

## What is Preference Mapping?

[demoPREFMAP.xls](#)

This tutorial was inspired by [Schlich P, McEwan J.A. (1992). *Cartographie des Préférences. Un outil statistique pour l'industrie agro-alimentaire. Sciences des aliments*, 12, pp 339-355] and by exchanges between the Addinsoft development team and some experts of sensory data analysis.

We distinguish two types of Preference Mapping methods:

### **Internal Preference Mapping**

This method corresponds to MDPREF (Multidimensional Analysis of Preference Data) and is based on a PCA (Principal Component Analysis) performed on preference data with as observations the products, and as variables the consumers. The data are ratings given by the consumers for each product. The preference map is the biplot (two or three dimensional) of the observations and the variables.

As the capacity to summarize of the preference map diminishes with the number of consumers (the number of axes to interpret increases), a non-metric PCA is sometimes used to reduce the number of necessary axes. The non-metric PCA consists of a monotone transform of the data so that the variability explained by the  $k$  ( $k=2$  or  $3$ ) first axes is maximized. This transform implies that we consider the ratings have an ordinal meaning, and that the distances or ratios between the ratings are not important. To reduce the number of axes you might also want to group the consumers and to perform the PCA with the groups as variables.

Internal preference mapping allows to generate a map on which one can identify the consumers' or groups of consumers preferences represented as vectors.

### **External Preference Mapping**

This method allows to relate the preferences shown by the consumers to some physico-chemical, sensory or economical characteristics of the products. This approach is essential as it gives a reliable basis to the marketing and R&D teams for adapting or creating products that will correspond to the consumers' expectations.

This method requires an additional table that describes the products with a series of criteria. Contrary to what is done with internal preference mapping, the first step consists in mapping the products on the basis of their characteristics. This can be obtained by running a [Principal Component Analysis](#) (PCA), a Correspondence Analysis (CA) or a [Generalized Procrustes Analysis](#) (GPA). This first visualization is named sensory map. By applying the PREFMAP method, we then model for each consumer (or group of consumers) the ratings that have been

given to the products, using as explanatory variables the characteristics of the products, with the aim of representing the consumers on the sensory map. The full model writes:

$$Y = S_i a_i X_i + S_i b_i X_i^2 + S_{ij} c_{ij} X_i X_j$$

The PREFMAP method uses four sub-models:

- Vector model:  $b_i$  and  $c_{ij}$  are null. The model is a hyperplan. This model allows to display the observations on a sensory map as vectors. The size of the vectors can be related to the  $R^2$  of the model ; in that case the longer the vector, the better the underlying model. The consumer's preference increases the further you go in the direction of the vector. Interpreting the preference can be done by projecting the products on the various vectors (product preference). The drawback of that model is that it neglects that for some criteria such as saltiness or temperature, the preference can increase until an optimal value, and then decrease;
- Circular ideal point model: The  $b_i$  are equal and the  $c_{ij}$  are null. The model corresponds to a hyperquadric hypersurface. If the surface has a maximum in terms of preference we will talk of an ideal point. If this surface has a minimum we will talk of anti-ideal point. With the circular model it is possible to trace isopreference circular lines around the ideal or anti-ideal point.
- Elliptical ideal point model: The  $c_{ij}$  are null. The model corresponds to a hyperquadric hypersurface. In this case the isopreference lines are ellipses, which makes the interpretation of the distances of the products to the ideal or anti-ideal points more complex. If the  $b_i$  have opposite signs, there is no ideal or anti-ideal point but only a saddle point which interpretation is tricky.
- Quadratic surface model: This model corresponds to the complete model which shape is a hypersurface. This model allows to take into account interactions between the characteristics (the  $c_{ij} X_i X_j$ ).

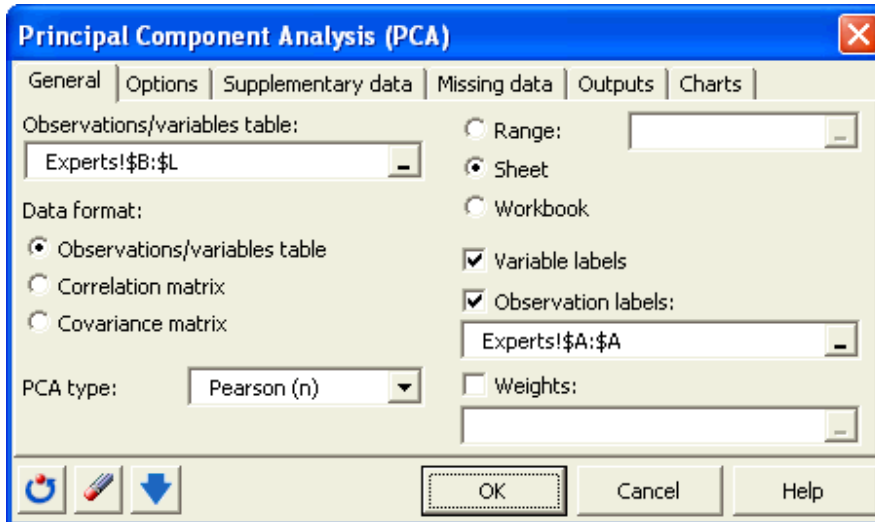
### **How do I run Preference Mapping with XLSTAT-MX?**

The following example shows how one can create a preference map using the PREFMAP method. An Excel sheet with both the data and the results can be downloaded by clicking [here](#). The data used here are composed by:

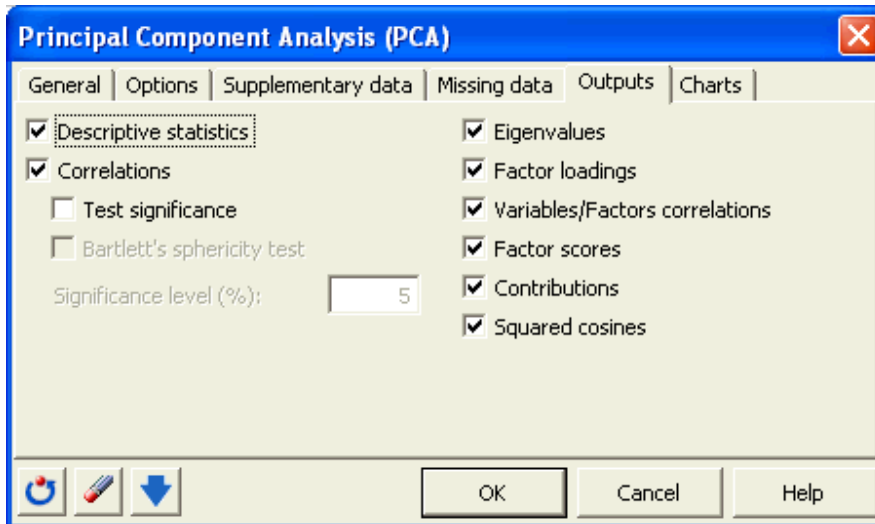
- The consumer acceptability data: 99 consumers rated 10 different commercial samples of potato crisps. These data have been obtained from the article by Schlich and McEwan (1992). The ratings have been discretized from 1 to 30 (30 corresponds to the highest acceptability). These data are stored in a 99 x 10 table.
- The average ratings computed from the ratings given by 8 experts to the 10 crisps samples for 4 texture attributes and 7 flavor attributes. These data, simulated for the purpose of teaching by the author of this tutorial on the basis of the article by Schlich and McEwan (1992), make up a 10 x 11 table.

## Step 1: Creating the sensory map

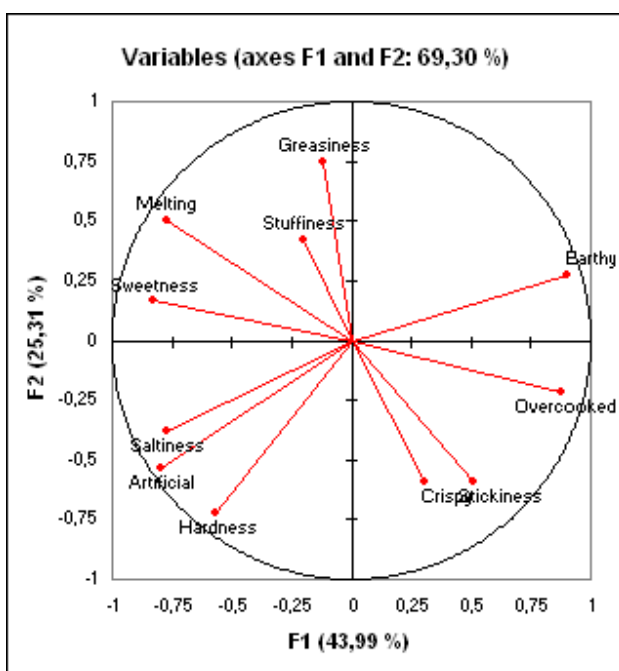
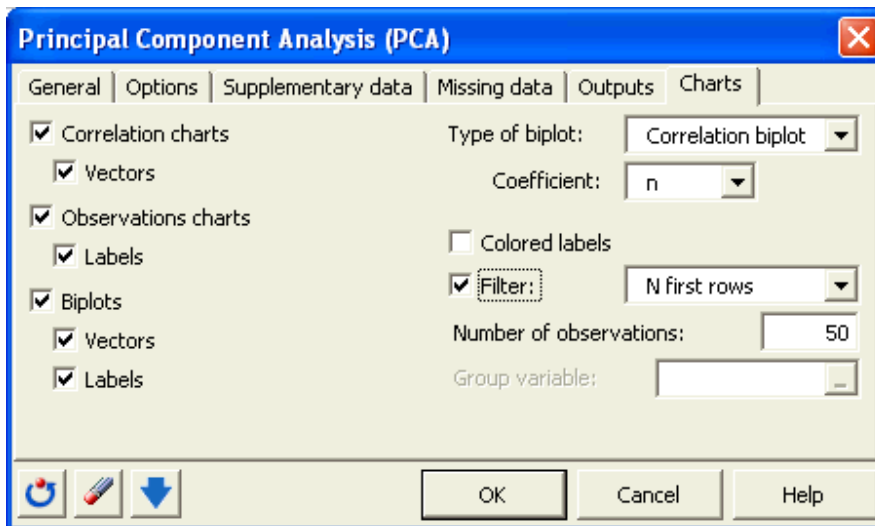
We first create the sensory map by applying a PCA on the 10 x 11 table. This gives us a two-dimensional visualization of the crisps. As a [tutorial](#) is dedicated to PCA, we do not elaborate on that subject here. The dialog box of the PCA has been filled in as shown below.



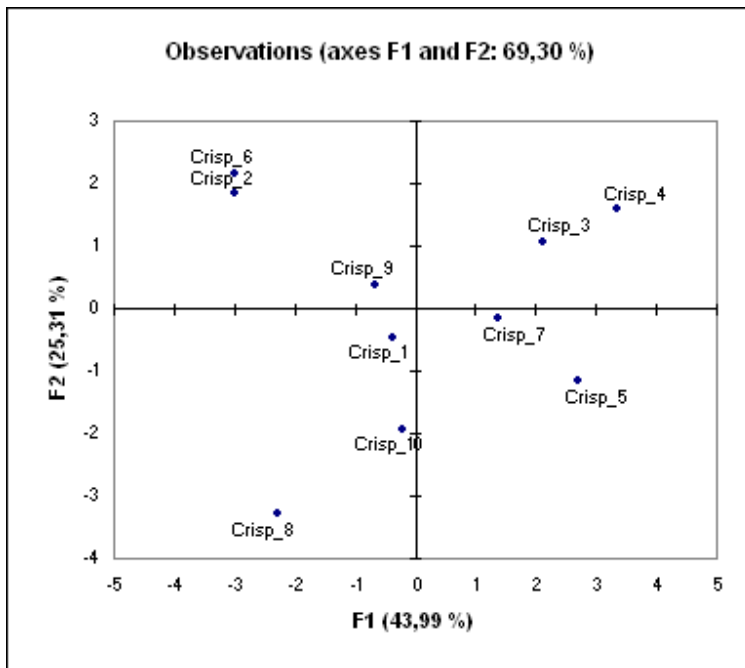
The display options have been set as follows:



Some criteria used by the experts seem to be little redundant given their dispersion on the correlations circle (sweetness is negatively correlated with overcooked, and saltiness is positively correlated with artificial).



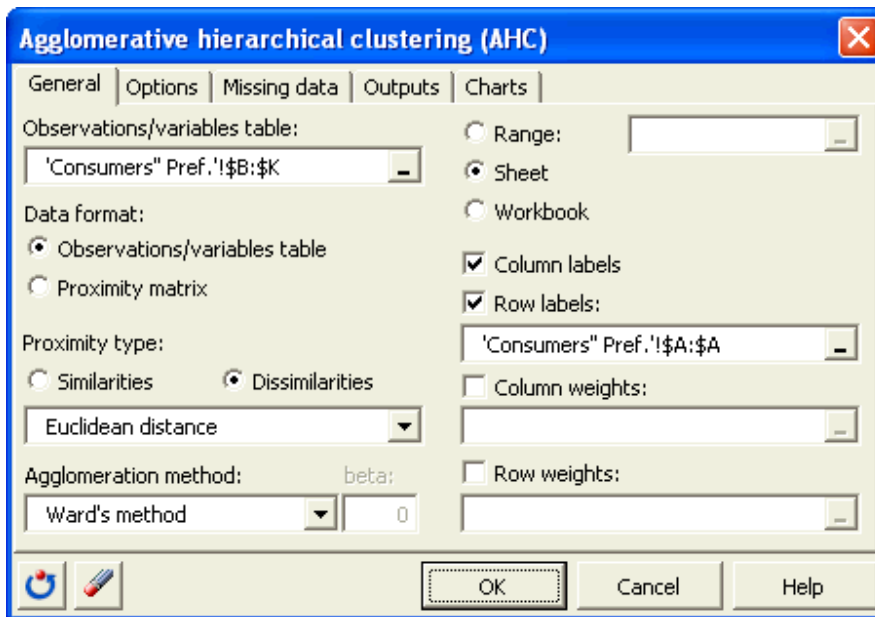
However, the map obtained of the chips, which quality is fine (69.3% of the variability is displayed on the first two dimensions), allows us to notice that the products are well differentiated by the experts.



With the PCA we projected the chips on a 2-dimensional sensory map. Before being able to position the consumers on the map to know which consumer prefers what type of chips, we want to group the consumers into homogeneous clusters in order to obtain a map that is visually satisfactory.

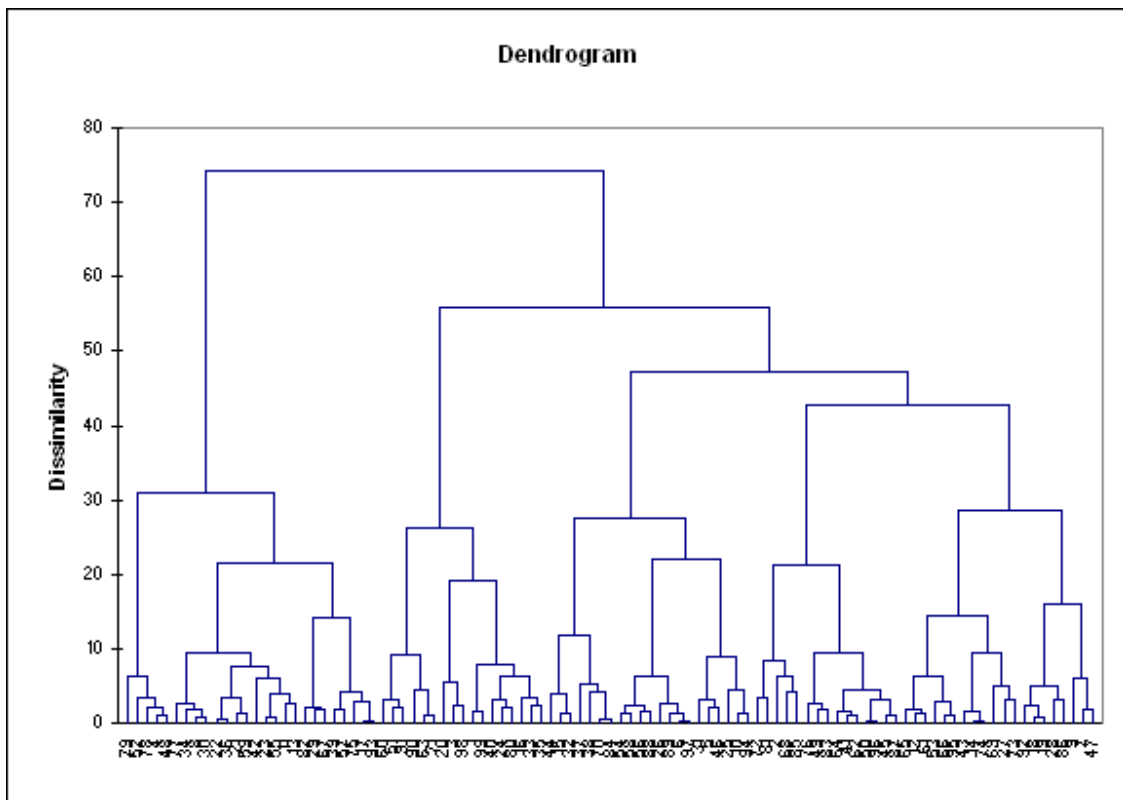
## Step 2. Grouping the consumers

We now focus on the ratings given by the 99 consumers. As the number of consumers is significant, we decided to group them into homogeneous groups in order to make the PREFMAP results easier to interpret. We chose the Agglomerative Hierarchical Clustering (AHC). As a [tutorial](#) is dedicated to AHC, we do not elaborate on that subject here. The dialog box of the AHC has been filled in as shown below.

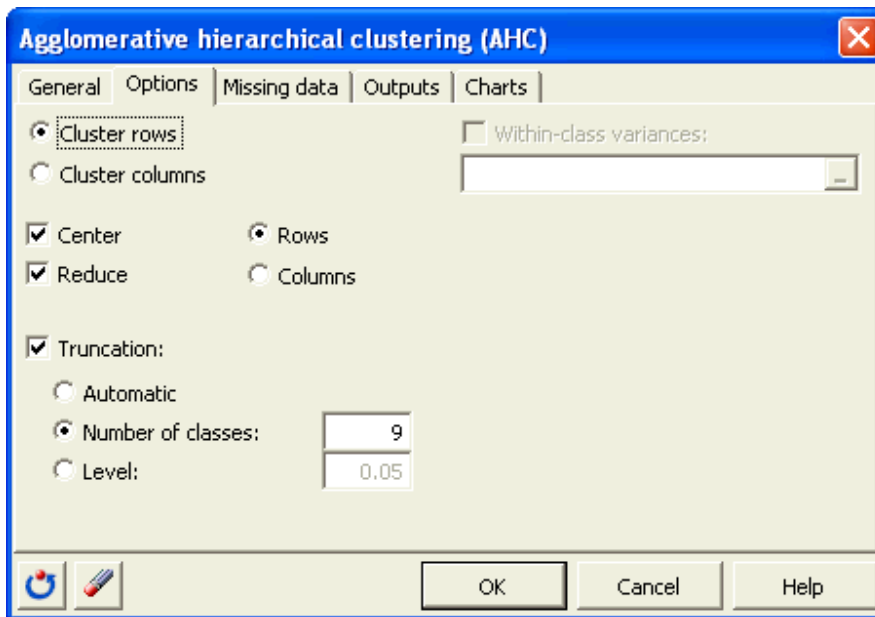


In the Options tab, the "Center / Reduce" option applied to "Rows" has been activated in order to diminish the differences between the judgment scales of the consumers. The truncation option was not activated.

By looking at the dendrogram, it makes sense to decide to work with 9 groups.



We then re-run the AHC while specifying that we request 9 clusters. The dialog box has been filled in as displayed below. The only difference with the previous case is that we ask for a truncation.



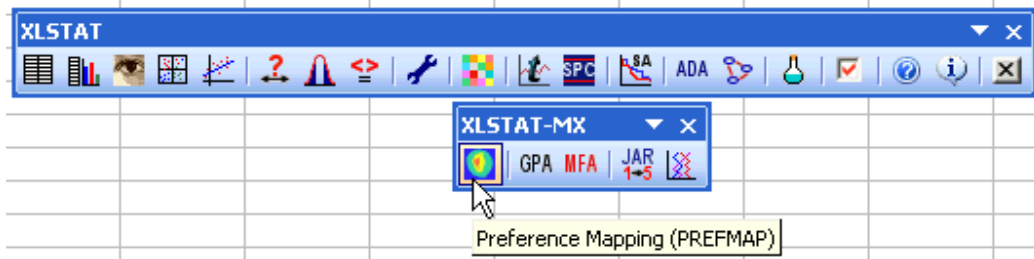
We then store the centroids of the clusters for the last step of the analysis. The table is copied and paste (Edit / Paste special with Transposed option) in a new sheet named "Clusters' pref.".

Class centroids:										
Class	Crisp_1	Crisp_2	Crisp_3	Crisp_4	Crisp_5	Crisp_6	Crisp_7	Crisp_8	Crisp_9	Crisp_10
1	14,091	20,182	5,727	3,545	20,545	22,000	22,545	8,455	21,636	23,545
2	11,000	21,600	6,933	8,867	14,333	25,267	8,533	25,333	3,800	18,133
3	23,000	1,200	7,800	20,800	25,800	3,600	19,200	1,000	17,600	25,200
4	23,929	23,786	2,000	7,143	15,786	13,786	9,071	27,071	22,357	20,429
5	18,125	7,125	21,750	1,750	17,250	10,625	23,125	27,500	10,750	3,500
6	16,190	6,381	8,810	4,286	22,048	5,952	10,095	25,619	20,381	23,762
7	13,333	20,833	3,000	4,167	23,417	8,833	22,750	23,500	7,500	17,333
8	18,857	15,000	2,286	25,000	19,429	14,286	18,571	27,000	25,571	6,143
9	14,667	26,500	25,000	2,500	20,167	26,833	15,833	5,333	13,833	3,667

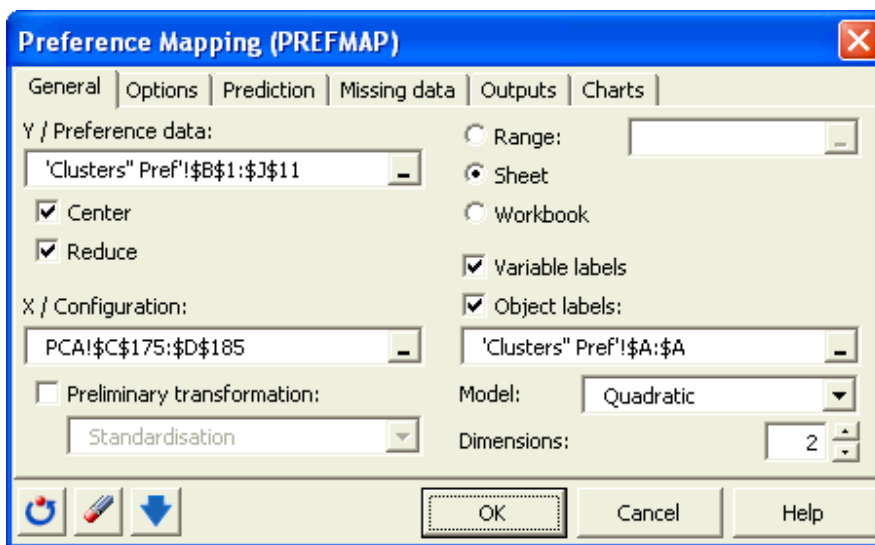
### Step 3: Creating the preference map using the PREFMAP method

In this section we apply the PREFMAP method, using the coordinates of the crisps in the two-dimensional factor space and the ratings given by the consumers summarized by the standardized ratings for the 9 clusters.

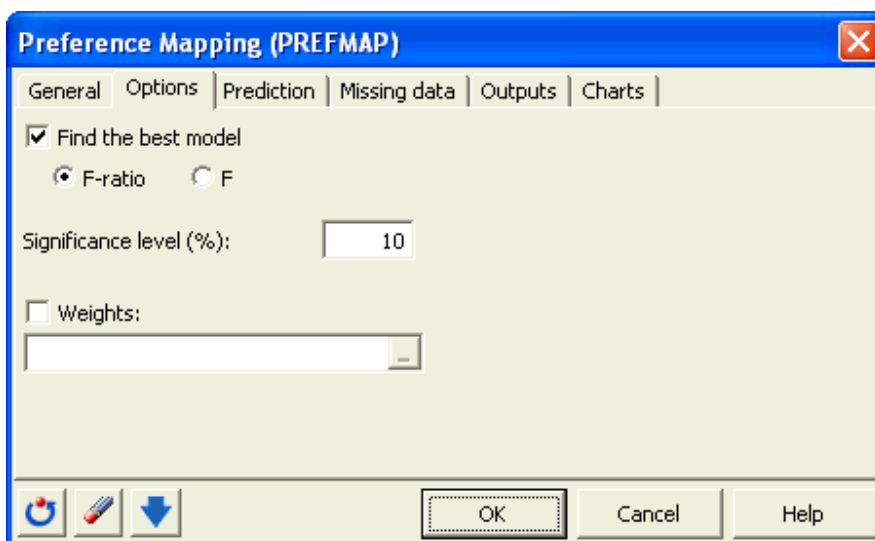
To activate the Preference Mapping dialog box, start XLSTAT, then select the XLSTAT/XLSTAT-MX/Preference Mapping command, or click on the corresponding button of the XLSTAT-MX toolbar (see below).



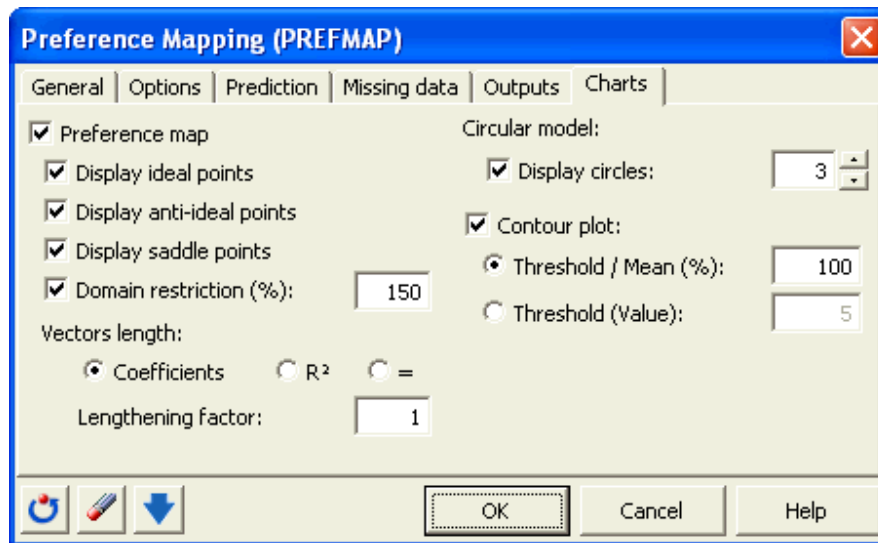
When you click on the button, the Preference Mapping dialog box will appear. Select the data on the Excel sheet. To the "Y / Preference data" correspond the ratings of the 8 clusters. The "X / Configuration" corresponds to the factor scores of the crisps obtained from the PCA. The "preliminary transformation" option allows to directly do a PCA on the raw data. As we did first a PCA, this is not necessary here. We decided to use the up to the quadratic model.



In the "Options tab" to select the best model among the four possible models, we used an F-ratio test, with a 0.1 (10%) significance level. That means that if a more complex model does not give an F-ratio with a p-value lower than 0.1 the more complex model is rejected.



We selected the coefficients of the model to determine the length of the vectors on the preference map.



The results we obtain (see below) show that the vector model is the best model for the clusters 1, 5, 8 and 9. For cluster 4 and 7, we do not obtain an ideal point but a saddle point. The saddle point represents a threshold where the variability of the preference is low, before inscreasing (or decreasing in the opposit direction) quicker. For cluster 2 we have an anti-ideal point. The anti-ideal point corresponds to the lowest preference for the group. For clusters 3 and 6 we have an ideal point. Ideal points correspond to the highest preference in the two-dimensional space for the consumers in that cluster.

Model selection:				
Y	Model	Point type	F1	F2
Cluster1	Quadratic	Saddle	-0.622	0.550
Cluster2	Circular	Anti-ideal	1.214	0.754
Cluster3	Vector	-		
Cluster4	Vector	-		
Cluster5	Vector	-		
Cluster6	Vector	-		
Cluster7	Vector	-		
Cluster8	Vector	-		
Cluster9	Vector	-		

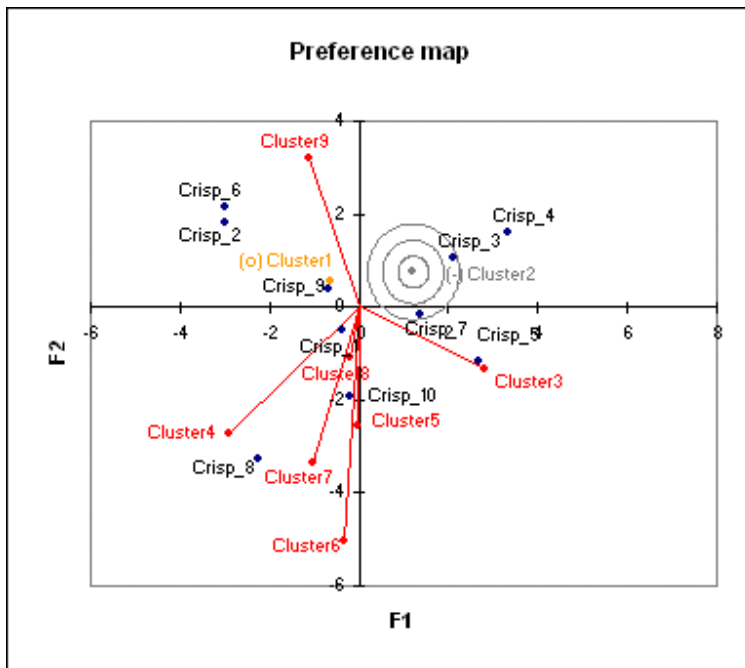
Looking at the analysis of variance table we see that the models are significant only for clusters 2, 3, 4 and 6. For cluster 1, the quadratic model was retained but it is however not significant with a 10% significance level. For cluster 2, the best model is the circular model. It is significant and gives an anti-ideal point. This point corresponds to the worst preference for the consumers of cluster 2.

Analysis of variance:								
Y	DF	Sum of squares	Mean square	R <sup>2</sup>	F	Pr > F	F-ratio	Pr > F
Cluster1	5	7.412	1.482	0.824	3.734	0.113	5.144	<b>0.074</b>
Cluster2	3	7.514	2.505	0.835	10.113	<b>0.009</b>	11.247	<b>0.015</b>
Cluster3	2	4.494	2.247	0.499	3.491	<b>0.089</b>		
Cluster4	2	6.301	3.150	0.700	8.171	<b>0.015</b>		
Cluster5	2	1.917	0.958	0.213	0.947	0.432		
Cluster6	2	7.380	3.690	0.820	15.942	<b>0.002</b>		
Cluster7	2	3.821	1.911	0.425	2.582	0.145		
Cluster8	2	0.366	0.183	0.041	0.148	0.865		
Cluster9	2	3.574	1.787	0.397	2.305	0.170		

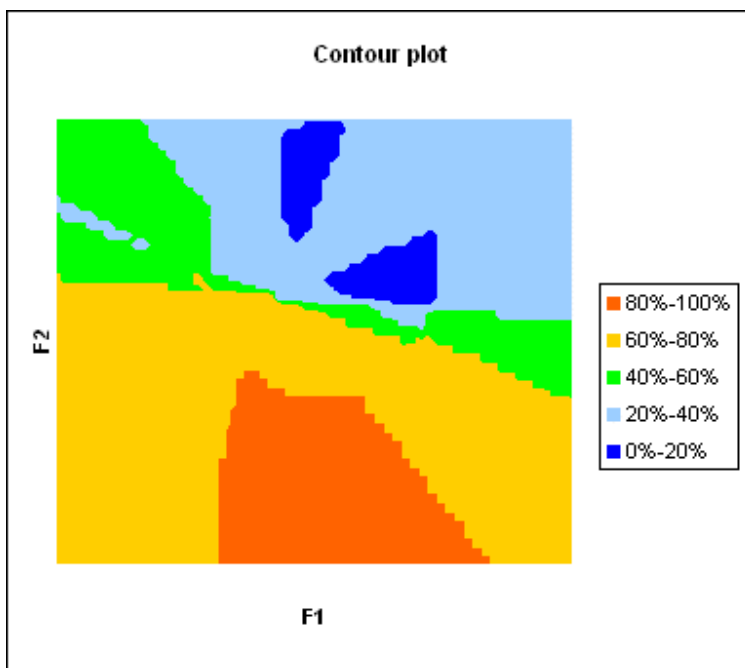
Model coefficients:						
Y	Intercept	F1	F2	F1^2	F2^2	F1*F2
Cluster1	0.328	0.020	-0.177	-0.086	0.031	-0.230
Cluster2	-0.743	-0.237	-0.147	0.097	0.097	0.000
Cluster3	0.000	0.287	-0.136	0.000	0.000	0.000
Cluster4	0.000	-0.294	-0.276	0.000	0.000	0.000
Cluster5	0.000	-0.001	-0.262	0.000	0.000	0.000
Cluster6	0.000	-0.034	-0.513	0.000	0.000	0.000
Cluster7	0.000	-0.105	-0.343	0.000	0.000	0.000
Cluster8	0.000	-0.021	-0.111	0.000	0.000	0.000
Cluster9	0.000	-0.112	0.327	0.000	0.000	0.000

The preference map allows to quickly interpret the results. When looking at both the preference map and the correlations circle of the PCA, we see that the consumers of cluster 3 prefer chips that are not melting, but not too overcooked and crispy. Consumers from cluster 6 like crispy chips and dislike greasy chips. For cluster 2 it is hard to know what they prefer, but it seems that they strongly dislike (anti-ideal point) a chip that would be salty and average for all the criteria (the point is close to the origin).

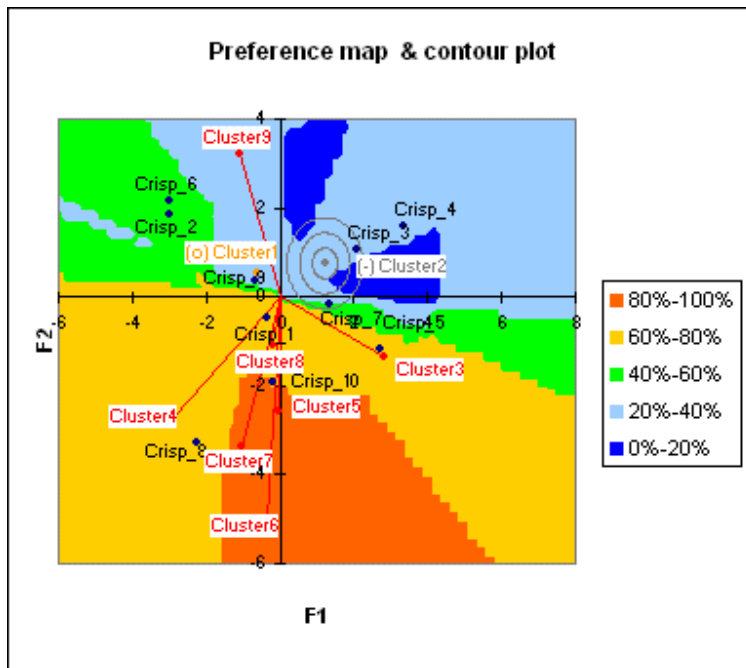
Note: If an ideal/anti-ideal/saddle point is not displayed because out of the plotting area, to extend the mapping area, you can increase the value of the "Domain restriction" in the "Charts" tab of the dialog box.



The contour plot allows to see how many clusters have a preference above average in a given region of the preference map (this is determined using all the fitted models).



Both maps can be superimposed. To do that, copy the preference map, then paste it over the contour plot. Then right click the preference map, change the Format/Patterns/Area and Border color to none. Then drag the preference map over the contour plot, and change the size until the border of the plot area matches the limits of the contour plot.



The preference orders for the various groups of consumers are displayed.

Objects sorted by increasing preference order:

Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9
Crisp_4	Crisp_7	Crisp_6	Crisp_4	Crisp_6	Crisp_6	Crisp_4	Crisp_4	Crisp_8
Crisp_8	Crisp_3	Crisp_2	Crisp_3	Crisp_2	Crisp_4	Crisp_3	Crisp_6	Crisp_5
Crisp_3	Crisp_9	Crisp_9	Crisp_5	Crisp_4	Crisp_2	Crisp_6	Crisp_3	Crisp_10
Crisp_9	Crisp_1	Crisp_8	Crisp_7	Crisp_3	Crisp_3	Crisp_2	Crisp_2	Crisp_7
Crisp_7	Crisp_4	Crisp_1	Crisp_9	Crisp_9	Crisp_9	Crisp_7	Crisp_9	Crisp_1
Crisp_1	Crisp_5	Crisp_10	Crisp_1	Crisp_7	Crisp_7	Crisp_9	Crisp_7	Crisp_3
Crisp_2	Crisp_10	Crisp_7	Crisp_6	Crisp_1	Crisp_1	Crisp_5	Crisp_1	Crisp_4
Crisp_10	Crisp_2	Crisp_3	Crisp_2	Crisp_5	Crisp_5	Crisp_1	Crisp_5	Crisp_9
Crisp_5	Crisp_6	Crisp_4	Crisp_10	Crisp_10	Crisp_10	Crisp_10	Crisp_10	Crisp_2
Crisp_6	Crisp_8	Crisp_5	Crisp_8	Crisp_8	Crisp_8	Crisp_8	Crisp_8	Crisp_6

We notice that the crisp 4, characterized by an earthy, not sweet and not salty taste, is not liked at all by clusters 1, 4, 6, 7 and 8. Chips 8 is the preferred one for most clusters, except by cluster 9. The marketing and R&D teams will be able to take this information into account to direct their creation of new crisps towards the right directions.

Click [here](#) for other tutorials.