Food Quality and Preference 39 (2015) 156–166

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model

Per Bruun Brockhoff [a,*], Pascal Schlich [b], Ib Skovgaard [c]

[a] DTU Informatics, Statistics Section, Technical University of Denmark, Matematiktorvet, Building 303b, DK-2800 Kongens Lyngby, Denmark
[b] INRA, UMR1324 Centre des Sciences du Goût et de l'Alimentation, F-21000 Dijon, France
[c] Department of Mathematical Sciences, Copenhagen University, Universitetsparken 5, 2100 København Ø, Denmark

A R T I C L E I N F O

Article history:
Received 9 July 2013
Received in revised form 28 June 2014
Accepted 2 July 2014
Available online 14 July 2014

Keywords:
Sensory profile data
Analysis of variance
Mixed model
Assessor differences
Scaling differences
Disagreement

A B S T R A C T

Scale range differences between individual assessors will often constitute a non-trivial part of the assessor-by-product interaction in sensory profile data (Brockhoff, 2003, 1998; Brockhoff and Skovgaard, 1994). We suggest a new mixed model ANOVA analysis approach, the Mixed Assessor Model (MAM) that properly takes this into account by a simple inclusion of the product averages as a covariate in the modeling and allowing the covariate regression coefficients to depend on the assessor. This gives a more powerful analysis by removing the scaling difference from the error term and proper confidence limits are deduced that include scaling difference in the error term to the proper extent. A meta study of 8619 sensory attributes from 369 sensory profile data sets from SensoBase (www.sensobase.fr) is conducted. In 45.3% of all attributes scaling heterogeneity is present (*P*-value <0.05). For the 33.9% of the attributes having a product difference *P*-value in an intermediate range by the traditional approach, the new approach resulted in a clearly more significant result for 42.3% of these cases. Overall, the new approach claimed significant product difference (*P*-value <0.05) for 66.1% of the attributes compared to the 60.3% of traditional approach. Still, the new, and non-symmetrical, confidence limits are more often wider than narrower compared to the classical ones: in 72.6% of all cases.

Introduction

Replicated quantitative descriptive sensory data, where I assessors scored J products in K replications, is frequently univariately analyzed by 2-way fixed analysis of variance (ANOVA) corresponding to the model

$$Y_{ijk} = \mu + \alpha_i + v_j + \gamma_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2) \quad (1)$$

where α_i is the assessor main effect, $i = 1, \ldots, I$, v_j the product main effect, $j = 1, \ldots, J$ and γ_{ij} the panelist-by-product interaction effect and $\varepsilon_{ijk}, k = 1, \ldots, K$ the random residual error with variance σ^2. Often the random panelist version of this, the mixed model ANOVA,

$$\begin{aligned} &Y_{ijk} = \mu + a_i + v_j + g_{ij} + \varepsilon_{ijk} \\ &\varepsilon_{ijk} \sim N(0, \sigma^2), \quad a_i \sim N(0, \sigma^2_{PAN}), \quad g_{ij} \sim N(0, \sigma^2_{PAN \times PROD}) \end{aligned} \quad (2)$$

is used for doing statistical inference regarding product differences and similarities (Lawless & Heymann, 2010; Lea, Næs, & Rødbotten, 1997; Næs, Tomic, & Brockhoff, 2010; Schlich, 1998). Here the roman letters are used for the random effects and σ^2_{PAN} is the variance of (centered) panelist effects and $\sigma^2_{PAN \times PROD}$ is the variance of panelist-by-product interaction effects. Essentially, this amounts to using the interaction mean square as error term instead of the residual error for product information inference. The choice between having effects related to the assessors as fixed or random in the model has been discussed in the sensory literature (Lawless, 1998; Lundahl & MacDaniel, 1988; O'Mahony, 1986). In Hunter (1996) as in Lawless and Heymann (2010) the general approach is that of the random assessor effect. Whereas both types of analysis certainly can be done with the proper interpretations of the results, we believe that most often in sensory QDA applications the random assessor effect type of interpretations resembles better the usual purpose of performing the experiment: to achieve at some results for the products in question that may be generalized to a larger setting than merely the assessors entering the used panel. It is clear that for this approach to be strictly valid, certain assumptions on how the replications were actually carried out have to be imposed. If these are not fulfilled, one would have to go for more complicated ANOVAs with e.g. a random effect of session and more interactions in play (see Næs et al., 2010). For this paper we restrict ourselves to

* Corresponding author.
E-mail addresses: perbb@dtu.dk (P.B. Brockhoff), pascal.schlich@dijon.inra.fr (P. Schlich), ims@life.ku.dk (I. Skovgaard).

http://dx.doi.org/10.1016/j.foodqual.2014.07.005
0950-3293/

discussing the 2-way setting as expressed here. We leave for future work to extend these ideas to situations calling for more complex structured models. Everything will be expressed for the typical design in sensory where products have been evaluated more than once by each assessor, that is, the replicated case. It will be clear that the non-replicated case can be handled similarly.

The interaction term of the model(s) captures the deviations from additivity (Panelist + Product) in the data. In other words the interaction term is modeling the potential individual differences between the panelists in their scorings of the product differences. This includes as well differences in individual ranges of scale use, the scale effect, as real differences in perception of product differences (disagreement effect).

In the model given in (1) these individual difference effects enter as many different fixed contributions γ_{ij} – one for each combination of panelist and product with certain technical restrictions due to the estimability/uniqueness of these. As such, this model does not assume any specific structure of these effects. In the model given in (2) on the other hand, it is assumed that these individual contributions g_{ij} follow a normal distribution with variance $\sigma^2_{PAN\times PROD}$.

So, for the univariate statistical analysis of sensory profile data with the purpose of achieving product information, this has been used for ages in sensory science and applications. Simultaneously, there is a large literature on studying and monitoring individual differences in sensory profile data. Also, there exist various ad-hoc procedures for how to proceed in a complete analysis process (see Schlich, 1997). There has been little or no attempts, however, to bring together in a common model framework the detailed modeling of individual differences with the "random-panelist-approach" of the typical mixed model ANOVA. It is the aim of this paper to create the modeling basis for this by means of well known and not too complicated statistical models. The models used are simple to the degree that they are within the classes of linear and mixed linear models for normally distributed data.

The benefits of this will be to be able to do improved product analysis – with higher power of detecting product difference and more correct product difference confidence bands. It also offers the basis for a merging of the panel performance analysis with the product difference analysis. This is beyond the scope of the this paper but has been pursued in detail in Peltier, Brockhoff, Visalli, and Schlich (2014).

In section 'Individual differences and statistical models', we introduce and discuss the key statistical models. Next, in section 'ANOVA decompositions', we show how simple ANOVA decompositions can form the basis of the computations as well for the overall analysis as for the panelist performance task. In section 'Statistical inference', the actual statistical inference procedures are presented including estimation, overall hypothesis testing and posthoc hypothesis testing. In section 'The MAM analysis approach', the computational challenges for the MAM analysis approach as such is discussed together with the presentation of two adaptations of the scale correction approach of the MAM. In section 'Example results', a detailed analysis example is provided using the data from the original assessor model paper (Brockhoff & Skovgaard, 1994) and in section 'SensoBase investigation', the results of the SensoBase meta study are reported. Finally section 'Summary and discussion' includes a summary and discussion of the entire paper including the overall recommendation for how to analyze sensory data. A number more technical details is given in three appendices. In Appendix A, more details on the statistical models and their relations to two other assessor model-like models are given. Appendix B contains small proofs of the contrast variance expressions from section 'Statistical inference', and Appendix C contains all the detailed development of the novel confidence band procedure.

Individual differences and statistical models

The model that we will use as the basis for (the main part of) the analysis, and the model we give the acronym MAM (Mixed Assessor Model) is the following:

$$\begin{aligned} &Y_{ijk} = \mu + a_i + v_j + \beta_i x_j + d_{ij} + \varepsilon_{ijk} \\ &a_i \sim N(0, \sigma^2_{PAN}), \quad \varepsilon_{ijk} \sim N(0, \sigma^2), \quad d_{ij} \sim N(0, \sigma^2_D) \end{aligned} \tag{3}$$

where $x_j = \bar{y}_{\cdot j \cdot} - \bar{y}_{\cdots}$ are the centered product averages inserted as a covariate, and hence β_i (with $\sum_{i=1}^{I} \beta_i = 0$) is the individual (scaling) slope. With the x_js considered fixed this is a simple mixed linear model, where the fixed part of the model is a regression line for each assessor versus the consensus product pattern. As will be clear the covariate simply has the effect to identify and remove the scaling heterogeneity from the interaction term, and it should not be seen as an analysis of covariance in the usual meaning, where one should be careful about including covariates that depend on treatment groups. In the MAM the term d_{ij} has taken the role of the random interaction term g_{ij} in the standard mixed model in (2), and we use the "d or D" to emphasize that now the term captures interactions that are not scale differences, hence "disagreements".

As it will be clear in the next sections, this model will produce valid and improved hypothesis tests for as well overall product differences as post hoc product difference testing. And the actual product differences are exactly those of the standard mixed model given in (2), only the inference part (P-values) has changed. However, it will also become clear that the model cannot in general produce valid post hoc product difference confidence intervals. We devote the Appendix C on details on a suggestion for obtaining better product difference confidence intervals.

We see the MAM as an approximation of the following more properly specified mixed model accounting for scaling heterogeneity. If the philosophy behind using the standard ANOVA mixed model given in (2) is accepted then the consequence would be that also the scale panelist effects should be considered random as the level panelist effect. Brockhoff (1994) presented one version of such a model. Another version would the following model:

$$\begin{aligned} &Y_{ijk} = \mu + a_i + v_j + b_i v_j + d_{ij} + \varepsilon_{ijk} \\ &a_i \sim N(0, \sigma^2_{PAN}), \quad b_i \sim N(0, \sigma^2_{SCALE}), \quad d_{ij} \sim N(0, \sigma^2_D), \\ &\varepsilon_{ijk} \sim N(0, \sigma^2) \end{aligned} \tag{4}$$

where the scaling differences b_i are now modeled as random effects and where the covariate is the true model product effects v_j instead of the x_j. This is a non-linear mixed model, a so-called mixed multiplicative model, and hence not in the class of linear mixed models. The main effects of interest: the unknown product values v_j enter both the expectation and the variance structure of this model. Multiplicative mixed models were introduced in the sensory context in Smith, Cullis, Brockhoff, and Thompson (2003) but there only in versions where the assessor effects were fixed and the product effect random. Whereas there could certainly be situations where the product effects could be assumed random, e.g. if products represent the variability of a certain food like a fruit type of fish type, we believe that the assessor effect should be random. And there is definitely also the need to be able to handle fixed product effects, which is more or less consensus, when performing sensory profile data ANOVAs.

No doubt, the model given in (4) could be handled by likelihood methods using techniques developed for non-linear mixed models. This is not the aim of this paper however, instead we provide a simple linear model approximation of this, and leave it for future research to investigate what potential improvements could be obtained by the full likelihood approach. The benefit of the approximative approach is threefold: (1) the theory and computations simplify immensely; (2) the insight and transparency provided by the simplified theory