



UNIWIN VERSION 10.6.0

PROCESSUS GAUSSIENS

Révision : 07/07/2026

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Classement binaire - Fichier Infarct2	3
L'option Rapports	6
L'option Graphiques	9
Exemple 2 : Classement multi-classes - Fichier Iris3	13
Exemple 3 : Régression - Fichier Boston	21
Enregistrement des résultats	27
Références	28

Définition

La procédure Processus gaussiens implémente une méthode d'apprentissage machine pour prévoir des observations à partir de données.

Les processus gaussiens reposent sur l'hypothèse que les observations adjacentes s'informent mutuellement. Plus précisément, il est supposé que les variables observées suivent une loi normale et que leur couplage s'effectue par l'intermédiaire de la matrice de covariance d'une distribution normale.

La procédure crée des modèles de deux formes :

1. Modèles de classement qui divisent des observations en classes en se basant sur les caractéristiques observées.
2. Modèles de régression qui prévoient la valeur d'une variable cible.

Les observations sont classiquement divisées en trois jeux : un jeu d'apprentissage utilisé pour construire le modèle, un jeu de validation, pour lequel la classe ou la valeur est connue, utilisé pour valider le modèle et un jeu de prévision, pour lequel la classe ou la valeur n'est pas connue, utilisé pour faire les prévisions désirées.

La variable cible et les caractéristiques prédictives peuvent être qualitatives ou quantitatives.

Les variables quantitatives sont automatiquement centrées et réduites.

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

Entrée des données

Cliquons sur l'icône GP dans le ruban Expliquer pour afficher la boîte de dialogue montrée ci-dessous.

Processus gaussiens

Variable à expliquer :

Variables explicatives quantitatives :

Variables explicatives qualitatives :

Méthode :
☒ Classement ☐ Régression

Optimisation des paramètres du noyau
☒ Automatique ☐ Manuelle

Tolérance d'arrêt : 0.001

Nombre de validations croisées : 5

Racine aléatoire : 16807

(Libellés des variables quantitatives :)

(Libellés modalités des variables qualitatives :)

(Libellés des observations :)

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de préciser la variable à expliquer ainsi que les variables explicatives quantitatives (s'il y en a) et qualitatives (s'il y en a) à utiliser.

Elle permet également, en option, d'indiquer le nom de la variable contenant les libellés des variables quantitatives, le nom de la variable contenant les libellés des modalités des variables qualitatives ainsi que le nom de la variable contenant les libellés des observations.

Le choix de la méthode est proposé : Classement (si la variable à expliquer est qualitative) ou Régression (si la variable à expliquer est quantitative).

Les paramètres pour l'analyse sont :

- Optimisation automatique ou manuelle des paramètres du noyau RBF
- Tolérance d'arrêt
- Nombre de validations croisées
- Racine aléatoire

Données manquantes

Les observations ayant des données manquantes pour les variables explicatives sont automatiquement éliminées des calculs.

Les observations ayant des données manquantes pour la variable à expliquer constituent le jeu de prévision.

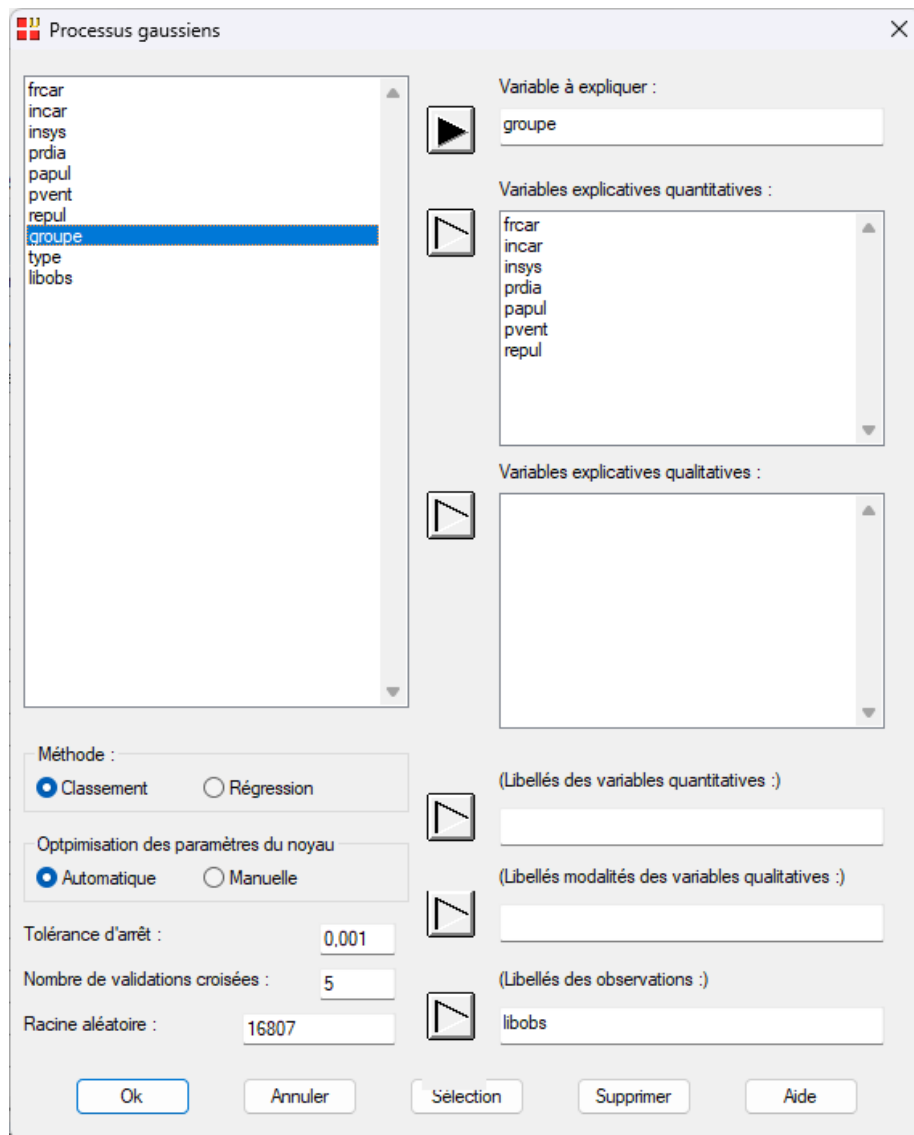
Exemple 1 : Classement binaire - Fichier Infarct2

Ce fichier contient des informations concernant 101 victimes d'un infarctus du myocarde. Les variables mesurées sont :

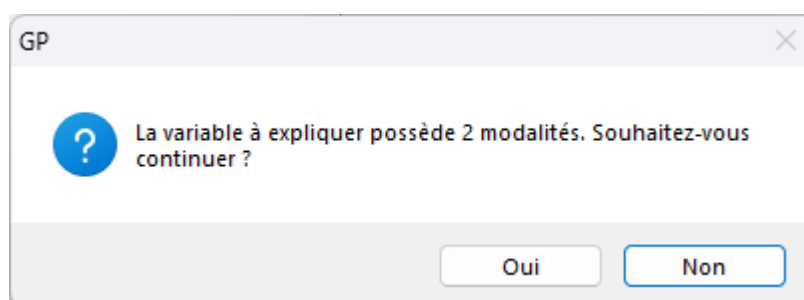
<i>frcar</i>	fréquence cardiaque	<i>incard</i>	index cardiaque
<i>insys</i>	index systolique	<i>prdia</i>	pression diastolique
<i>papul</i>	pression artérielle pulmonaire	<i>pvent</i>	pression ventriculaire
<i>repul</i>	résistance pulmonaire		

La variable *libobs* contient les libellés des observations et la variable qualitative *groupe* indique par ses deux codes les personnes décédées ou vivantes.

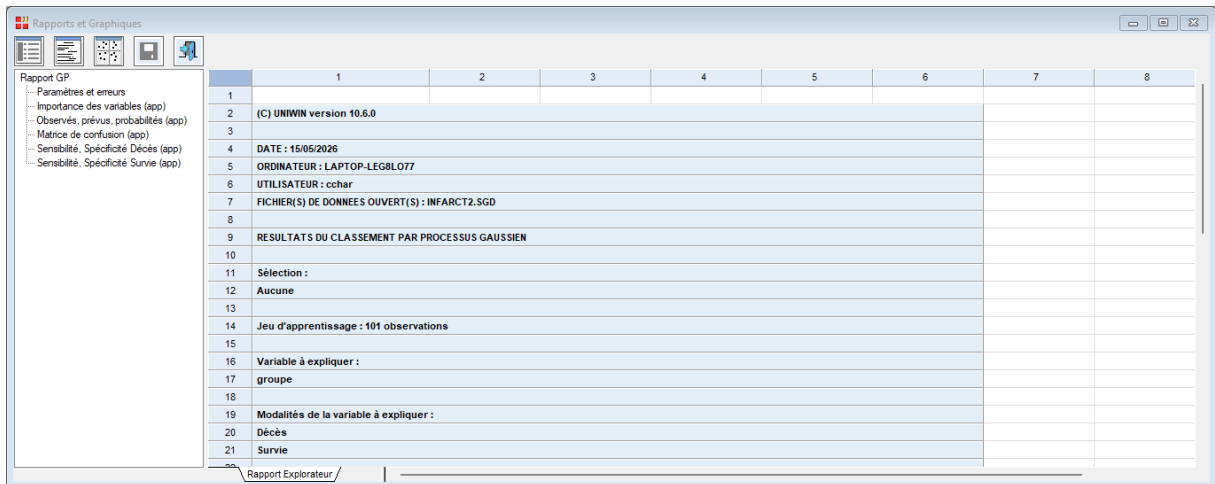
Sélectionnons la variable *groupe* comme variable à expliquer, les variables *frcar* à *repul* comme variables explicatives quantitatives, la variable *libobs* pour les libellés des observations, choisissons la méthode *Classement* et laissons les autres paramètres aux valeurs par défaut.



Cliquons sur OK et confirmons que la méthode de classement utilise une variable à 2 modalités.

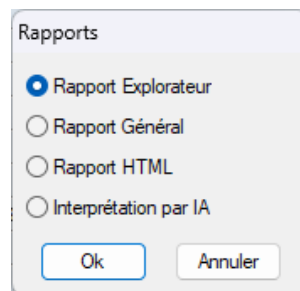


Après quelques instants, la fenêtre suivante s'affiche :

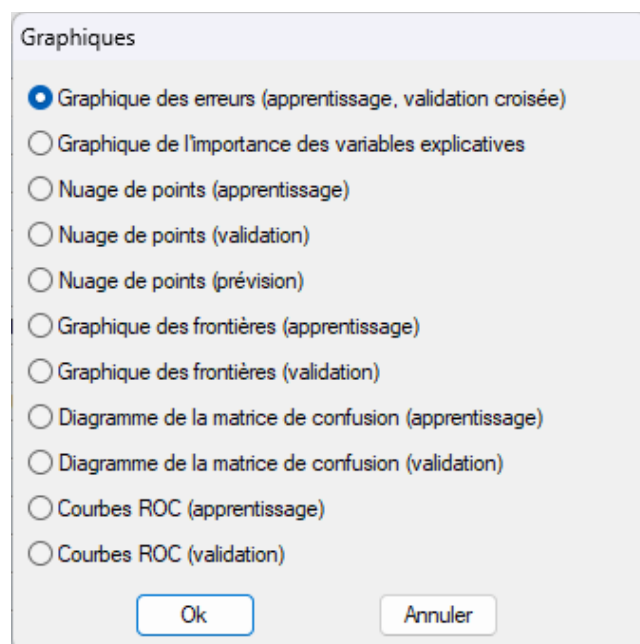


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

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.

Enregistrement des résultats (1/2)

Enregistrer

- ☐ Valeurs du paramètre sigma
- ☐ Erreurs d'apprentissage
- ☐ Erreurs de validation croisée
- ☐ Variables explicatives
- ☐ Importance des variables
- ☐ Libellés des observations (apprentissage)
- ☐ Probabilités (apprentissage)
- ☐ Classes observées (apprentissage)
- ☐ Classes prévues (apprentissage)
- ☐ Seuils (apprentissage)

Noms attribués aux variables cibles

- sigma
- erreurapp
- erreurcv
- varexpli
- impvarexpli
- libobsapp
- probapp_1
- obsapp
- vpapp
- seuilA

Ok Plus Tout Annuler

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme Explorateur, Général ou HTML et d'accéder à l'aide à l'interprétation par Intelligence Artificielle.

Le premier tableau indique que le jeu d'apprentissage est composé de 101 observations et rappelle les modalités de la variable à expliquer, les variables explicatives utilisées et les paramètres de l'analyse : standardisation des données, optimisation automatique, tolérance d'arrêt, nombre de validations croisées et racine aléatoire.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.6.0							
3								
4	DATE : 15/05/2026							
5	ORDINATEUR : LAPTOP-LEG8L077							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : INFARCT2.SGD							
8								
9	RESULTATS DU CLASSEMENT PAR PROCESSUS GAUSSIEN							
10								
11	Sélection :							
12	Aucune							
13								
14	Jeu d'apprentissage : 101 observations							
15								
16	Variable à expliquer :							
17	groupe							
18								
19	Modalités de la variable à expliquer :							
20	Décès							
21	Survie							

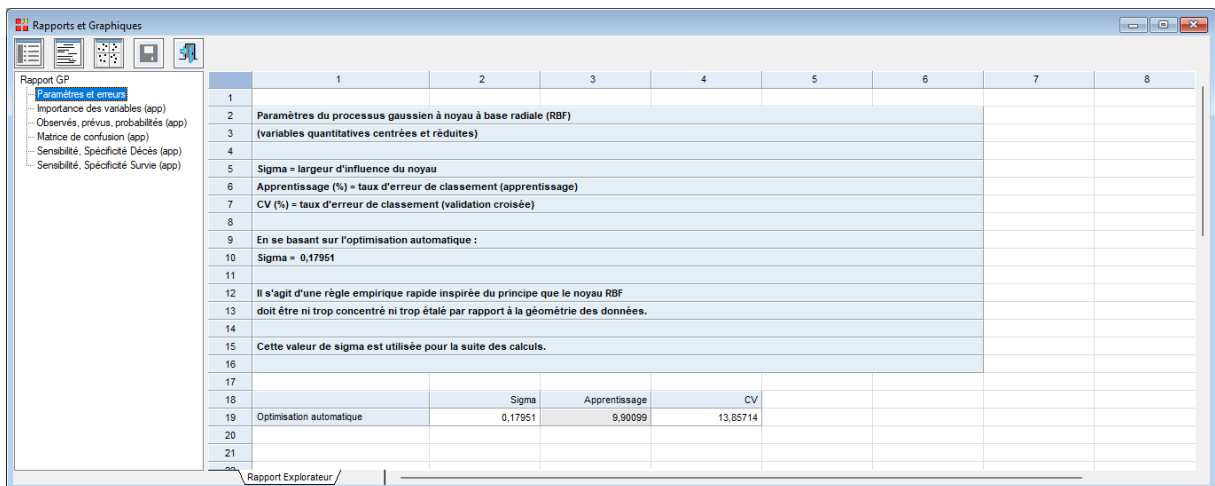
Rapport Explorateur /

Le deuxième tableau indique la valeur optimale du paramètre sigma du noyau à base radiale (RBF) ainsi que les erreurs d'apprentissage et de validation croisée.

Dans cette procédure, la valeur du noyau pour deux observations x et y est définie par :

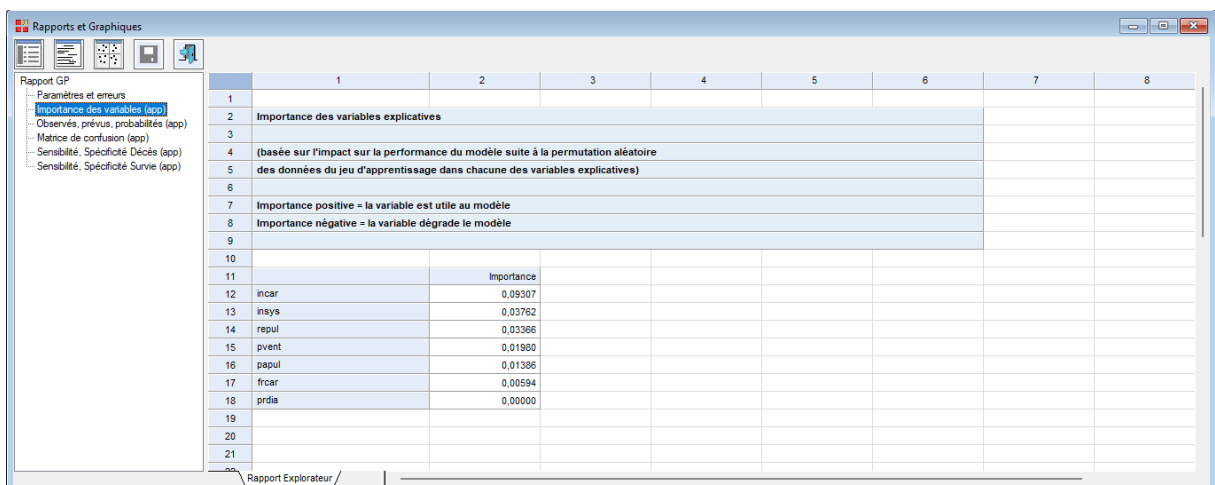
$$\text{rbf}(x,y) = \exp(-\text{sigma} * ||x-y||^2)$$

Cette valeur optimale de sigma est utilisée pour la suite des calculs.



	1	2	3	4	5	6	7	8
1								
2	Paramètres du processus gaussien à noyau à base radiale (RBF)							
3	(variables quantitatives centrées et réduites)							
4								
5	Sigma = largeur d'influence du noyau							
6	Apprentissage (%) = taux d'erreur de classement (apprentissage)							
7	CV (%) = taux d'erreur de classement (validation croisée)							
8								
9	En se basant sur l'optimisation automatique :							
10	Sigma = 0,17951							
11								
12	Il s'agit d'une règle empirique rapide inspirée du principe que le noyau RBF							
13	doit être ni trop concentré ni trop étalé par rapport à la géométrie des données.							
14								
15	Cette valeur de sigma est utilisée pour la suite des calculs.							
16								
17								
18								
19	Optimisation automatique	Sigma	Apprentissage	CV				
20		0,17951	9,90099	13,85714				
21								

Le troisième tableau affiche l'importance des variables explicatives dans le modèle.



	1	2	3	4	5	6	7	8
1								
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire							
5	des données du jeu d'apprentissage dans chacune des variables explicatives)							
6								
7	Importance positive = la variable est utile au modèle							
8	Importance négative = la variable dégrade le modèle							
9								
10								
11		Importance						
12	incar	0,09307						
13	insys	0,03762						
14	repul	0,03366						
15	pvent	0,01980						
16	papul	0,01386						
17	frcar	0,00594						
18	prdis	0,00000						
19								
20								
21								

L'importance est basée sur l'impact sur la performance du modèle à la suite de la permutation aléatoire des données du jeu d'apprentissage dans chacune des variables explicatives.

Le quatrième tableau affiche pour chaque observation du jeu d'apprentissage la classe observée, la classe prévue et les probabilités associées.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	Jeu d'apprentissage - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Décès							
7	2 = Survie							
8								
9								
10	Observation	Classe observée	Classe prévue	Décès	Survie			
11	Obs011 (*)	1	2	0,47318	0,52682			
12	Obs012	1	1	0,71812	0,28188			
13	Obs013	1	1	0,56938	0,43062			
14	Obs014	1	1	0,84474	0,15526			
15	Obs015	1	1	0,85637	0,14363			
16	Obs016	1	1	0,89499	0,10501			
17	Obs017	1	1	0,76624	0,23376			
18	Obs018	1	1	0,63049	0,36951			
19	Obs019	1	1	0,87336	0,12664			
20	Obs020	1	1	0,85771	0,14229			
21	Obs021	1	1	0,70445	0,29555			

Rapport Explorateur

Le tableau suivant affiche la matrice de confusion pour le jeu d'apprentissage.

Environ 90% des observations sont bien classées par le modèle construit.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu d'apprentissage							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 9,901 %							
10	Pourcentage de bien classés (exactitude) : 90,099 %							
11								
12	Précision = VP / (VP + FP)							
13	Rappel = VP / (VP + FN)							
14	Score F1 = 2 x (Précision x Rappel) / (Précision + Rappel)							
15								
17	Observé \ Prévu	Taille	Décès	Survie	Précision	Rappel	Score F1	
18	Décès	51	45	6	0,91837	0,88235	0,90000	
19	Survie	50	4	46	0,88462	0,92000	0,90196	
20								
21								

Rapport Explorateur

Les deux tableaux suivants affichent les sensibilités et spécificités ainsi que l'aire sous la courbe ROC pour chacune des classes. Ces tableaux ne sont disponibles que pour un classement binaire (deux modalités pour la variable à expliquer).

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Classe : Décès							
5								
6	Sensibilité en %							
7	Spécificité en %							
8								
9	Aire sous la courbe (AUC) = 0,976							
10								
12		Seuil	Sensibilité	Spécificité				
13	1	-Infini	100,00000	0				
14	2	0,07766	100,00000	0				
15	3	0,07808	100,00000	2				
16	4	0,08464	100,00000	4				
17	5	0,08724	100,00000	6				
18	6	0,09108	100,00000	8				
19	7	0,09543	100,00000	10				
20	8	0,11592	100,00000	12				
21	9	0,11835	100,00000	14				

Rapport Explorateur

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Sensibilité, Spécificité Décès (app)
- Sensibilité, Spécificité Survie (app)

	1	2	3	4	5	6	7	8
1								
2	SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Classe : Survie							
5								
6	Sensibilité en %							
7	Spécificité en %							
8								
9	Aire sous la courbe (AUC) = 0,976							
10								
11								
12		Seuil	Sensibilité	Spécificité				
13	1	-infini	0	100,00000				
14	2	0,07766	0	100,00000				
15	3	0,07808	2	100,00000				
16	4	0,08464	4	100,00000				
17	5	0,08724	6	100,00000				
18	6	0,09108	8	100,00000				
19	7	0,09543	10	100,00000				
20	8	0,11592	12	100,00000				
21	9	0,11835	14	100,00000				

Rapport Explorateur

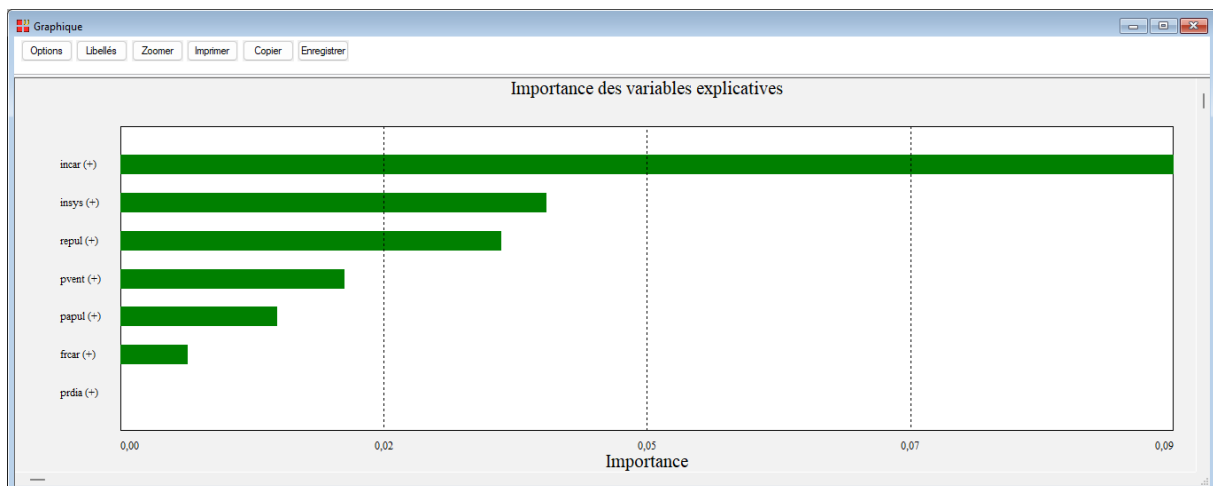
L'option Graphiques

- Graphique des erreurs (apprentissage et validation croisée)

Ce graphique n'est disponible que s'il y a plusieurs valeurs de sigma testées.

Ce n'est pas le cas dans ce premier exemple.

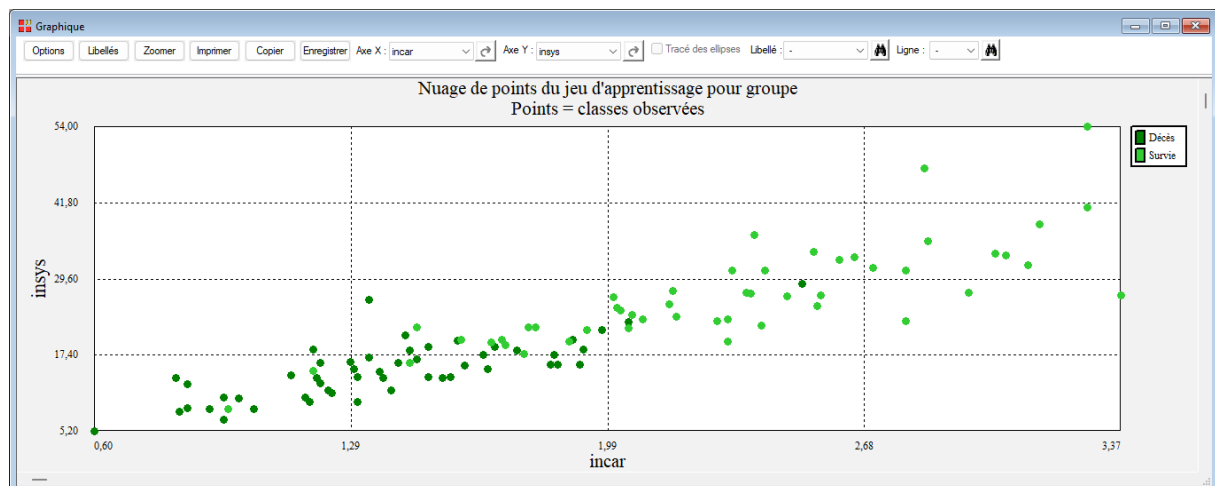
- Graphique de l'importance des variables explicatives



Les deux variables les plus importantes sont *incar* (index cardiaque) et *insys* (index systolique).

- Graphique Nuage de points (apprentissage)

Cette option permet d'afficher un nuage des observations pour deux variables sélectionnées. Il est possible d'afficher les points avec ou sans libellés, avec les codes des classes observées ou avec les codes des classes prévues. Dans le graphique, la barre d'outils permet de modifier les variables affichées.



- Graphique Nuage de points (validation)

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

- Graphique Nuage de points (prévision)

Non disponible dans cet exemple car il n'y a pas de jeu de prévision.

- Graphique des frontières (apprentissage)

Cette option permet d'afficher un nuage des observations pour deux variables explicatives quantitatives sélectionnées et de préciser les niveaux auxquels les autres variables sont maintenues. Par défaut les deux variables automatiquement sélectionnées sont les deux variables explicatives quantitatives les plus importantes.

Pour préciser les niveaux des autres variables, il faut cliquer sur le bouton 'Niveaux des autres variables' dans la boîte de dialogue 'Choix des variables'.

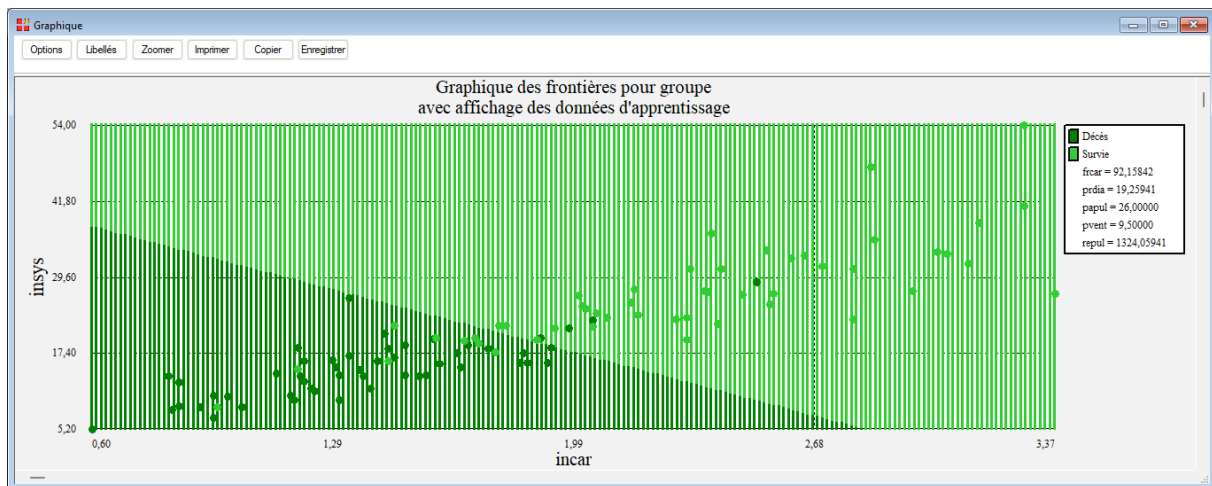
Par défaut les autres variables quantitatives sont maintenues aux valeurs moyennes des données du jeu d'apprentissage.

L'option 'Résolution graphique' permet de préciser le nombre d'évaluations de chacune des variables.

	Variable	Niveau
1	frcar	92,15842
2	prdia	19,25941
3	papul	26,00000
4	pvent	9,50000
5	repul	1324,05941

Niveaux

Ok

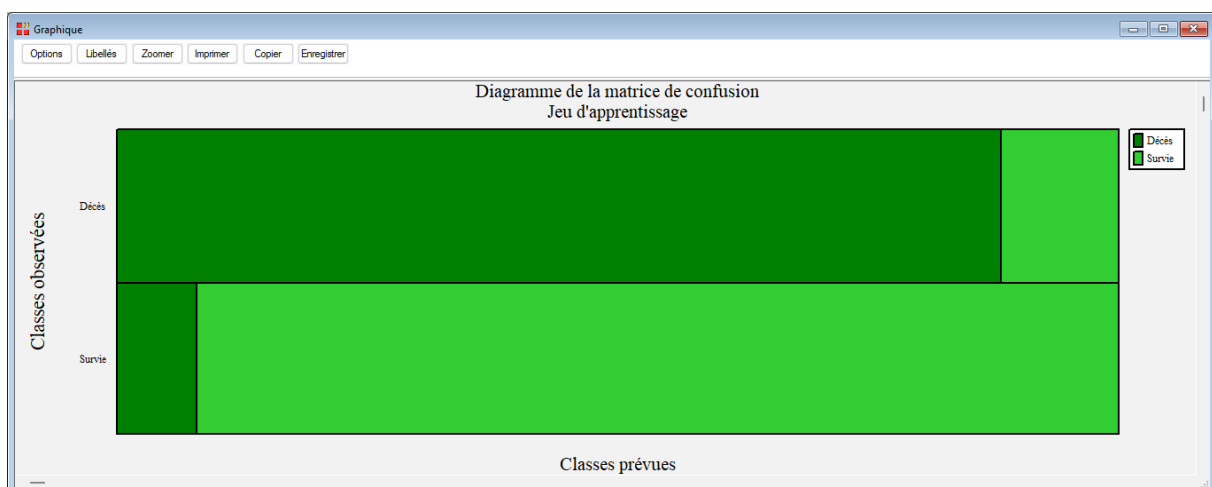


- Graphique des frontières (validation)

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

- Diagramme de la matrice de confusion (apprentissage)

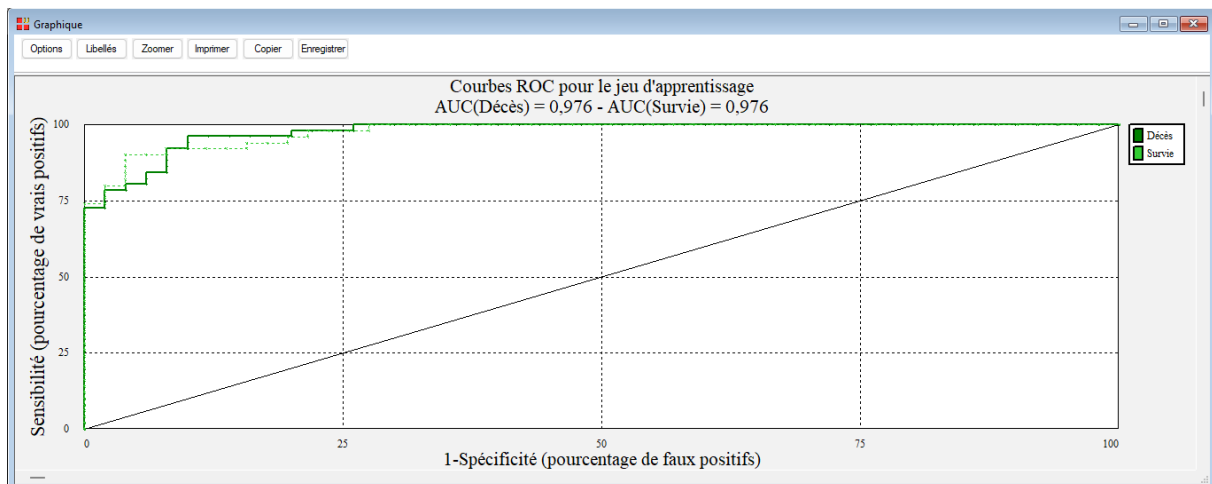
Ce diagramme permet de visualiser les résultats obtenus dans le tableau 'Matrice de confusion'.



- Diagramme de la matrice de confusion (validation)

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

- Graphique de la courbe ROC (apprentissage)



L'aire sous la courbe (AUC) nous indique l'efficacité du modèle construit.

Plus la valeur de cette aire est élevée, meilleures sont les performances du modèle pour faire la distinction entre les classes *Survie* et *Décès*.

- Graphique de la courbe ROC (validation)

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

Exemple 2 : Classement multi-classes - 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, largeur du pétale.

Les trois espèces sont : Setosa, Versicolor et Virginica.

Ce fichier contient 6 iris pour lesquels les classes d'appartenance sont inconnues. Ils définiront le jeu de prévision.

Sélectionnons la variable *codesp2* comme variable à expliquer, les variables *lonsepal* à *larpetal* comme variables explicatives, la variable *mesures* pour les libellés des variables explicatives et la variable *numiris* pour les libellés des observations.

Laissons les autres paramètres aux valeurs par défaut et cliquons sur le bouton *Sélection* pour définir le jeu d'apprentissage.

Processus gaussiens

type
lonsepal
larsepal
lonpetal
larpetal
codesp1
codesp2
numiris
mesures
nomesp

Variable à expliquer :
codesp2

Variables explicatives quantitatives :
lonsepal
larsepal
lonpetal
larpetal

Variables explicatives qualitatives :

Méthode :
☒ Classement ☐ Régression

Optimisation des paramètres du noyau
☒ Automatique ☐ Manuelle

Tolérance d'arrêt : 0,001

Nombre de validations croisées : 5

Racine aléatoire : 16807

(Libellés des variables quantitatives :)
mesures

(Libellés modalités des variables qualitatives :)

(Libellés des observations :)
numiris

Ok Annuler Sélection Supprimer Aide

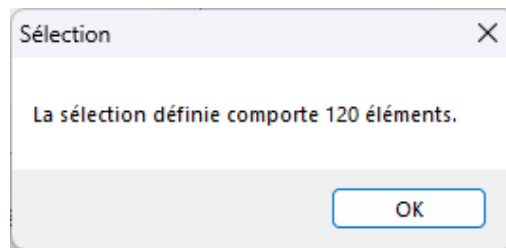
Définition de la sélection

Et type = A

Liaison	Variable	Relation	Valeur ou variable
Et	larpetal	=	codesp1
Et non	larsepal	<>	codesp2
Ou	lonpetal	<	larpetal
Ou non	lonsepal	<=	larsepal
	mesures	>	lonpetal
	nomesp	>=	lonsepal
	numiris	début	mesures
	type		nomesp

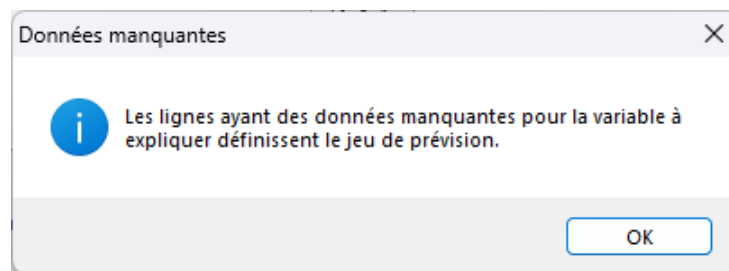
☒ Trier alphabétiquement les noms des variables

Ok Annuler Ajouter Aide

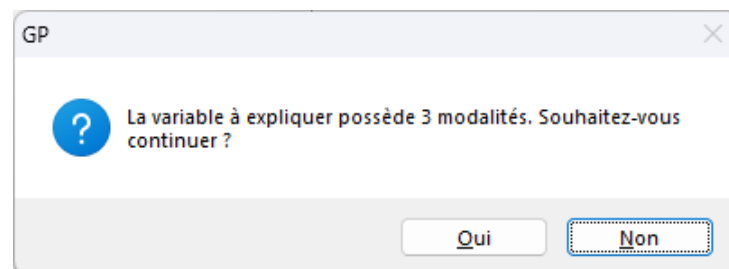


120 iris définissent le jeu d'apprentissage.

Cliquons sur le bouton OK.

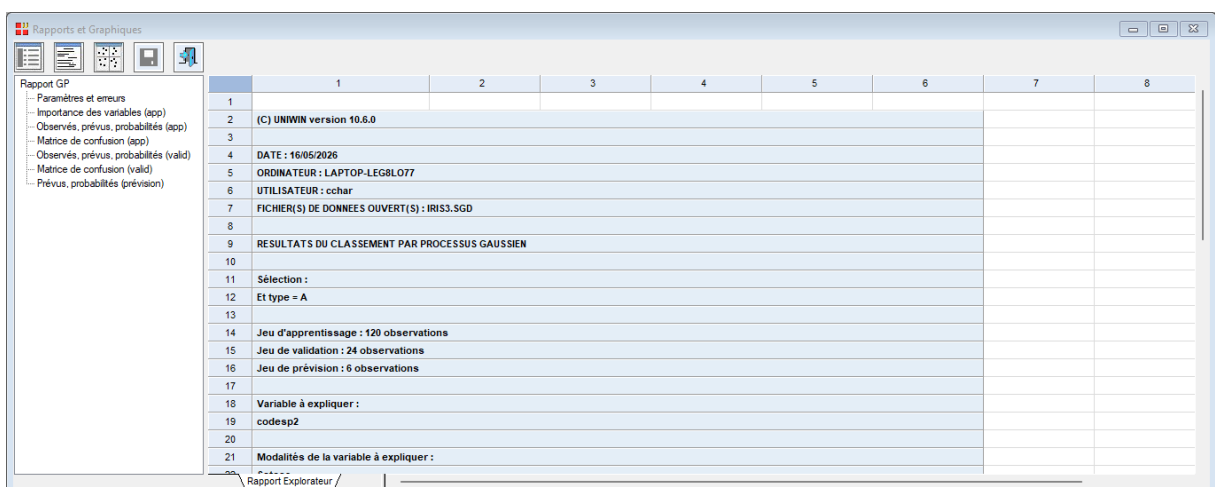


Un message nous informe que les lignes ayant des données manquantes pour la variable à expliquer définissent le jeu de prévision. Cliquons sur le bouton OK.



Confirmons que la méthode de classement utilise une variable à 3 modalités.

Après quelques instants, la fenêtre suivante s'affiche :



Le premier tableau, affiché ci-dessus, indique que le jeu d'apprentissage est composé de 120 observations, le jeu de validation de 24 observations et le jeu de prévision de 6 observations.

Il indique les paramètres de l'analyse : standardisation des données, optimisation automatique, tolérance d'arrêt, nombre de validations croisées et racine aléatoire.

Le deuxième tableau indique la valeur optimale du paramètre sigma du noyau à base radiale (RBF) ainsi que les erreurs d'apprentissage et de validation croisée.

	1	2	3	4	5	6	7	8
1								
2		Paramètres du processus gaussien à noyau à base radiale (RBF)						
3		(variables quantitatives centrées et réduites)						
4								
5		Sigma = largeur d'influence du noyau						
6		Apprentissage (%) = taux d'erreur de classement (apprentissage)						
7		CV (%) = taux d'erreur de classement (validation croisée)						
8								
9		En se basant sur l'optimisation automatique :						
10		Sigma = 1,04538						
11								
12		Il s'agit d'une règle empirique rapide inspirée du principe que le noyau RBF						
13		doit être ni trop concentré ni trop étalé par rapport à la géométrie des données.						
14								
15		Cette valeur de sigma est utilisée pour la suite des calculs.						
16								
17								
18			Sigma	Apprentissage	CV			
19		Optimisation automatique	1,04538	0	2,5			
20								
21								

Ces valeurs ont été obtenues par optimisation automatique.

Revenons à la boîte de dialogue d'entrée des données en cliquons sur l'icône , demandons une optimisation manuelle et saisissons des valeurs pour la largeur d'influence dans le voisinage de la valeur optimale obtenue précédemment.

Optimisation des paramètres du noyau RBF

Entrer des valeurs voisines de celle obtenue par l'optimisation automatique pour la largeur d'influence du noyau basée sur la géométrie des données.

Largeur d'influence : 0,8 0,9 1 1,1 1,2

Variance du bruit : 0,001

Ok Défaut Annuler

Cliquons sur Ok pour exécuter à nouveau l'analyse et affichons le tableau 'Paramètres et erreurs'.

Il indique que la valeur optimale obtenue est de 1,2.

Cette valeur sera utilisée pour la suite des calculs.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Paramètres du processus gaussien à noyau à base radiale (RBF)							
3	(variables quantitatives centrées et réduites)							
4								
5	Sigma = largeur d'influence du noyau							
6	Apprentissage (%) = taux d'erreur de classement (apprentissage)							
7	CV (%) = taux d'erreur de classement (validation croisée)							
8								
9	En se basant sur l'optimisations manuelle (minimisation CV) :							
10	Sigma = 1,20000							
11	Cette valeur de sigma est utilisée pour la suite des calculs.							
12								
13								
14								
15								
16			Sigma	Apprentissage	CV			
17	Optimisation automatique	1,04538	0,00000	2,50000				
18	Optimisation manuelle	1,20000	0,00000	2,50000				
19	Optimisation manuelle	0,80000	0,83333	3,33333				
20	Optimisation manuelle	0,90000	0,00000	3,33333				
21	Optimisation manuelle	1,00000	0,00000	3,33333				
22	Optimisation manuelle	1,10000	0,00000	3,33333				

Rapport Explorateur /

Le troisième tableau affiche l'importance des variables explicatives dans le modèle.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire							
5	des données du jeu d'apprentissage dans chacune des variables explicatives)							
6								
7	Importance positive = la variable est utile au modèle							
8	Importance négative = la variable dégrade le modèle							
9								
10								
11			Importance					
12	larpetal	0,31167						
13	lonpetal	0,24833						
14	larsepal	0,07000						
15	lonsepal	0,03833						
16								
17								
18								
19								
20								
21								

Rapport Explorateur /

Le quatrième tableau affiche pour chaque observation du jeu d'apprentissage les valeurs observées et prévues de la variable à expliquer ainsi que les probabilités d'affectation à chacune des classes.

Rapports et Graphiques

Rapport GP

- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Jeu d'apprentissage - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe observée	Classe prévue	Setosa	Versicolor	Virginica		
12	1	1	1	0,84891	0,07554	0,07555		
13	2	1	1	0,76085	0,12028	0,11887		
14	4	1	1	0,79282	0,10374	0,10344		
15	5	1	1	0,83429	0,08284	0,08287		
16	6	1	1	0,74515	0,12744	0,12741		
17	7	1	1	0,79433	0,10281	0,10286		
18	8	1	1	0,85289	0,07356	0,07355		
19	9	1	1	0,67695	0,16213	0,16092		
20	11	1	1	0,80050	0,09875	0,09874		
21	12	1	1	0,83730	0,08134	0,08136		

Rapport Explorateur /

Le tableau suivant affiche la matrice de confusion pour le jeu d'apprentissage.

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs								
- Importance des variables (app)								
- Observés, prévus, probabilités (app)								
- Matrice de confusion (app)								
- Observés, prévus, probabilités (valid)								
- Matrice de confusion (valid)								
- Prévus, probabilités (prévision)								
1	1	2	3	4	5	6	7	8
2	Matrice de confusion pour le jeu d'apprentissage							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 0,000 %							
10	Pourcentage de bien classés (exactitude) : 100,000 %							
11								
12	Précision = $VP / (VP + FP)$							
13	Rappel = $VP / (VP + FN)$							
14	Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
15								
16								
17	Observé \ Prévu	Taille	Setosa	Versicolor	Virginica	Précision	Rappel	Score F
18	Setosa	43	43	0	0	1	1	
19	Versicolor	38	0	38	0	1	1	
20	Virginica	39	0	0	39	1	1	
21								

Les deux tableaux suivants affichent ces mêmes informations pour le jeu de validation.

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs								
- Importance des variables (app)								
- Observés, prévus, probabilités (app)								
- Matrice de confusion (app)								
- Observés, prévus, probabilités (valid)								
- Matrice de confusion (valid)								
- Prévus, probabilités (prévision)								
1	1	2	3	4	5	6	7	8
2	Jeu de validation - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe observée	Classe prévue	Setosa	Versicolor	Virginica		
12	10	1	1	0,79654	0,10192	0,10153		
13	25	1	1	0,81794	0,09107	0,09100		
14	41	1	1	0,84740	0,07629	0,07631		
15	48	1	1	0,80796	0,09602	0,09602		
16	49	1	1	0,81460	0,09270	0,09270		
17	54	2	2	0,14695	0,66540	0,18764		
18	55	2	2	0,09140	0,56851	0,34009		
19	57	2	2	0,13645	0,58324	0,28031		
20	68	2	2	0,10372	0,77611	0,12917		
21	78 (*)	2	3	0,09054	0,37659	0,53287		

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs								
- Importance des variables (app)								
- Observés, prévus, probabilités (app)								
- Matrice de confusion (app)								
- Observés, prévus, probabilités (valid)								
- Matrice de confusion (valid)								
- Prévus, probabilités (prévision)								
1	1	2	3	4	5	6	7	8
2	Matrice de confusion pour le jeu de validation							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 8,333 %							
10	Pourcentage de bien classés (exactitude) : 91,667 %							
11								
12	Précision = $VP / (VP + FP)$							
13	Rappel = $VP / (VP + FN)$							
14	Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
15								
16								
17	Observé \ Prévu	Taille	Setosa	Versicolor	Virginica	Précision	Rappel	Score F
18	Setosa	5	5	0	0	1,00000	1,00000	1,00000
19	Versicolor	10	0	9	1	0,90000	0,90000	0,90000
20	Virginica	9	0	1	8	0,88889	0,88889	0,88888
21								

Le dernier tableau affiche les résultats pour le jeu de prévision.

Rapports et Graphiques

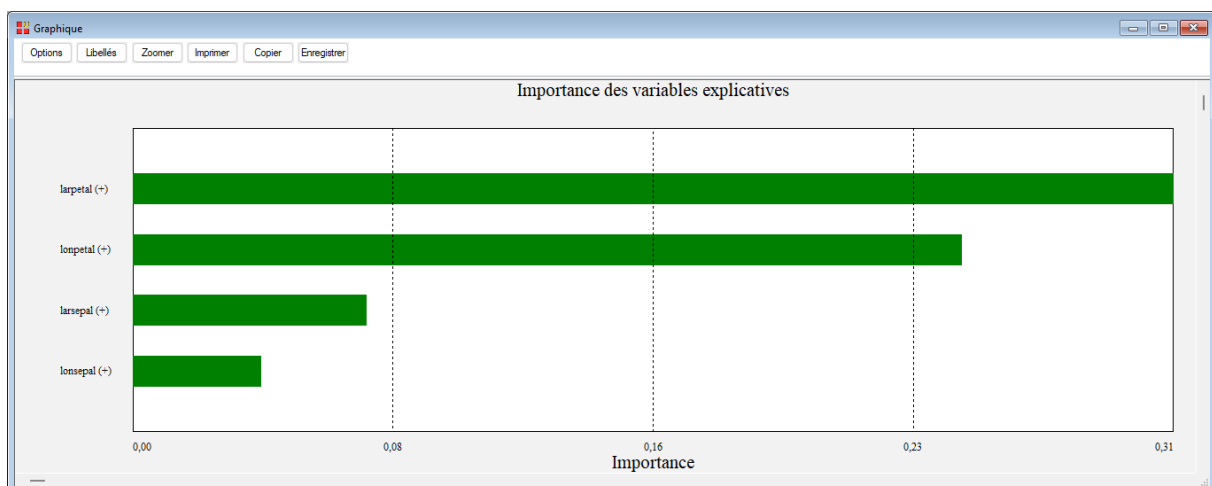
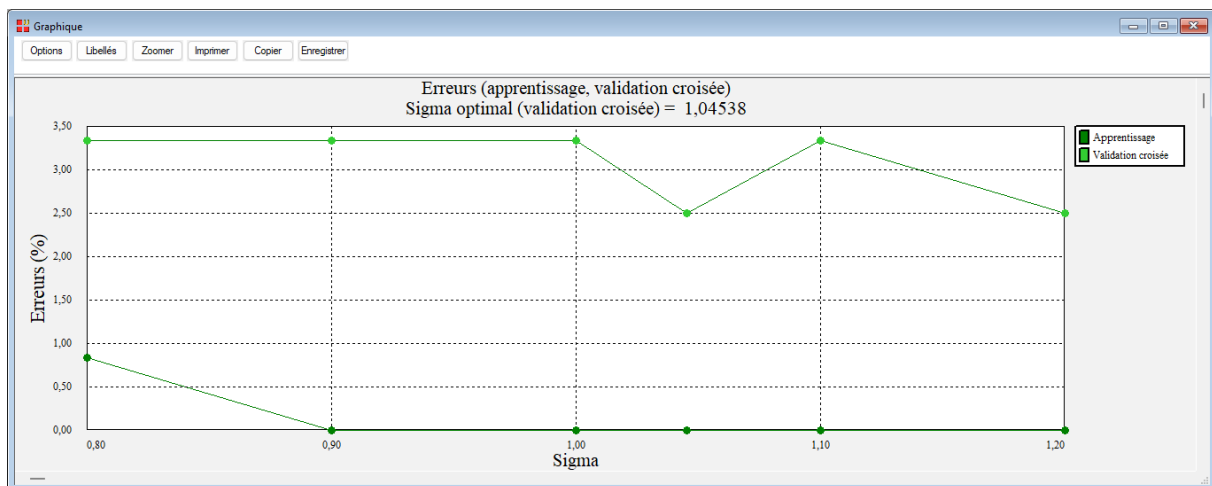
Rapport GP

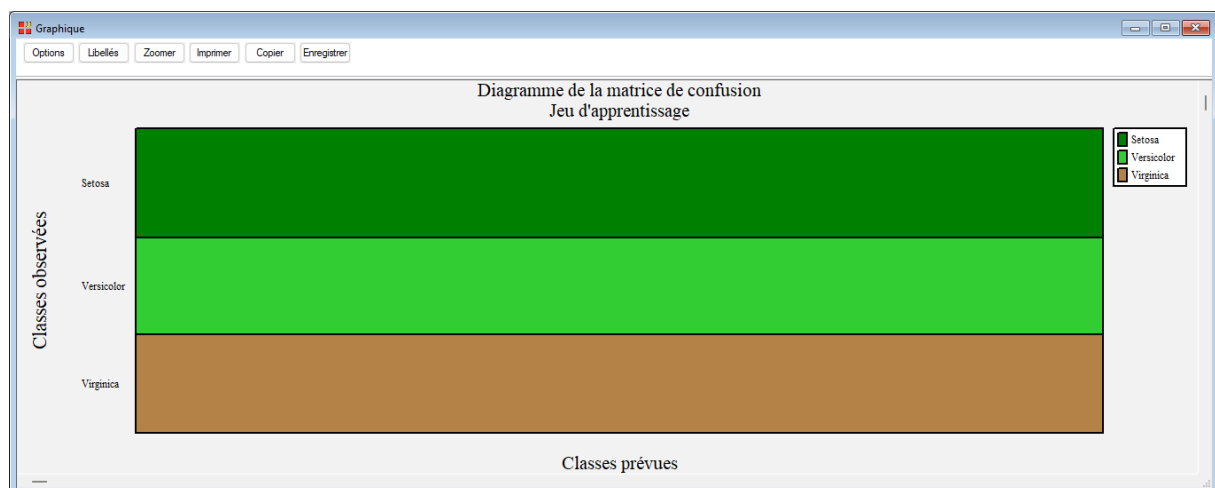
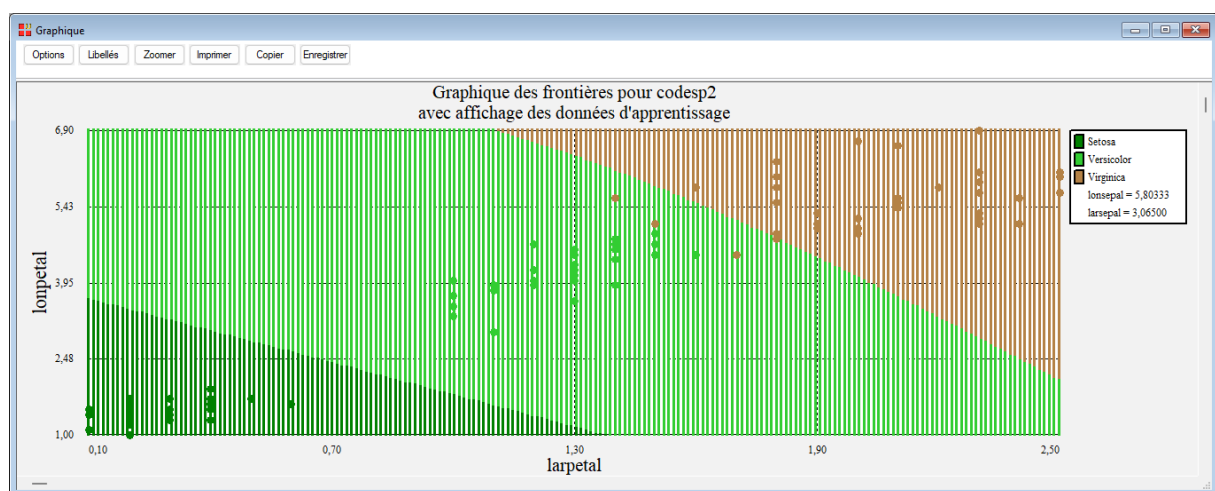
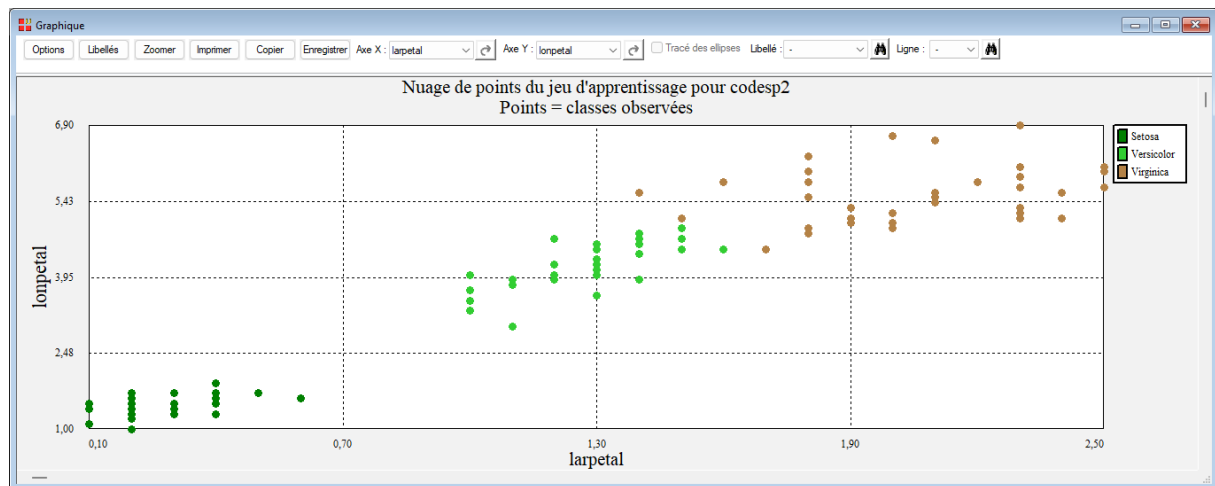
- Paramètres et erreurs
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Jeu de prévision - classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe prévue	Setosa	Versicolor	Virginica			
12	3	1	0,82207	0,08896	0,08896			
13	36	1	0,81948	0,09031	0,09020			
14	62	2	0,08655	0,73069	0,18277			
15	84	3	0,08926	0,34642	0,56432			
16	104	3	0,07861	0,16391	0,75748			
17	125	3	0,11670	0,15862	0,72467			
18								
19								
20								
21								

Rapport Explorateur

Voici quelques graphiques issus de cette analyse.





Exemple 3 : Régression - Fichier Boston

Ce fichier contient des informations collectées par le « U.S Census Service » concernant la valeur des habitations dans l'agglomération de Boston (source : <http://lib.stat.cmu.edu/datasets/boston>).

Les caractéristiques contenues dans le fichier sont les suivantes :

crim	taux de criminalité par habitant par ville
zn	proportion de terrains résidentiels zonés pour des lots de plus de 25 000 pieds carrés.
indus	proportion d'acres non commerciales par ville
chas	indicatrice de proximité à la Charles River (1 = proche, 0 = sinon)
nox	concentration d'oxydes nitriques (partie pour 10 millions)
rm	nombre moyen de pièces par habitation
age	proportion des habitations construites avant 1940
dis	distance pondérée à cinq lieux d'emplois de Boston
rad	indice d'accessibilité aux autoroutes
tax	taux d'imposition foncière par \$ 10.000
ptratio	ratio élèves / enseignants par ville
black	$1000(Bk-0,63)^2$ où Bk = proportion de noirs par ville
lstat	% statut inférieur de la population
medv	valeur médiane de l'habitation (en milliers de \$)
type	indique si l'observation est utilisée pour l'apprentissage (A) ou la validation (V)

Le but est de prévoir la variable *medv* en utilisant ces données.

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

La variable à expliquer est *medv* et les variables explicatives sont *crim*, *zn*, *indus*, *chas*, *nox*, *rm*, *age*, *dis*, *rad*, *tax*, *ptratio*, *black*, *lstat*.

Cliquons sur le bouton 'Sélection' pour sélectionner les données du jeu d'apprentissage (type=A).

391 observations sont sélectionnées, 105 observations définissent le jeu de validation et 10 le jeu de prévision.

Cliquons sur Ok.

Processus gaussiens

obs
crim
zn
indus
chas
nox
rm
age
dis
rad
tax
ptratio
black
lstat
medv
type

Variable à expliquer :
medv

Variables explicatives quantitatives :
crim
zn
indus
chas
nox
rm
age
dis
rad
tax
ptratio

Variables explicatives qualitatives :

Méthode :
☐ Classement ☒ Régression

Optimisation des paramètres du noyau
☒ Automatique ☐ Manuelle

Tolérance d'arrêt : 0,001

Nombre de validations croisées : 5

Racine aléatoire : 16807

(Libellés des variables quantitatives :)

(Libellés modalités des variables qualitatives :)

(Libellés des observations :)
obs

Ok Annuler Sélection Supprimer Aide

Définition de la sélection

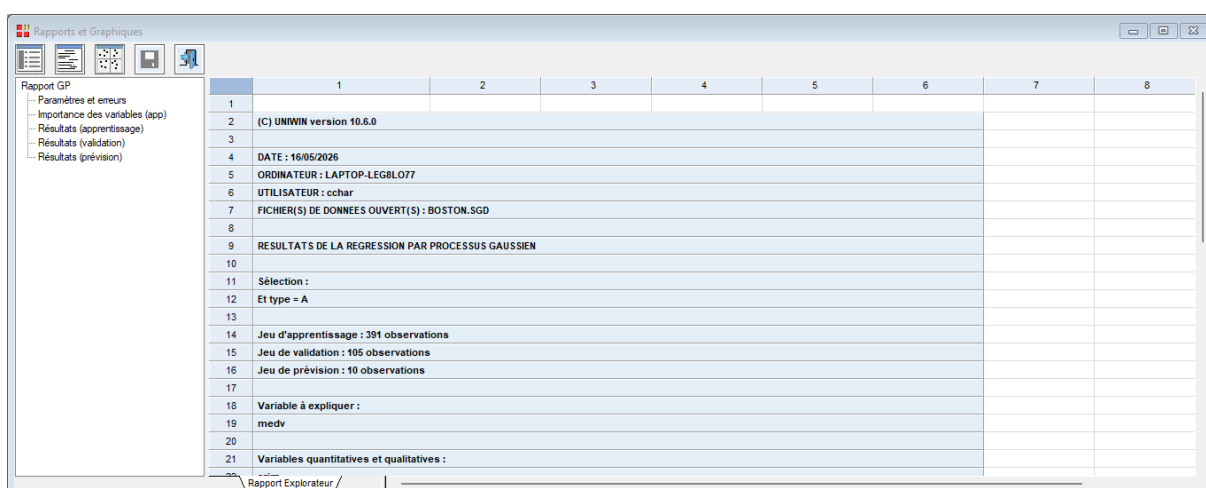
Et type = A

Liaison	Variable	Relation	Valeur ou variable
Et	nox	=	age
Et non	obs	<>	black
Ou	ptratio	<	chas
Ou non	rad	<=	crim
	rm	>	dis
	tax	>=	indus
	type	débute	lstat
	zn		medv

☒ Trier alphabétiquement les noms des variables

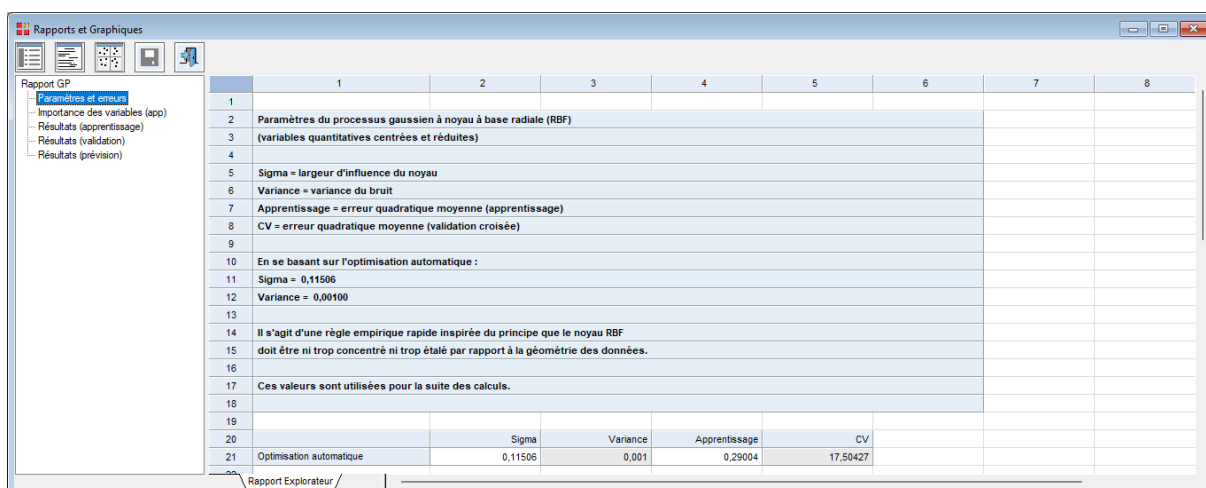
Ok Annuler Ajouter Aide

Après quelques instants, le tableau suivant s'affiche :



	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.6.0							
3								
4	DATE : 16/05/2026							
5	ORDINATEUR : LAPTOP-LEG8LO77							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : BOSTON.SGD							
8								
9	RESULTATS DE LA REGRESSION PAR PROCESSUS GAUSSIEN							
10								
11	Sélection :							
12	Et type = A							
13								
14	Jeu d'apprentissage : 391 observations							
15	Jeu de validation : 105 observations							
16	Jeu de prévision : 10 observations							
17								
18	Variable à expliquer :							
19	medv							
20								
21	Variables quantitatives et qualitatives :							

Le deuxième tableau indique la valeur optimale du paramètre sigma du noyau à base radiale (RBF), la variance du bruit (var) autour des valeurs observées des données ainsi que les erreurs d'apprentissage et de validation croisée.



	1	2	3	4	5	6	7	8
1								
2	Paramètres du processus gaussien à noyau à base radiale (RBF)							
3	(variables quantitatives centrées et réduites)							
4								
5	Sigma = largeur d'influence du noyau							
6	Variance = variance du bruit							
7	Apprentissage = erreur quadratique moyenne (apprentissage)							
8	CV = erreur quadratique moyenne (validation croisée)							
9								
10	En se basant sur l'optimisation automatique :							
11	Sigma = 0,11506							
12	Variance = 0,00100							
13								
14	Il s'agit d'une règle empirique rapide inspirée du principe que le noyau RBF							
15	doit être ni trop concentré ni trop étalé par rapport à la géométrie des données.							
16								
17	Ces valeurs sont utilisées pour la suite des calculs.							
18								
19								
20		Sigma	Variance	Apprentissage		CV		
21	Optimisation automatique	0,11506	0,001	0,29004		17,50427		

Dans cet exemple, l'optimisation automatique donne :

Sigma = 0,11506

Variance = 0,001

Le troisième tableau affiche l'importance des variables explicatives dans le modèle.

Les deux tableaux suivants affichent les résultats (valeurs observées, prévues, résidus et intervalle de confiance) pour les jeux d'apprentissage et de validation.

Enfin, le dernier tableau affiche les résultats (valeurs prévues et intervalle de confiance) pour le jeu de prévision.

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs - Importance des variables (app) - Résultats (apprentissage) - Résultats (validation) - Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire							
5	des données du jeu d'apprentissage dans chacune des variables explicatives)							
6								
7	Importance positive = la variable est utile au modèle							
8	Importance négative = la variable dégrade le modèle							
9								
10								
11		Importance						
12	lstat	8,12977						
13	rm	7,79720						
14	age	7,32479						
15	dis	6,28128						
16	nox	6,23555						
17	rad	4,88994						
18	tax	4,10176						
19	crim	3,72602						
20	ptratio	3,31602						
21	indus	3,11462						

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs - Importance des variables (app) - Résultats (apprentissage) - Résultats (validation) - Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Résultats pour le jeu d'apprentissage							
3								
4	R-carré = 99,709 %							
5								
6	IC inférieur = borne inférieure de l'intervalle de confiance à 95%							
7	IC supérieur = borne supérieure de l'intervalle de confiance à 95%							
8								
9								
10	Observation	Observé	Prévu	Résidu	IC inférieur	IC supérieur		
11	o1	24,0	23,95615	0,04385	-20,62675	68,53904		
12	o2	21,6	21,81400	-0,21400	-22,74060	66,36859		
13	o3	34,7	34,43159	0,26841	-10,13828	79,00145		
14	o4	33,4	33,65275	-0,25275	-10,89730	78,20280		
15	o5	36,2	35,93635	0,26365	-8,61363	80,48654		
16	o6	28,7	28,22077	0,47923	-16,33957	72,78111		
17	o7	22,9	22,08338	0,81662	-22,45347	66,62024		
18	o8	27,1	25,66615	1,43385	-18,88421	70,21652		
19	o11	15,0	15,98635	-0,98635	-28,58188	60,55458		
20	o12	18,9	19,88741	-0,98741	-24,68866	64,44348		
21	o13	21,7	21,83086	-0,13086	-22,74891	66,41063		

Rapports et Graphiques								
Rapport GP								
- Paramètres et erreurs - Importance des variables (app) - Résultats (apprentissage) - Résultats (validation) - Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Résultats pour le jeu de validation							
3								
4	R-carré = 75,989 %							
5								
6	IC inférieur = borne inférieure de l'intervalle de confiance à 95%							
7	IC supérieur = borne supérieure de l'intervalle de confiance à 95%							
8								
9								
10	Observation	Observé	Prévu	Résidu	IC inférieur	IC supérieur		
11	o10	18,9	20,96012	-2,06012	-24,43676	66,35900		
12	o18	17,5	19,15852	-1,65852	-25,36520	63,68224		
13	o30	21,0	20,34058	0,65942	-24,84919	65,33036		
14	o31	12,7	12,40745	0,29255	-33,00324	57,81813		
15	o33	13,2	17,39024	-4,19024	-33,11701	67,89749		
16	o43	25,3	24,58435	0,71565	-20,48978	69,65848		
17	o48	16,6	20,65722	-4,05722	-25,99682	67,31126		
18	o49	14,4	29,21794	-14,81794	-24,21229	82,64816		
19	o57	24,7	24,32407	0,37593	-21,85576	70,50389		
20	o66	23,5	27,02919	-3,52919	-19,18458	73,24296		
21	o68	22,0	21,97759	0,02241	-22,94996	66,90513		

Rapports et Graphiques							
Rapport GP							
Paramètres et erreurs							
Importance des variables (app)							
Résultats (apprentissage)							
Résultats (validation)							
Résultats (prévision)							
1	1	2	3	4	5	6	7
2	Résultats pour le jeu de prévision						
3							
4	IC inférieur = borne inférieure de l'intervalle de confiance à 95%						
5	IC supérieur = borne supérieure de l'intervalle de confiance à 95%						
6							
7							
8	Observation	Prévu	IC inférieur	IC supérieur			
9	o9	45,09052	-7,69753	97,87858			
10	o25	17,18981	-27,31255	61,69217			
11	o60	21,25144	-23,68193	66,18481			
12	o97	20,33964	-24,85627	65,53555			
13	o141	13,67893	-32,61104	59,96890			
14	o174	27,71695	-17,23777	72,67167			
15	o202	22,39165	-25,00346	69,78676			
16	o239	31,02993	-14,01307	76,07294			
17	o351	28,64532	-19,69350	76,98414			
18	o495	25,45301	-19,52257	70,42860			
19							
20							
21							

Les graphiques disponibles pour un modèle de régression sont :

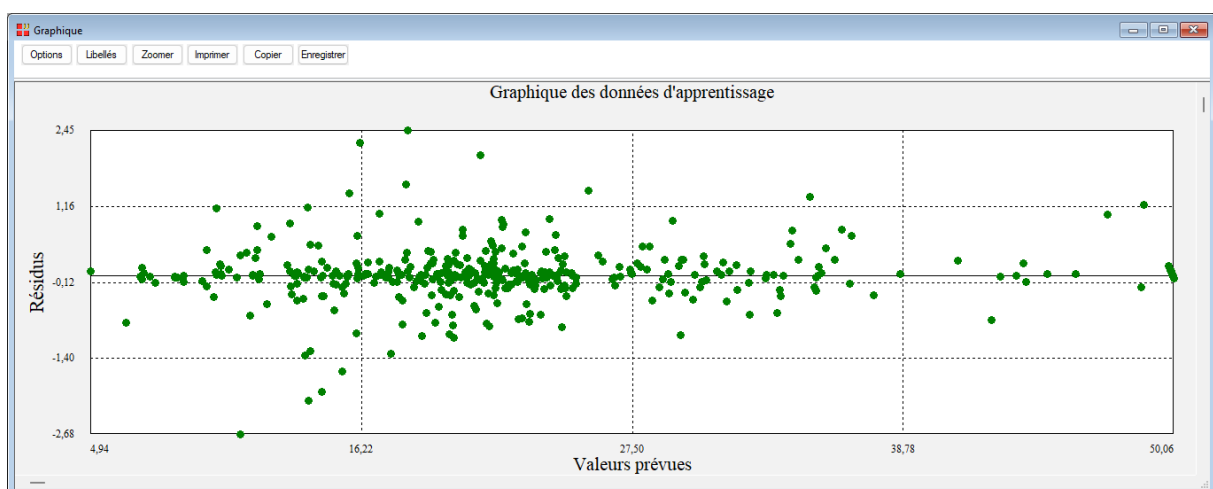
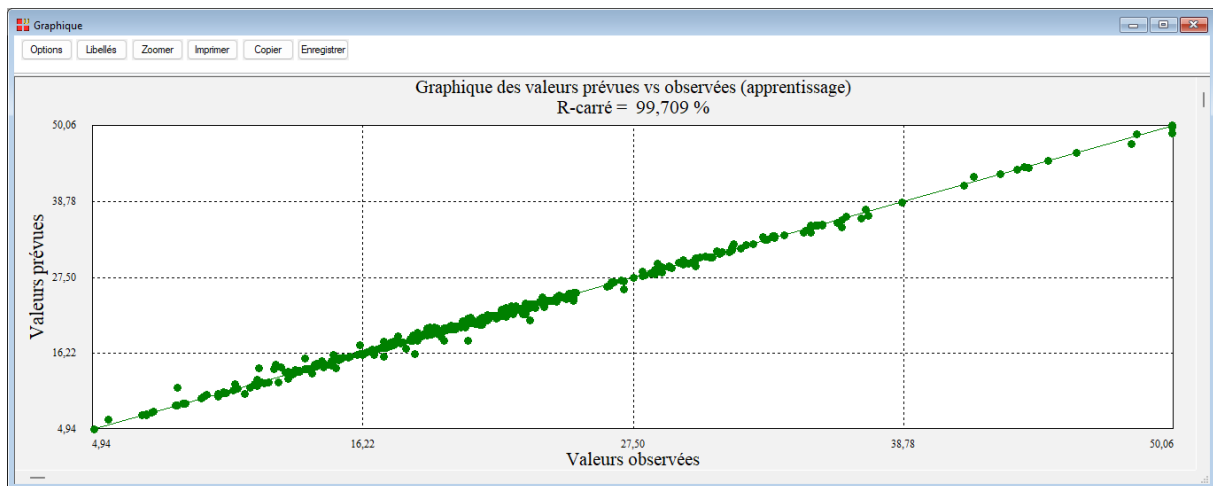
Graphiques

☒ Graphique des erreurs (apprentissage, validation croisée)
☐ Graphique de l'importance des variables explicatives
☐ Nuage de points (apprentissage)
☐ Nuage de points (validation)
☐ Nuage de points (prévision)
☐ Graphique des frontières (apprentissage)
☐ Graphique des frontières (validation)
☐ Prévus versus observés (apprentissage)
☐ Prévus versus observés (validation)
☐ Résidus versus prévus (apprentissage)
☐ Résidus versus prévus (validation)
☐ Prévus versus X avec IC (apprentissage)
☐ Prévus versus X avec IC (validation)
☐ Prévus versus X avec IC (prévision)

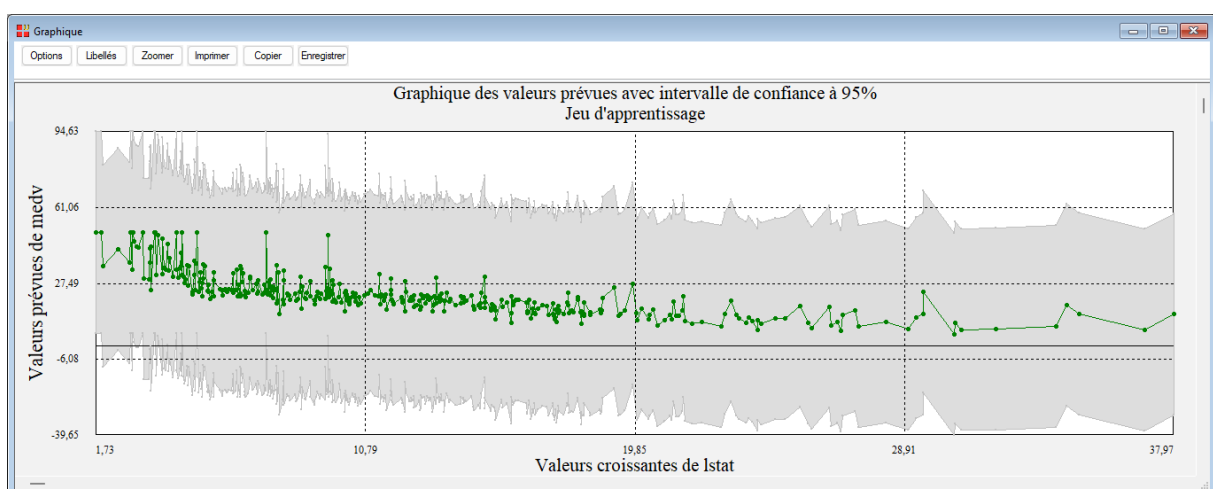
Ok
Annuler

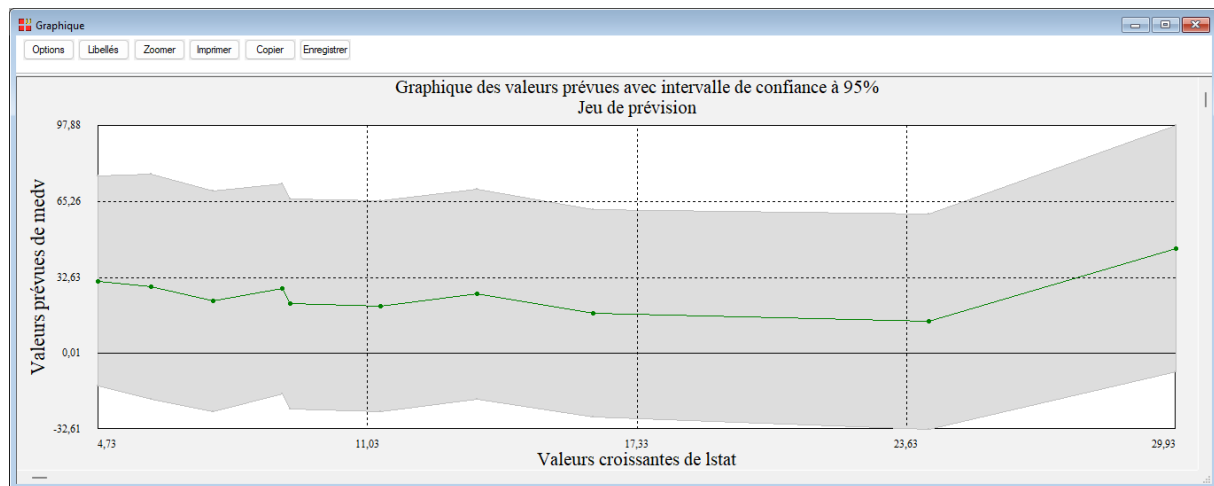
Visualisons ceux spécifiques à un modèle de régression :

- Prévus versus observés
- Résidus versus prévus
- Prévus versus X avec IC



Après avoir choisi la variable X à utiliser, ici *lestât*, le graphique affiche les valeurs prévues de la variable à expliquer *medv* en fonction de X avec un intervalle de confiance à 95%.





Enregistrement des résultats

Voici la liste des variables créées par la procédure.

<i>Variable</i>	<i>Contenu</i>
sigma	Valeurs du paramètre sigma
var	Valeurs de paramètre var (régression)
erreurapp	Erreurs d'apprentissage
erreurcv	Erreurs de validation croisée
varexpli	Variables explicatives
impvarexpli	Importance des variables explicatives
libobsapp	Libellés des observations (apprentissage)
probapp	Probabilités (classement, apprentissage)
obsapp	Classes ou valeurs observées (apprentissage)
vpapp	Classes ou valeurs prévues (apprentissage)
ectapp	Ecart-types (apprentissage, régression)
icinfapp	Borne inférieure IC (apprentissage, régression)
icsupapp	Borne supérieure IC (apprentissage, régression)
residapp	Résidus (apprentissage, régression)
libobsvalid	Libellés des observations (validation)
probvalid	Probabilités (classement, validation)
obsbvalid	Classes ou valeurs observées (validation)
vpvalid	Classes ou valeurs prévues (validation)
ectvalid	Ecart-types (validation, régression)
icinvalid	Borne inférieure IC (validation, régression)
icsupvalid	Borne supérieure IC (validation, régression)
residvalid	Résidus (validation, régression)

libobsprev	Libellés des observations (prévision)
probprev	Probabilités (classement, prévision)
vpprev	Classes ou valeurs prévues (prévision)
ectprev	Ecart-types (prévision, régression)
icinfprev	Borne inférieure IC (prévision, régression)
icsupprev	Borne supérieure IC (prévision, régression)
seuilA	Seuils (apprentissage, classement binaire)
specificiteA	Spécificités (apprentissage, classement binaire)
sensibiliteA	Sensibilités (apprentissage, classement binaire)
aireA	Aires sous les courbes (apprentissage, classement binaire)
seuilV	Seuils (validation, classement binaire)
specificiteV	Spécificités (validation, classement binaire)
sensibiliteV	Sensibilités (validation, classement binaire)
aireV	Aires sous les courbes (validation, classement binaire)

Références

Package R 'kernlab' (2026) :

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

Gaussian Processes for Machine Learning

C. E. Rasmussen & C. K. I. Williams, the MIT Press, 2006

Exemple 'Boston' : <http://lib.stat.cmu.edu/datasets/boston>