Proposal for addressing parameter effect significance in a factorial design of experiment with sensory data

Mireille Moser[1], Thorn Thaler[2], Nicolas Pineau[1], Andreas Rytz[1]

[1] *Nestlé Research Center, Lausanne, Switzerland*
E-mail: *mireille.moser@rdls.nestle.com*

[2] *Nestlé, PTC Singen, Germany*
E-mail: *thorn.thaler@rdsi.nestle.com*

Abstract

When applying some factorial design of experiment (DoE) approach in the food industry context, the designs are often close to saturation (resolution III or IV) in order to optimize the cost efficiency of the experiments. One of the drawbacks of these almost saturated designs is the very low number of residual degrees of freedom to estimate the error term, which means a potentially poor estimate of the "true" residual variability. On top of this, the error term is actually the sum of the interactions supposed to be negligible, and in case some of them are not, the error term will not really represent the variability of the measure (over-estimation of the error term).

Focusing on the case where the output measurement is sensory profiling, it is possible to use the raw data (at panelist level) to get an error term based on the panel variability. But building a model on these data to estimate the parameter effects of the DoE will drive to an error that is a mix between the error from the production variability, the method (sensory evaluation) and the remaining interactions between the parameters of the DoE. In order to give conclusions on the parameter effects based on the sensory relevance, it is therefore proposed to run the analysis in two steps:

- Run the usual model for sensory data, "forgetting" the DoE structure (e.g. 2-way ANOVA) and calculate the Least Significant Difference (LSD) accordingly
- Use the LSD of this first model to estimate the significance of the effects of the DoE parameters, and accounting for the higher number of values when estimating a parameter effect than when estimating single product difference (LSD/√n, n being the number of samples under each level of the parameter)

It is finally proposed to represent the parameters effects in a bar chart with a 3-level color code to differentiate three cases for significance:

- No parameter effect at all (effect < LSD/√n)
- Significant parameter effect, but not big enough to be detected by the sensory panel if this is the only change in the product (in between)
- Significant parameter effect that could be directly detected by the sensory panel if it is the only change between two products (effect > LSD).

Keywords: Least Significant Difference, Design of Experiment, Sensory

1. Introduction

In the context of an Innovation/Renovation project in Nestlé R&D, sensory properties of the samples are almost always measured (monadic profiling) and communicated to a very large audience, i.e. from the project manager or the scientists working on the project to the business partners. Communication of the sensory results, in particular regarding the significance of the difference between samples on each sensory dimension, must therefore be very clear and understandable by everybody. In the very same context, Design of Experiments (DoE) are also very often used, in order to optimize the number of samples to be produced and tasted while getting clear estimates of the effects of recipe or process parameters on the final product. In this framework, and in order to communicate as clear and actionable results as possible to all partners, choices have been made internally to present sensory results, and in particular in a DoE context.
This paper is therefore sharing:

- The reasons why the Least Significant Difference (LSD) has been chosen as the reference post-hoc test for sensory studies in Nestlé R&D context
- The proposed approach to represent the effects from a DoE with sensory outcomes and a way to deal with the significance of these effects

2. Reasons for Least Significance difference post-hoc test

As proposed in many textbooks dealing with the analysis of sensory data (e.g. Piggot,), the ANOVA model used to analyze monadic profiling sensory data is the 2-way ANOVA with the product (fix) and subject (random) effects, and their interactions in case data contains repetitions. Most of the post-hoc tests are then using the Mean Square of the error (MSE) when there is no repetitions, or the Mean Square of the Interaction (MSI) when there are repetitions, as the basis to determine which product averages are significantly different from each other. To simplify the reading of the paper, and because the initial model does not impact the following, both MS discussed above will not be differentiated and called MSE, and refer to the "error" in a general way.

Among the post-hoc tests, the LSD is one of the simplest one, and can be computed as follow:

$$LSD = qt(0.975; df_{\ error}) * \sqrt{\frac{2 * MSE}{n_{eval}}}$$

With qt, the quantile of the Student law, df_{error}; the degrees of freedom of the error term in the ANOVA model and n_{eval}, the number of evaluations for each product.

The LSD value can be interpreted as the minimum difference between two product averages so that they can be declared as significantly different from each other. This corresponds to a post-hoc comparison without correction for multiplicity.

This choice was done because it is not conservative and in Nestlé R&D context, monadic profiling are used to describe products and not to prove/support "claims" on products. There is therefore no reason to be strict on the comparisons, and to use a test that would penalize the high number of products tested (and so high number of possible "pairs" to test). In addition, the LSD is an easy way to communicate about the significance of the differences, since it can be interpreted as a threshold value to reach significance valid for any pair of products, whereas some other procedures give one p-value for each paired comparison.

3. Representation of DoE parameter effect with sensory data

The two most common modelling approaches that can be used when the output of a Design of Experiments is sensory data are:

- Modelling on raw data (i.e. with individual panelist scores), accounting for the DoE parameters and the subject effect, and eventually interactions. The significance of the DoE parameters effects is then assessed based on the model error.
- Modelling on mean data (i.e. mean sensory profiles acoss subjects), accounting only for the DoE parameter effects. The significance of the DoE parameters effects is then assessed based on the model error as well.

However in the first case, the error from the raw data model corresponds to a mix between different sources of error, in particular the sample heterogenity, the measurement variability from the sensory evaluation, and non-estimated interactions between DoE parameters supposed to be negligible. The interpretation of the error used to test the significance of the parameter effects is therefore difficult.

In the second case, since DoE are often saturated (or close to saturation), only a low number of degrees of freedom remains to estimate the error, which makes it not particularly robust. In addition, this error is actually the sum of the interactions not accounted for in the DoE. So depending on the interactions considered in the model, the error term can vary significantly. Or if the error contains interactions that are actually non negligible, the error term is over-estimated.

For these reasons, it is proposed to perform the analysis in two steps:

- Run the usual 2-way ANOVA model on the raw data as described in section 2. (i.e. ignoring the DoE structure) and compute the LSD based on this model
- Use this LSD to determine the significance of the effect of the parameters (parameters' effect size remain unchanged whatever the model in a DoE context)

In this way, the significance of the effects will be assessed according to the ability of the panel to detect a sensory difference between samples, and this will not by biased by questions around the DoE parameter interactions to be considered in the model.

Following on this approach, it is proposed to use two "cut off" values to assess the importance of a DoE parameter effect:

- The **LSD** from the 2-way ANOVA model. In this regard, a DoE parameter effect greater than the LSD can be interpreted as an effect that can be perceived by the sensory panel if the only change between two products is related to the given DoE parameter
- The **LSD√n**, i.e. LSD divided by the root square of the number of products per level of a DoE parameter. This is based on the idea that the effect is not only estimated on the difference between 2 products but between 2 means of several samples. For instance, for a 2-level DoE parameter in a design with 8 samples, the estimate of the DoE parameter effect is actually the difference between the estimates for the two levels, both being calculated as the average of 4 of the samples. In this case, the threshold proposed here would be LSD/√4. A DoE parameter with an effect size greater than this threshold cannot be directly detected by a sensory panel if this is the only difference between two products, but it can contribute to a significant change if associated with other effects.

It is worth mentioning that in case of independent measurements (physical/chemical) the gain between estimating the effect of a DoE parameter and the effect due to a single product is proportional to $\sqrt{n}$. In case of sensory measurement, this gain will still hold true if we can assume the hypothesis of independence between product measurements is valid. If not, the true gain would be smaller than $\sqrt{n}$.

As displayed in Figure 1, it is finally proposed to represent the parameters effects in a bar chart with a 3-level color code to differentiate three cases for significance:

- No parameter effect at all (effect < LSD/$\sqrt{n}$) – grey/neutral
- Significant parameter effect, but not big enough to be detected by the sensory panel if this is the only change in the product (in between) – light color
- Significant parameter effect that could be directly detected by the sensory panel if it is the only change between two products (effect > LSD) – dark color.

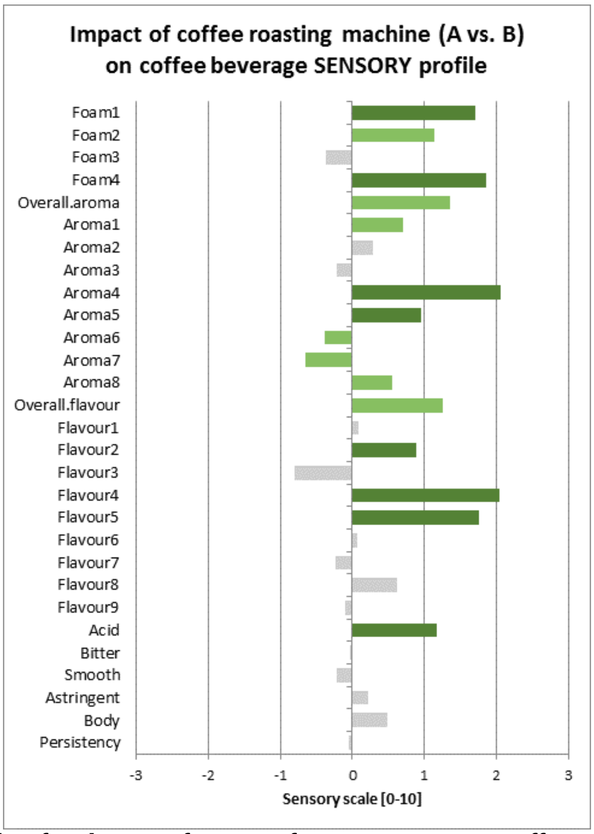

Figure 1: example of display for the significance of a DoE parameter effect with sensory data

Thanks to this kind of output, it is easy to read the impact of a given parameter (type of coffee roasting machine) on the sensory profile. In this example, foam properties are highly impacted by the roasting machine (Foam1 and Foam4 strongly impacted, Foam2 impacted to a lower extent), as well as the overall aroma and flavor, in particular regarding attributes 4 and 5.

References

J.R. Piggot, Statistical procedures in food research (1987). Elsevier applied science, p.39-52