
Relating instrumental and sensory data for German Riesling wines by means of PLS

Anita S. Busch¹, Michael Meyners¹, Ulrich Fischer²

¹Fachbereich Statistik, Universität Dortmund, 44221 Dortmund, Germany

²Staatliche Lehr- und Forschungsanstalt für Landwirtschaft, Weinbau und Gartenbau,
67435 Neustadt a.d. Weinstraße, Germany

We consider a data set containing the observations for 56 German Riesling wines originating from two different German viticultural areas. These wines have been either conventionally or ecologically produced. The data set is a part of a research project assigned from the ministry of economics, transportation, agriculture and viticulture of the state of Rheinland-Pfalz to the SLFA Neustadt (Dupin *et al.* 1999).

The data consists of instrumental as well as sensory profiling data: On the one hand, the concentration of 55 aroma compounds has been analysed by means of gas chromatography. On the other hand, all wines have been assessed by descriptive analysis using 15 sensory variables and a total of 22 judges. The intensely trained assessors developed a set of relevant sensory attributes for this particular set of wines based on group consensus. The wines were presented twice to the assessors. However, the design is not perfectly balanced because some assessors did not taste each wine. To avoid systematical errors, the order of presentation varied among the assessors. For our analysis, we confined ourselves to the observed mean values, which allows us to neglect the details of the experimental design.

The objective of our analysis was to relate instrumental to sensory profiling data and to differentiate between ecological and conventional produced wines with respect to both the sensory profile and the chemical composition. Furthermore, we stress the question whether the differences found can be reasonably interpreted. A linear regression model may be taken as a simple starting point, but the problem of multicollinearity emerges when 55 regressors and 15 response variables are to be linked by the information of just 56 objects.

Partial Least Squares (PLS, cf. Frank and Friedman, 1993) is intuitively build up on the base of latent variables in both the independent and the response variables. Frank and Friedman (1993) state that the univariate version has proven a good prediction quality. The multivariate method might be used to subsume the information in a data mining way when the response variables are highly correlated and of similar type, as can be seen in the analysis of preference data (Helgesen *et al.*, 1997).

Starting with a Principal Component Analysis (PCA) of the mean sensory data, the first two components, which explain together about 56% of the variance, seem to represent fruity and unpleasant (bitter, sour, aftertaste, not sweet) perceptions. A slight separation between ecologically and conventionally produced wines can be found within the space of this two components, in which the conventional wines are rated higher than the ecological ones.

Analysing the data by means of multivariate PLS-Regression, we find that the first two PLS-components are related to the sensory descriptors in a similar manner as the first two principal components are. The corresponding biplots additionally show that the conventionally produced wines vary much more in the space spanned by the first four component loadings. Hence we conclude that their characteristic spread is wider than the one of the ecologically produced wines, whose diversity is possibly limited by the production regulations.

Admittedly the prediction quality measured by means of the cross-validated coefficient of determination is extremely low for most response variables (descriptors). Only the descriptors “bitter” and “fruity by mouth” can be modelled satisfying to some extent. For these, univariate PLS-regressions with the same influence variables are performed. Model selection by means of the cross-validated coefficient of determination leads to the use of two components for “bitter” and only one for “fruity by mouth”. Similar results like the wider spread of conventionally produced wines arise here, which is reasonable since these variables are the most important descriptors in the multivariate PLS-analysis. The most important instrumental variables can be classified into subgroups (alcohols, acids etc.). We discuss whether the relation between these chemicals and the perceptions can be reasonably interpreted.

References

- Dupin, I., Schlich, P., and Fischer, U. (1999). *Differentiation of wines produced by organic or conventional viticulture according to their sensory profiles and aroma composition*. Proceedings of the 24th Weltkongreß für Rebe und Wein, 128-135.
- Frank, I.E. and Friedman, J.H. (1993). *A Statistical View of some Chemometrics Regression Tools*, Technometrics 35(2), 109-135.
- Helgesen, H., Solheim, R. and Næs, T. (1997). *Consumer Preference Mapping of Dry Fermented Lamb Sausages*, Food Quality and Preference 8(2), 97-109.