

ConsumerCheck: PCA and preference mapping

Kuznetsova Alexandra

Statistics and Data analysis Section, DTU compute

Overview

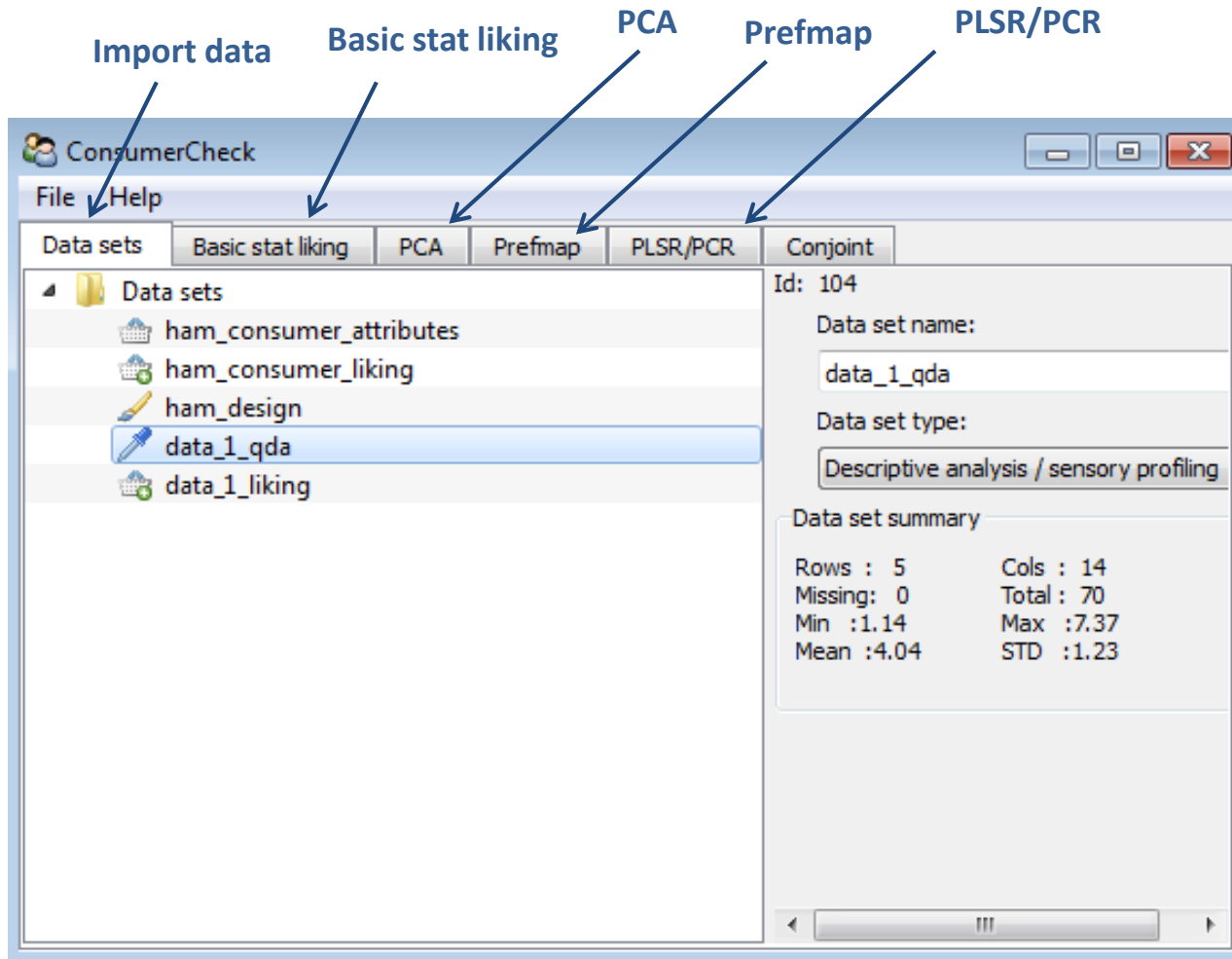
- **ConsumerCheck** software
 - Overview
 - Data sets for PCA and preference mapping
 - PCA
 - Preference mapping
 - Example (apple data)

ConsumerCheck



- Standalone software dedicated for analysis of consumer data
- **PanelCheck**-like software
 - easy-to-use
 - Flexible
 - dedicated for sensory practitioners
- Visualize and analyze your data fast and efficient!
- Classical and advanced statistical methods:
 - Basic statistics (plots/tables)
 - **PCA**
 - **Preference mapping (PLSR, PCR)**
 - Conjoint analysis (mixed effects models)

ConsumerCheck GUI





Example: Apple data

5 apples

108 consumers

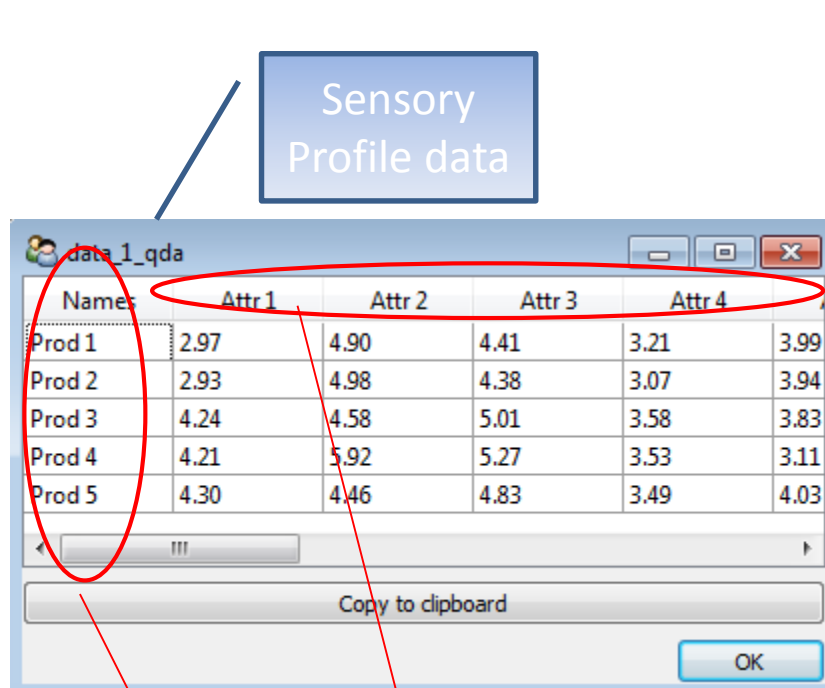
Consumer liking data

Assessors scored
the same 5 apples
according to 14
attributes

Sensory profiling data

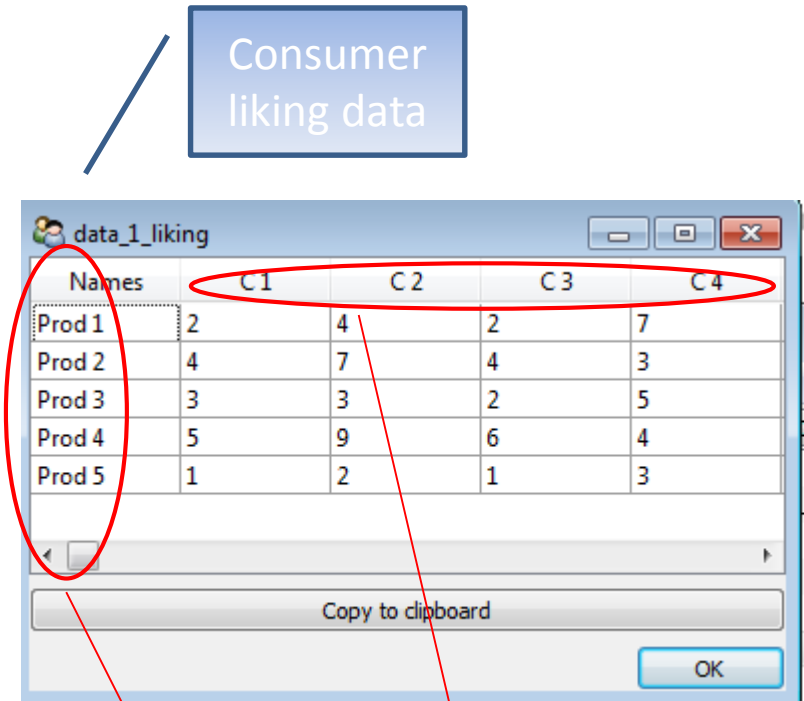
PCA / Prefmap data sets

Sensory Profile data



Names	Attr 1	Attr 2	Attr 3	Attr 4	
Prod 1	2.97	4.90	4.41	3.21	3.99
Prod 2	2.93	4.98	4.38	3.07	3.94
Prod 3	4.24	4.58	5.01	3.58	3.83
Prod 4	4.21	5.92	5.27	3.53	3.11
Prod 5	4.30	4.46	4.83	3.49	4.03

Consumer liking data



Names	C1	C2	C3	C4
Prod 1	2	4	2	7
Prod 2	4	7	4	3
Prod 3	3	3	2	5
Prod 4	5	9	6	4
Prod 5	1	2	1	3

Products

Attributes

Product

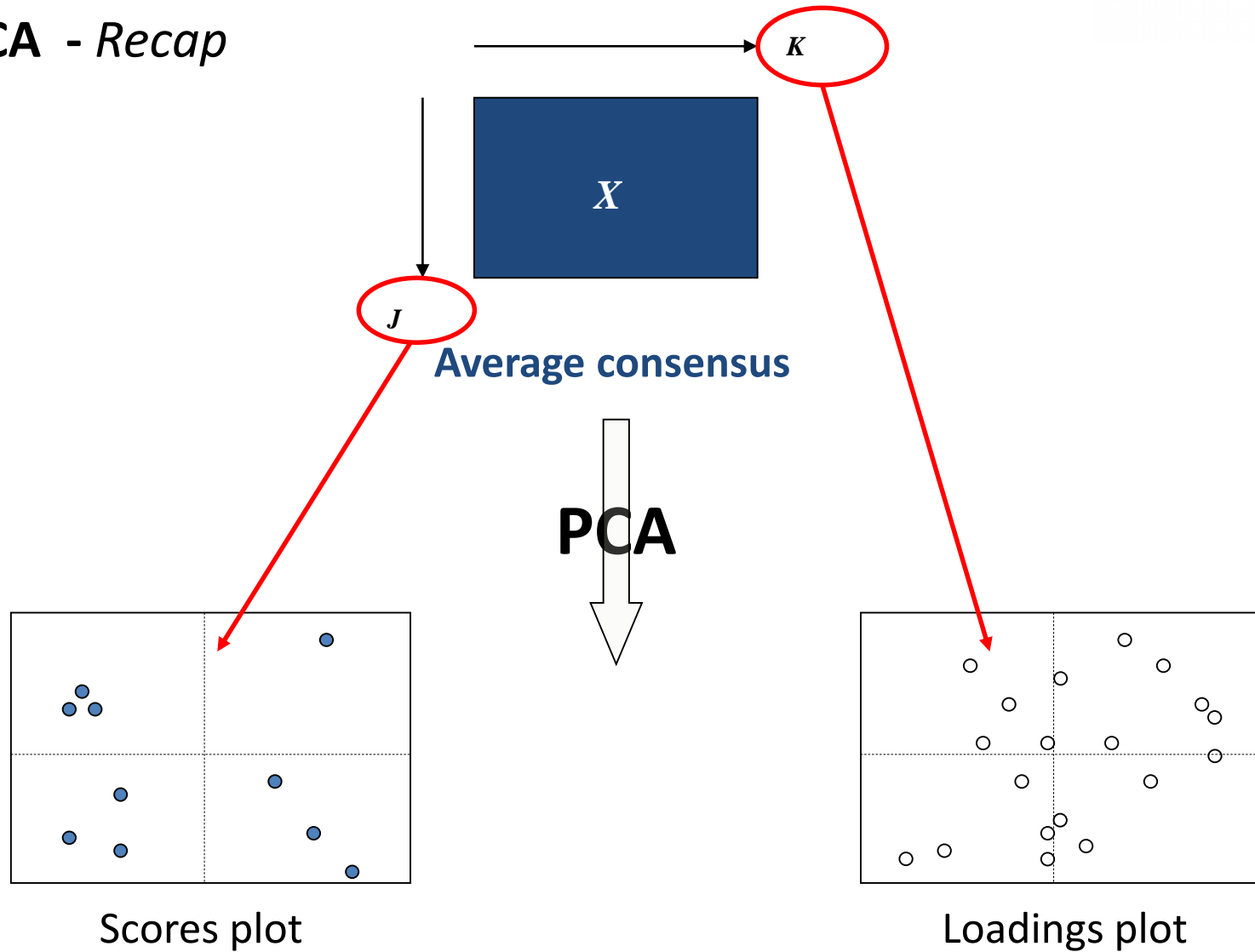
Consumers

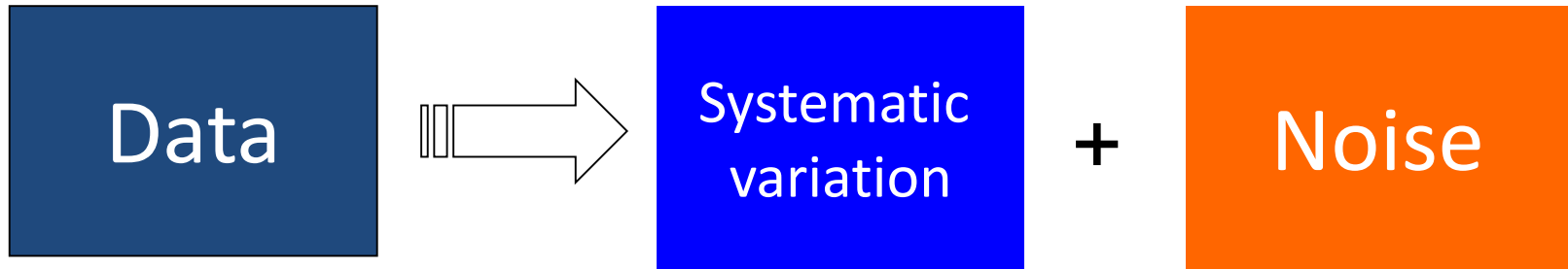
Data need to be balanced:

ALL assessors need to have tested **ALL** products

ALL consumers need to have tested **ALL** products

PCA - Recap





$$X = TP' + E$$



X: data matrix
T: PCA scores
P: PCA loadings
E: residuals /noise

Principal Components (PC's) describing
the systematic variation in the data

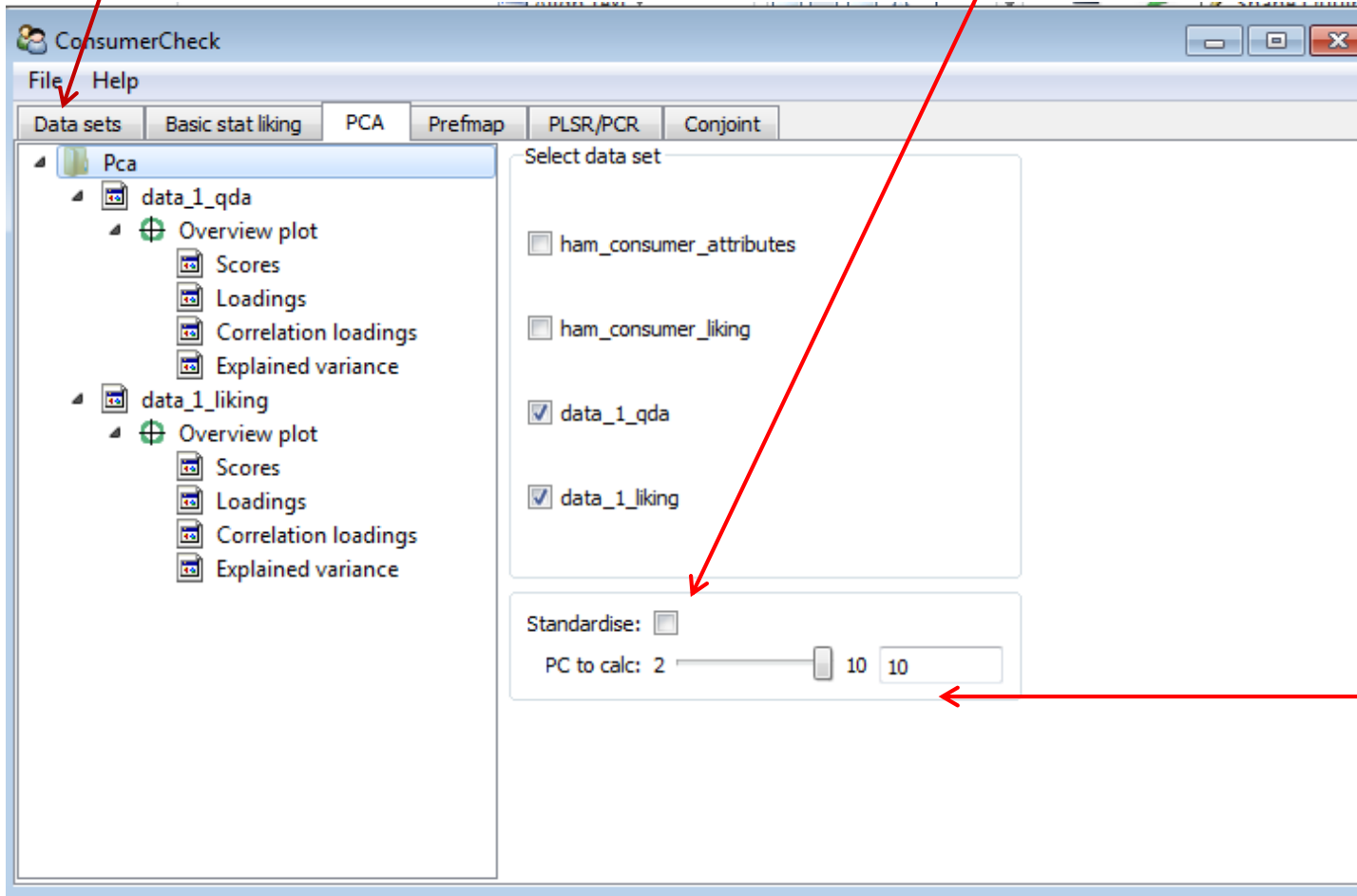
ConsumerCheck PCA GUI

Tree control for the plots

By checking the Standardize checkbox
All variables in the data are standardised
Such that they have zero mean and a standard deviation
Equals to one. If unchecked, then only mean centered

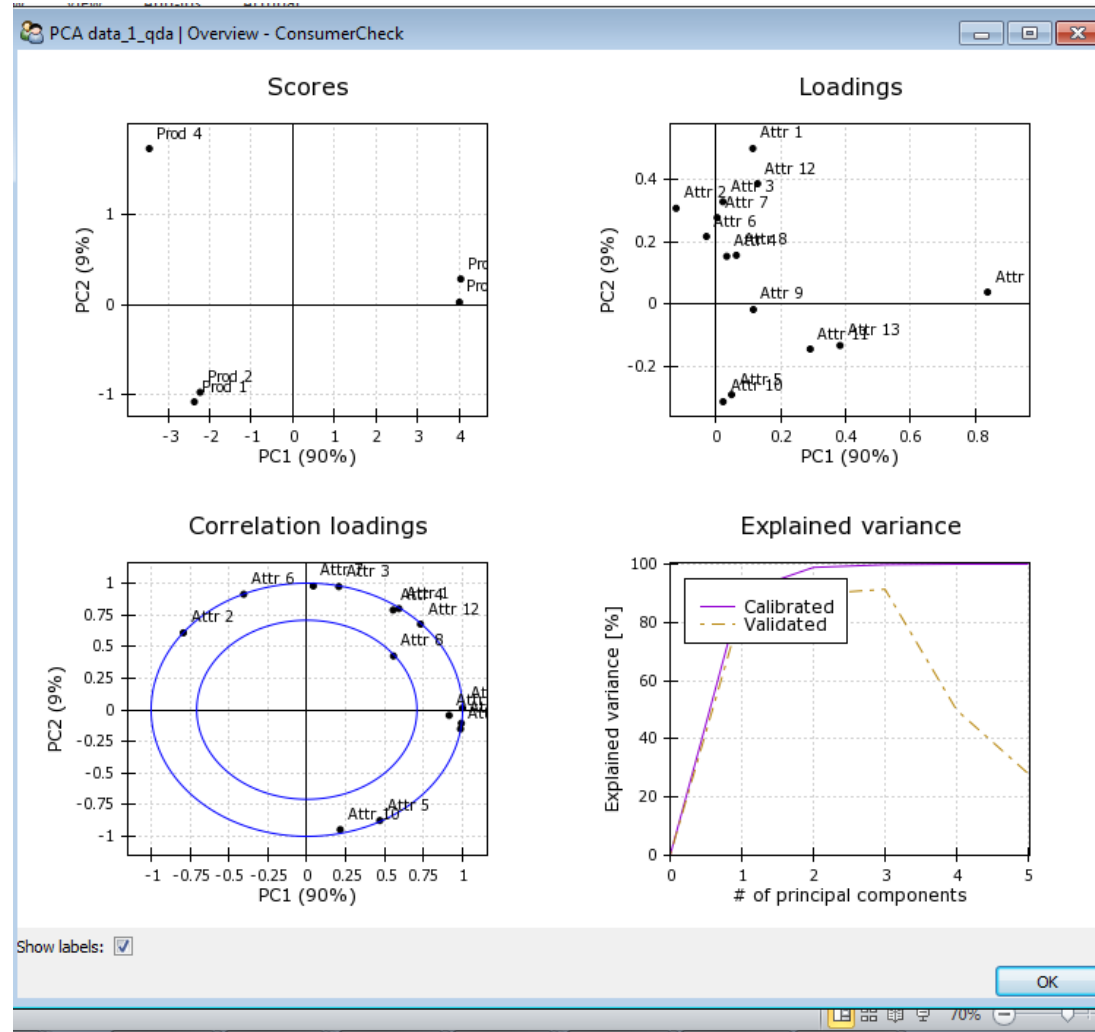
Variables with
zero variance
Across objects
(rows)
Are left out

How many PC to
compute?



PCA (sensory profiling data)– Overview plot

- PCA Scores
- Loadings
- Correlation loadings
- Explained variance

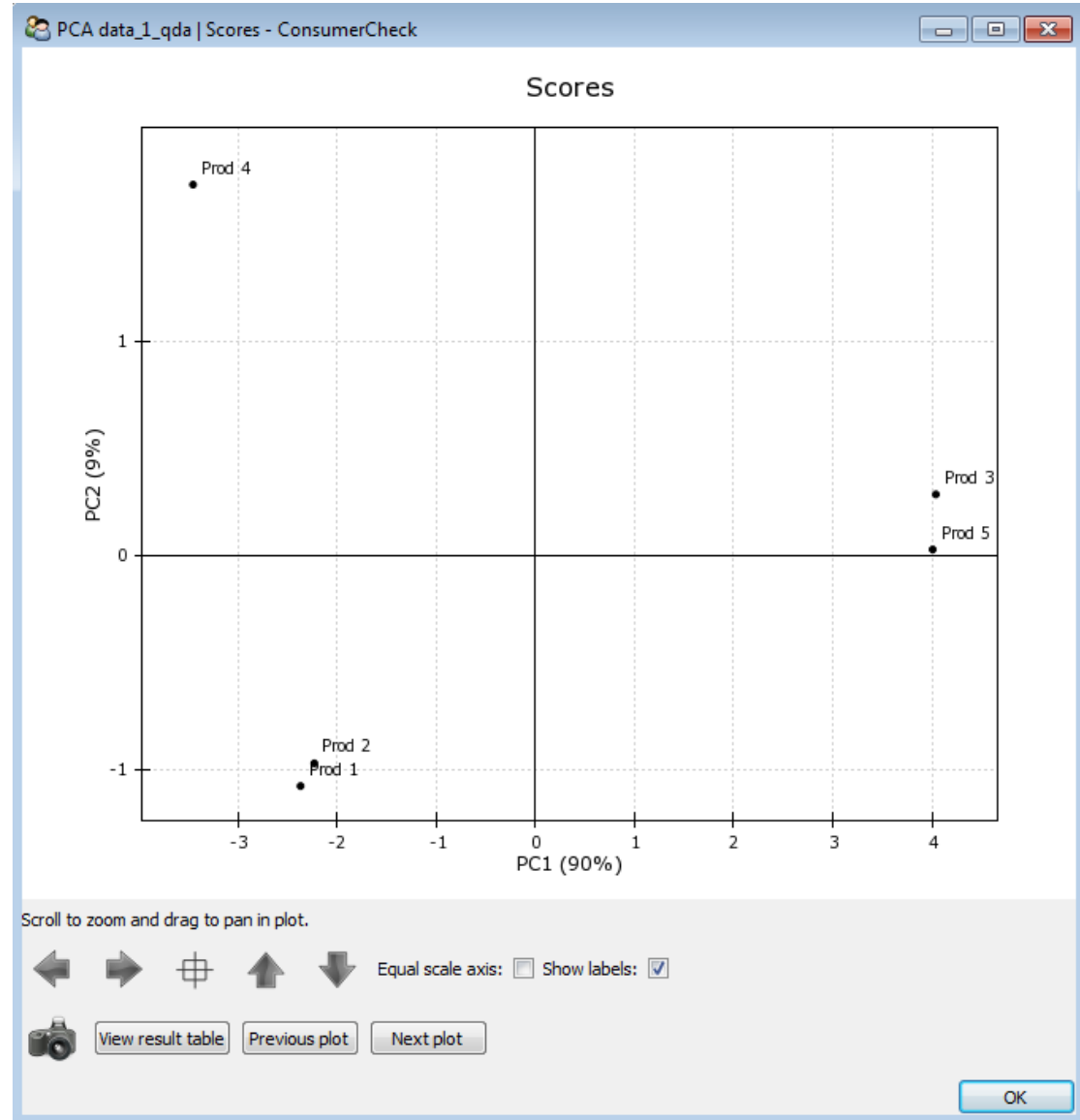


PCA (sensory profiling data) – Scores

Visualizes how objects (products) are distributed across space spanned by two principal components (PC)

PC1 and PC2 describe 99% of the variation in the data

Product 3 and 5 are very similar
Products 1,2,4 are very different from 3 and 5

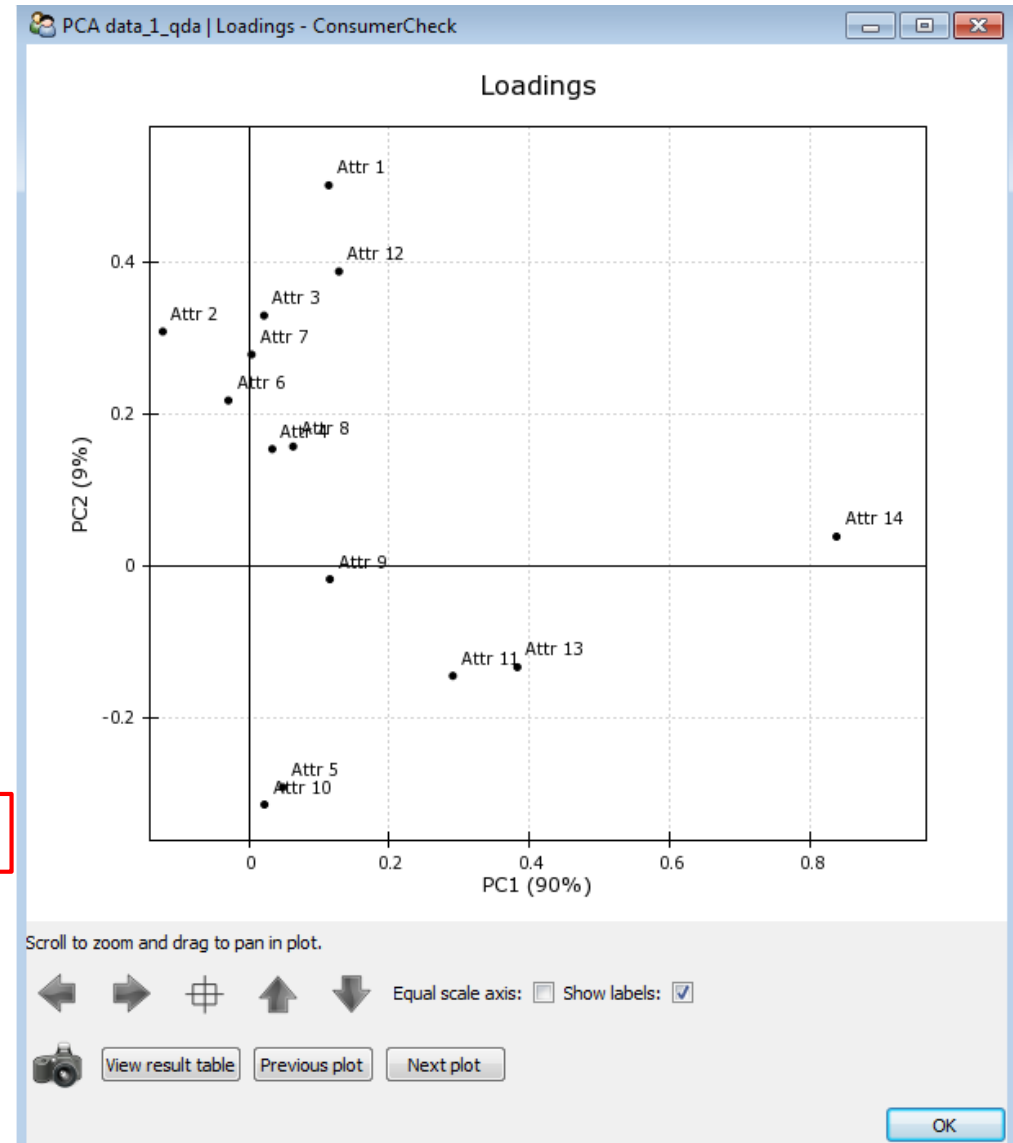


PCA (sensory profiling data) – loadings

Visualizes how the variables (sensory attributes) contribute to the variation in the data

Attribute 4 contributes much to the variation explained by PC1

PC2: attributes 5,10 versus 1,2,3,12



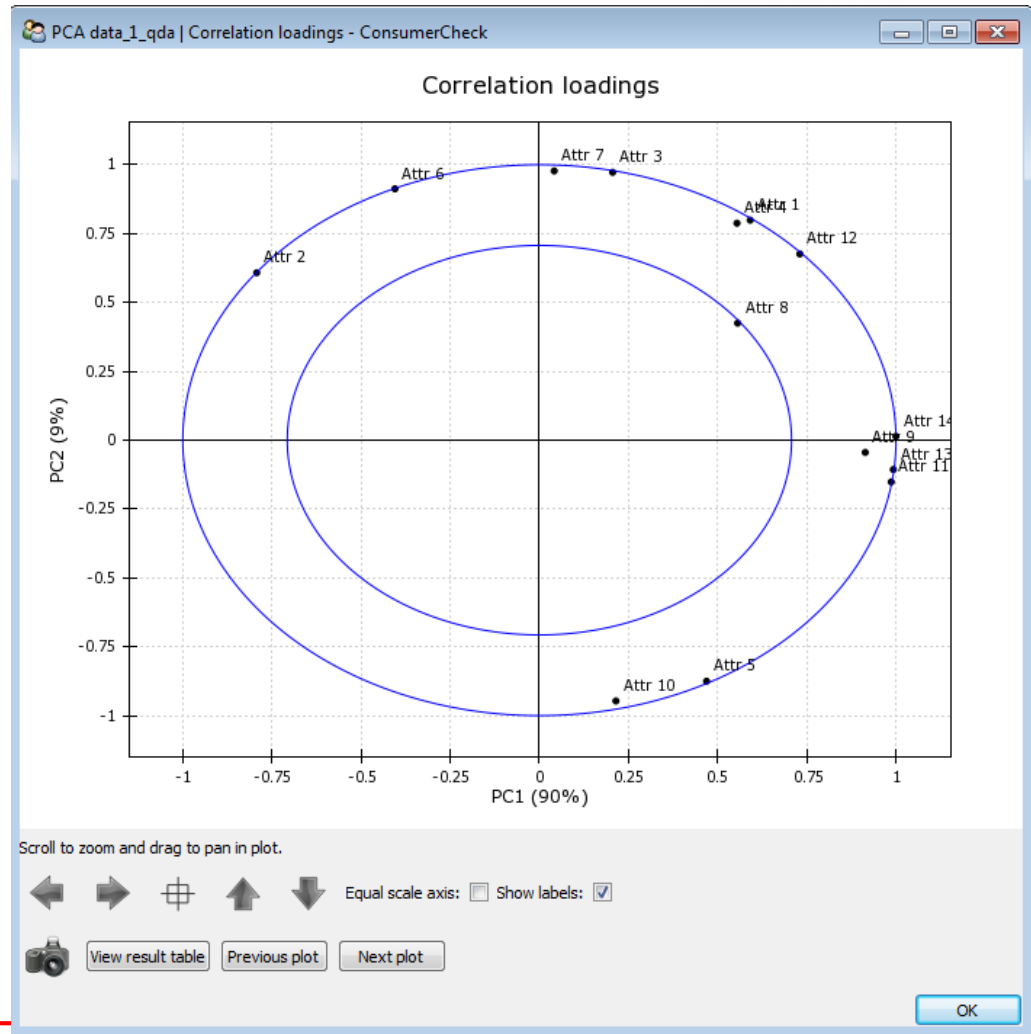
PCA – correlation loadings

Provides information on how systematic the variance of a variable is with regard to the computed PC's

Correlation loading = correlation between the original data of a specific variable and the scores of a specific PC

The outer ring represents 100% of the variance
The inner ring represents 50%

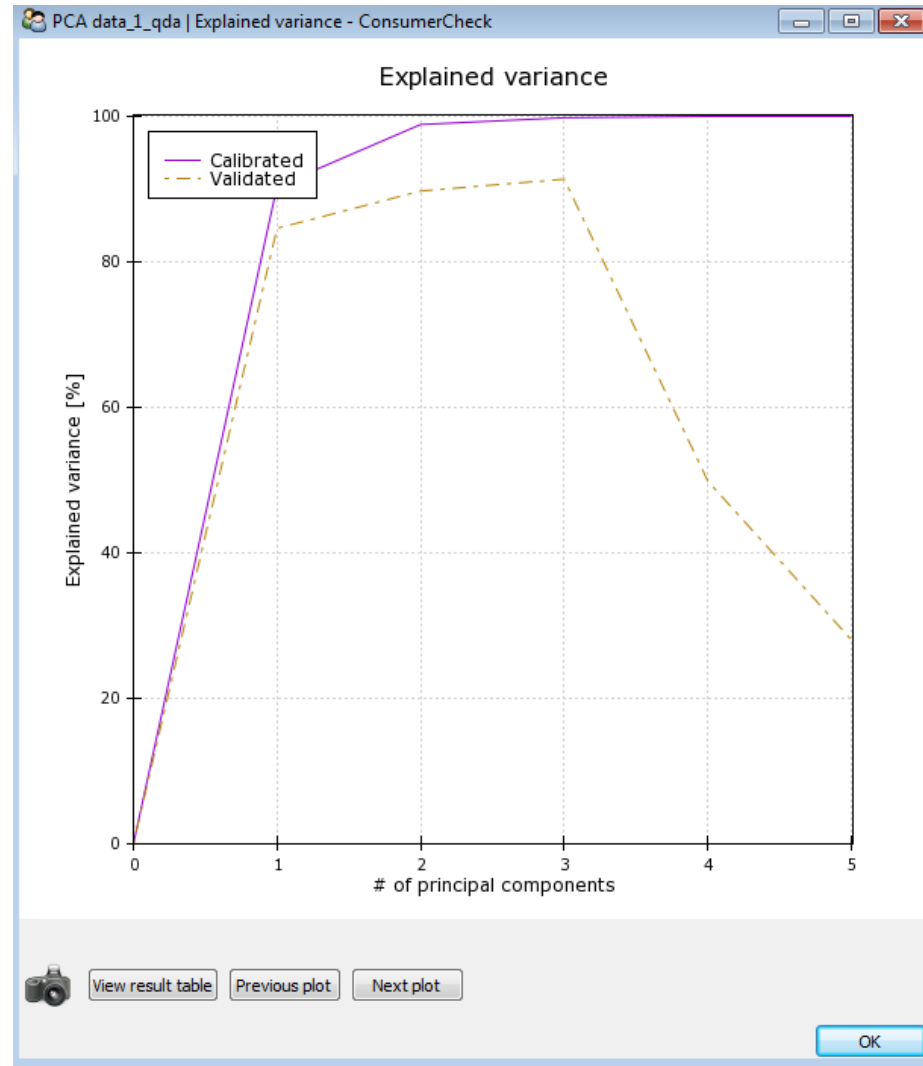
The two rings in the correlation loadings plot indicate amounts of explained variance for the attribute at hand



PCA – explained variances

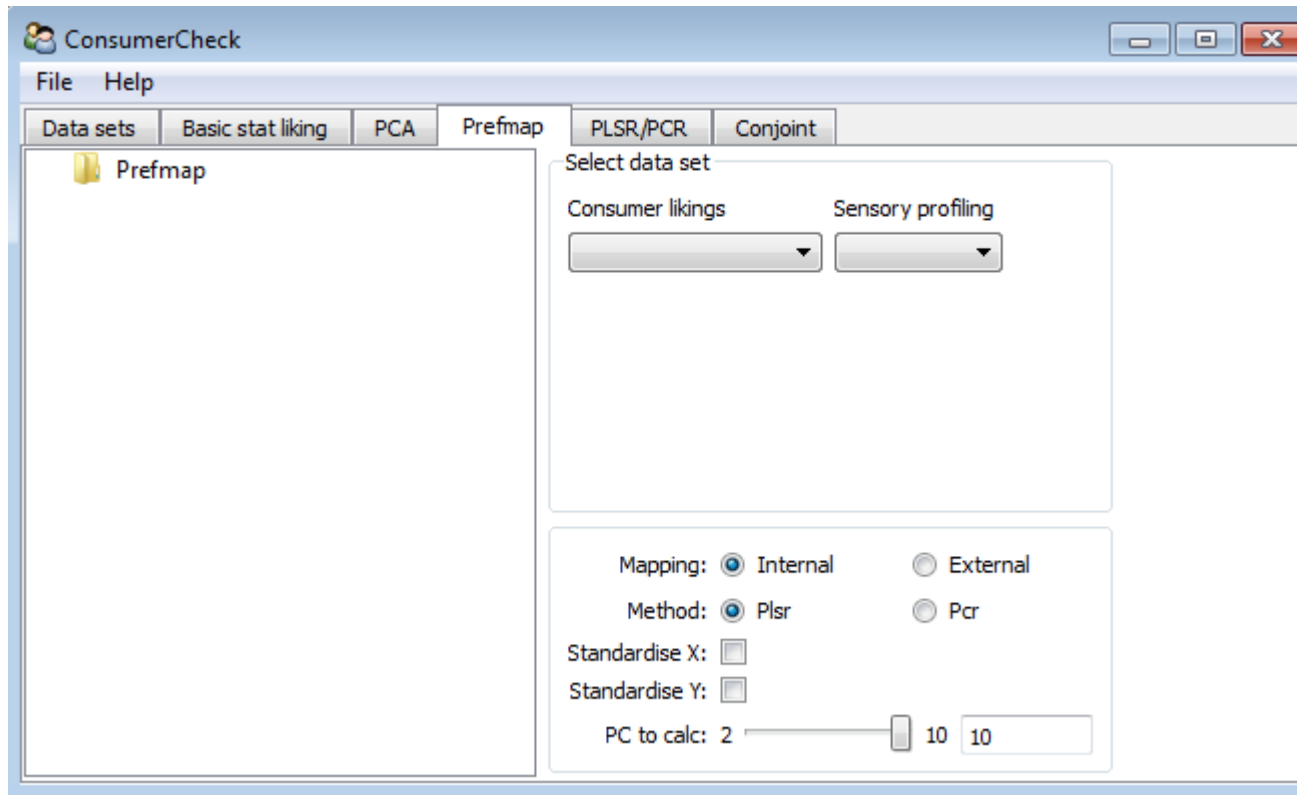
Visualizes calibrated and v validated explained variances

The validated explained variance is computed by systematically leaving out objects/rows from the data, then computing new PCA models and using the new loadings to predict values of the data that were left out. The closer the predictions of the left out data are to the real values of the left out data, the more robust the model.

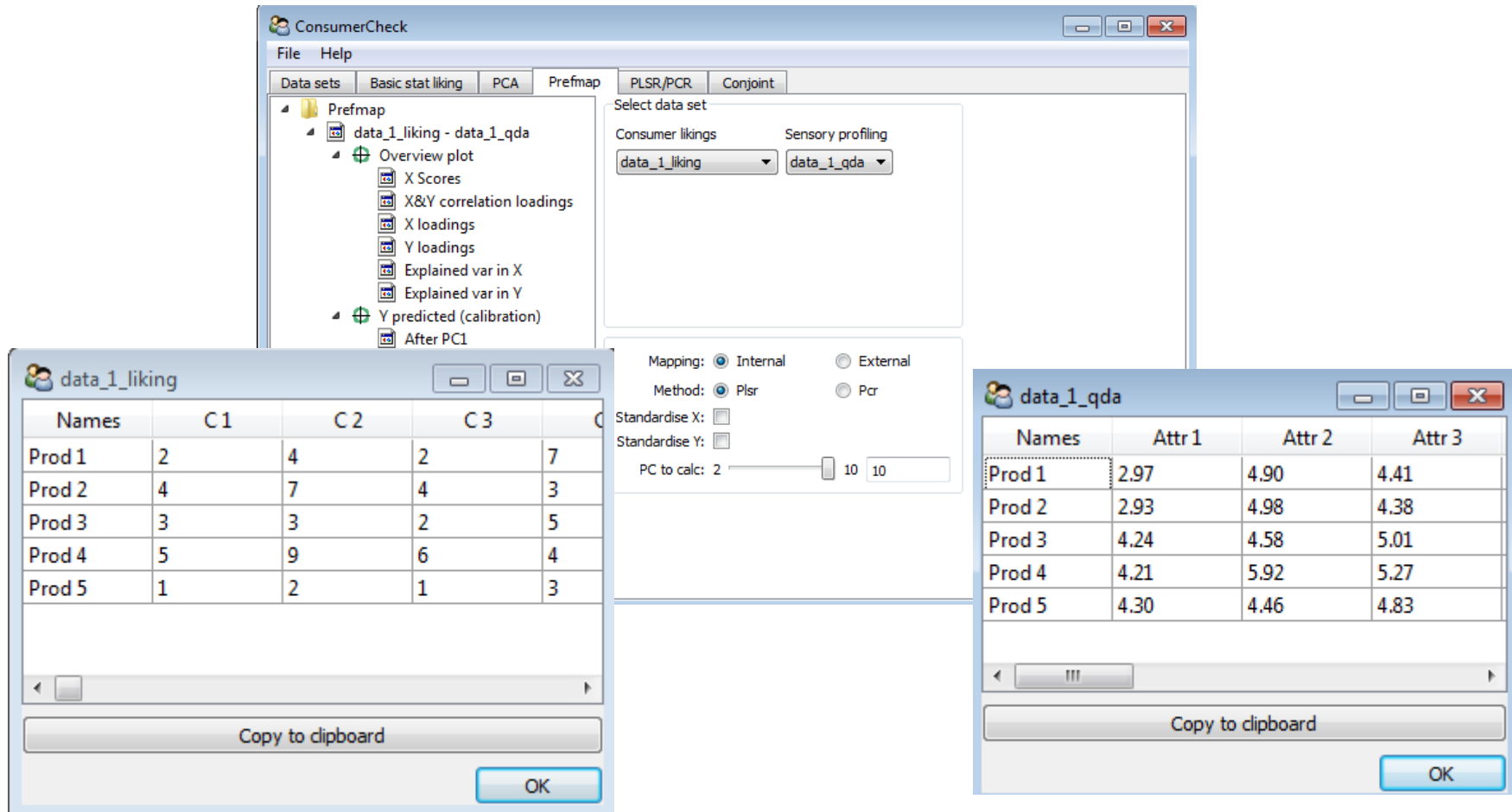


Preference mapping

Preference mapping (Greenho and MacFie 1994; McEwan 1996) is a much used statistical method in the field of sensometrics that analyses *consumer liking* and *sensory profiling data* together.



Preference mapping GUI



The screenshot shows the ConsumerCheck software interface. The main window is titled "ConsumerCheck" and has a menu bar with "File" and "Help". Below the menu bar are several tabs: "Data sets", "Basic stat liking", "PCA", "Prefmap", "PLSR/PCR", and "Conjoint". The "Prefmap" tab is selected, showing a tree view of analysis options under "data_1_liking - data_1_qda". The options include "Overview plot", "X Scores", "X&Y correlation loadings", "X loadings", "Y loadings", "Explained var in X", "Explained var in Y", "Y predicted (calibration)", and "After PC1".

In the center of the main window, there is a "Select data set" section with two dropdown menus: "Consumer likings" (set to "data_1_liking") and "Sensory profiling" (set to "data_1_qda"). Below this are settings for "Mapping" (Internal selected, External unselected), "Method" (Plsr selected, Pcr unselected), and "PC to calc" (2, 10, 10). There are also checkboxes for "Standardise X" and "Standardise Y".

Two smaller windows are open in the foreground:

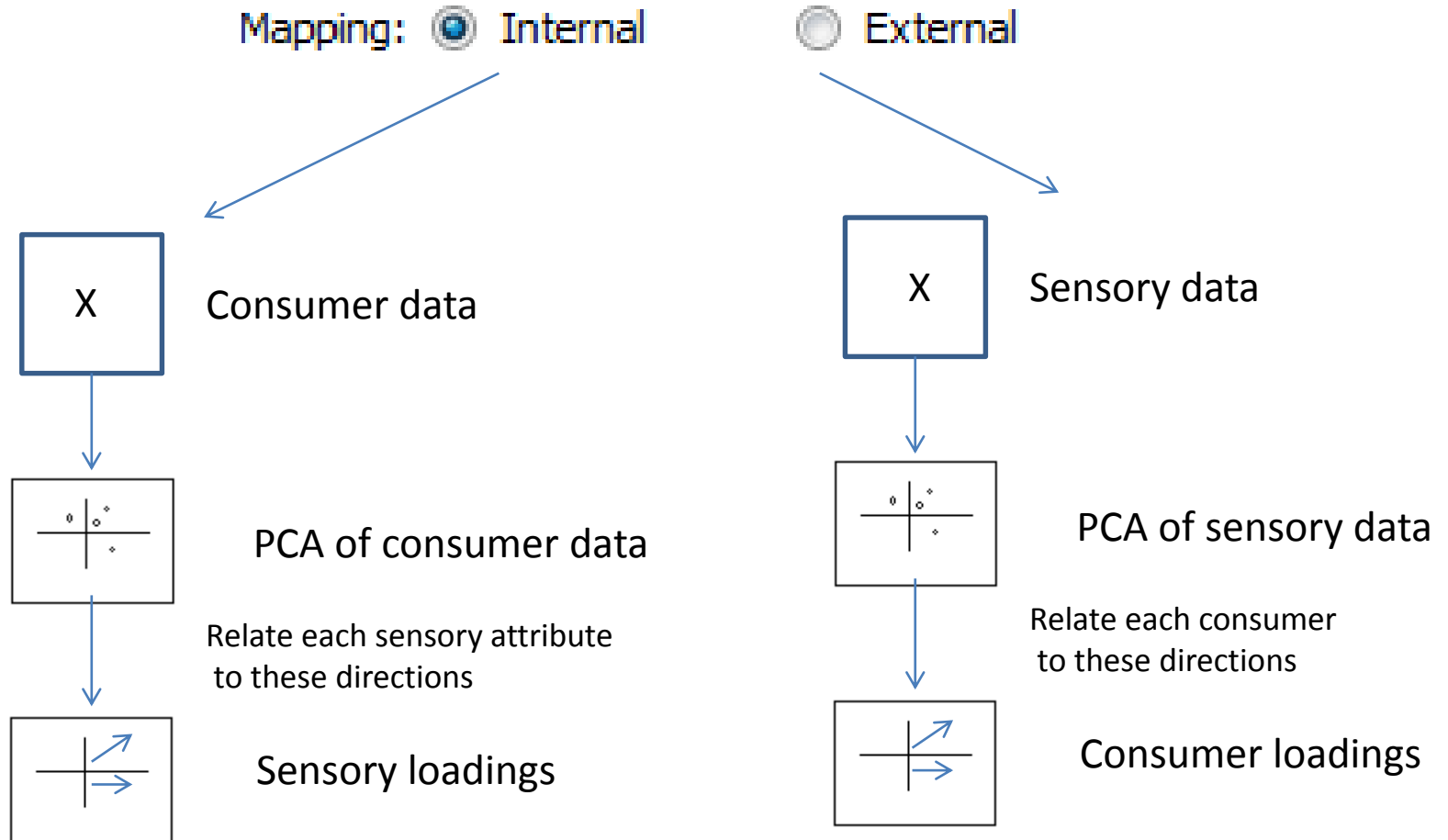
- data_1_liking**: A table with columns "Names", "C 1", "C 2", "C 3", and "C 4". The data is as follows:

Names	C 1	C 2	C 3	C 4
Prod 1	2	4	2	7
Prod 2	4	7	4	3
Prod 3	3	3	2	5
Prod 4	5	9	6	4
Prod 5	1	2	1	3
- data_1_qda**: A table with columns "Names", "Attr 1", "Attr 2", and "Attr 3". The data is as follows:

Names	Attr 1	Attr 2	Attr 3
Prod 1	2.97	4.90	4.41
Prod 2	2.93	4.98	4.38
Prod 3	4.24	4.58	5.01
Prod 4	4.21	5.92	5.27
Prod 5	4.30	4.46	4.83

Row order of the products needs to be the same

Preference mapping

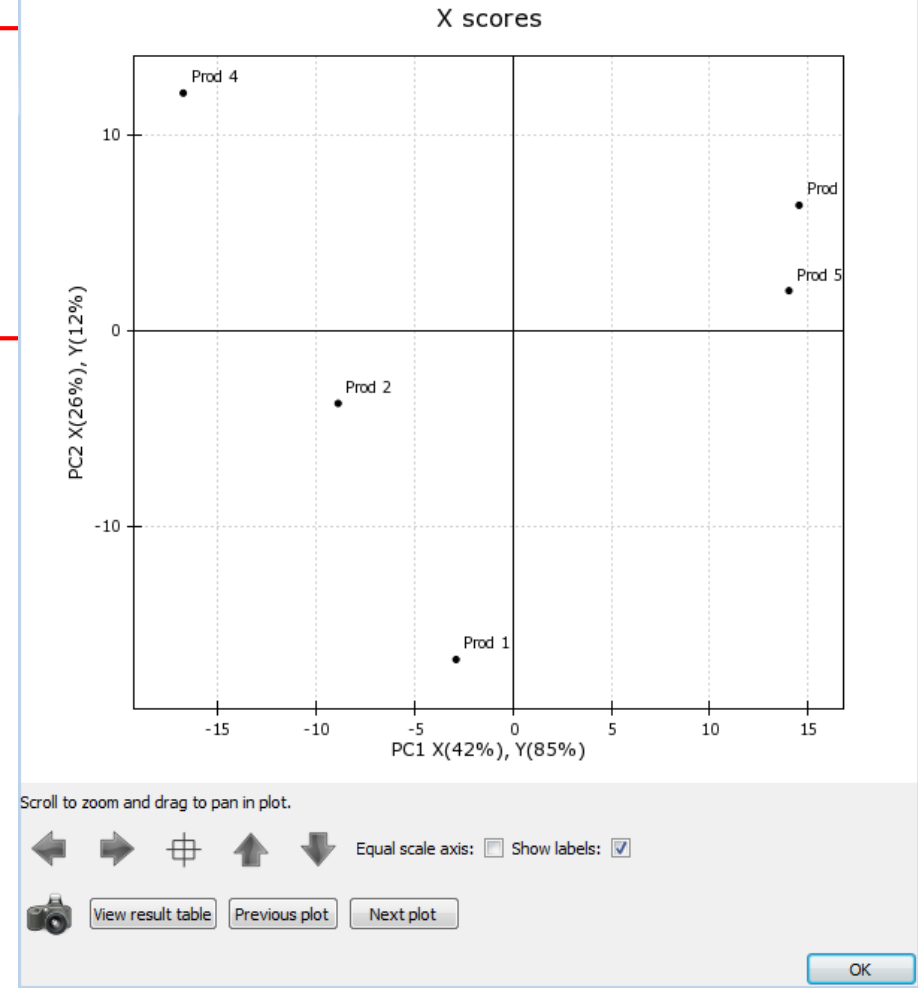


Preference mapping – X scores

Here:
internal preference mapping
PLSR
X and Y not standardized
(all variables are based on the same scale)

X-scores: visualizes how the products related to each other in the space spanned by the first principal components

Only 68% from consumer liking data

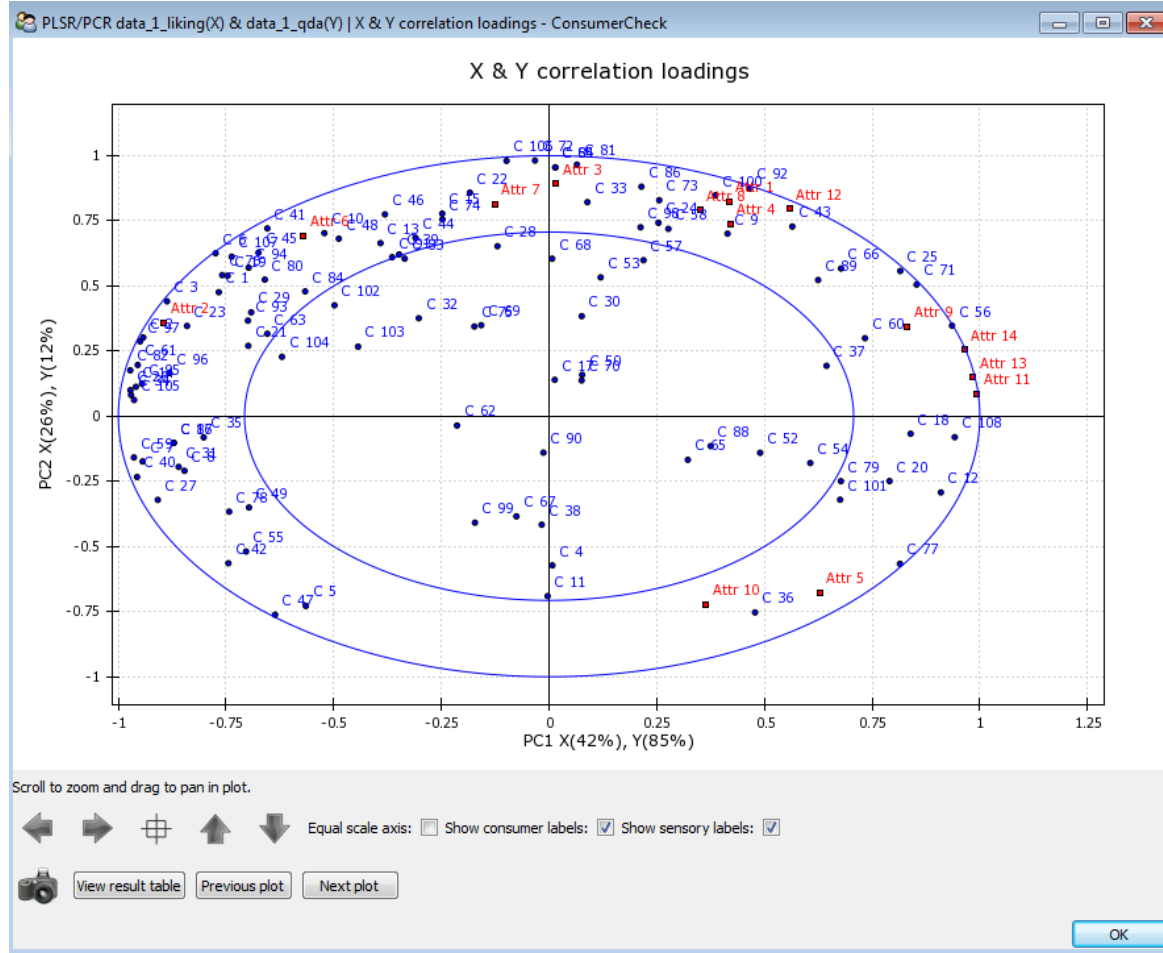


Preference mapping – X and Y correlation loadings

Correlation loadings from both X and Y are visualized

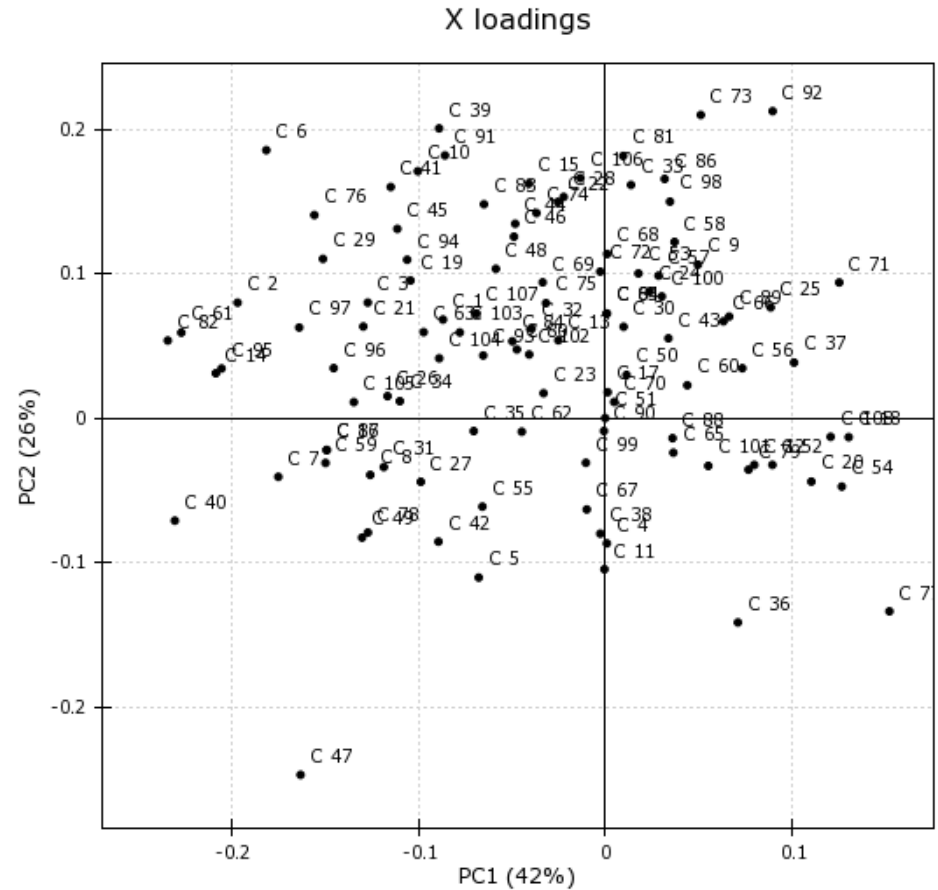
Correlation loadings belonging to X are blue

Many consumers prefer Products with high intensities of Attribute 2 and 6



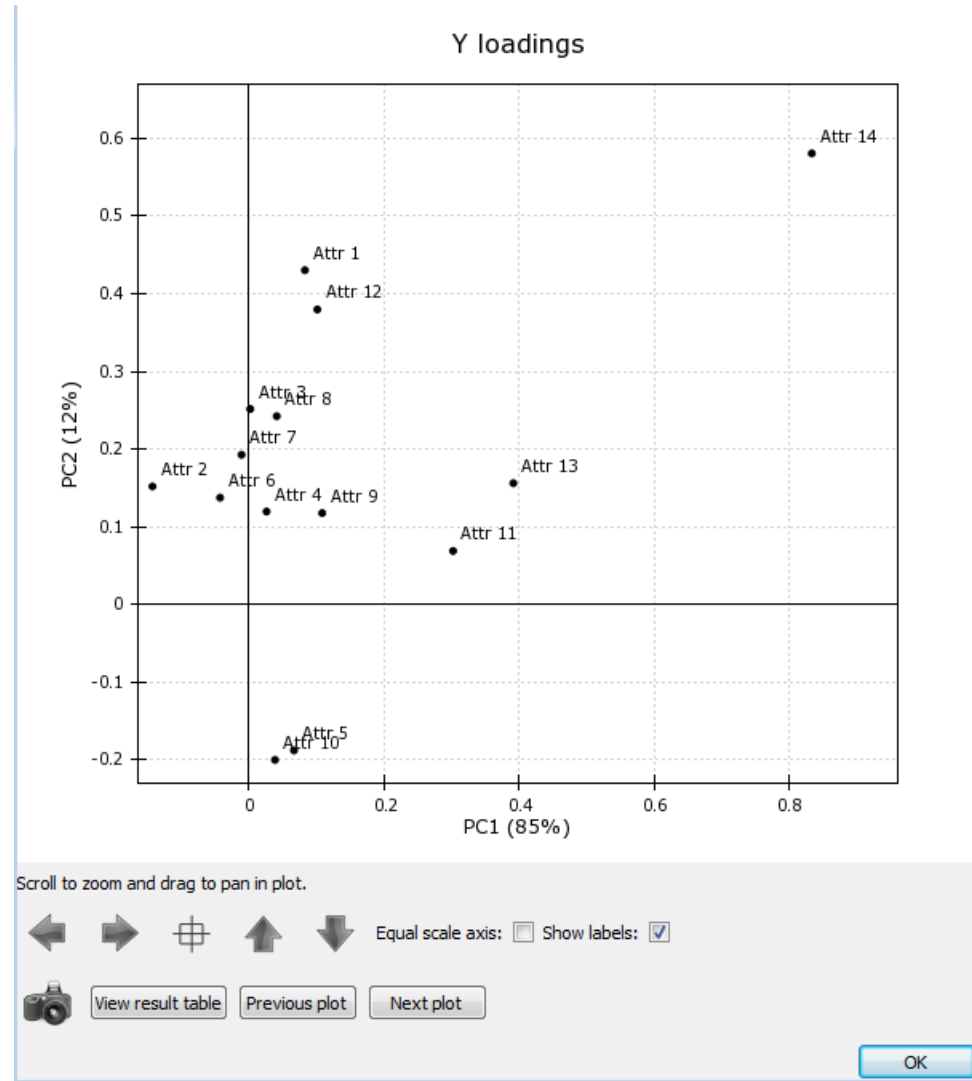
Preference mapping – X loadings

X loadings: show how the variables of the X matrix (here consumer data) contribute to the common variation between X and Y for each PC



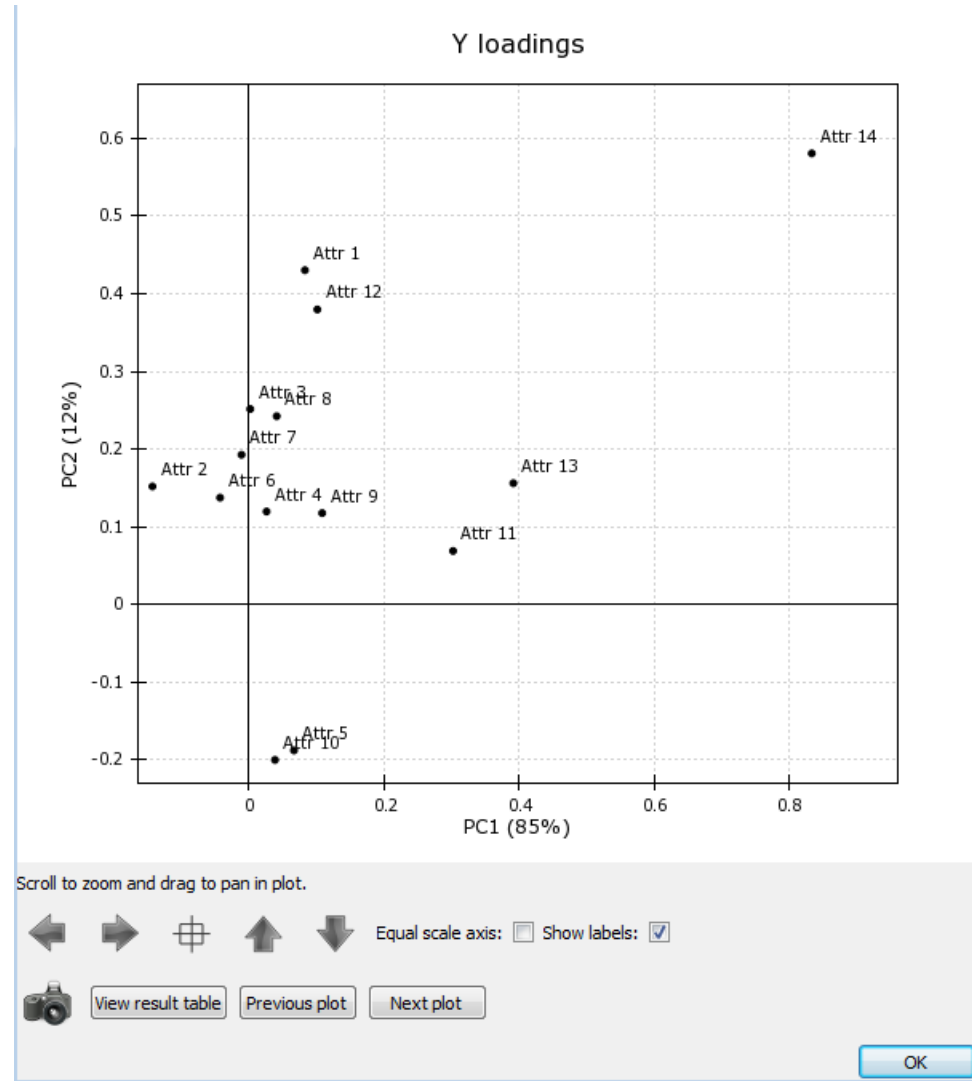
Preference mapping – Y correlation loadings

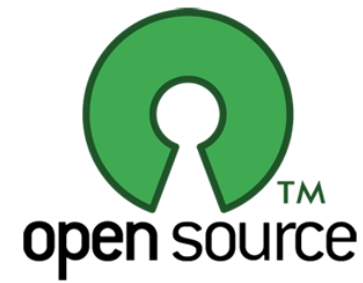
Y loadings: show how the variables of the Y matrix contribute to the common variation between X and Y for each PC



Preference mapping – Y correlation loadings

Y loadings: show how the variables of the Y matrix contribute to the common variation between X and Y for each PC





Summary **ConsumerCheck**

- Easy-to-use software for non-statisticians
- Proposes classical as well as advanced tools for analysis of consumer data
- Important methods:
 - + PCA
 - + Preference mapping
 - + Conjoint analysis