

Lecture 00 – Exercise 01

PCA on consumer liking data

Background information for analysis

You are presented consumer liking data from a study on apple juices. Six different apple juices were designed with different levels of sugar and acid (high acid level is often perceived as freshness in fruits). See Fig. 1 below for a more detailed description of the apple juices.

- 2 design variables:

- **Sugar level:** low medium high
- **Acid level:** low high

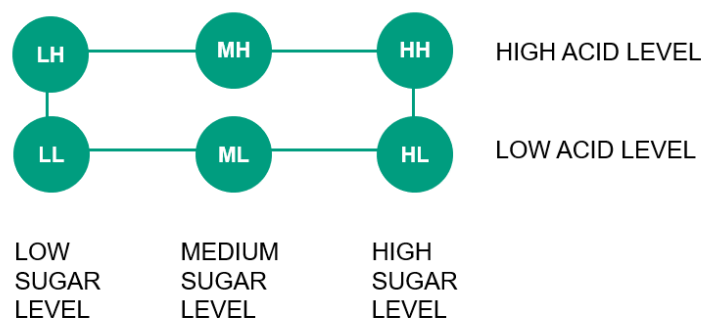


Fig. 1: Note that unique combinations of sugar and acid levels allowed for six different apple juices (indicated by green circles). The first letter of the apple juice name represents the sugar level. The second letter represents the acid level. Example: apple juice ML would have medium sugar level (M) and low acid level (L), hence the name ML.

125 consumers tasted each of the apple juices and rated their liking from 1 (dislike very much) to 7 (like very much). The consumer liking data are stored in file "AppleJuice_consumer.txt".

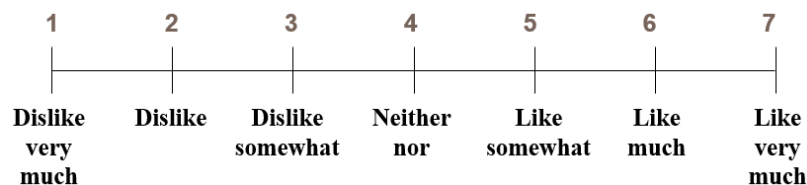


Fig. 2: For every apple juice the consumers were asked to rate their liking on the above scale.

Run PCA on the consumer liking data” with the Hoggorm package. Visualise results from your PCA with the HoggormPlot package. (Hint: set input parameter `cvType=[“loo”]` when computing the PCA model).

Questions / tasks

1. Do you think the variables should be centred or standardised? Why would you make this choice?
2. What is the maximum number of components you could compute for the data at hand?
3. Generate these plots with HoggormPlot: score plot, loading plot, explained variance plot, bi-plot, correlation loading plot. (study the HoggormPlot documentation)
4. How many components would you use for interpretation of the PCA model?
5. Based on what you see in the score plot, what do you think component 1 represents? What do you think component 2 represents?
6. Which products do most of the consumers prefer?
7. Why are consumers C26, C34 and C40 close to the origin in the correlation loadings plot?
8. How much of the total variance do component 1 and component 2 explain together?
9. How would you extract the scores T and loadings P from the computed PCA model (study the Hoggorm documentation)?
10. How would you extract the settings under which the PCA model was computed?