

The use of analysis of variance and three-way factor analysis methods for studying the quality of a sensory panel

Rosaria Romano
Dep. of Food Science
University of Copenhagen
Copenhagen, Denmark
rro@life.ku.dk

Tormod Næs
Dep. of Mathematics
University of Oslo
Oslo, Norway
tormod.næs@matforsk.no

Per B. Brockhoff
Inf. and Mat. Modelling
Tech. Univ. of Denmark
Lyngby, Denmark
pbb@imm.dtu.dk

Abstract:

In sensory analysis a panel of assessors evaluate a collection of samples/products with respect to a number of sensory characteristics. Assessments are collected in a three-way data matrix crossing products, attributes and assessors. The main objective of the experiment is to evaluate products. However, the performance of each assessor and of the panel as a whole is of crucial importance for a successful analysis. At this aim univariate analysis for each sensory attribute as well as multi-way analysis considering all directions of information are usually performed. The present work studies the quality of a panel using both methods. The basic idea is to compare results and investigate relations between the two different analytical approaches.

Keywords: two-way anova, interaction, assessor model, parafac model, sensory analysis

1 Introduction

In Sensory descriptive analysis a group of trained assessors, the so-called panel, gives scores on a continuous scale to evaluate different products with respect to a certain number of sensory attributes. Despite the training sessions there will be significant differences between assessors to be taken into account in the analysis of this type of data. For instance, assessors may differ in the use of scale (level effect, range effect, disagreement effect) or in the replication of their scores (reproducibility error).

Various methods have been proposed for the evaluation of assessor/panel performance (Næs (1990); Brockhoff and Skovgaard (1994)). They use the basic principles of univariate Analysis of Variance (ANOVA) for modeling individual differences. Nevertheless, differences between assessors may also be investigated by using three-way factor analysis (TWFA) methods (Brockhoff, Hirst, Naes (1996)).

A question to be raised, which is the main objective of the present paper, is how results from the two different approaches are related. In a recent paper (Romano *et al.* (2007)) a comparative analysis of univariate and three-way analysis of assessors differences has been presented, with focus on scaling effects. However, Tucker1 model was used in the comparison, which is an unfolding method rather than a three-way method, i.e. it does not fully take into account the really multivariate error structure. For this reason, the present paper focus on methods accounting for the three dimensions of information. Specifically, an application to real sensory data will be presented to discuss similarities and the differences between ANOVA and TWFA from an applied point of view.

2 Materials and methods

Seven varieties of milk were profiled by a panel of 9 assessors over 12 descriptors (green odor, yellow appearance, creamy flavor, boiled milk flavor, sweet flavor, stald feed flavor, bitter flavor, metallic flavor, sourness flavor, fatness after taste, astringent0 after taste, astringent20 after taste). The samples were evaluated in 3 replicates according to a continuous scale anchored at 0 and 15. The data were collected in a three-way table (samples \times assessors \times attributes) with the I assessors as one of the *ways*, the J^*M products (J products in M replicates) as the second *way* and the K attributes as the third *way*.

2.1 Univariate modeling

Let Y_{ijm}^k denote the score of assessor i on attribute k of the r th replicate of the j th product. Data can be then described by an ANOVA model including two main effects (samples and assessors) and the interactions (samples \times assessors):

$$Y_{ijm}^k = \mu^k + \alpha_i^k + v_j^k + \delta_{ij}^k + e_{ijm}^k \quad e_{ijm}^k \sim N(0, \sigma^2) \quad (1)$$

Here, μ^k is the grand mean for the attribute k . The assessor main effects α_i^k represent differences in scoring level between the assessors. The product main effects v_j^k represent the differences between the average score for the different products. The assessor-product interaction δ_{ij}^k expresses differences between assessors in measuring differences between products. The error term e_{ijm}^k represents the residual variation due to replicates. It is natural to consider assessor and interaction effects as *random*, since assessors are considered *random* representatives for a population. This assumption leads to the so-called Mixed Model ANOVA (MMA), where only product main effects are *fixed*. Note that the *assessor-product* interaction is the largest noise contributor in sensory data. All information about systematic individual differences, except eventual differences in level, lie in this term. Model (1) does not take into account differences in variability (reproducibility error), due to the homogeneity variance assumption.

A model accounting for all individual differences, apart disagreement, is the *assessor model* (Brockhoff and Skovgaard, 1994):

$$Y_{ijm}^k = \mu^k + \alpha_i^k + \beta_i v_j^k + e_{ijm}^k \quad e_{ijm}^k \sim N(0, \sigma_a^2) \quad (2)$$

where the error variance (σ_a^2) allows for different assessors' variability. Comparing data fits for model (1) with model (2) a test for *disagreements* is given. Note that the scaling effects (β_i), usually included in the interaction (δ_{ij}^k) (1), are here explicitly modeled as multiplicative terms. Assessor model provides tests for the main effects and estimates for individual scalings, variabilities and disagreements for each attributes.

2.2 Three-way modeling

Three-way factor analysis methods are extensions of Principal Components Analysis to three-way data matrix (Tucker (1977); Kroonenberg and De Leeuw (1980)). In sensory analysis they are used to investigate relations between samples, attributes and assessors at the same time (Brockhoff *et al.* (1996)).

For sake of space focus is given to PARAFAC model, but other three-way methods as Tucker2 and Tucker3 need to be considered. PARAFAC model may be written as:

$$Y_{ijk} \approx \sum_{l=1}^L a_{il} b_{jl} c_{kl} \quad (3)$$

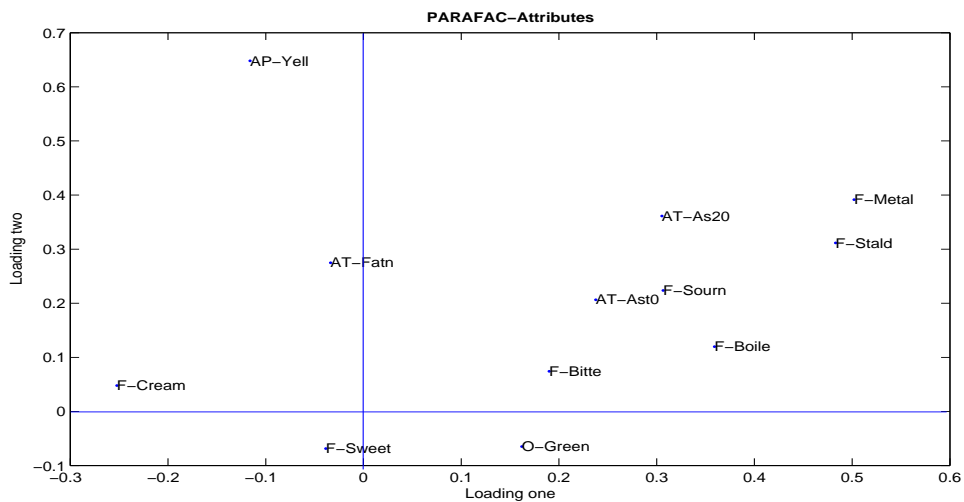
where L is the number of components. Using a PARAFAC model for sensory data means to assume that assessors perceive the same latent variables, but in different proportions. Thus, this model is useful when there is no agreement between assessors on which are the most important attributes for describing differences between assessors.

3 Results

Table 1: *P-values from Mixed Model Anova.*

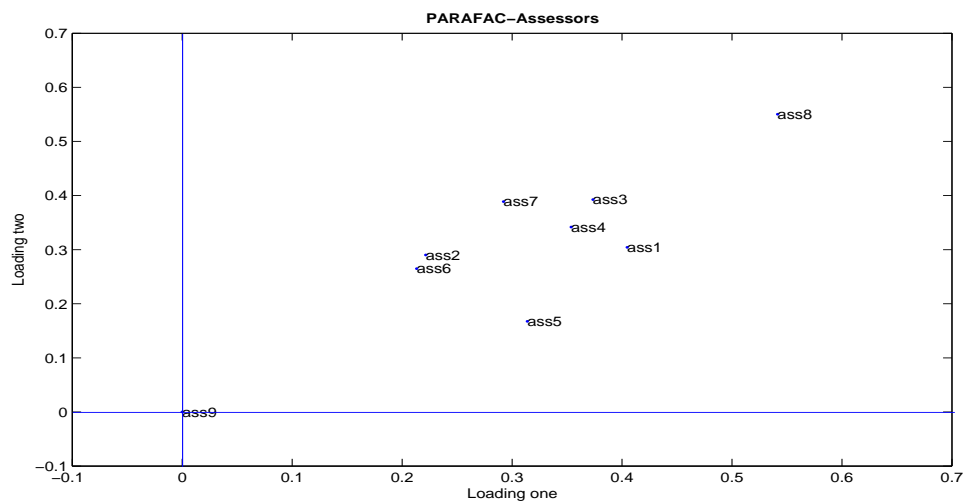
attributes	assessor effects	products effects	ass*product interactions
O-green	< .0001	< .0001	0.7803
AP-Yellow	< .0001	< .0001	0.0005
F-Creamy	< .0001	< .0001	0.0539
F-BoiledMilk	< .0001	0.0005	0.0238
F-Sweet	< .0001	0.7605	0.2598
F-Bitter	< .0001	0.0340	0.0175
F-metallic	< .0001	0.1377	< .0001
F-Sourness	< .0001	0.1691	0.0022
F-StaldFeedRelat	< .0001	< .0001	0.0665
AT-Astringent0	< .0001	0.3621	0.4047
AT-Fatness	< .0001	< .0001	0.0183
AT-Astringent20	< .0001	0.5926	0.0524

Figure 1: *PARAFAC on raw data.*



First, main effects and the interactions are estimated by means of MMA (tab.1). Successively, results from *assessor model* are considered. The idea is that *assessor model* corresponds to a PARAFAC model if the product mean structure is modeled by a PCA. Hence, *assessor model* is performed on the scores from a PCA. Finally, a PARAFAC model is performed both on raw data and interactions estimated by MMA. It is shown that results from PARAFAC on the raw data are related to results from MMA. In fact, the less significant attributes for the MMA (i.e, F-sweet) are positioned close to the origin of the axes in the loadings plot (Fig.1), whereas the most significant ones (i.e., AP-Yellowness) are spread out. There is a relation also between PARAFAC and *assessor model*. In fact, results from *assessor model* show significant scaling effects on the first factor, where Ass8 and Ass9 present highest and lowest values, respectively. These are the most extreme assessors in the PARAFAC loading plots on the interactions (Fig.2).

Figure 2: PARAFAC on interactions.



References

- Brockhoff P.B., Hirst D. and Næs T. (1996) Analysing individual profiles by three-way factor analysis, in: *Multivariate Analysis of Data in Sensory Science*, Naes T. and Hirst D., eds., Elsevier Science, Amsterdam, The Netherlands, 307–342.
- Brockhoff P.B. and Skovgaard I. (1994) Modelling individual differences between assessors in sensory evaluation, *Food Quality and Preference*, 5, 215–224.
- Kroonenberg P.M. and De Leeuw J. (1980) Principal component analysis of three mode data by means of alternating least squares algorithm, *Psychometrika*, 45(1), 69–97.
- Næs T. (1990) Handling individual differences between assessors in sensory profiling, *Food Quality and Preference*, 2, 187–199.
- Romano R., Brockhoff P.B., Hersleth M., Tomic O. and Næs T. (2007) Correcting for different use of the scale and the need for further analysis of individual differences in sensory analysis, *Food Quality and Preference*, in press.
- Tucker L.R. (1977) Some mathematical notes on three-mode factor analysis, *Psychometrika*, 31, 279–311.