

## UNIWIN VERSION 10.2.0

# REGRESSION TYPOLOGIQUE

Révision : 25/03/2025

Définition.....	1
Entrée des données .....	2
Données manquantes .....	3
Remarques concernant l'algorithme utilisé .....	3
Exemple 1 : Fichier Regtypo1.....	3
L'option Rapports .....	6
L'option Graphiques .....	8
Exemple 2 : Fichier Regtypo2.....	10
Les variables internes créées par la procédure .....	15
Références .....	15

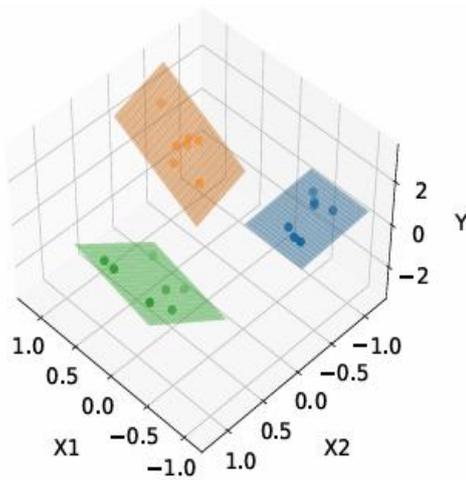
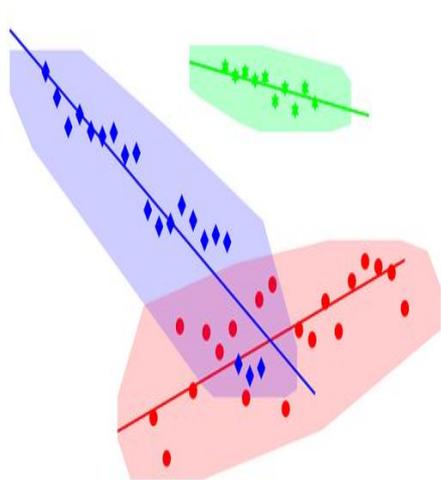
### Définition

La régression linéaire typologique (RTYPO) est une technique statistique qui s'attaque à une limitation fondamentale de la régression linéaire conventionnelle : l'incapacité à saisir des relations complexes au sein d'un ensemble de données contenant des sous-groupes distincts.

RTYPO y parvient en partitionnant l'ensemble de données en plusieurs sous-ensembles, chacun doté de sa propre fonction de régression linéaire entre variables indépendantes et dépendantes.

Une description courante, et souvent plus technique, de ce problème consiste à partitionner un ensemble de données hétérogène en sous-groupes d'observations homogènes plus petits, tout en ajustant simultanément des modèles de régression linéaire adaptés à chaque sous-groupe.

RTYPO est une extension naturelle de la formulation conventionnelle de la régression linéaire multivariée à la modélisation de données contenant des structures de sous-groupes inconnues ou à des problèmes de régression non linéaire se prêtant à une approximation linéaire par morceaux.



## Entrée des données

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

RTYPO Régression typologique
✕

Variable à expliquer :

Variables explicatives quantitatives :

(Libellés des variables explicatives :)

(Libellés des observations :)

Nombre demandé de classes :

Nombre maximum de classes à tester :

Nombre maximum d'itérations :

Seuil pour la convergence :

Nombre de tirages aléatoires :

Ok
Annuler
Sélection
Supprimer
Aide

Cette boîte de dialogue permet de définir la variable quantitative à expliquer, les variables explicatives quantitatives et les libellés associés ainsi que les libellés des observations.

Le nombre demandé de classes, le nombre maximum de classes à tester, le nombre maximum d'itérations de l'algorithme, le seuil pour convergence de l'algorithme et le nombre de tirages aléatoires pour l'initialisation des appartenances des observations aux classes peuvent être précisés.

## Données manquantes

Dans cette procédure les données manquantes ne sont pas permises pour les variables explicatives mais le sont pour la variable à expliquer.

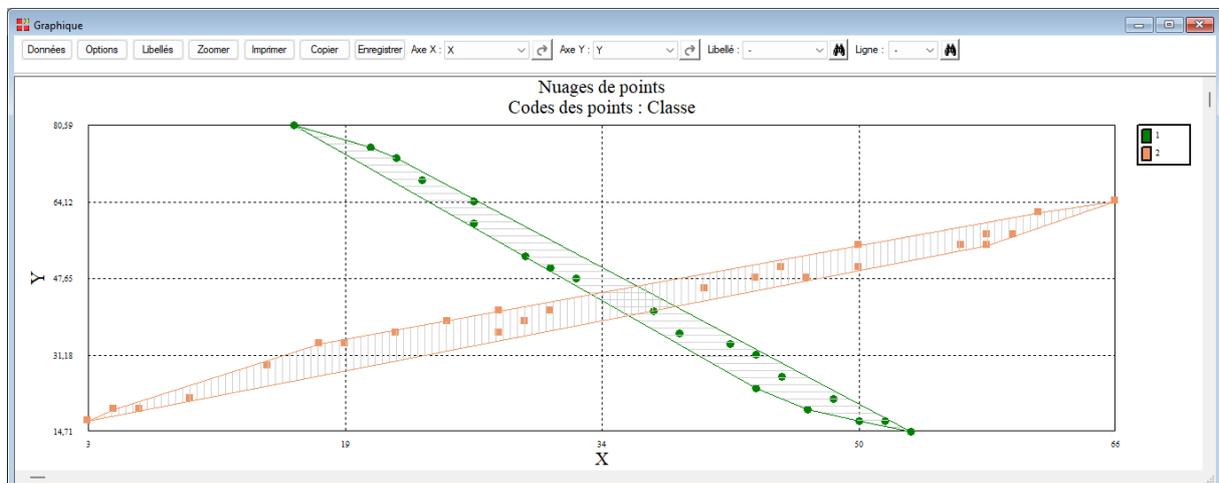
Les observations pour lesquelles la valeur de la variable à expliquer est manquante définissent le jeu de prévision.

## Remarques concernant l'algorithme utilisé

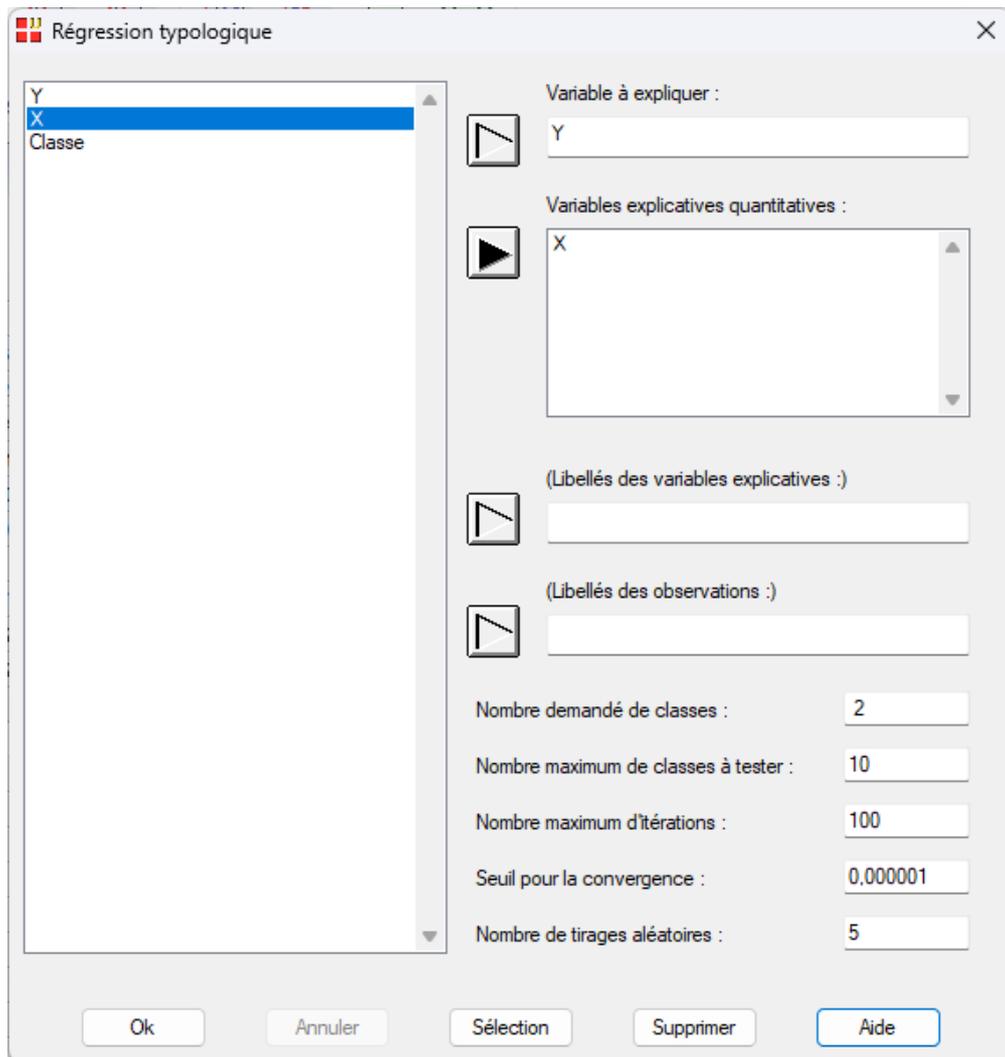
- Si lors d'une itération, le nombre d'observations d'une classe devient inférieur au nombre de coefficients à estimer du modèle de régression, ceux-ci sont générés aléatoirement.
- Si lors d'une itération, une classe devient vide alors un certain nombre d'observations de la plus grande classe sont affectées à cette classe vide.

## Exemple 1 : Fichier Regtypo1

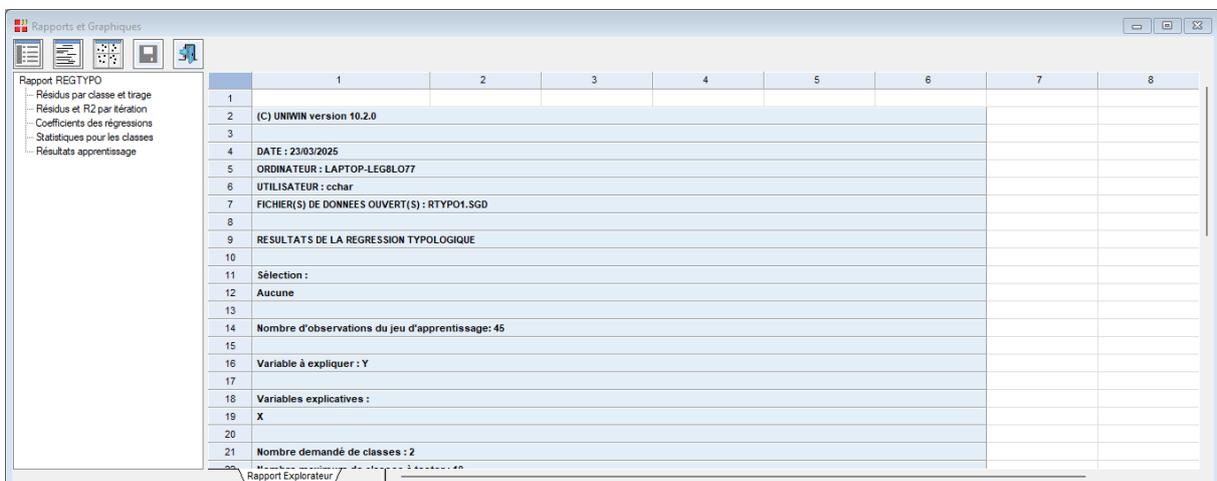
Pour illustrer cette procédure, nous utiliserons le fichier Regtypo1 contenant les données utilisées par Hennig (2000). Représentons graphiquement ces données.



Renseignons la boîte de dialogue de la régression typologique comme montré ci-dessous en précisant que nous souhaitons 2 modèles de régression (2 classes).

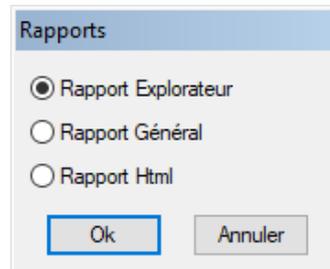


Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse. Après quelques instants, l'écran suivant s'affiche :

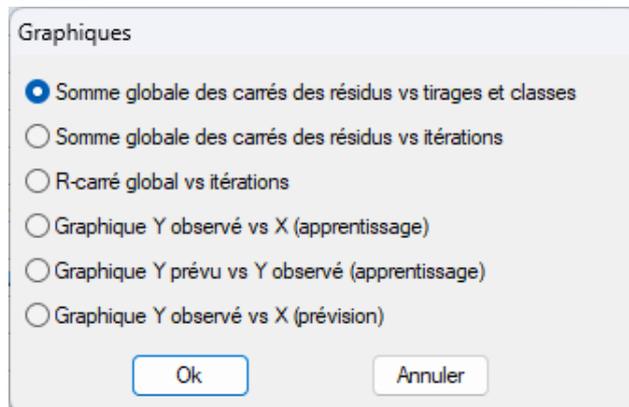


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

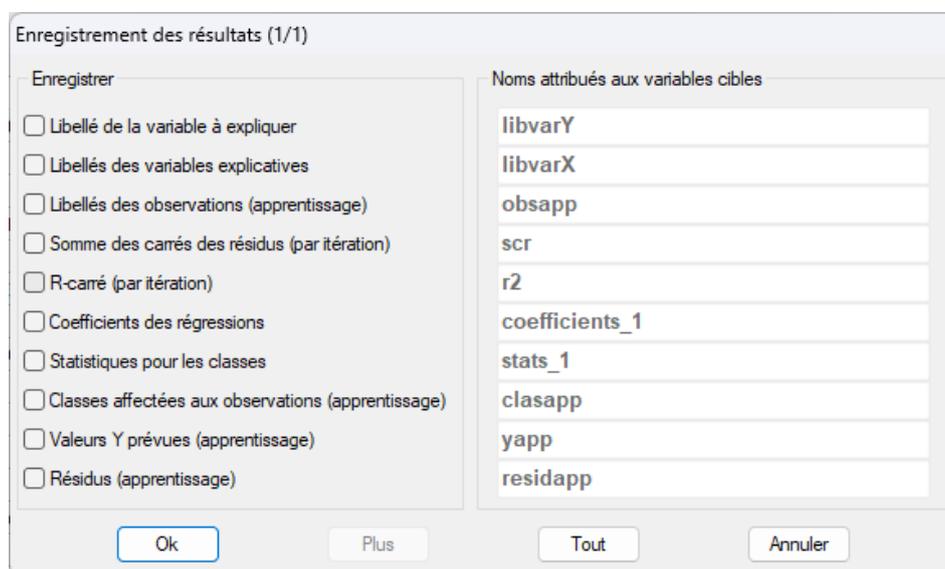
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



Note : le bouton 'Plus' permet d'afficher la suite de la liste des variables, si cela est nécessaire.

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

## L'option Rapports

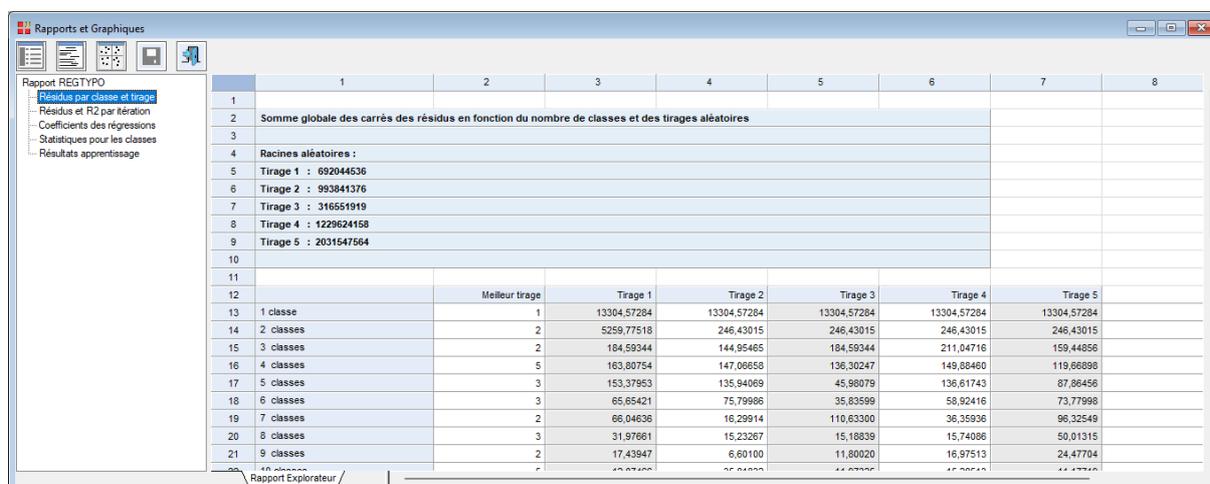
Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableau ou au format HTML.

L'impression des rapports fait appel à la procédure 'Aperçu avant impression'. Pour des informations sur cette procédure, voir le 'Manuel de l'Utilisateur'.

Ce rapport nous donne les informations suivantes :

- Résidus par classe et tirage

Ce tableau affiche pour les 5 tirages aléatoires demandés et pour chaque nombre de classes jusqu'au nombre maximum demandé de 10, la somme globale des carrés des résidus et indique le meilleur tirage obtenu et sa racine aléatoire.



	1	2	3	4	5	6	7	8
1								
2	Somme globale des carrés des résidus en fonction du nombre de classes et des tirages aléatoires							
3								
4	Racines aléatoires :							
5	Tirage 1 : 692044536							
6	Tirage 2 : 993841376							
7	Tirage 3 : 316551919							
8	Tirage 4 : 1229624158							
9	Tirage 5 : 2031547564							
10								
11								
12		Meilleur tirage	Tirage 1	Tirage 2	Tirage 3	Tirage 4	Tirage 5	
13	1 classe	1	13304,57284	13304,57284	13304,57284	13304,57284	13304,57284	
14	2 classes	2	5259,77518	246,43015	246,43015	246,43015	246,43015	
15	3 classes	2	184,59344	144,95465	184,59344	211,04716	159,44856	
16	4 classes	5	163,80754	147,06658	136,30247	149,86460	119,66898	
17	5 classes	3	153,37953	135,94069	45,98079	136,61743	87,86456	
18	6 classes	3	65,65421	75,79986	35,83599	58,92416	73,77998	
19	7 classes	2	66,04636	16,29914	110,63300	36,35936	96,32549	
20	8 classes	3	31,97861	15,23267	15,18839	15,74086	50,01315	
21	9 classes	2	17,43947	6,60100	11,80020	16,97513	24,47704	

Pour notre recherche de 2 classes, le meilleur tirage aléatoire est le n° 2. Ce tirage aléatoire sera automatiquement utilisé pour la suite de l'analyse.

- Résidus et R2 par itération

Ce tableau affiche pour le meilleur tirage, l'évolution de la somme globale des carrés des résidus et le R-carré global associé en fonction des itérations de l'algorithme.

Dans notre exemple, la convergence de l'algorithme a été obtenu après 5 itérations avec un R-carré de 98,18 %.

Rapports et Graphiques

Rapport REGTYPO

- Résidus par classe et tirage
- Résidus et R2 par itération
- Coefficients des régressions**
- Statistiques pour les classes
- Résultats apprentissage

	1	2	3	4	5	6	7	8
1								
2	<b>Somme globale des carrés des résidus et R-carré global par itération</b>							
3								
4	<b>Nombre de classes = 2</b>							
5	<b>Racine aléatoire utilisée = 993841376 (tirage 2)</b>							
6	<b>Convergence atteinte après 5 itérations.</b>							
7								
8	<b>Somme des carrés = somme globale des carrés des résidus</b>							
9	<b>R2 = R-carré global (%)</b>							
10								
11								
12		Somme des carrés	R2					
13	Itération 1	9857,89464	27,18183					
14	Itération 2	4976,57711	63,23908					
15	Itération 3	1477,56947	89,08551					
16	Itération 4	261,06851	98,07139					
17	Itération 5	246,43015	98,17967					
18								
19								
20								
21								

Rapport Explorateur

- Coefficients des régressions

Ce tableau affiche les coefficients des régressions pour les 2 classes formées.

Rapports et Graphiques

Rapport REGTYPO

- Résidus par classe et tirage
- Résidus et R2 par itération
- Coefficients des régressions**
- Statistiques pour les classes
- Résultats apprentissage

	1	2	3	4	5	6	7	8
1								
2	<b>Coefficients des régressions pour les classes</b>							
3								
4								
5		Constante	X					
6	Classe 1	110,64208	-1,86641					
7	Classe 2	18,26343	0,67978					
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

Rapport Explorateur

- Statistiques pour les classes

Ce tableau affiche pour chacune des classes les effectifs, les R-carrés et les erreurs quadratiques moyennes.

Rapports et Graphiques

Rapport REGTYPO

- Résidus par classe et tirage
- Résidus et R2 par itération
- Coefficients des régressions
- Statistiques pour les classes**
- Résultats apprentissage

	1	2	3	4	5	6	7	8
1								
2	<b>Statistiques pour les classes</b>							
3								
4	<b>R2 = R-carré (%)</b>							
5	<b>Erreur = Erreur quadratique moyenne</b>							
6								
7								
8		Effectif	R2	Erreur				
9	Classe 1	20	98,61954	6,23565				
10	Classe 2	25	97,29300	4,86869				
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

Rapport Explorateur

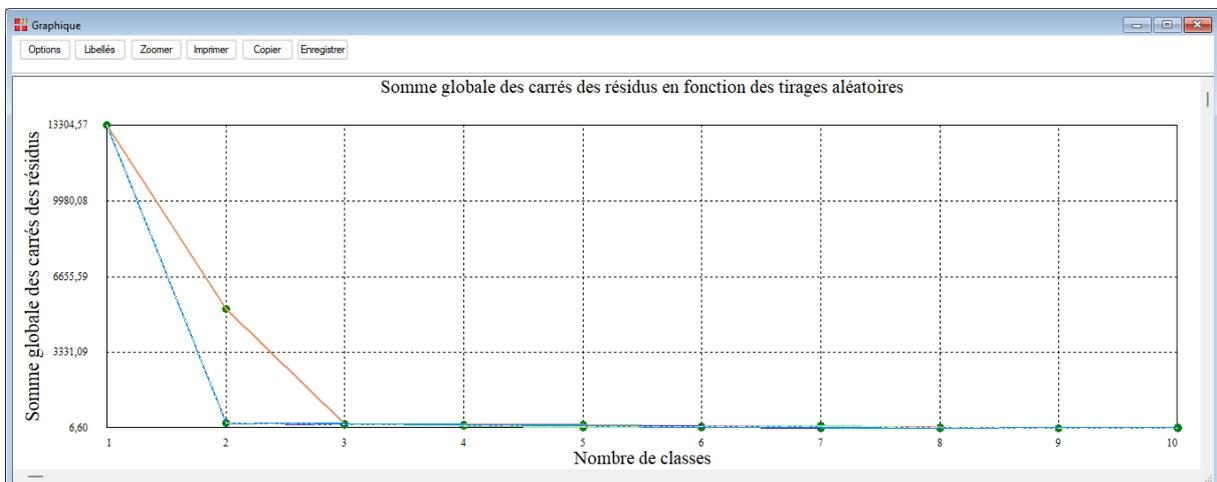
- Résultats apprentissage

Ce tableau affiche pour chaque observation sa classe d'affectation, les valeurs prévue et observée de la variable à expliquer ainsi que le résidu.

	1	2	3	4	5	6	7	8
1								
2	<b>Résultats pour les données d'apprentissage</b>							
3								
4								
5		Classe affectée	Y prévu	Y observé	Résidu			
6	i1	1	81,88292	80,58823	1,29469			
7	i2	1	73,07908	75,88235	-2,80327			
8	i3	1	70,14448	73,52941	-3,38493			
9	i4	1	67,20987	68,82353	-1,61366			
10	i5	2	20,18733	17,05882	3,12851			
11	i6	2	21,25816	19,41177	1,84639			
12	i7	2	22,32499	19,41177	2,91322			
13	i8	2	24,46265	21,76471	2,69794			
14	i9	2	27,86914	28,82353	-1,15439			
15	i10	2	29,80679	33,52941	-3,72262			
16	i11	2	30,87563	33,52941	-2,65378			
17	i12	2	50,11457	47,64706	2,46751			
18	i13	2	52,25222	50,00000	2,25222			
19	i14	2	56,52754	54,70588	1,82166			
20	i15	2	57,59637	54,70588	2,89049			
21	i16	2	57,59637	57,05882	0,53755			

## L'option Graphiques

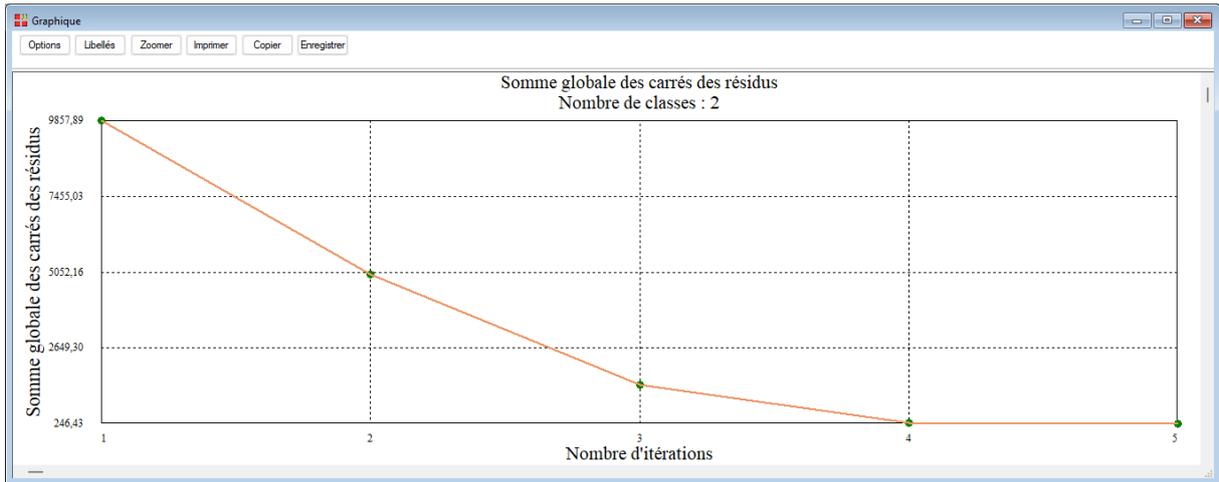
- Somme globale des carrés des résidus vs tirages et classes



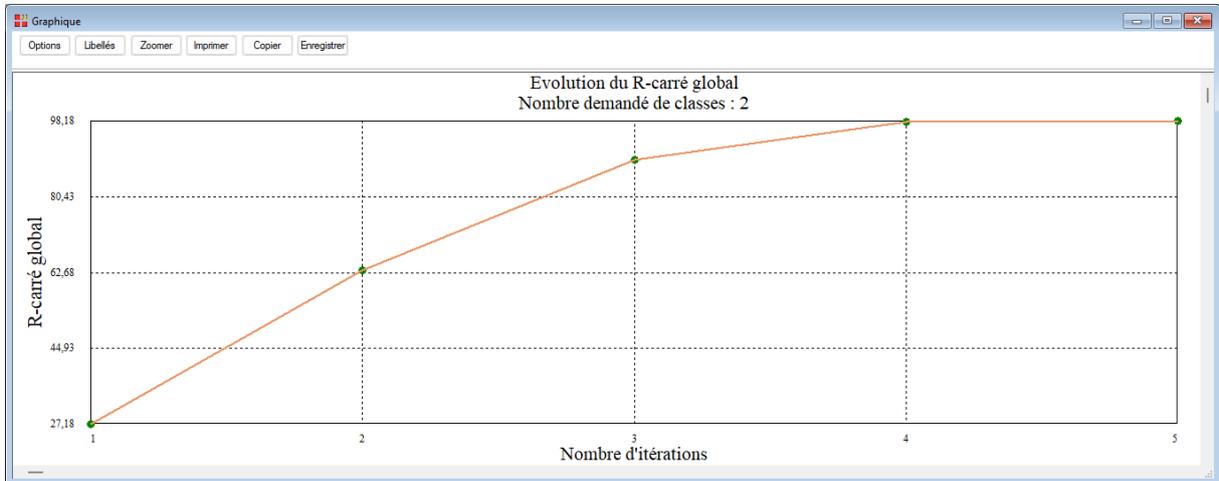
Ce graphique affiche l'évolution de la somme globale des carrés des résidus par rapport aux tirages et aux nombres de classes formées.

- Somme globale des carrés des résidus vs itérations

Ce graphique affiche l'évolution de la somme globale des carrés des résidus par rapport aux itérations pour le meilleur tirage.

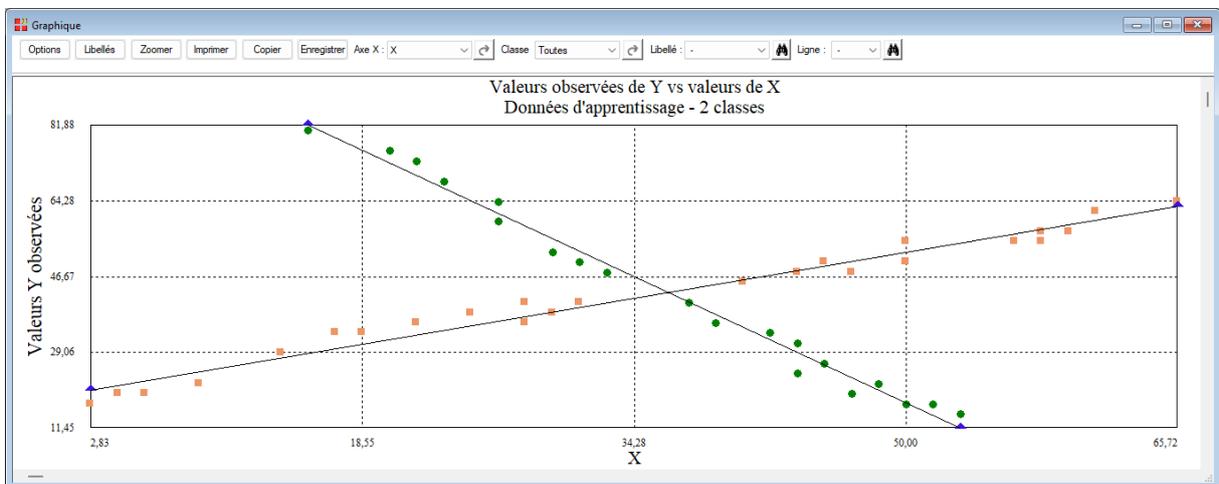


- R-carré global vs itérations



Ce graphique affiche l'évolution du R-carré global par rapport aux itérations pour le meilleur tirage.

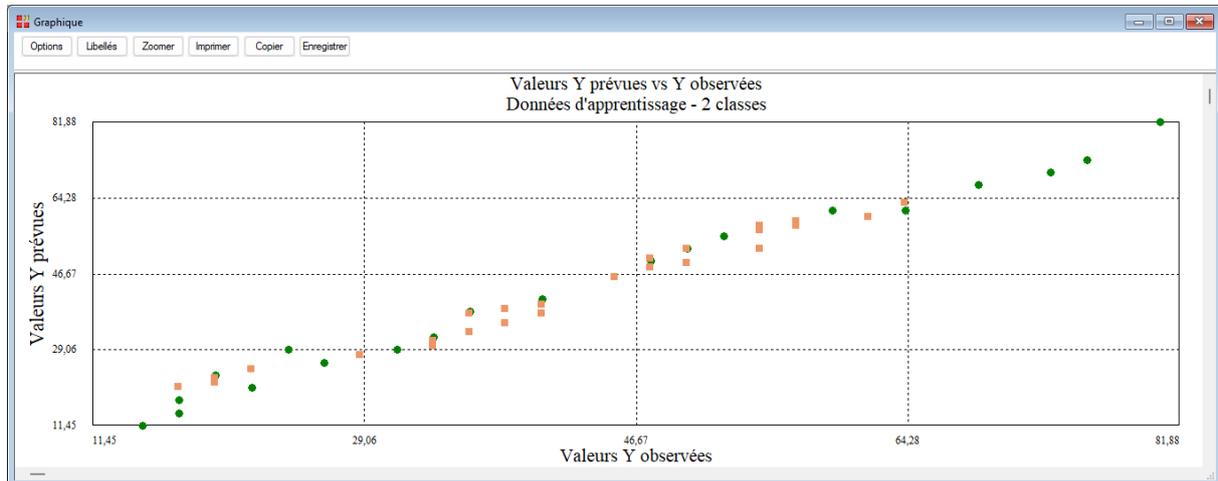
- Graphique Y observé vs X (apprentissage)



Ce graphique affiche les valeurs observées de Y par rapport aux valeurs de X et les régressions pour les deux classes formées avec les données d'apprentissage.

- Graphique Y prévu vs Y observé (apprentissage)

Ce graphique affiche les valeurs Y prévues par rapport aux valeurs Y observées pour chacune des classes formées.



- Graphique Y observé vs X (prévision)

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

## Exemple 2 : Fichier Regtypo2

Pour illustrer ce deuxième exemple, nous utiliserons le fichier Regtypo2.

Il contient 3 ensembles de données Y, X1 et X2 générées aléatoirement.

Il est constitué de 306 observations pour le jeu d'apprentissage et de 6 observations pour le jeu de prévision.

Renseignons la boîte de dialogue comme montré ci-après.

Cliquons sur le bouton 'Sélection' pour définir le jeu d'apprentissage.

Régression typologique

type  
Y  
X1  
X2  
libobs  
libX  
classe

Variable à expliquer :  
Y

Variables explicatives quantitatives :  
X1  
X2

(Libellés des variables explicatives :)  
libX

(Libellés des observations :)  
libobs

Nombre demandé de classes : 3  
Nombre maximum de classes à tester : 10  
Nombre maximum d'itérations : 100  
Seuil pour la convergence : 0,000001  
Nombre de tirages aléatoires : 10

Ok Annuler Sélection Supprimer Aide

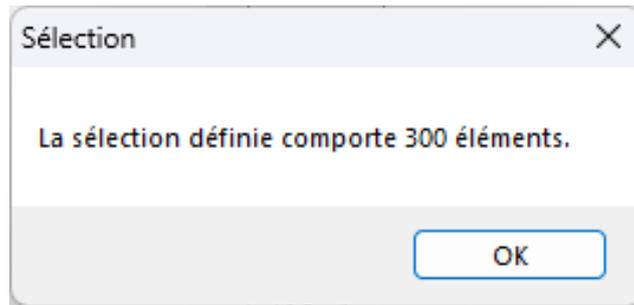
Définition de la sélection

Et type = A

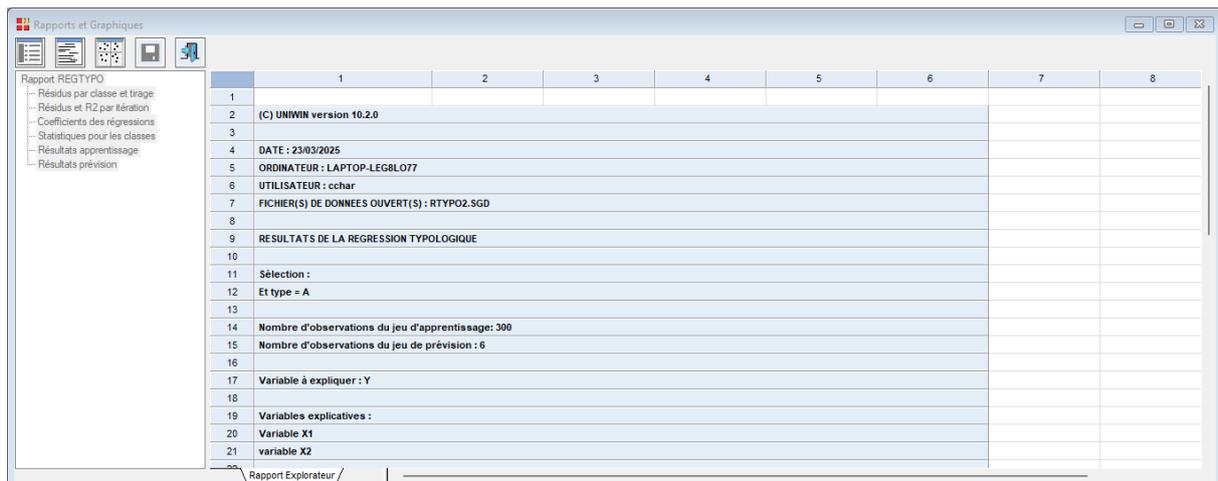
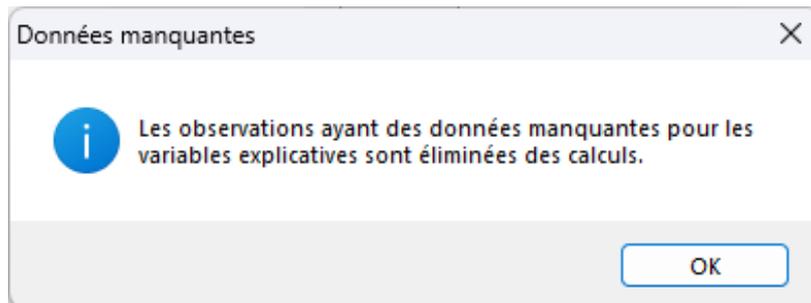
Liaison	Variable	Relation	Valeur ou variable
Et	classe	=	classe
Et non	libobs	<>	libobs
Ou	libX	<	libX
Ou non	type	<=	type
	X1	>	X1
	X2	>=	X2
	Y	débute	Y

Trier alphabétiquement les noms des variables

Ok Annuler Ajouter Aide



Cliquons sur Ok pour exécuter l'analyse.



Coefficients des régressions pour les classes			
	Constante	Variable X1	variable X2
Classe 1	8,94197	-0,91659	5,13639
Classe 2	-3,52991	5,18330	1,11548
Classe 3	1,95834	2,99422	-2,05087

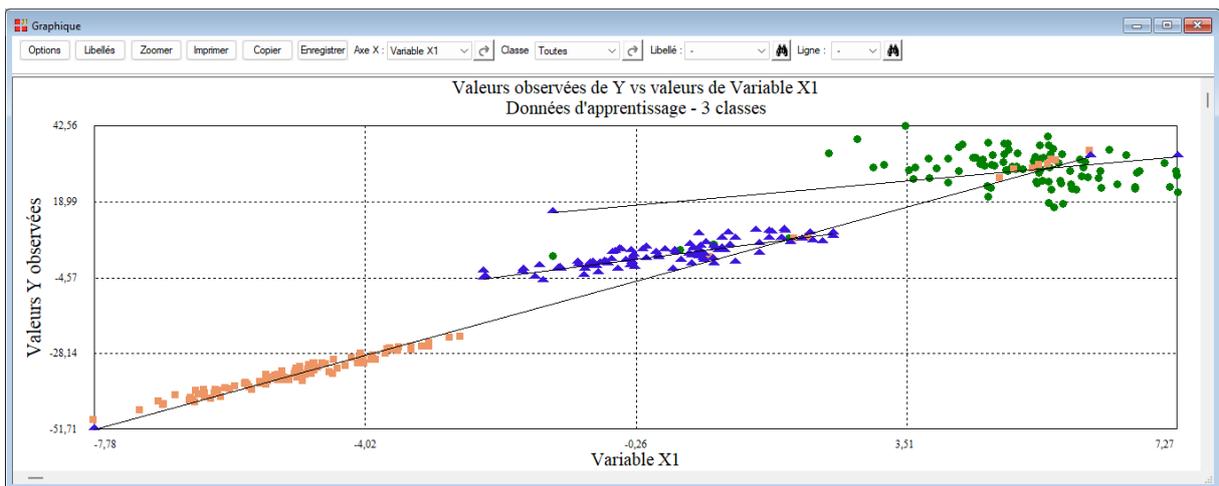
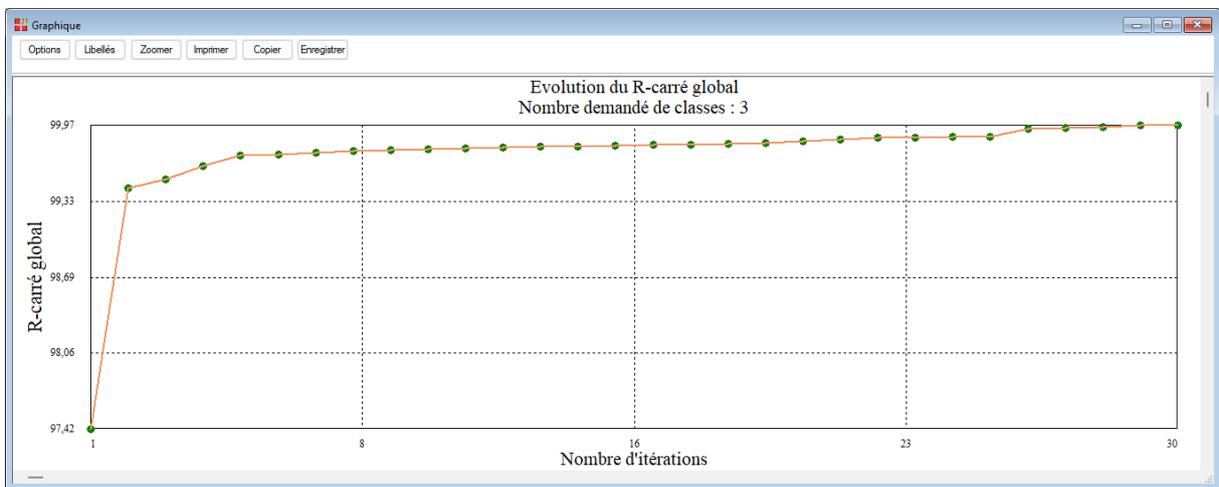
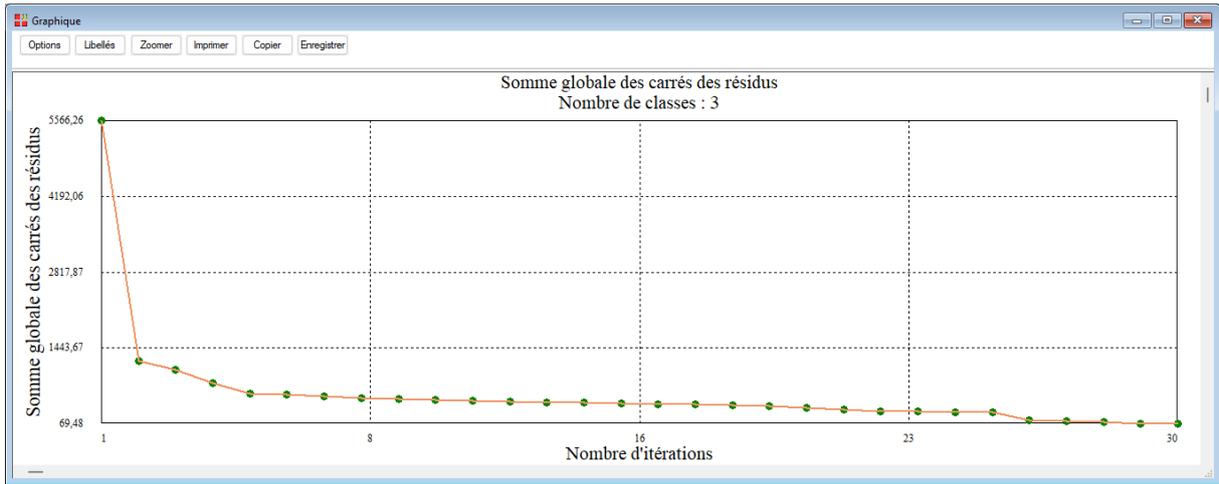
Statistiques pour les classes			
R2 = R-carré (%)			
Erreur = Erreur quadratique moyenne			
	Effectif	R2	Erreur
Classe 1	97	99,49787	0,26066
Classe 2	111	99,93107	0,23666
Classe 3	92	98,62440	0,19482

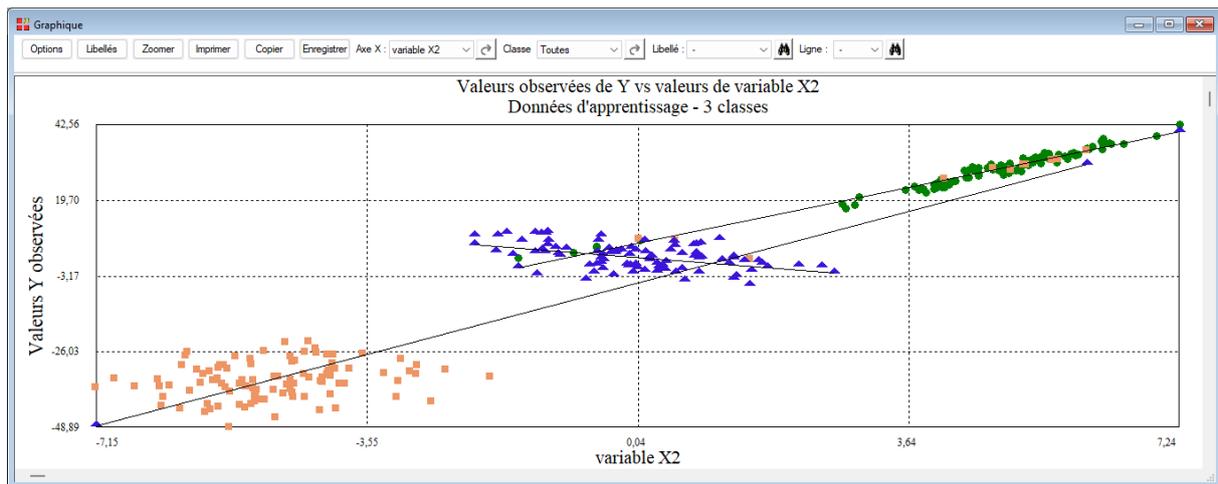
Résultats pour les données d'apprentissage				
	Classe affectée	Y prévu	Y observé	Résidu
Obs1	3	3,25229	2,59066	0,66163
Obs2	3	6,45390	6,12621	0,32768
Obs3	3	0,76473	0,94901	-0,18428
Obs4	3	3,31737	3,81820	-0,50083
Obs5	3	3,48315	3,95117	-0,46802
Obs6	3	-2,38566	-2,32917	-0,05649
Obs7	3	4,48405	3,64265	0,84140
Obs8	3	-2,06000	-1,61815	-0,44185
Obs9	3	2,02636	2,09214	-0,06578
Obs10	3	-1,45290	-1,43858	-0,01432
Obs11	3	4,13319	4,60758	-0,47439
Obs12	3	10,11524	9,67043	0,44481
Obs13	3	0,48143	0,44861	0,03282
Obs14	3	0,81040	1,95531	-1,14491

Résultats pour les données de prévision		
	Classe affectée	Y prévu
Pobs001	3	3,25229
Pobs100	3	-0,14858
Pobs150	1	24,03957
Pobs200	1	29,70056
Pobs250	2	-34,92606
Pobs300	2	-29,75301

Les 6 observations non sélectionnées sont utilisées comme jeu de prévision.

Chaque observation est affectée à la classe dont le barycentre est le plus proche et la valeur prévue de la variable à expliquer est calculée en utilisant le modèle ajusté de régression pour cette classe.





## Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure. Ces variables peuvent notamment être utilisées avec l'option 'Sélection'. A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

<i>Variable</i>	<i>Contenu</i>
libvarY	Libellé de la variable à expliquer
libvarX	Libellés des variables explicatives
obsapp	Libellés des observations (apprentissage)
scr	Somme des carrés des résidus (par itération)
r2	R-carré (par itération)
coefficients	Coefficients des régressions
stats	Statistiques pour les classes
clasapp	Classes affectées aux observations (apprentissage)
yapp	Valeurs Y prévues (apprentissage)
residapp	Résidus (apprentissage)
obsprev	Libellés des observations (prévision)
clasprev	Classes affectées aux observations (prévision)

## Références

CHARLES, C. (1977) : Régression Typologique et Reconnaissance des Formes. Thèse de doctorat, Université Paris IX. <https://theses.hal.science/tel-01497620/>

CHARLES, C (1977) : Régression typologique. [Rapport de recherche] INRIA-RR-257, INRIA. 1977, pp.44. hal-04716445

CHARLES, C, LECHEVALLIER, Y (1979) : Pattern recognition by a piecewise polynomial approximation with variable joints. [Research Report] INRIA-RR-338, INRIA. 1979, pp.23. hal-04716617

DIDAY, E. (1974) : Introduction à l'analyse factorielle typologique, Revue de Statistique Appliquée, 22, 4, pp.29-38

HENNIG, C. (1999) : Models and methods for clusterwise linear regression. In: Classification in the Information Age, Springer, pp.179-187.

HENNIG, C. (2000) : Identifiability of models for Clusterwise linear regression. Journal of Classification, 17, pp.273-296.

NIANG-KEITA, N, SAPORTA G. (2014) : Régression typologique pour données multi blocs, 46 èmes journées de statistique, Rennes

SPÄTH, H. (1979) : Clusterwise linear regression, Computing, 22, pp.367-373