Food Quality and Preference 32 (2014) 126–131

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

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

Consideration of sample heterogeneity and in-depth analysis of individual differences in sensory analysis

Cécile Bavay [a,*], Per Bruun Brockhoff [b], Alexandra Kuznetsova [b], Isabelle Maître [a], Emira Mehinagic [a], Ronan Symoneaux [a]

[a] LUNAM Université, SFR QUASAV 4207, Groupe ESA, UPSP GRAPPE, 55 Rue Rabelais BP 30748, F-49007 Angers Cedex 01, France
[b] DTU Informatics, Statistical Section, Technical University of Denmark, Richard Petersens Plads, Building 305, DK-2800 Kongens Lyngby, Denmark

ARTICLE INFO

Article history:
Received 3 October 2012
Received in revised form 20 February 2013
Accepted 7 June 2013
Available online 14 June 2013

Keywords:
Assessor model
Mixed model
Sample heterogeneity
Sensory profile
Variability

ABSTRACT

In descriptive sensory analysis, large variations may be observed between scores. Individual differences between assessors have been identified as one cause for these variations. Much work has been done on modeling these differences and accounting for them through analysis of variance (ANOVA). When the products studied are prone to biological heterogeneity (e.g. fruits, vegetables, cheeses, etc.), variations in the data may be due to assessor differences and/or product heterogeneity. The present paper proposes an approach for quantifying these two sources of variation. For individual differences, an extended version of the assessor model approach is applied. The data set used in the paper is based on sensory evaluations of three apple samples scored by a panel of 19 assessors using seven descriptors in four replicates. The application of the extended assessor model approach to unbalanced data provides more insight into assessor differences and a better test for product differences. These results demonstrate the importance of choosing the right model and taking all potential sources of variation into account.

1. Introduction

Sensory quality is commonly assessed through conventional sensory profiling methods. The data resulting from such methods present variations that may be due to assessor differences and/or sample heterogeneity. On one hand, individual differences between assessors are an inherent source of variation. For example, assessors may vary in both their perception and their use of the intensity scale. They may differ in their average (level), in their dispersion on the scale (scaling), in their repeatability (variability) and even in their ranking of the products (disagreement) (Brockhoff, 2003). Training may reduce, but not erase, these effects. This issue of assessor differences has been investigated by many authors (Brockhoff, 1998, 2003; Brockhoff & Skovgaard, 1994; Næs, 1990, 1998; Næs & Solheim, 1991; Romano, Brockhoff, Hersleth, Tomic, & Næs, 2008; Schlich, 1994, 1996). On the other hand, products such as fruits, vegetables, cheeses, etc. are prone to biological heterogeneity. Many authors have highlighted this issue in studies about apples. For example, in a study to develop a specific sensory methodology for the assessment of Cox's Orange Pippin apples, Williams and Carter (1977) reported difficulties in drawing conclusions due to the uncertainty in determining the sources of the variations in the results. Clearly, assessor differences or apple heterogeneity could be responsible for these variations. Moreover, according to Hampson et al. (2000) who studied genotype differences from a sensory point of view, apple heterogeneity may cause difficulty in differentiating samples. In fact, real variations within a given genotype may make differences among genotypes harder to detect.

To study sample differences in sensory evaluation, a version of a mixed model analysis of variance (ANOVA) is commonly performed for each attribute. ANOVA models have been discussed and debated to take into account the particular nature of sensory data better, such as assessor differences, replicates, etc. Despite the development of specific analyses that meet the special requirements of sensory data, such as the assessor model (Brockhoff, 2003; Brockhoff & Skovgaard, 1994) and its extended version (Brockhoff, Schlich, & Skovgaard, 2013), the standard model is most often used. This analysis includes a fixed sample effect, a random assessor effect and the interaction between sample and assessor. This model is applied in order to obtain information about the samples while accounting properly for possible assessor differences. However, the interaction term accounts for both disagreement and scaling differences and is generally falsely interpreted as only disagreement. Moreover, in statistical data analysis, it is good practice to model all the different effects involved. So, regarding sensory data, the scaling effect and the unit effect, in the

* Corresponding author. Address: GROUPE ESA, Unité de Recherche GRAPPE, Ecole Supérieure d'Agriculture d'Angers' 55 Rue Rabelais, BP 30748, F-49007 Angers Cedex 01, France. Tel.: +33 2 41 23 55 55; fax: +33 2 41 23 55 00.
E-mail address: c.bavay@groupe-esa.com (C. Bavay).

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.06.003

specific case of a heterogeneous product, should be included in the analysis.

The focus of the present paper is the modeling of the variability of the sensory response in descriptive sensory analysis in order to understand the results better. Data are analyzed using three models: the standard model, a model considering sample heterogeneity and the extended mixed assessor model approach. The contributions of the model considering heterogeneity are investigated in comparison with the standard model. Then, the contributions of the extended assessor model approach, in combination with accounting for the complex replication structure, are studied. Inclusion of within-sample heterogeneity may, with good reason, affect other noise parts of the model, e.g. the important assessor-by-sample interaction. The main goal of the scaling correction, which comes from using the assessor model approach, is to affect this interaction. We demonstrate the possibility and the advantage of using the assessor model approach in combination with the inclusion of other effects, such as within-sample heterogeneity. Using the assessor model in combination with accounting for this more complex and unbalanced sample replication structure is novel.

2. Materials and methods

2.1. Models

Model analyses are run with the step function of the lmerTest package (Kuznetsova, Christensen, Bavay, & Brockhoff, 2013; Kuznetsova, Christensen, & Brockhoff, 2012) using the R software (version 2.14.2) (R Core Team, 2012). The *F* test and log likelihood ratio test are applied to test for fixed effects and for random effects, respectively. R codes are provided in the appendix for useRs.

2.1.1. Standard model

The standard model for analyzing sensory data is:

$$X_{asr} = \upsilon_s + \alpha_a + \gamma_{as} + \varepsilon_{asr} \qquad (1)$$

where $\alpha_a \sim N(0,\ \sigma^2_{assessor})$, $\gamma_{as} \sim N(0,\ \sigma^2_{sample*assessor})$ and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

The model includes a fixed sample effect υ_s, a random assessor effect α_s and the interaction between sample and assessor γ_{as}. The *r* subscript accounts for random replicates. This model is applied in order to obtain information about the samples while accounting properly for possible assessor differences. The assessor effect accounts for level differences while the interaction term accounts for both disagreement and scaling differences.

2.1.2. Model considering within-sample heterogeneity

Consider a sensory experiment with no session effect and a simple one-way sample structure. Each sample is made up of several units (e.g. individual fruits within an apple cultivar). These units may present differences and we therefore want to take into account the main unit effect. With that aim, a random unit effect nested within the sample effect $\delta_{u(asr)}$ (e.g. a fruit nested within an apple cultivar) is introduced into the model:

$$X_{asr} = \upsilon_s + \alpha_a + \gamma_{as} + \delta_{u(asr)} + \varepsilon_{asr} \qquad (2)$$

where $\alpha_a \sim N(0,\ \sigma^2_{assessor})$, $\gamma_{as} \sim N(0,\ \sigma^2_{assessor*sample})$, $\delta_{u(asr)} \sim N(0,\ \sigma^2_{unit})$ and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

As in model (1), the *r* subscript accounts for replications of the measurement. In our example, replicates consist of pieces of apple (apples are units and each apple is cut into pieces). In the subscript $u(asr)$, u accounts for the actual unit (a fruit) and *asr* indicates the numbering of the actual observation.

Model (2) takes into account assessor level differences, assessor scaling differences and disagreement included in the interaction term and actual unit differences. The introduction of the unit effect is made possible by having a single unit rated by several assessors.

2.1.3. Assessor model approach

The original assessor model for sensory data was proposed by Brockhoff and Skovgaard (1994) and only included the scaling differences (as fixed effects):

$$X_{asr} = \alpha_a + \upsilon_s \cdot \beta_a + \varepsilon_{asr} \qquad (3)$$

where $\varepsilon_{asr} \sim N(0,\ \sigma^2_a)$; all terms are independent.

The model comprises a fixed sample effect υ_s, a fixed assessor effect α_a and the individual scaling coefficient β_a.

In Brockhoff (2003), this model together with other different models were further developed into an approach for univariate assessor performance investigations. In Brockhoff et al. (2013), an extended version of the assessor model including a random interaction (disagreement) effect was presented using the centered sample means m_s as covariates:

$$X_{asr} = \alpha_a + \upsilon_s + m_s \cdot \beta_a + \gamma_{as} + \varepsilon_{asr} \qquad (4)$$

where $\gamma_{as} \sim N(0,\ \sigma^2_{assessor*sample})$, and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

This model includes a fixed sample effect υ_s, a random assessor effect α_s, the interaction between sample and assessor γ_{as}, the individual scaling coefficient β_a and the centered sample means m_s. In their paper, the authors show how proper hypothesis testing for sample comparisons can be based on this model, which amounts to simply removing the scaling part of the interaction by linear regression (although proper confidence bands would require more complicated computations).

To investigate the consequence of both the scaling/disagreement decomposition and the unit effect, we would like to extend the model in (5) with the random unit effect:

$$X_{asr} = \alpha_a + \upsilon_s + m_s \cdot \beta_a + \gamma_{as} + \delta_{u(asr)} + \varepsilon_{asr} \qquad (5)$$

where $\gamma_{as} \sim N(0,\ \sigma^2_{assessor*sample})$, $\delta_{u(asr)} \sim N(0,\ \sigma^2_{unit})$ and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

It is beyond the scope of the present paper to provide a full methodological treatment of this model (5) applied to unbalanced data, which has not been presented in the literature with its extension. We apply here a simple approach to investigate both effects. We construct a processed version of the data where we have removed (additively) the scaling part of the interaction, similar to the "additive approach" suggested and discussed in Romano et al. (2008). Here, this is done by applying the version of the assessor model (4), where the scaling effects are estimated based on the centered sample means m_s and then subtracted from the data:

$$X_{asr} - m_s \cdot (\beta_a - \overline{\beta}) \qquad (6)$$

2.1.4. Random component models

To study the relative sizes of the various effects, a version of models (1) and (2) above, where all effects are considered random (also the sample effect), is applied:

$$X_{asr} = \upsilon_s + \alpha_a + \gamma_{as} + \varepsilon_{asr} \qquad (7)$$

where $\upsilon_s \sim N(0,\ \sigma^2_{sample})$, $\alpha_a \sim N(0,\ \sigma^2_{assessor})$, $\gamma_{as} \sim N(0,\ \sigma^2_{assessor*sample})$ and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

$$X_{asr} = \upsilon_s + \alpha_a + \gamma_{as} + \delta_{u(asr)} + \varepsilon_{asr} \qquad (8)$$

where $\upsilon_s \sim N(0,\ \sigma^2_{sample})$, $\alpha_a \sim N(0,\ \sigma^2_{assessor})$, $\gamma_{as} \sim N(0,\ \sigma^2_{assessor*sample})$, $\delta_{u(asr)} \sim N(0,\ \sigma^2_{unit})$ and $\varepsilon_{asr} \sim N(0,\ \sigma^2)$; all terms are independent.

The last model, including the unit effect, is then applied to both the original data and the data with the scaling removed. For the