

# TUTORIAL

## PLS Regression

- **Dataset**

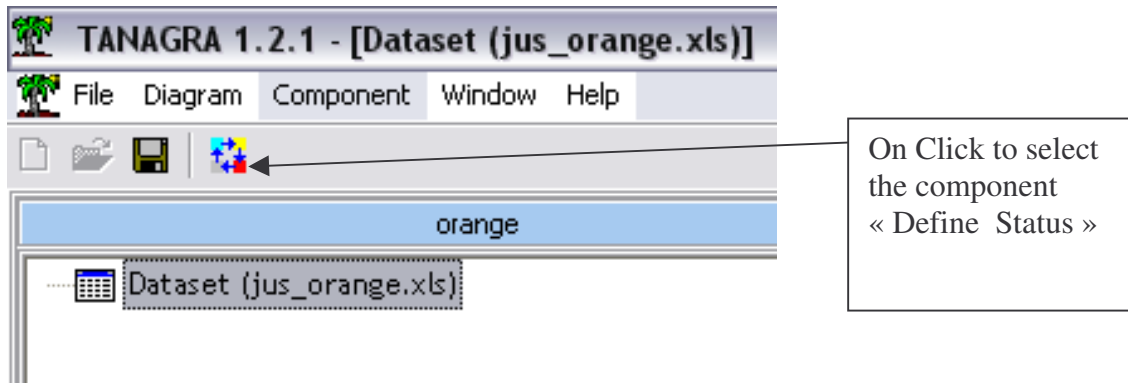
The data used correspond to 6 orange juices described by 16 physico-chemical descriptors and evaluated by 96 judges.

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

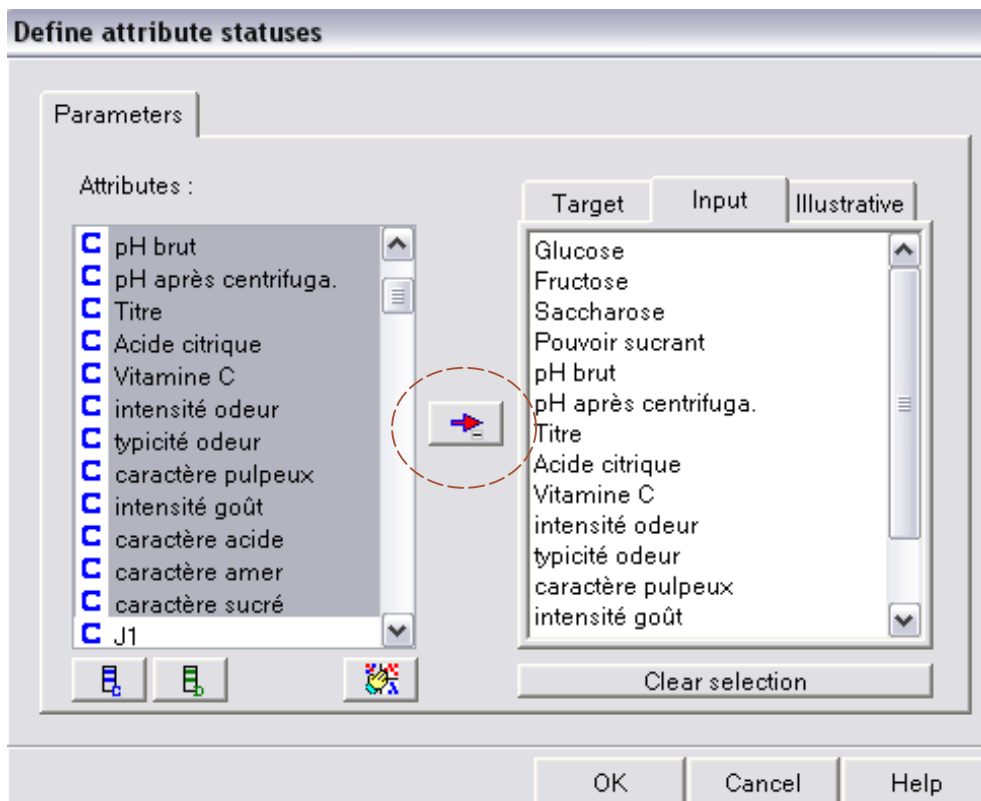
PLS regression is going to allow us to show which descriptors are related to their preferences, and to predict them.

- **PLS Regression Implementation**

After having loaded the dataset (« orange.bdm »), you have to select a component « Define Status » to could chosen the attributes on which we want to make a study.

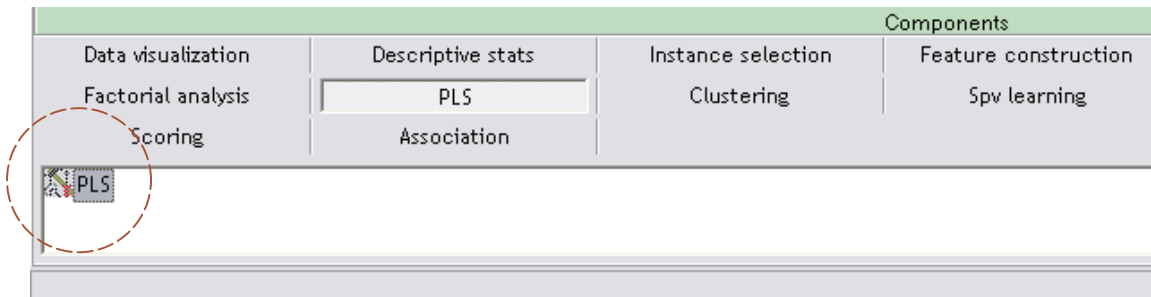


Next :



We have to select Input (X : descriptors) and Output (Y : judges) attributes.

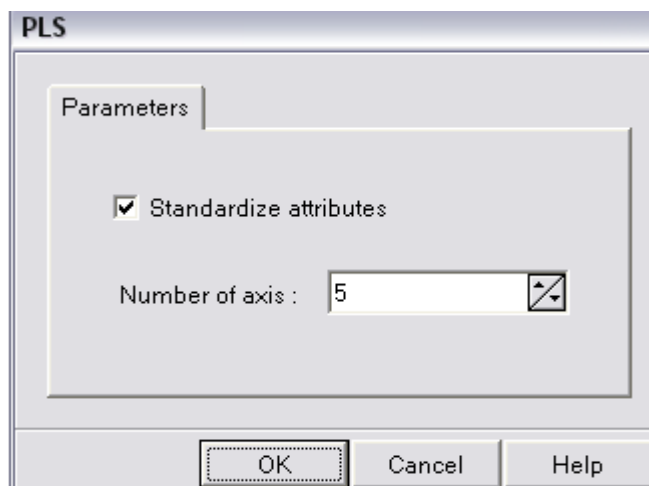
To perform the PLS regression we have to select the PLS component which is in the PLS repertory.



Then we have to move the PLS component to the « Define Status » component.

Then on right click, we can select « parameters » to show a new window.

We can choose to use the standardization and the number of axis wanted.



Then, On double Click you can perform the calculations.

**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

## Results :

- Regression coefficients
- R2 coefficients
- Redundancy coefficients \*100 (%)
- VIP coefficients (Variable Importance in the Projection).
- Wh, Ch, Ph Matrix

- Ch : Regression coefficient of h axis in the regression of Y(h-1) attribute on h axis.

- Ph : Regression coefficient of h axis in the regression of Xh attribute on h axis.

- Wh : Regression coefficient of Uh in the regression of X(h-1) attribute on Uh. Uh being the regression coefficient of Ch in the regression of the nth line of Y(h-1) attribute on Ch.