

UNIWIN VERSION 10.3.0

REGRESSION PLS

Révision : 09/06/2025

Définition	1
Entrée des données	2
Données manquantes ou non sélectionnées	3
Exemple 1 : Fichier Octane (PLS1)	3
L'option Rapports	6
L'option Graphiques	11
Exemple 2 : Fichier Octane2 (PLS1)	17
Exemple 3 : Fichier Thé (PLS2)	21
Les variables internes créées par la procédure	25
Références	26

Définition

La méthode Régression PLS (partial least squares ou moindres carrés partiels) est conçue pour ajuster un modèle statistique reliant un ensemble de variables explicatives X à une variable à expliquer Y (PLS1) ou à plusieurs variables à expliquer Y (PLS2). La procédure est utile lorsqu'il y a de nombreux X et que le but principal est de prévoir simultanément les variables Y. Elle est recommandée dans le cas où un grand nombre de variables X est utilisé, lorsqu'il y a de fortes colinéarités entre ces variables X, lorsque le nombre de variables X est supérieur au nombre d'observations et lorsqu'il y a des données manquantes. La méthode PLS est notamment largement utilisée par les chimistes et les chimiométriciens pour l'étalonnage en spectrométrie.

Un rapport général de synthèse est proposé contenant notamment les PRESS, R2, Q2, R2X, R2Y, les poids w et w*, les scores t et u, les poids des variables X et Y, les corrélations des variables avec les composantes, les valeurs observées, ajustées et résidus du modèle, les distances au modèle en X et Y, les T2 de Hotelling et les VIP. Les graphiques des R2X, R2Y, Q2, des coefficients standardisés, des cercles des corrélations, des plans factoriels, des poids des variables, des T2, des distances au modèle en X et Y, des VIP, des valeurs observées vs estimées et des résidus sont également disponibles.

La procédure implémentée est basée sur le package R 'plsdepot'.

Entrée des données

Régression PLS	×
	Variables à expliquer quantitatives :
	×
	Variables explicatives quantitatives :
	×
	(Libellés des variables à expliquer :)
	(Libellés des variables explicatives :)
×	
Nombre de composantes à extraire :	(Libellés des observations :)
✓ Validation croisée	
Racine aléatoire : 2107043751	
Ok Annuler	Sélection Supprimer Aide

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

Cette boîte de dialogue permet de définir la ou les variables à expliquer, les variables explicatives quantitatives, les libellés optionnels des variables à expliquer et explicatives, les libellés optionnels des observations, le nombre de composantes à extraire, si la validation croisée est mise en œuvre on non et la racine aléatoire pour la validation croisée.

Les données à expliquer et explicatives sont automatiquement centrées et réduites.

Si le nombre de composantes à extraire n'est pas précisé, il est déterminé par la validation croisée. Un minimum de deux composantes est extrait. A noter que la validation croisée n'est pas mise en œuvre s'il y a des données manquantes dans les variables explicatives ou si le nombre d'observations est inférieur à 10.

Données manquantes ou non sélectionnées

Les données manquantes ne sont pas autorisées dans les variables à expliquer. Les lignes contenant une ou des données manquantes ne participent donc pas aux calculs et définissent le jeu de prévision. Les lignes non sélectionnées définissent le jeu de validation. Les données manquantes ne sont pas autorisées dans les variables explicatives en régression PLS2.

Exemple 1 : Fichier Octane (PLS1)

Pour illustrer ce premier exemple, nous utiliserons le fichier OCTANE (données de Cornell augmentées de deux observations pour lesquelles la valeur d'octane est non présente). Ce fichier contient les données suivantes :

distil	Distillation directe
reformat	Réformat
naphtat	Naphta de craquage thermique
naphtac	Naphta de craquage catalytique
polymère	Polymère
alkylat	Alkylat
essence	Essence naturelle
octane	indice d'octane

La variable octane est la variable Y à expliquer.

De Ba	²² Données										
	distil	reformat	naphtat	naphtac	polymère	alkylat	essence	octane			
	Distillation directe	Réformat	Naphta craquage thermique	Naphta craquage cataly.	Polymère	Alkylat	Essence naturelle				
	Type = Numérique	Type = Numérique	Type = Numérique	Type = Numérique	Type = Numérique	Type = Numérique	Type = Numérique	Type = Numérique			
	Longueur = 14	Longueur = 14	Longueur = 14	Longueur = 14	Longueur = 14	Longueur = 14	Longueur = 14	Longueur = 12			
1	0,00	0,23	0,00	0,00	0,00	0,74	0,03	98,7			
2	0,00	0,10	0,00	0,00	0,12	0,74	0,04	97,8			
3	0,00	0,00	0,00	0,10	0,12	0,74	0,04	96,6			
4	0,00	0,49	0,00	0,00	0,12	0,37	0,02	92,0			
5	0,00	0,00	0,00	0,62	0,12	0,18	0,08	86,6			
6	0,00	0,62	0,00	0,00	0,00	0,37	0,01	91,2			
7	0,17	0,27	0,10	0,38	0,00	0,00	0,08	81,9			
8	0,17	0,19	0,10	0,38	0,02	0,06	0,08	83,1			
9	0,17	0,21	0,10	0,38	0,00	0,06	0,08	82,4			
10	0,17	0,15	0,10	0,38	0,02	0,10	0,08	83,2			
11	0,21	0,36	0,12	0,25	0,00	0,00	0,06	81,4			
12	0,00	0,00	0,00	0,55	0,00	0,37	0,08	88,1			
13	0,00	0,20	0,00	0,40	0,00	0,50	0,06				
14	0,17	0,40	0,10	0,40	0,10	0,10	0,08				

Renseignons la boîte de dialogue comme montré ci-dessous.

Nous ne précisons pas le nombre de composantes à extraire et laissons la validation croisée le déterminer.

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

Régression PLS		×
distil reformat naphtat naphtac polymère alkylat essence octane		Variables à expliquer quantitatives :
		Variables explicatives quantitatives : distil reformat naphtat naphtac polymère alkylat essence
		(Libellés des variables à expliquer :)
~	\square	(Libellés des variables explicatives :)
Nombre de composantes à extraire :		(Libellés des observations :)
Racine aléatoire : 12345 Ok Annuler	Sélectio	n Supprimer Aide

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

Rapports et Graphiques									
Rapport PLS		1	2	3	4	5	6	7	8
PRESS, R2, Q2	1								
R2X, R2Y	2	(C) UNIWIN version 10.2.0							
Coeffs and régression	3								
Coefts non std régression	4	DATE: 29/12/2024							
Poids w	5	ORDINATEUR : LAPTOP-LEG8L077							
Poids w*	6	UTILISATEUR : cchar							
Scores X (t)	7	FICHIER(S) DE DONNEES OUVERT(S) :	OCTANE.SGD						
Poids variables X et Y	8								
Corrélations var comp.	9	RESULTATS DE LA REGRESSION PLS		1					
- octane (apprentissage)	10								
- Distances au modèle en X	11	Sélection :							
T2 de Hoteling	12	Aucune							
octane (validation, prévision)	13								
	14	Jeu d'apprentissage : 12 observation							
	15	15 Jeu de validation : 0 observation							
	16	Jeu de prévision : 2 observations							
	17								
	18	Variables à expliquer :							
	19	octane							
	20								
	21	Variables explicatives :							
	22	distil							
		Rapport Explorateur /							

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

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

Rapports							
Rapport Explorateur							
O Rapport Général							
◯ Rapport Html							
Ok Annuler							

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

Graphiques
Diagramme des R2X cumulés
O Diagramme des R2Y cumulés (PLS2)
O Diagramme des Q2 cumulés
O Diagramme des coefficients standardisés
O Cercle des corrélations (points)
O Cercle des corrélations (points + lignes)
O Plan factoriel des observations (t vs t)
O Plan factoriel des observations (u vs t)
O Plan factoriel des observations (u vs u)
◯ Diagramme des poids des variables X et Y
⊖ Graphique des T2 de Hotelling (PLS1)
⊖ Graphique des distances au modèle en X
⊖ Graphique des distances au modèle en Y
⊖ Graphique des VIP (PLS2)
O Graphique observés vs estimés (apprentissage)
O Graphique des résidus (apprentissage)
◯ Graphique observés vs estimés (validation)
O Graphique des résidus (validation)
Ok Annuler

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

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

Enregistrement des résultats (1/2)							
Enregistrer	Noms attribués aux variables cibles						
Scores X	xscores_1						
Poids variables X	xloads_1						
Scores Y	yscores_1						
Poids variables Y	yloads						
Corrélations variables - composantes	corxyt_1						
Poids calcul scores	rawwgs_1						
Poids modifiés calcul scores	modwgs_1						
Coefficients standardisés régression	stdcoefs						
Coefficients non standardisés régression	regcoefs						
Libellés des observations (apprentissage)	libobsapp						
Ok Plus	Tout Annuler						

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

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML. Voici des exemples du rapport pour notre analyse.

Rapports et Graphiques									
Rapport PLS		1	2	3	4	5	6	7	8
PRESS, R2, Q2	1								
	2	PRESS, RSS, R2, R2 cumulé, Q2, LIMI	TE Q2, Q2 CUMULE						
R2X, R2Y cumules	3								
- Coeffs non std régression	4	PRESS = somme des carrés des err	reurs de prévision						
- Poids w	5	RSS = somme des carrés résiduelle	e						
Poids w*	6	R2 = pourcentage de la somme des	carrés des X expliqué	9					
Scores X (t)	7	Q2 = pourcentage de la variation tot	ale des X et de Y prévu	e					
- Scores f (u)	8	Limite Q2 : seuil de significativité de							
Corrélations var comp.	9	***** Un modèle à 3 composante(s) l		1					
octane (apprentissage)	10								
 Distances au modèle en X 	11								
- Distances au modèle en Y	12		Composante 1	Composante 2	Composante 3				
octane (validation, prévision)	13	PRESS	1,14570	0,69611	0,19347				
	14	RSS	11,00000	0,84047	0,26023				
	15	R2	0,57361	0,15252	0,19213				
	16	R2 cumulé	0,57361	0,72613	0,91826				
	17	02	0,89585	0,17176	0,25653				
	18	Limite Q2	0,09750	0,09750	0,09750				
	19	Q2 cumulé	0,89585	0,91373	0,93586				
	20								
	21								
	22								
		Rapport Explorateur /							

Ce premier tableau est important. Il n'est affiché que si la validation croisée a été mise en œuvre. Il donne les informations suivantes pour chaque composante :

- PRESS (PRediction Error Sum of Squares): somme des carrés des erreurs de prévision
- RSS (Residual Sum of Squares) : somme des carrés résiduelle
- R2 : pourcentage de la somme des carrés des X expliquée

- Q2 : pourcentage de la variation totale des X et de Y prévue
- Limite Q2 : seuil de significativité de la composante égal à (1-0,95²) = 0,0975

L'en-tête de ce tableau indique le nombre optimal de composantes déterminé par la validation croisée, ici 3.

<u>Note</u> : si le nombre de composantes à extraire a été précisé dans la boîte de dialogue initiale et que la validation croisée a été demandée, alors il est conseillé de mettre en œuvre la méthode en précisant le nombre de composantes déterminé par la validation croisée.

Le deuxième tableau affiche les corrélations entre chaque X et le Y avec les composantes extraites. Il décrit le pouvoir explicatif de chaque composante.

Le troisième tableau affiche ces informations cumulées.

Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	~
PRESS, R2, Q2	1									
	2	TABLEAU DES R2X ET R2Y CUMULES								
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 compo	osantes							
Coeffs std régression	4	R2X = corrélation cumulée entre cha	aque X et les composar	ntes						
- Poids w	5	R2Y = corrélation cumulée entre cha	aque Y et les composar	ntes						
Poids w*	6									
Scores X (t)	7									
Scores Y (u) Reide unrichten X at X	8		Composante 1	Composante 2	Composante 3					
Corrélations var comp.	9	(X) distil	0,81769	0,81897	0,91866					
octane (apprentissage)	10	(X) reformat	0,00399	0,71575	0,98584					
 Distances au modèle en X 	11	(X) naphtat	0,82059	0,82166	0,91972					
Distances au modèle en Y T2 de Hatelling	12	(X) naphtac	0,50389	0,55857	0,93889					
octane (validation, prévision)	13	(X) polymère	0,34429	0,34575	0,67395					
	14	(X) alkylat	0,84830	0,98494	0,99980					
	15	(X) essence	0,67649	0,83725	0,99094					
	16	(Y) octane	0,92359	0,97634	0,99056					
	17									_
	18									
	19									
	20									
	21									
	22									~
	< > \	Rapport Explorateur / <								>

Le quatrième tableau affiche les coefficients standardisés de la régression pour le modèle à 3 composantes.

Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	~
PRESS, R2, Q2	1									
R2X, R2Y	2	COEFFICIENTS STANDARDISES DE LA								
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coefts std régression	4									
Coefts non std régression	5									
Poids w*	6		octane							<u> </u>
Scores X (t)	7	distil	-0.13909							
Scores Y (u)	8	reformat	-0 20869							
Poids variables X et Y	9	naphtat	-0 13756							
octane (apprentissage)	10	naphtac	-0 29317							
- Distances au modèle en X	11	polymère	-0.03843							
- Distances au modèle en Y	12	alkylat	0.45639							
- 12 de Hoteling	13	essence	-0 14338							
	14		-0,11000							
	15									
	16									
	17									
	18									
	10									
	20									
	20									
	21									
	< > \	Rapport Explorateur / <								>

Le cinquième tableau affiche ces informations non standardisées c'est-à-dire dans les unités des variables X et de la variable Y.

Rapports et Graphiques										×
	4									
Bapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
	2	COEFFICIENTS NON STANDARDISES D	E LA REGRESSION							
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coefts std régression	4									
- Poids w	5									
Poids w*	6		octane							
Scores X (t)	7	constante	92,67599							
Scores Y (u)	8	distil	-9.82832							
Corrélations var a comp	9	reformat	-6.96018							
- octane (apprentissage)	10	naphtat	-16 66624							
Distances au modèle en X	11	naphtac	-8 42180							
Distances au modèle en Y	12	polymère	-4.38893							
- 12 de Hoteling	13	alkylat	10 16130							
Containe (Validation, previaion)	14	essence	-34 52896							
	15									
	16									
	17									
	18									
	10									
	10									
	20									
	21									
	< > \	Rapport Explorateur / <								>

Les deux tableaux suivants affichent les poids w et w^{*} (w modifiés pour tenir compte des nombres de degrés de liberté) qui sont utilisés pour le calcul des scores des observations.

Rapports et Graphiques										×
	s									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	POIDS POUR LE CALCUL DES SCORES	S (W)							
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coefts std régression	4									
Coeffs non std regression	5									
- Poids w*	6		w1	w2	w3					
- Scores X (t)	7	distil	-0.43700	0.16432	0.28970					
Scores Y (u)	8	reformat	-0.03696	-0.68100	0.45133					
Corrélations var comp.	9	naphtat	-0,43734	0,16885	0,29112					
octane (apprentissage)	10	naphtac	-0,36884	-0,12694	-0,56729					
Distances au modèle en X	11	polymère	0,25772	-0,36294	-0,44472					
Distances au modèle en Y	12	alkylat	0,51412	0,51735	0,10897					
octane (validation, prévision)	13	essence	-0,38680	0,25540	-0,31010					
	14									
	15									
	16									
	17									
	18									
	19									
	20									
	21									
	22	· · · · · · · · · · · · · · · · · · ·								×
	<>/	Rapport Explorateur / <								>

Le tableau des scores X affiche les coordonnées des observations sur les composantes t. Le tableau des scores Y affiche ces coordonnées sur les composantes u. En pratique seule la composante u₁ est interprétable et donc utilisée.

Rapports et Graphiques										
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	SCORES X (t)								
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coeffs non std regression	4									
- Poids w	5									
Poids w*	6		Composante 1	Composante 2	Composante 3					
Scores X (t)	7	01	2,05132	0,82180	1,58205					
	8	02	2,47466	0,64882	0,10940					
Corrélations var comp.	9	03	2,33108	0,92670	-0,17292					
octane (apprentissage)	10	04	2,03715	-1,59574	-0,50154					-
Distances au modèle en X	11	05	-0,06811	-0,21782	-2,95590					
T2 de Hoteling	12	06	1,61381	-1,42276	0,97111					
octane (validation, prévision)	13	07	-2,20425	-0,17812	0,23753					-
	14	08	-1,99355	0,10060	0,11842					-
	15	09	-2,08759	0,14859	0,35461					
	16	010	-1,91577	0,31841	0,19648					
	17	011	-2,07628	-0,46064	1,03118					
	18	012	-0,16247	0,91014	-0,97041					-
	19									
	20									
	21									
	22									~
	< > /	Rapport Explorateur / <								>

Le tableau suivant affiche les poids (loadings) des variables X et Y.

Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	POIDS DES VARIABLES EXPLICATIVE	S X ET A EXPLIQUER Y							
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coeffs std regression	4									
- Poids w	5									
- Poids w*	6		Composante 1	Composante 2	Composante 3					
Scores X (t)	7	(X) distil	-0,45356	-0,04251	0,27297					
Scores Y (u)	8	(X) reformat	0,03169	-1,00323	0,44932					
- Corrélations var comp.	9	(X) naphtat	-0,45436	-0,03900	0,27073					
octane (apprentissage)	10	(X) naphtac	-0,35605	0,27807	-0,53317					_
Distances au modèle en X	11	(X) polymère	0,29431	-0,04544	-0,49530					_
	12	(X) alkylat	0,46197	0,43955	0,10541					
octane (validation, prévision)	13	(X) essence	-0,41254	0,47679	-0,33893					_
	14	(Y) octane	0,48204	0,27311	0,10307					
	15									
	16									
	17									
	18									
	19									
	20									
	21									
	22	Rapport Explorateur /								~ ~

Le tableau suivant affiche les corrélations entre les variables X et Y et les composantes principales.

Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	CORRELATIONS DES VARIABLES X ET	Y AVEC LES COMPOSAN	ITES						
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 compo	osantes							
Coeffs std regression	4									-
- Poids w	5									
Poids w*	6		Composante 1	Composante 2	Composante 3					-
Scores X (t)	7	(X) distil	-0,90426	-0,03575	0,31573					-
Scores Y (u)	8	(X) reformat	0,06317	-0,84366	0,51971					-
Corrélations var comp.	9	(X) naphtat	-0,90586	-0,03279	0,31315					-
octane (apprentissage)	10	(X) naphtac	-0,70985	0,23384	-0,61670					
Distances au modèle en X	11	(X) polymère	0,58676	-0,03821	-0,57289					
	12	(X) alkylat	0,92103	0,36964	0,12192					1
octane (validation, prévision)	13	(X) essence	-0,82249	0,40095	-0,39202					
	14	(Y) octane	0,96104	0,22967	0,11922					1
	15									
	16									
	17									
	18									
	19									
	20									
	21									
	22	l ,								¥
	< > \[Rapport Explorateur / <							>	

Les valeurs observées, estimées par le modèle à 3 composantes ainsi que les résidus sont ensuite affichés.

Rapports et Graphiques										
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	Y OBSERVE, Y ESTIME, RESIDU POUR	LES DONNEES NON STANE	DARDISEES						
	3	Résultats pour le modèle à 3 comp	osantes							
Coefts non std régression	4	Données d'apprentissage								
Poids w	5									
Poids w*	6									
Scores X (t)	7		Y observé	Y estimé	Résidu					
Poids variables X et Y	8	01	98,7	97,55864	1,14136					
Corrélations var comp.	9	02	97,8	97,59151	0,20849					
	10	o3	96,6	97,44534	-0,84534					
Distances au modèle en X	11	04	92,0	91,80793	0,19207					
Distances au modele en 1	12	o5	86,6	85,99452	0,60548					
octane (validation, prévision)	13	06	91,2	91,77507	-0,57507					
	14	07	81,9	81,49670	0,40330					
	15	08	83,1	82,57542	0,52458					
	16	09	82,4	82,52399	-0,12399					
	17	010	83,2	83,26027	-0,06027					
	18	o11	81,4	81,92924	-0,52924					
	19	012	88,1	89,04136	-0,94136					
	20									
	21									
	22									~
	< > /	Rapport Explorateur / <								>

Les deux tableaux suivants affichent les distances des observations au modèle en X et au modèle en Y. Plus la distance est grande, moins l'observation caractérisée par ses valeurs en X et en Y est bien reconstituée par le modèle.

Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	DISTANCES AU MODELE EN X (DONNE	ES STANDARDISEES)							
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coeffs std regression	4									
- Poids w	5									
Poids w*	6		Distance							
Scores X (t)	7	01	0,50340							
Scores Y (u)	8	02	0,83906							
Corrélations var comp.	9	03	0,70064							
octane (apprentissage)	10	04	0,30646							
Distances au modèle en X	11	05	0,02565							
Distances au modele en Y	12	o6	1,13914							
octane (validation, prévision)	13	07	0,02044							
	14	08	0,09945							
	15	09	0,00573							
	16	o10	0,12324							
	17	011	0,25436							
	18	012	2,27659							
	19									
	20									
	21									
	22									~
	< > \	Rapport Explorateur / <								>

Le tableau suivant affiche les T2 de Hotelling. Il permet de détecter d'éventuelles observations atypiques.

Rapports et Graphiques										
	4									
Rapport PLS		1	2	3	4	5	6	7	8	~
PRESS, R2, Q2	1									
R2X, R2Y	2	T2 DE HOTELLING								
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coeffs std regression	4									
- Poids w	5									
Poids w*	6		Composante 1	Composante 2	Composante 3					
Scores X (t)	7	Limite à 95%	4,74968	8,68913	13,35085					
Scores Y (u)	8	Limite à 99%	9,33408	15,72636	23,59279					-
Corrélations var comp.	9	01	1,15487	2,19667	4,23754					
octane (apprentissage)	10	o2	1,68073	2,33012	2,33988					+
Distances au modèle en X	11	03	1,49136	2,81613	2,84051					+
Distances au modele en Y	12	04	1,13898	5,06705	5,27216					+
octane (validation, prévision)	13	05	0,00127	0,07446	7,19900					
	14	06	0,71478	3,83740	4,60638					-
	15	07	1,33349	1,38243	1,42844					
	16	08	1,09074	1,10635	1,11779					
	17	09	1,19608	1,23014	1,33267					
	18	o10	1,00729	1,16369	1,19517					
	19	011	1,18315	1,51047	2,37753					
	20	012	0,00724	1,28509	2,05295					
	21									
	22	· · · · · · · · · · · · · · · · · · ·								~
	< > \	Rapport Explorateur / <								>

Le dernier tableau affiche les valeurs observées, estimées et les résidus pour les données de validation et de prévision.

Rapports et Graphiques										×
	3									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	Y OBSERVE, Y ESTIME, RESIDU POUR	LES DONNEES NON STAN	DARDISEES						
R2X, R2Y cumulés	3	Résultats pour le modèle à 3 comp	osantes							
Coeffs std regression	4	Données de validation (V) et de pré	vision (P)							
- Poids w	5									
Poids w*	6									
- Scores X (t)	7		Y observé	Y estimé	Résidu					
- Scores Y (u)	8	o13 (P)		90,92415						
Comélations var - comp	9	o14 (P)		81.00068						
- octane (apprentissage)	10			,						_
Distances au modèle en X	11									
Distances au modèle en Y	12									
12 de Hoteling	13									
	14									
	15									
	16									
	17	-								
	18									
	19	-								
	20									
	21									
	22									~
	< > \	Rapport Explorateur / <								>

L'option Graphiques

• Diagrammes des R2X cumulés, des R2Y cumulés et des Q2 cumulés

Le diagramme des R2Y cumulés n'est proposé que pour si une régression PLS2 a été mise en œuvre.

Le diagramme des Q2 cumulés n'est proposé que si la validation croisée a été mise en œuvre.





• Diagramme des coefficients standardisés

Une boîte de dialogue permet de choisir la variable Y (ici une unique variable octane) et de préciser si les libellés des variables X doivent être affichés.

Options pour les g	raphiques PL	S				×
Axe horizontal	Axe vertical		Variable à expliquer			
1 ^ 2 3 4 5 6 7 v	1 3 4 5 6 7	^ ~	octane			^
Réutiliser les titres	Oui	Non				~
Libellés des variables	Oui	◯ Non	Times New Ron Nor	mal 12	Police	Couleur
Libellés des observatio	o Oui	Non	Times New Ron Nor	mal 12	Police	Couleur
Défaut		Ok	Séle	ction	Annuler	

💾 Grap	hique											- • •
Option	ns I	Libellés	Zoom	Imprimer	Copier	Enregistrer						
							Diagramme des	coefficients de rég Modèle à 3 c	ression standardis	sés pour octane		I
	0,40										Valkylat	
	0,27											
oefficients	0,08											
Ŭ	0.11									•polymère		
	0,11			•	distil			naphtat				•essence
							ereformat					
	0,29	L							maphtac			
-								Variables e	explicatives			

• Cercle des corrélations (points ou points + lignes)

Une boîte de dialogue permet d'indiquer les composantes PLS (axes) à représenter et de préciser si les libellés des variables X et Y doivent être affichés.

La barre d'outils affichée au-dessus du graphique permet de modifier le choix des axes et de localiser des variables par libellés ou lignes (numéros).

Cela est utile si le graphique a été affiché sans les libellés des variables dans le cas où ces variables sont nombreuses.

Options pour les gr	aphiques PL	S		×
Axe horizontal	Axe vertical		Variable à expliquer	
1 2 3	1 2 3	~	octane	^
Réutiliser les titres	Oui	Non		~
Libellés des variables	Oui	Non	Times New Ron Normal 12 Police Cou	lleur
Libellés des observation	Oui	Non	Times New Ron Normal 12 Police Course	lleur
Défaut		Ok	Sélection Annuler	



• Plan factoriel des observations (t vs t) - espace des X

Une boîte de dialogue permet d'indiquer les composantes PLS (axes) à représenter et de préciser si les libellés des variables X et Y doivent être affichés.

Options pour les g	raphiques PLS		×
Axe horizontal	Axe vertical	Variable à expliquer	
1 2 3	1 2 3	octane	^
Réutiliser les titres	Oui Non		~
Libellés des variables	Oui Non	Times New Ron Normal 12	Police Couleur
Libellés des observatio	Oui ONon	Times New Ron Normal 12	Police Couleur
Défaut	Ok	Sélection	Annuler



• Plan factoriel des observations (u vs t) - liaison entre l'espace des X et celui des Y



Ce graphique montre la forte corrélation R2Y entre Y et la première composante.

- 💾 Graphique Options Libe Copier Enregistrer Axe X : 1 v C Axe Y : 2 v C Tracé des ellipses Libellé : V 🏘 Ligne : 🔸 ~ # Plan factoriel des observations (u vs u) 2,06 1,13 Composante u2 **•**o2 ***8**810 03 0,19 -0,74 05 006 -1,68 Composante u1 -2.29 -0.91 1,84 3.22
- Plan factoriel des observations (u vs u) espace des Y



• Diagramme des poids des variables X et Y

• Graphique des T2 de Hotelling



• Graphiques des distances au modèle en X et Y





• Graphique observés vs estimés (apprentissage)



• Graphique des résidus (apprentissage)



• Graphique observés vs estimés (validation)

Ce graphique n'est disponible que s'il y a des données de validation, ce qui n'est pas le cas dans cet exemple.

• Graphique des résidus (validation)

Ce graphique n'est disponible que s'il y a des données de validation, ce qui n'est pas le cas dans cet exemple.

Exemple 2 : Fichier Octane2 (PLS1)

Ce deuxième exemple illustre l'étalonnage multidimensionnel sur un exemple de données de spectroscopie.

Il montre l'intérêt de la régression PLS qui permet de prendre en compte un grand nombre de variables explicatives sur un petit nombre d'échantillons.

La variable Y représente l'indice d'octane et les 225 variables des valeurs d'absorbance à différentes longueurs d'onde.

Le modèle est construit à partir d'un ensemble de 39 échantillons (26 d'étalonnage et 13 de validation) d'essence recueillis sur une période suffisamment longue pour être représentatif de la dispersion de l'ensemble de la production.



Renseignons la boîte de dialogue de la régression PLS comme montré ci-après en sélectionnant OCTANE comme variable à expliquer, les colonnes V1100 à V1550 comme variables explicatives et la colonne IDENTIFIANT comme libellés des observations.

Demandons 3 composantes et la validation croisée et sélectionnons les échantillons d'étalonnage (TYPE débute par E) en cliquant sur le bouton 'Sélection' :

P Régression PLS		×
V1490 V1492		Variables à expliquer quantitatives :
V1494 V1496 V1498 V1500		
V1502 V1504 V1506 V1508		×
V1500 V1510 V1512		Variables explicatives quantitatives :
V 1512 V1514 V1516 V1518 V1520 V1522 V1522 V1524 V1526 V1528 V1530 V1532 V1532 V1534		V1100 V1102 V1104 V1106 V1108 V1110 V1112 V1116 V1118 V1120 V1122 V1122
V 1536 V1538 V1540 V1542 V1544		(Libellés des variables à expliquer :)
V1546 V1548 V1550 ALCOOL		(Libellés des variables explicatives :)
Nombre de composantes à extraire : 3		(Libellés des observations :) IDENTIFIANT
Racine aléatoire : 12345		
Ok Annuler	Sélectio	on Supprimer Aide

Definitio	on de la selection				×
Et	TYPE		débute	E	
Liaison	Variable		Relation	Valeur ou variable	
Et Et non Ou Ou non	ALCOOL IDENTIFIANT OCTANE TYPE V1100 V1102 V1104	~	= <> < = > = débute	ALCOOL IDENTIFIANT OCTANE TYPE V1100 V1102 V1104	~
	Ok	Annuler	A	jouter	Aide

Après quelques instants, le rapport s'affiche.

W Rapports et Graphiques										×
	4									
Rapport PLS		1	2	3	4	5	6	7	8	^
PRESS, R2, Q2	1									
R2X, R2Y	2	PRESS, RSS, R2, R2 cumulé, Q2, LIMITE Q2, Q2 CUMULE								
R2X, R2Y cumulés	3									
Coeffs std régression	4	PRESS = somme des carrés des erreurs de prévision								
- Poids w	5	RSS = somme des carrés résiduel	e							
Poids w*	6	R2 = pourcentage de la somme des	a carrés des X expliquée	,						
Scores X (t)	7	Q2 = pourcentage de la variation totale des X et de Y prévue								
	8	Limite Q2 : seuil de significativité de la composante : Q2 >= 0,0975 = (1-0,95*2)								
Corrélations var comp.	9	**** Un modèle à 3 composante(s) PLS semble adéquat.								
OCTANE (apprentissage)	10									
Distances au modèle en X	11									
Distances au modèle en Y	12		Composante 1	Composante 2	Composante 3					
OCTANE (validation, prévision	13	PRESS	7,81316	1,59065	0,56516					
	14	RSS	25,00000	2,71585	1,66104					
	15	R2	0,30589	0,43650	0,14861					
	16	R2 cumulé	0,30589	0,74239	0,89099					
	17	02	0,68747	0,41431	0,65976					
	18	Limite Q2	0,09750	0,09750	0,09750					
	19	Q2 cumulé	0,68747	0,81696	0,93772					
	20									
	21									
	22	· · · · · · · · · · · · · · · · · · ·								~
< >	< > (Rapport Explorateur / <								>

Affichons le plan factoriel des observations (t1 vs t2) et demandons le tracé des ellipses de Hotelling à 95% et 99%.



Il y a trois niveaux d'octane.

Les identifiants débutant par 'L' indique un niveau bas, ceux débutant par 'M' un niveau moyen et ceux débutant par 'H' un niveau élevé.

Dans ce graphique, les trois niveaux d'octane sont bien séparés et l'échantillon 'H59' avec addition d'alcool y apparaît comme un point atypique.

L'échantillon 'M52' (niveau moyen) également avec addition d'alcool est proche du groupe 'H' (niveau élevé).

L'addition d'alcool a pour effet d'augmenter l'indice d'octane.

Le plan t2 vs t3 montre bien que ces deux échantillons avec addition d'alcool sont différents des autres échantillons.



Le graphique observés vs estimés (apprentissage) confirme la qualité du modèle.

Le diagramme des coefficients de régression standardisés permet de visualiser les longueurs d'onde les plus utiles pour prévoir l'indice d'octane. Il est possible en cliquant sur un point dans ce graphique d'afficher la longueur d'onde correspondante. Le bouton Zoom dans la barre d'outils permet de zoomer de diverses façons dans le graphique.







Les deux longueurs d'onde 1214 et 1366 sont donc les deux variables principales pour la prévision de l'indice d'octane.

Le graphique observés vs estimés (validation et prévision) affiche les données des 13 spectres de validation.





Ce troisième exemple illustre l'usage de la régression PLS en analyse conjointe. Il s'agit de relier les préférences de consommateurs aux caractéristiques des produits évalués.

Des scénarios sont construits représentant des tasses de thé hypothétiques caractérisées par quatre facteurs :

1. Température	(1 = chaud, 2 = tiède, 3 = froid)
----------------	-----------------------------------

- 2. Sucre (1 = sans sucre, 2 = un sucre, 3 = deux sucres)
- 3. Force (1 = fort, 2 = moyen, 3 = léger)
- 4. Citron (1 = avec du citron, 2 = sans citron)

Parmi les 3x3x3x2 = 54 scénarios possibles, on sélectionne 18 combinaisons formant un plan orthogonal (toutes les paires de variables indicatrices de modalités n'appartenant pas aux mêmes facteurs sont non corrélées). On demande ensuite à chacun des 6 juges interrogés nommés J1 à J6 de classer par ordre de préférence les 18 scénarios.

Les variables X sont constituées des variables indicatrices des modalités des facteurs. Elles sont nommées : Chaud, Tiède, Froid, Sucre0, Sucre1, Sucre2, Fort, Moyen, Léger, Citron0 et Citron1.

Les variables Y1 à Y6 sont obtenues en inversant les rangs fournis par les juges de manière à obtenir des corrélations positives entre les classements et les variables indicatrices des caractéristiques préférées.

La régression PLS de Y sur X permet une modélisation des classements à l'aide des caractéristiques des scénarios.

Renseignons la boîte de dialogue de la régression PLS sans préciser le nombre de composantes PLS qui sera ainsi déterminé par la validation croisée.

	Régression PLS			×
Г	Y1		Variables à expliquer quantitatives :	
	Y2 Y3 Y4 Y5 Y6		Y1 Y2 Y3 Y4	
	Chaud Tiède Froid		Y6 V	
	Sucre0 Sucre1		Variables explicatives quantitatives :	
	Sucre2 Fort Moyen Léger Citron0 Citron1 J1 J2 J3 J4 J5 J6 Température Sucre Force Citron		Chaud Tiède Froid Sucre0 Sucre1 Sucre2 Fort Moyen Léger Citron0 Citron1	
	~		(Libellés des variables explicatives :)	
1	lombre de composantes à extraire :]	(Libellés des observations :)	
6	⊇ Validation croisée	\square		
F	łacine aléatoire : 12345			
	Ok Annuler	Sélecti	ion Supprimer Aide	

Après quelques instants, le rapport nous indique qu'un modèle à 3 composantes semble adéquat. Exécutons donc à nouveau la régression PLS en précisant que 3 composantes doivent être extraites.

Affichons le cercle des corrélations pour visualiser les poids des variables X et Y et préciser les caractéristiques importantes pour chacun des juges.

Le tableau des coefficients standardisés de régression et les graphiques associés indiquent également les caractéristiques importantes (en positif ou en négatif) pour chaque juge. Par exemple, Chaud et Tiède pour le juge 1 et Citron0 et Citron1 pour le juge 5.





Le diagramme des poids des variables X et Y pour la composante 1 rassemble les juges 1, 2, 4 et 6.

Ils apprécient le thé chaud ou froid, sans sucre et rejettent le thé tiède ou très sucré.



La deuxième dimension regroupe les juges 3 et 5.

Ils apprécient le thé chaud, fort ou moyen, sans citron. Ils rejettent le thé tiède, léger avec du citron.



Le graphique des VIP pour la dernière composante classe les modalités par ordre d'importance dans la construction de l'ensemble des préférences.

Les modalités essentielles ont un VIP supérieur ou voisin de 1 : Tiède, Chaud, Sucre0, Sucre2, Léger, Citron0 et Citron1.

Les autres modalités exercent une influence moindre sur les préférences.



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.

Variable	Contenu
cortu	Corrélations entre t et u (PLS2)
corxt	Corrélations entre X et t (PLS2)
corxu	Corrélations entre X et u (PLS2)
corxyt	Corrélations entre X et Y (PLS1)
coryt	Corrélations entre Y et t (PLS2)
coryu	Corrélations entre Y et u (PLS2)
distx	Distances au modèle en X
disty	Distances au modèle en Y
modwgs	Poids modifiés pour le calcul des scores
rawwgs	Poids pour le calcul des scores
regcoefs	Coefficients non standardisés de régression
stdcoefs	Coefficients standardisés de régression
T2	T2 de Hotelling (PLS1)
VIP	VIP - Variable Importance in the Projection - (PLS2)
xloads	Poids des variables X
xscores	Scores sur les composantes (espace des X)
yloads	Poids des variables Y
yscores	Scores sur les composantes (espaces des Y)

libobsapp	Libellés des observations (apprentissage)
prevapp	Valeurs estimées des Y (apprentissage)
residapp	Résidus (apprentissage)
libobsvp	Libellés des observations (validation et prévision)
prevvp	Valeurs estimées des Y (validation et prévision)

Références

Documentation du package R 'plsdepot' (2016)

http://www.gastonsanchez.com/plsdepot

Exemple 1 : Octane1 – données de Cornell (1990)

La régression PLS – Théorie et Pratique – Michel Tenenhaus – Editions Technip

Exemple 2 : Octane2

La régression PLS – Théorie et Pratique – Michel Tenenhaus – Editions Technip UOP Guided Wave, Inc. – Esbensen, Schönkopf, Midtgaard (1994)

Exemple 3 : Thé

La régression PLS – Théorie et Pratique – Michel Tenenhaus – Editions Technip