Food Quality and Preference 63 (2018) 129–134

Contents lists available at ScienceDirect

Food Quality and Preference

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

Do we need to replicate in sensory profiling studies?

C. Peltier[a,b,*], N. Mammasse[a], M. Visalli[a], S. Cordelle[a], P. Schlich[a]

[a] Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France
[b] Agrosup Dijon, F-21000 Dijon, France

A R T I C L E I N F O

Keywords:
Sensory profiling
Replicate
Correlation
ANOVA
MANOVA
RV
CVA
Meta-analysis
SensoBase

A B S T R A C T

Several evaluations are necessary to test individual repeatability in sensory profiling. But do additional replicates really provide more information than the first evaluation? Statistical analysis of 339 studies with at least two evaluations showed that when using the first evaluation alone:

- Significance of product discrimination was identical for 6 out of 7 attributes,
- 85% of the significant attributes had a similar vector of mean scores *ie* a correlation coefficient higher than 0.9,
- Significance of multivariate product discrimination assessed by MANOVA was identical for 97% of the datasets,
- 93% of the datasets had a similar product configuration (RV > 0.9),
- the CVA product maps were similar (according to several criteria including product configurations, attribute positions…).

Further, the second evaluation was found less relevant in studies including a small number of products (< 5). Finally, a third evaluation seemed irrelevant for the improvement of the statistical analysis.

1. Introduction

Sensory profiling has been widely used in sensory analysis. The aim of the method is to describe the sensory characteristics of food products using a list of defined attributes. Panelists involved in a sensory profiling task are required to rate the perceived intensities of a number of attributes such as: sweetness, saltiness, crunchiness, etc. A classical sensory descriptive profile consists of two main phases: the training phase and the testing phase (also known as the measurement phase).

During these phases, the panelists often evaluate the set of products more than once. In this paper, we state that there are k replicates when a product is scored k times by a panelist. Thus, a dataset with only one evaluation per 3-tuple (product; panelist; attribute) corresponds to a dataset with one replicate.

Replicates allow panel performance to be tested. According to Stone and Sidel (2004), it represents an internal check of the response system, enabling the panel leader to determine individual subject and panel consistency and to establish a numerical measure of performance. According to the same paper, from an empirical point of view, about four replicates are required to be optimal; but this can depend on the magnitude of differences among products, subject sensitivity or objective of the test. Fewer replicates could be sufficient to provide directions in the early stages of product development. However, it would be extremely risky to rely on descriptive information for which there are no repeated judgments (Stone & Sidel, 2004). Besides, Naes, Brockhoff, and Tomic (2010) claimed that during a profile sensory study, each sample is tested in duplicate or triplicate.

Choosing the number of replicates involves some practical considerations, such as the potential differences within replicated samples, the choice of the number of samples to be tested by assessors without causing sensory fatigue, or the increase of the financial cost of the study. Further, replications obviously extend the time required to complete a test. Consequently, by choosing the number of replicates in his study, the panel leader must find a compromise between the reliability of his data and the cost of the study.

Pineau et al. (2013) claimed that replicates are not needed in descriptive sensory studies once the panel is well trained. Their study was based on several datasets, where averages of scores (for each pair product*attribute) with or without replicate were graphically compared, and the discriminative power were also graphically compared. The exact number of used datasets was not specified, but graphically appeared as lower than 10.

Learning from actual data is essential in order to improve sensory methodologies and to investigate different points of view, such as the

* Corresponding author at: Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France.
E-mail address: caroline.peltier@agrosupdijon.fr (C. Peltier).

http://dx.doi.org/10.1016/j.foodqual.2017.09.001
Received 21 December 2016; Received in revised form 1 September 2017; Accepted 1 September 2017
Available online 06 September 2017
0950-3293/

usefulness of replicate in sensory profiling. In this paper, we will examine the impact of replicates on statistical outputs using a large number of sensory profile studies from the SensoBase, a database of sensory profiling. The key question in our research is: to what extent additional replicates would change the outcomes of the studies based on the first evaluation only?

For each study, univariate and multivariate statistical analyses were carried out to compare the full datasets (with the first two replicates, henceforth **2-rep datasets**) and the truncated datasets (subset of the data including only the first replicate, henceforth **1-rep datasets**). The impact of the sensory modality of the attributes (texture, visual, taste…) on the usefulness of the replicate will also be studied.

The same methodology was used to test whether the differences between 1-rep and 2-rep datasets increase with the number of products and subjects in the dataset. Furthermore, for datasets with three replicates, the differences between the datasets constituted of the first two evaluations (**2-rep datasets**) and the whole datasets (**3-rep datasets**) were also studied.

2. Material and methods

2.1. Procedure

SensoBase is a database of descriptive sensory studies, currently containing 1 169 datasets, and created and maintained at Centre des Sciences du Goût et de l'Alimentation. Each dataset contains the scores of a set of products of a given type profiled by a panel of assessors. The datasets were collected *via* Internet. In exchange for the inclusion of one dataset in SensoBase, the provider of the dataset received outputs from statistical analyses (CAPTable, FlashTable…) (Pineau, 2006; Porcherot & Schlich, 2000). The database was created in order to describe sensory analysis practices, to compare sensometric techniques of analysis and to benchmark panel and panelist performances.

Datasets were selected from SensoBase as satisfying the following conditions:

– containing at least two replicates
– containing between 3 and 20 products
– containing more than 4 panelists
– containing less than 100 attributes.

These datasets were chosen fully balanced to avoid potential statistical or computational issues. The restricted database resulted in 339 datasets, representing 2504 panelists, 1895 products and 7434 descriptors. 48 datasets contained more than 2 replicates. The distributions of the number of subjects, products and attributes are presented in Fig. 1.

The 1-rep and 2-rep datasets were compared with univariate and multivariate criteria.

Univariate criteria included a comparison of the discrimination of the products (detection of statistical differences between products) in 1-rep and 2-rep datasets, and the calculation of the degree of agreement between 1-rep and 2-rep datasets.

Multivariate criteria included product discrimination, agreement and dimensionality of the product structures (indicating the complexity of the product space). Product maps obtained with 1-rep or 2-rep datasets were also compared with several criteria presented in Peltier, Visalli, and Schlich (2015b).

2.2. Univariate analysis

2.2.1. Panel discrimination

A two-way analysis of variance (ANOVA) model with product and subject as factors was carried out on the intensity scores for each attribute. Analyses with a single replicate were done thanks to the additive ANOVA model, whereas those with more replicates used the mixed model with subject and product by subject interaction as random factor (as recommended by Næs and Langsrud (1998)). The significance of the product effect was calculated to attest the discrimination of products for different thresholds.

Note that performing an ANOVA on balanced data with the mixed two-way model with subject as a random effect is equivalent to averaging the replicates and then performing an additive two-way ANOVA (Peltier, Visalli, & Schlich, 2015a).

2.2.2. Agreement between replicates

For each attribute, the Pearson correlation coefficient was calculated between the vector of product mean scores for the 2-rep dataset and the vector of product mean scores for the 1-rep dataset. As the agreement between subjects about differences between products makes sense only if the products are significantly different, its distribution was presented for significant attributes ($p = 0.1$) only.

2.2.3. Impact of the sensory modality on the need to be replicated

The usefulness of the replicates could depend on the sensory modality of the scored attribute. To check this hypothesis, the distributions of attributes by their sensory modalities were calculated:

– for the whole set of attributes
– for the attributes being significant in the 2-rep dataset, but not in the 1-rep dataset.

The two different distributions were compared with a Chi^2 test in order to determine whether replicates are more useful for some sensory modalities than for others.

The effect of the sensory modality on the correlation coefficients presented in 2.2.2. was studied by an ANOVA based on the one-way model: correlation ~ sensory modality. Here, the correlation corresponds to the Fisher transformation of the Pearson coefficients: $f(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$. Thus, the transformed coefficients follow a normal distribution. This analysis was followed by a Tukey's Honestly Significant Difference (HSD) test ($p = 0.05$) in order to determine groups of sensory modalities according to these correlations coefficients. This showed the links between the agreement across replicates and the sensory modality of the descriptor.

2.3. Multivariate analysis

2.3.1. Agreement between replicates: RV-coefficient

For each study, the product spaces derived from the 1-rep and 2-rep datasets were compared. After averaging over assessors, the RV coefficient (Escoufier, 1973) was computed to measure the similarity between the two product configurations. The average of RV coefficients was calculated and named as multivariate agreement. As the RV coefficient can be very sensitive to the number of products in the dataset, we also calculated its significance with the Pearson type III approximation recommended in Josse, Pagès, and Husson (2008).

2.3.2. Panel discrimination

Multivariate analysis of variance (MANOVA) measures product discrimination when all sensory attributes are considered simultaneously (Mardia, Kent, & Bibby, 1989). A significant MANOVA F-test (with the Hotelling-Lawley trace) indicates that the products differ significantly in the space generated by the attributes.

Then, the ratio of the F obtained for the 2-rep dataset and the F obtained for the 1-rep dataset was calculated (noted F_{2-rep}/F_{1-rep} and also used in Peltier et al. (2015b)). A ratio higher than 1 means that the 2-rep dataset was more discriminative than the 1-rep dataset. As this ratio could be very high for some datasets, the median of the F-ratio distribution was chosen as the comparison indicator of the multivariate discrimination for several datasets (instead of the average, very