

TUTORIEL

La régression PLS

- **Fichier de données**

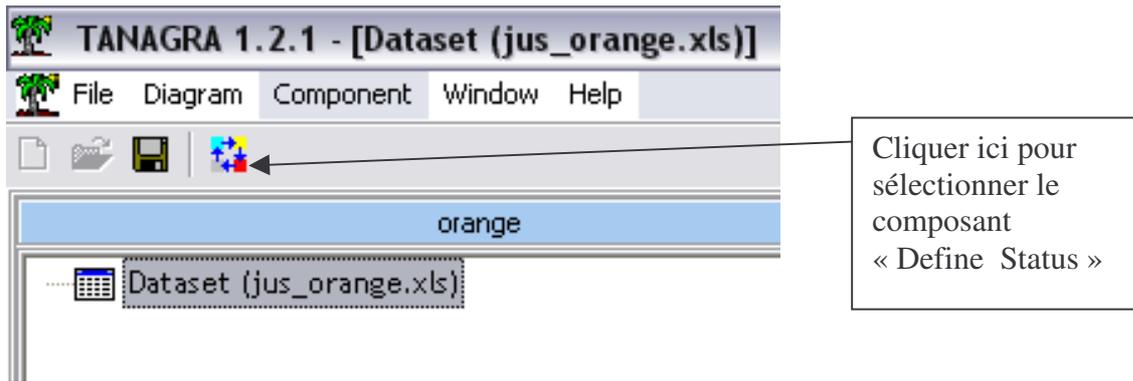
Les données utilisées correspondent à 6 jus d'orange évalués suivant 16 descripteurs physico-chimiques, et notés par 96 juges.

Source : Tenenhaus, M., Pagès, J., Ambroisine L. and & Guinot, C. (2005). PLS methodology for studying relationships between hedonic judgements and product characteristics. *Food Quality and Preference*. **16**, 4, pp 315-325.

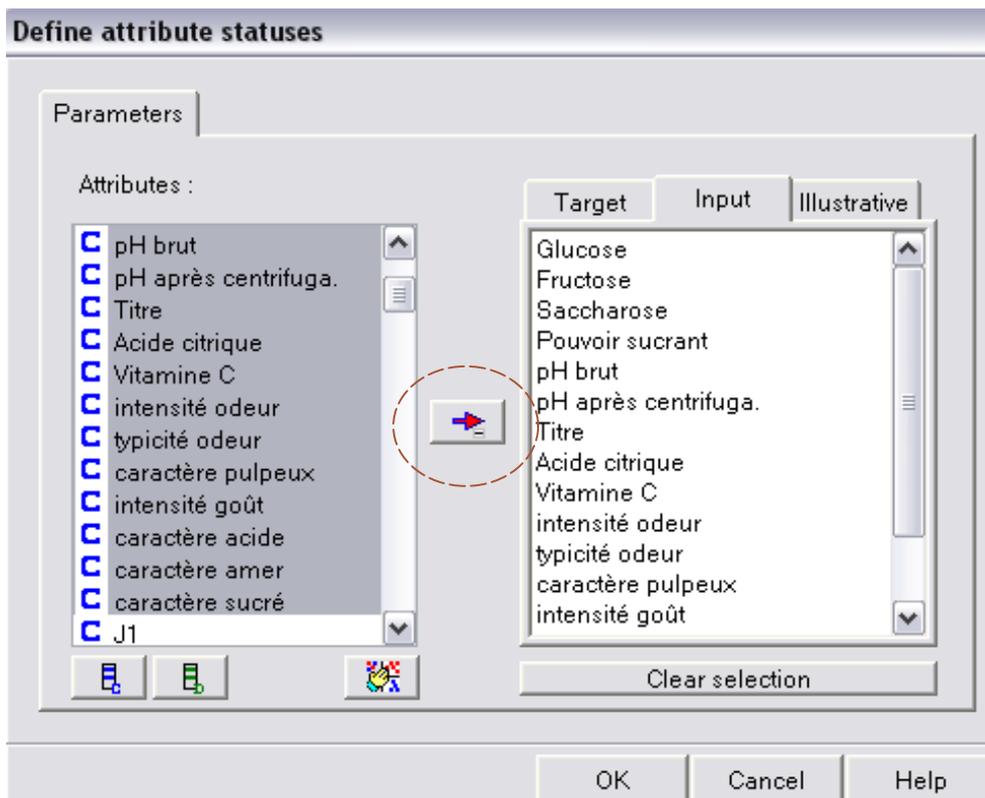
L'utilisation de la régression PLS va permettre de montrer quelles caractéristiques sous-tendent les préférences exprimées, et de prédire les jugements.

- **Mise en application de la méthode dans TANAGRA**

Après avoir chargé le fichier de données « orange.bdm », il faut sélectionner un composant « Define Status » pour permettre de choisir les variables sur lesquelles va porter l'étude.

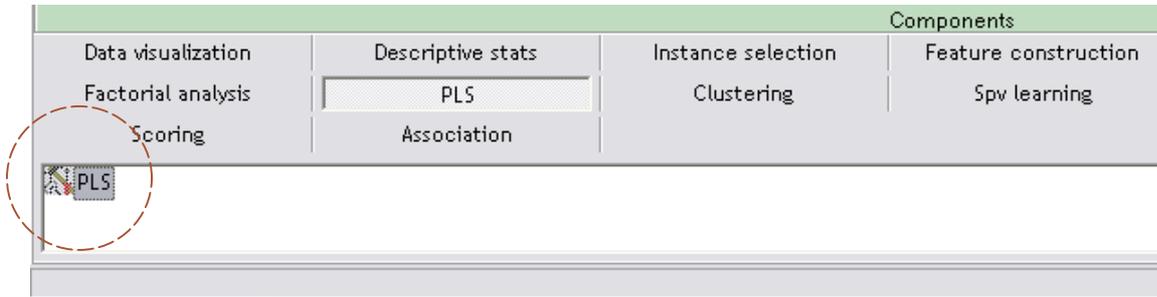


Ensuite la page suivante apparaît :



Sélectionner les variables que vous souhaitez en entrée (Input X : les descripteurs) et en sortie (Output Y : les juges), pour se faire il faut utiliser le bouton central.

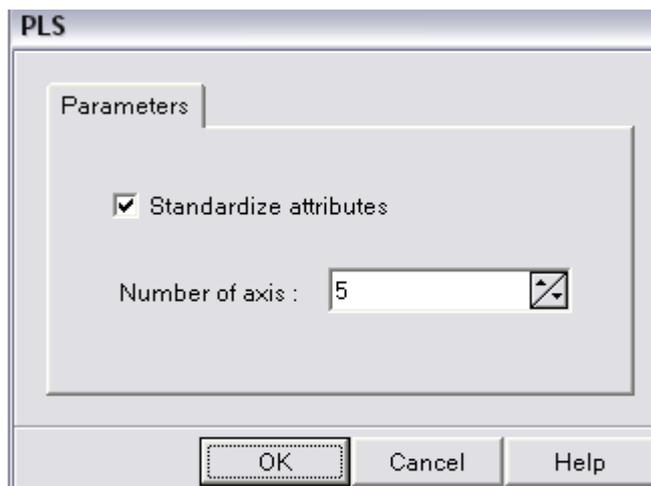
Pour lancer la régression PLS, il faut sélectionner le composant PLS qui se trouve dans le répertoire du même nom :



Puis avec la souris, faites le glisser jusqu'au composant « Define Status » que vous avez créé au départ.

Quand le composant PLS est installé, cliquer droit sur ce dernier et sélectionner « parameters » pour afficher une nouvelle fenêtre.

Ainsi vous pourrez choisir de standardiser les données ou non (division par l'écart type) ainsi que le nombre de composantes que vous souhaitez.



Ensuite vous n'avez qu'à lancer les calculs en double cliquant sur le composant PLS.

La page contenant les résultats s'affiche :

TANAGRA 1.2.1 - [PLS 1]

File Diagram Component Window Help

orange

- Dataset (jus_orange.xls)
 - Define status 1
 - PLS 1

PLS 1

Parameters

PLS parameters

Number of axis	2
Standardize	1

Results

R2 Coefficients

Attribute	Axis_1	Axis_2
Glucose	0,3880	0,4842
Fructose	0,3805	0,4908
Saccharose	0,7741	0,0401
Pouvoir sucrant	0,1729	0,5180
pH brut	0,7654	0,2144
pH après centrifuga.	0,7844	0,1779
Titre	0,7539	0,0995
Acide citrique	0,7571	0,1053
Vitamine C	0,0258	0,0001
intensité odeur	0,2894	0,3999
typicité odeur	0,9436	0,0005
caractère pulpeux	0,7016	0,2248

Components

Data visualization	Descriptive stats	Instance selection	Feature construction
Factorial analysis	PLS	Clustering	Spv learning
Scoring	Association		

PLS

Résultats :

- Les coefficients de la régression
- Les coefficients R^2
- Les coefficients de redondances *100 (%)
- Le coefficient VIP (Variable Importance in the Projection).
- Les matrices W_h , C_h , P_h
 - C_h : Coefficient de régression de la composante h dans la régression de la variable Y_{h-1} sur la composante h .
 - P_h : Coefficient de régression de la composante h dans la régression de la variable X_h sur la composante h .
 - W_h : Coefficient de régression de U_h dans la régression de la variable X_{h-1} sur U_h . U_h représentant le coefficient de régression de C_h dans la régression de la n -ième ligne de la matrice des Y_{h-1} sur C_h .