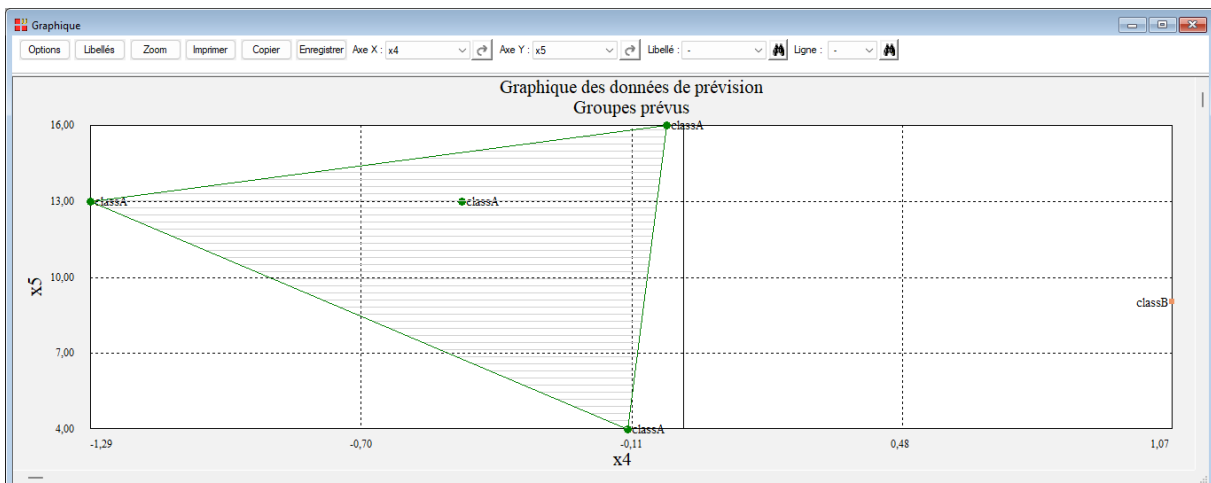


- Graphique des données de validation

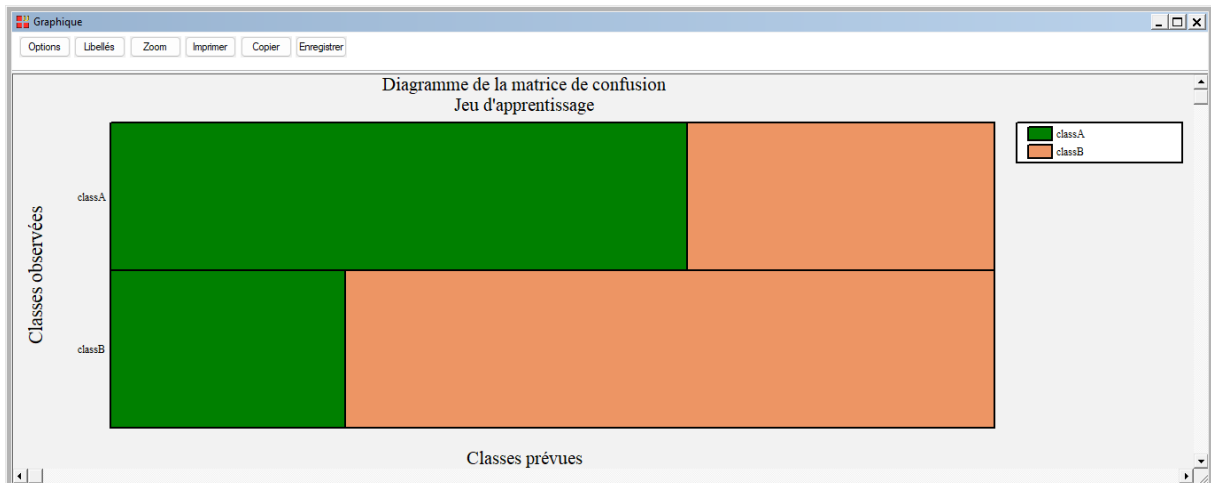
Non disponible dans cet exemple car il n'y a pas de données de validation.

- Graphique des données de prévision

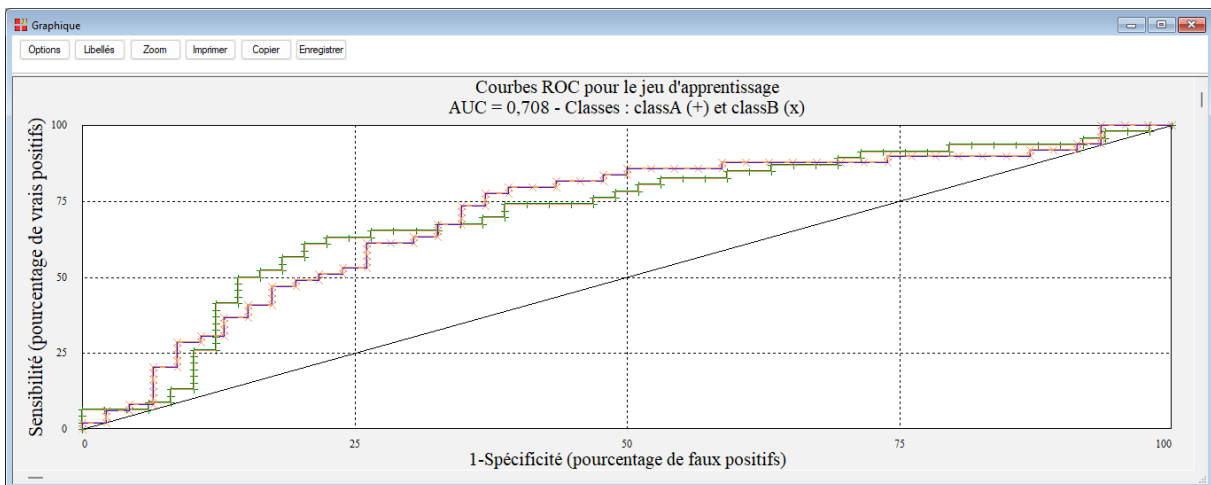
Il affiche les données pour lesquelles la classe d'appartenance n'est pas connue.



- Graphique de la matrice de confusion (apprentissage)



- Courbe ROC pour les données d'apprentissage



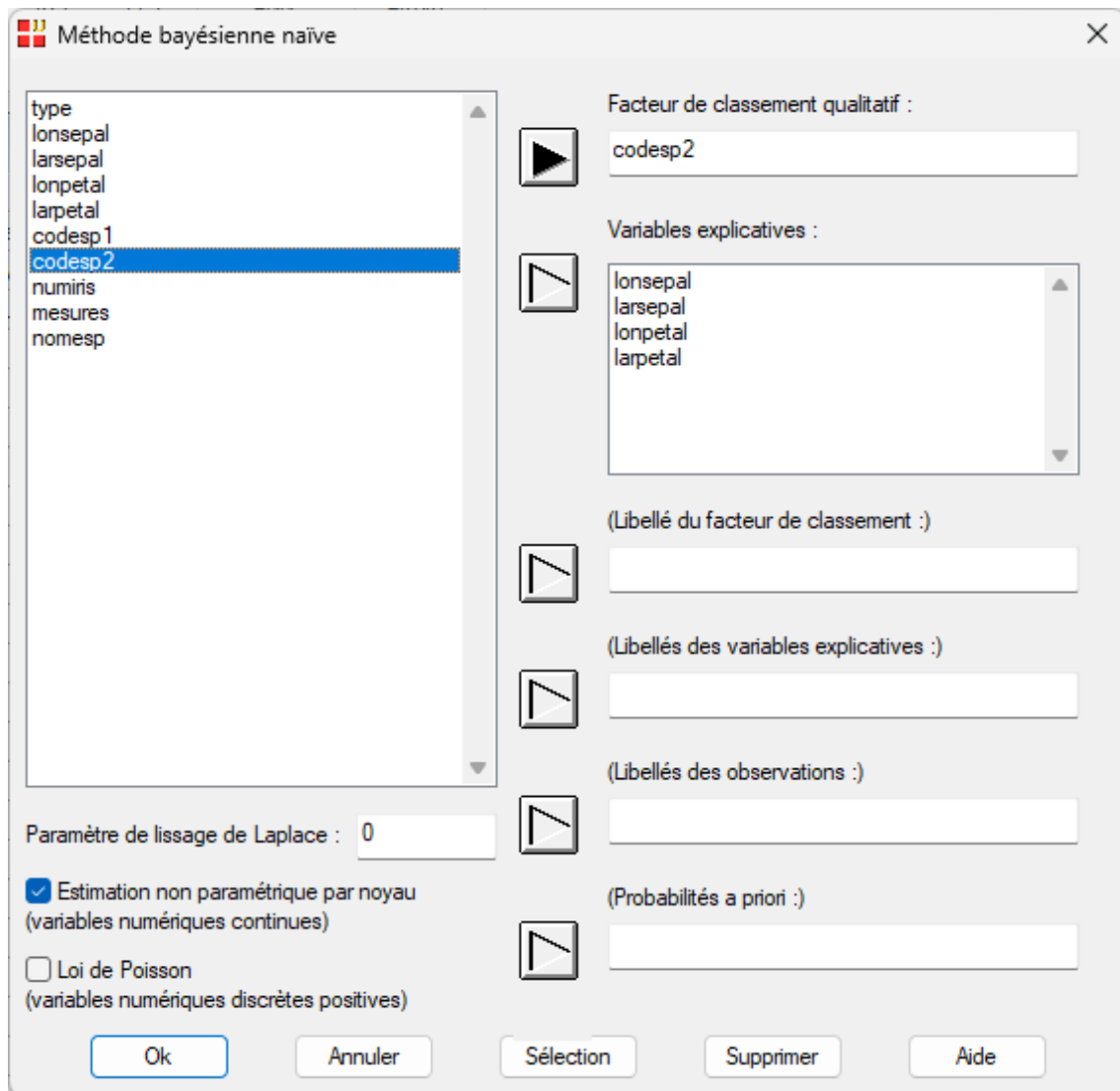
## Exemple 2 : Fichier IRIS3

Pour ce deuxième exemple, ouvrons le fichier Iris3. Ce fichier contient les données relatives à 150 iris de trois espèces : Iris Setosa, Iris Versicolor et Iris Virginica.

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

Ce fichier contient 6 iris pour lesquels les classes d'appartenance sont inconnues. Ils définissent l'échantillon de prévision.

Renseignons la boîte de dialogue comme montré ci-dessous.

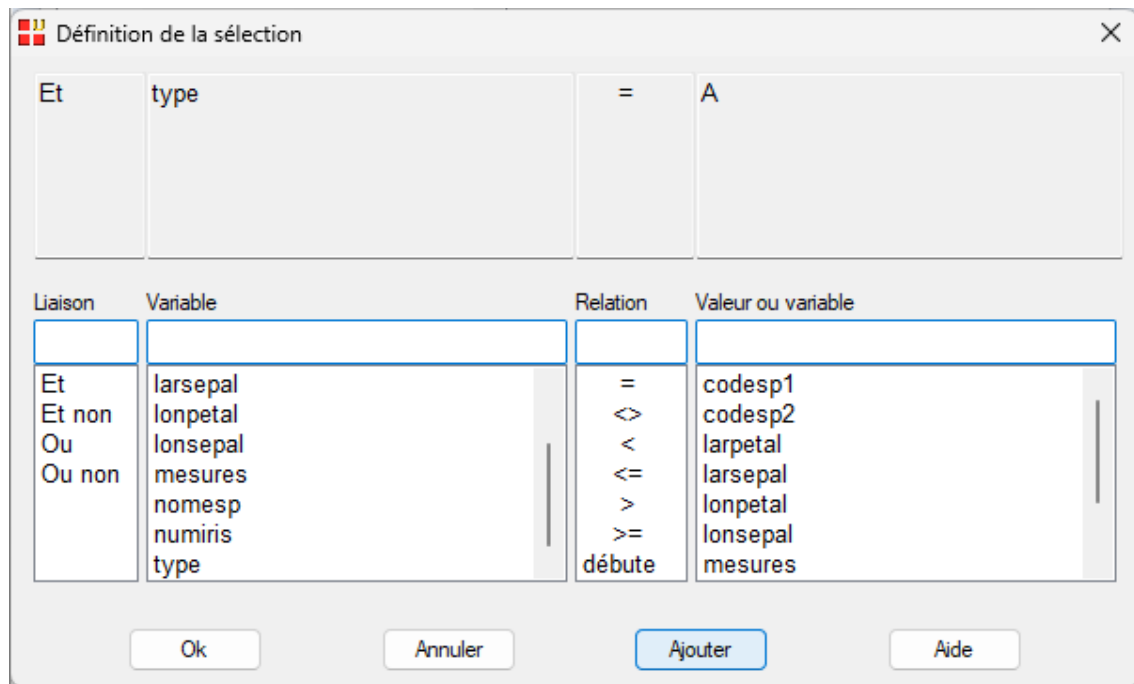


Sélectionnons la variable *codesp2* comme facteur de classement et les quatre variables *lonsepal*, *larsepal*, *lonpetal* et *lonsepal* comme variables explicatives.

Cochons la case '*Estimation non paramétrique par noyau*' pour estimer de façon non paramétrique les lois des quatre variables explicatives numériques continues.

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

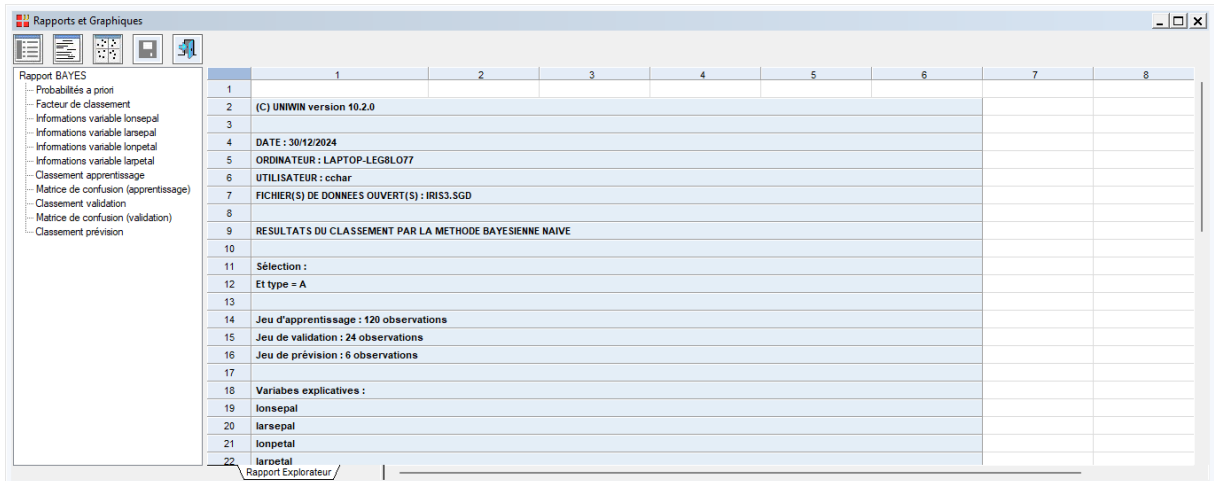
Les observations non sélectionnées pour lesquelles les valeurs du facteur de classement sont connues définissent le jeu de validation (24) et celles pour lesquelles les valeurs du facteur de classement sont inconnues le jeu de prévision (6).



La sélection comporte 120 observations.

Confirmons le nombre de modalités du facteur de classement.

Après quelques instants, les résultats de l'analyse s'affichent.



Les tableaux suivants affichent des informations concernant chacune des variables explicatives.

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable lonsepal
- Informations variable larsepal
- Informations variable lorpetal
- Informations variable larpetal
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Classement validation
- Matrice de confusion (validation)
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>PARAMETRES DE LA VARIABLE lonsepal</b>							
3								
4	<b>Estimation non paramétrique par noyau</b>							
5								
6	-----							
7	<b>Classe Setosa</b>							
8	Données : 43 obs.							
9	Largeur de bande = 0.1266							
10	-----							
11	<b>Classe Versicolor</b>							
12	Données : 38 obs.							
13	Largeur de bande = 0.219							
14	-----							
15	<b>Classe Virginica</b>							
16	Données : 39 obs.							
17	Largeur de bande = 0.2098							
18	-----							
19								
20								
21								
22								

Rapport Explorateur /

La matrice de confusion pour le jeu d'apprentissage indique un taux d'erreur de classement de 1,67%.

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable lonsepal
- Informations variable larsepal
- Informations variable lorpetal
- Informations variable larpetal
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Classement validation
- Matrice de confusion (validation)
- Classement prévision

	1	2	3	4	5	6	7	8
1								
2	<b>MATRICE DE CONFUSION DE LA POPULATION D'APPRENTISSAGE</b>							
3								
4	<b>En lignes, les classes observées</b>							
5	<b>En colonnes, les classes prévues</b>							
6								
7	Pourcentage de mal classés : 1.667 %							
8	Pourcentage de bien classés : 98.333 %							
9								
10								
11		Taille	Setosa	Versicolor	Virginica	Précision	Rappel	Score F
12	Setosa	43	43	0	0	1.00000	1.00000	1.0000
13	Versicolor	38	0	37	1	0.97368	0.97368	0.9736
14	Virginica	39	0	1	38	0.97436	0.97436	0.9743
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

Celui du tableau sur les données de validation est de 8,3%.

Le tableau du classement des données de prévision est montré ci-dessous.

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable lonsepal
- Informations variable larsepal
- Informations variable lorpetal
- Informations variable larpetal
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Classement validation
- Matrice de confusion (validation)
- Classement prévision

	1	2	3	4	5	6	7
1							
2	<b>RESULTATS DU CLASSEMENT DE LA POPULATION DE PREVISION</b>						
3							
4	<b>Individu - Classe prévue - Probabilités a posteriori</b>						
5							
6							
7		Setosa	Versicolor	Virginica			
8	Individu : o3 -> Prévu : Setosa	1.00000	0.00000	0.00000			
9	Individu : o36 -> Prévu : Setosa	1.00000	0.00000	0.00000			
10	Individu : o62 -> Prévu : Versicolor	0.00000	0.39585	0.00415			
11	Individu : o84 -> Prévu : Virginica	0.00000	0.26437	0.73563			
12	Individu : o104 -> Prévu : Virginica	0.00000	0.00019	0.99981			
13	Individu : o125 -> Prévu : Virginica	0.00000	0.00000	1.00000			
14							
15							
16							
17							
18							
19							
20							
21							
22							

Rapport Explorateur /

### Exemple 3 : Fichier TITANIC

Pour ce troisième exemple, nous utiliserons le fichier TITANIC.

Ce fichier contient des informations concernant 714 passagers :

Statut	Décès ou Survie
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 BAYES dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-dessous.

Méthode bayésienne naïve

Statut  
Age  
Tarif  
Nbfse  
Nbpe  
Classe  
Sexe  
Poids  
LibVarQuanti  
LibVarQuali  
LibObs

Facteur de classement qualitatif :  
Statut

Variables explicatives :  
Age  
Tarif  
Nbfse  
Nbpe

(Libellé du facteur de classement :)  
[ ]

(Libellés des variables explicatives :)  
[ ]

(Libellés des observations :)  
[ ]

(Probabilités a priori :)  
[ ]

Paramètre de lissage de Laplace : 0

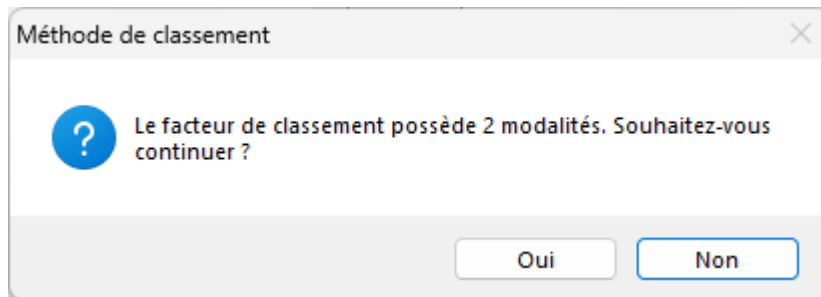
Estimation non paramétrique par noyau  
(variables numériques continues)

Loi de Poisson  
(variables numériques discrètes positives)

Ok Annuler Sélection Supprimer Aide

Un message informatif nous indique le nombre de modalités du facteur de classement.

Cliquons sur Oui.



Voici quelques résultats obtenus par cette analyse.

Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable Age
- Informations variable Tarif
- Informations variable Nbfse
- Informations variable Nbpse
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	<b>MATRICE DE CONFUSION DE LA POPULATION D'APPRENTISSAGE</b>							
3								
4	<b>En lignes, les classes observées</b>							
5	<b>En colonnes, les classes prévues</b>							
6								
7	Pourcentage de mal classés : 33.473 %							
8	Pourcentage de bien classés : 66.527 %							
9								
10								
11		Taille	Décès	Survie	Précision	Rappel	Score F1	
12	Décès	424	401	23	0.64992	0.94575	0.77041	
13	Survie	290	216	74	0.76289	0.25517	0.38243	
14								
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

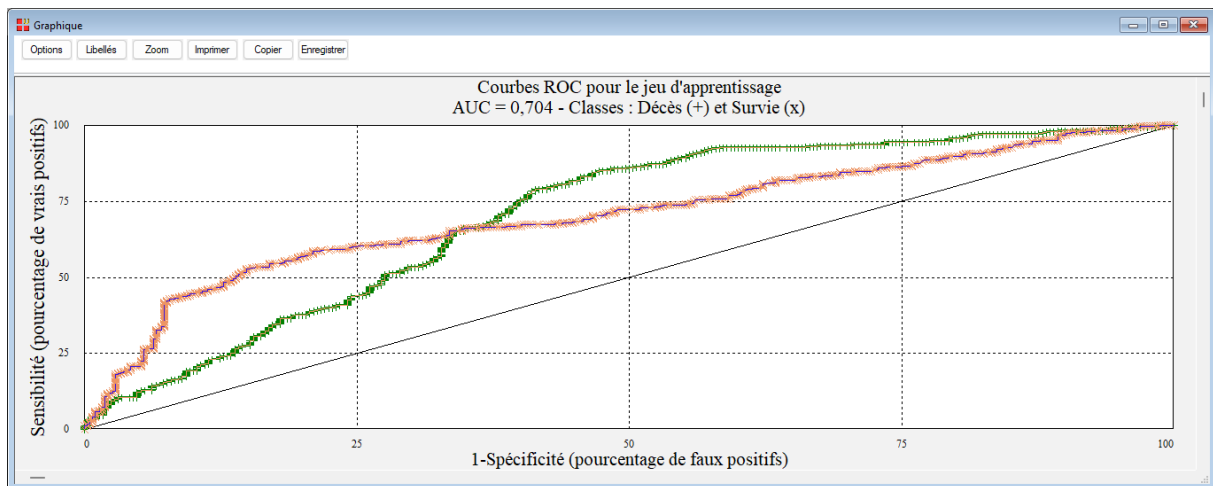
Rapports et Graphiques

Rapport BAYES

- Probabilités a priori
- Facteur de classement
- Informations variable Age
- Informations variable Tarif
- Informations variable Nbfse
- Informations variable Nbpse
- Classement apprentissage
- Matrice de confusion (apprentissage)
- Sensibilité, Spécificité (apprentissage)
- Score F1 (apprentissage)

	1	2	3	4	5	6	7	8
1								
2	<b>VP, FN, FP, VN, SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE</b>							
3								
4	<b>VP = Nombres de vrais positifs</b>							
5	<b>FN = Nombres de faux négatifs</b>							
6	<b>FP = Nombres de faux positifs</b>							
7	<b>VN = Nombres de vrais négatifs</b>							
8	Sensibilité en % = $100 \times VP / (VP + FN)$							
9	Spécificité en % = $100 \times VN / (VN + FP)$							
10								
11	<b>Événement : Survie</b>							
12	<b>Aire sous la courbe (AUC) = 0.704</b>							
13								
14								
15		Seuils	VP	FN	FP	VN	Sensibilité	Spécificit
16	1	-Infini	290	0	424	0	100.00000	0.0000
17	2	0.00004	290	0	424	0	100.00000	0.0000
18	3	0.00004	290	0	423	1	100.00000	0.2358
19	4	0.00005	290	0	422	2	100.00000	0.4717
20	5	0.00005	290	0	421	3	100.00000	0.7075
21	6	0.00006	290	0	420	4	100.00000	0.9434
22	7	0.00134	290	0	419	5	100.00000	1.1792

Rapport Explorateur /



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.

### 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  $2 * (précision * rappel) / (précision + rappel)$

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>
resappobs	libellés des observations de la population d'apprentissage
resappclas	classes prévues pour la population d'apprentissage
resappprob	probabilités d'affectation pour le jeu d'apprentissage
resvalobs	libellés des observations de la population de validation
resvalclas	classes prévues pour la population de validation
resvalprob	probabilités d'affectation pour le jeu de validation
specificiteA	Spécificité (apprentissage)
sensibiliteA	Sensibilité (apprentissage)
aireA	Aire sous la courbe (apprentissage)
specificiteV	Spécificité (validation)
sensibiliteV	Sensibilité (validation)
aireV	Aire sous la courbe (validation)
resprevobs	libellés des observations de la population de prévision
resprevclas	classes prévues pour la population de prévision
resprevprob	probabilités d'affectation pour le jeu de prévision

## Références

Documentation du package R 'naivebayes' (2024)

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