

Replicates in sensory profiling: one vs. two assessments

Mireille Moser, Mélissa Lepage, Nicolas Pineau & Andréas Rytz

Nestlé Research Center, Lausanne, Switzerland

E-mail : *mireille.moser@rdls.nestle.com, Melissa.lepage@rdls.nestle.com*

Abstract

Monadic (descriptive) profiling is a commonly used sensory method in food industry to describe product sensory characteristics. Usual recommendation is to use a panel of trained panelists, generally with around ten people and to assess the products at least in duplicate to get robust average estimates. In practice, in order to optimize the use of the panels, the necessity of duplicates was challenged. This work presents a comparison of results obtained with one vs. two assessments, based on 15 studies featuring at least 8 products, representing diverse product categories, diverse panels, a total of 380 attributes and 2836 comparisons of product pairs. It is shown that average estimates are very similar for one vs. two assessments: correlations are very high for all attributes that discriminate products. It is also shown that 85.5% of pairwise comparisons lead to the same conclusions for one and two assessments. The remaining 14.5% are showing no systematic trend (i.e. 5.4% more discriminating for one assessment vs. 8.3% more discriminant with two assessments). As a conclusion, one assessment is considered sufficient in various cases.

Keywords: monadic sensory profile, replicates, estimates, discrimination

1. Introduction

Monadic (descriptive) profiling is a commonly used sensory method in food industry to describe product sensory characteristics. Usual recommendation is to use a panel of trained panelists, generally with around ten people and to assess the products at least in duplicate to get robust average estimates [1]. In practice, in order to optimize the use of the panels, the necessity of duplicates was challenged. A first internal study [2] based on one study showed that product estimates based on one assessment were very close from the estimates based on two assessments. In addition the loss of power to detect significant differences was relatively low, even though not always negligible. To go further in this direction and quantify this loss of discrimination ability in a more robust way, we propose here to use two statistical indexes:

- the correlation between the estimated product averages with one assessment and the estimated product average with two assessments
- the contingency tables assessing the consistency of the conclusions for each pair of products (Is my product A significantly higher (resp. lower) than my product B for a given sensory attribute?) between the datasets with one vs. two assessments.

Results will be presented based on a benchmark of 15 studies featuring at least 8 products, representing diverse product categories (including beverages, food and petfood), diverse panels, a total of 380 attributes and 2836 comparisons of product pairs. Table 1 indicates the main characteristics of these 15 studies that all featured two assessments (i.e. it is possible to consider both of them or only the first assessment).

Study ID	Nb panelists	Nb products	Nb attributes	Nb rep
Average	11	17	25	2
1	12	8	17	2
2	11	9	20	2
3	12	9	30	2
4	11	11	23	2
5	11	11	23	2
6	8	12	41	2
7	9	12	21	2
8	11	12	14	2
9	12	13	22	2
10	12	15	17	2
11	12	18	23	2
12	9	24	59	2
13	12	26	24	2
14	12	36	24	2
15	11	42	22	2

Table 1: Description of the 15 studies (number of panelists, products, attributes, and replicates)

2. Comparing one vs. two assessments

2.1 Correlations between 1 vs. 2 assessments for average estimates

In figure 1, one can see that there are in majority high correlations between the estimated product averages for one assessment and the estimated product average for two assessments. Lower correlations (less than 0.8) are associated with small \sqrt{F} , in other terms with non-significant attributes. Therefore, having only one assessment enables then to draw similar conclusions on significant attributes than using two assessments.

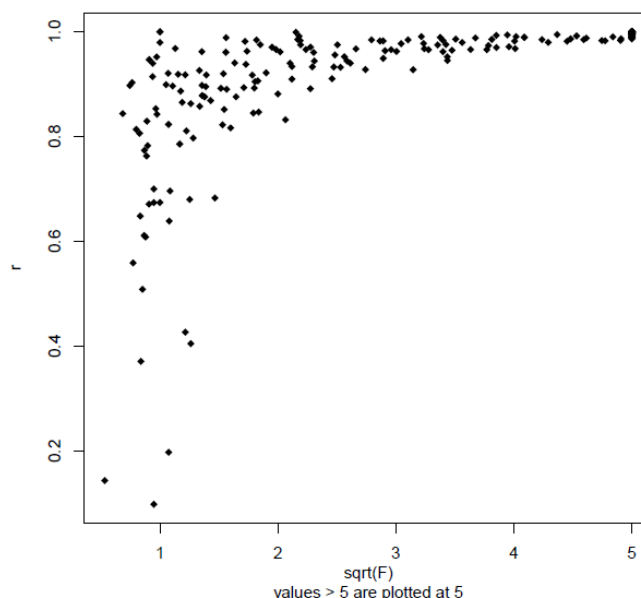


Figure 1: Association between the Pearson's correlations coefficients between the estimated product averages for one vs. two assessments (Y) and the square root of the Fisher value to evaluate the Product effect in the 2-way ANOVA with the product (fix) and subject (random) effects (X). Each dot corresponds to a sensory attribute (380 attributes in 15 total studies)

2.2 Consistencies between 1 vs. 2 assessments for pairwise product comparisons

Table 2 quantifies the consistency of the results obtained for each comparison of two products (namely A and B) between one and two assessments. Globally, in 85.5% of the cases, conclusions on the pairwise products are similar between the full dataset and the dataset with only the first replicate. Out of the 14% 5.4% are more discriminating for one assessment and 8.3% are more discriminant with two assessments

		2 assessments		
		A < B	n.s.*	A > B
1st assessment only	A < B	20.1%	2.7%	0.4%
	n.s.*	4.2%	46.0%	4.1%
	A > B	0.4%	2.7%	19.4%

Color code for comparison between 1 and 2 assess^{mt}

Consistent difference between the product pairs
Signif. with 1st assessment, not with 2
Signif. with 2 assessments, not with the 1st
Signif. in both case, but in opposite direction

Table 2: Contingency table based on 76116 pairwise comparisons coming from 15 studies *NS: Non significant result, significance level has been set to 5% (LSD post-hoc comparison...)

3. Conclusion

According to these results, one assessment is generally sufficient to have a robust conclusion. In the following 2 cases, we strongly recommend the usage of only one assessment in order to optimize the usage of the panels:

- Sensory profiles assessing products that were designed using sound Design of Experiment (DoE) techniques: the effects of factors are anyway based on several products (i.e. the difference between the means of product and not the difference between 2 products) and therefore the repetitions brings less information than the evaluation of a few product more.
- Sensory profiles are used to correlate with other factors (ex: consumers liking), as in this case the importance is the “story” the sensory tells and this is unchanged with 1 or 2 replicates as shown by the very high correlations between one or two assessment for average estimates.

References

- [1] AFNOR, Ed. (1993). *Guide général pour la sélection, l’entraînement et le contrôle des sujets. Partie 1 : Sujets qualifiés. Analyse Sensorielle*. Paris – La Défense, AFNOR.
- [2] Pineau, N., Moser, M., Graham, V., Arslanian, R., Fillion, L., Rytz, A. (2013). *No need for replicates in descriptive sensory studies once the panel is well trained*. Pangborn symposium, Rio, Brazil. Poster presentation.