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, Vers-chez-les-blanc - Lausanne, Switzerland
[2]Nestlé Product Technology Center, Singen Germany

Nestlé Research
AgroStat 2016
Lausanne, March 21-24 2016
SFdS Société Française de Statistique

Introduction

In the context of an Innovation/Renovation project in Nestlé R&D, sensory properties of the samples are almost always measured (monadic descriptive 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 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 poster 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

1. Reasons for Least Significance Difference (LSD) post-hoc test

ANOVA models generally used in Sensory context:

1 assess^mt^	model	obs ~	product + (fix)	subject + (random)	Error	
	df		p-1	s-1	(p-1)(s-1)	
					ǁ	
2 or more assess^mts^	model	obs ~	product + (fix)	subject + (random)	product x panelist +	Error
	df		p-1	s-1	(p-1)(s-1)	ps(r-1)**

Term used as the error to calculate the significance of the product effect*

* associated Mean Squares will be called "MSE" in both case in this poster
** r: number of assessments. df calculation if same # assessments per prod*subj

LSD formula

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

LSD = Minimum difference required between two sample means to reach significance

Reasons for this choice in Nestlé R&D context

- It is not conservative:
  - Monadic profiling are used to describe products and not to prove/support "claims" on products.
  - So there is 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).
- It makes communication easy:
  - LSD = threshold value to reach significance, valid for any pair of products.
  - Some other procedures give one p-value for each paired comparison.

2. Strategy to analyze sensory data in DoE context

Most common modelling approaches with DoE and sensory

- **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 across 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.

Issues with each approach

- **Modelling on raw data:** error = mix between different sources of error (sample heterogeneity, measurement variability from sensory evaluation, and non-estimated interactions between DoE parameters supposed to be negligible).
  → Difficult to interpret the error used to test the significance of the parameters
- **Modelling on mean data:** DoEs often close to saturation ⇨ low number of dfs to estimate the error, i.e. not very robust. + error depends on the selected interactions, and can be over-estimated (if non negligible interactions in error).

Proposed alternative in twos steps

- Run usual ANOVA, i.e. ignoring the DoE structure, and compute the LSD (1.)
- Use this LSD to determine the significance of the effect of the parameters

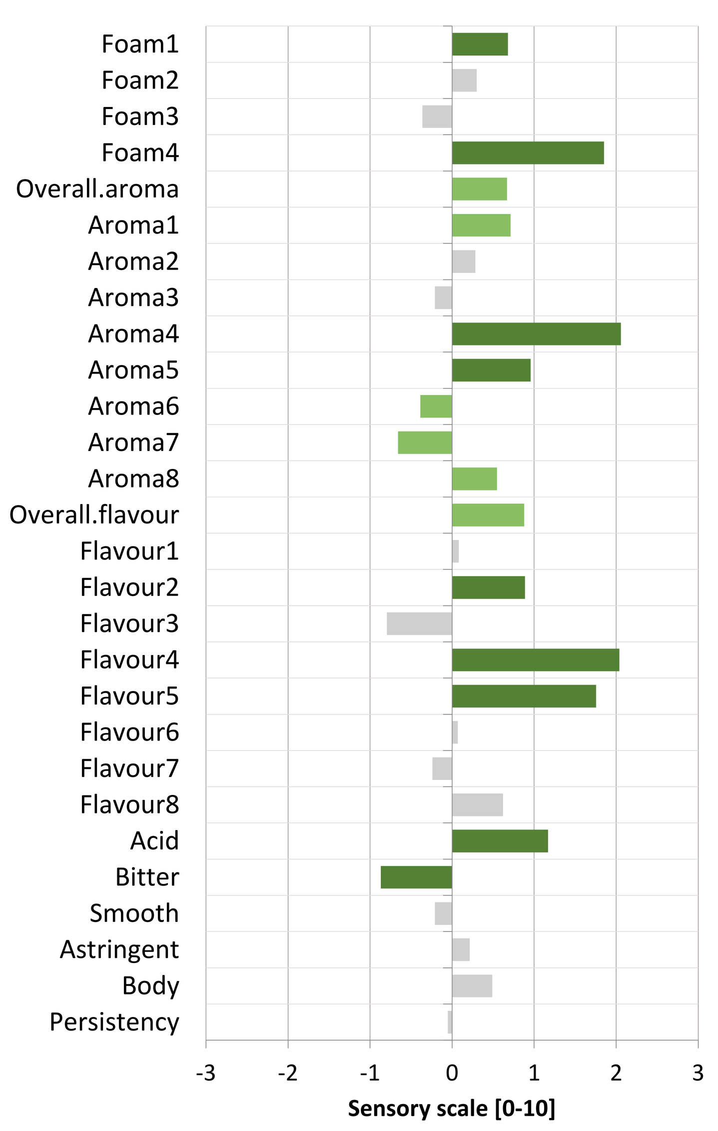
→ significance of the effects assessed based on panel ability to detect a sensory difference between samples (not biased by questions around DoE interactions).

Effect size interpretation using 2 thresholds (LSD-based)

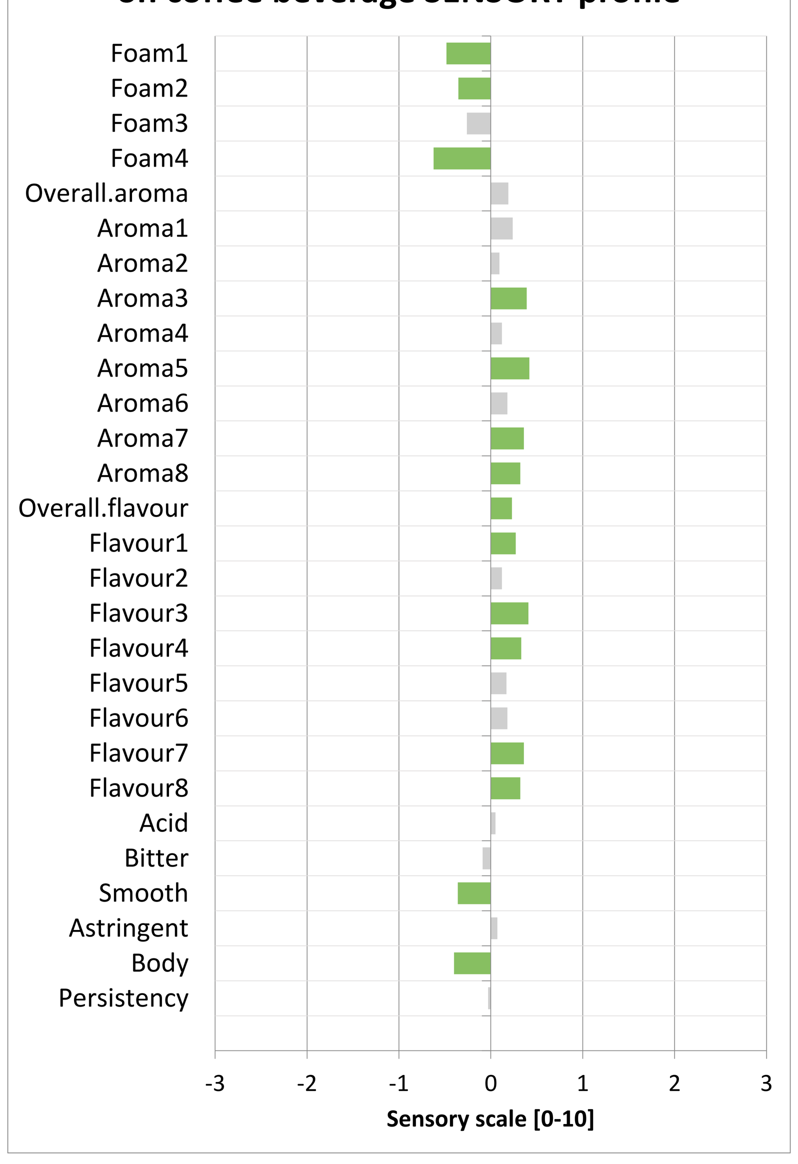
Effect Size	Color	Interpretation
$< \frac{LSD}{\sqrt{Nb\ products\ by\ factor\ level}}$	None	Only Noise
$[\frac{LSD}{\sqrt{Nb\ products\ by\ factor\ level}}\ ; LSD]$	Light	Effect to further investigate
$\geq LSD$	Dark	Important effect

3. Representation of DoE parameter effects with sensory data & interpretation

Impact of Coffee Origin (O1 vs. O2) on coffee beverage SENSORY profile



Impact of Roaster (R1 vs. R2) on coffee beverage SENSORY profile



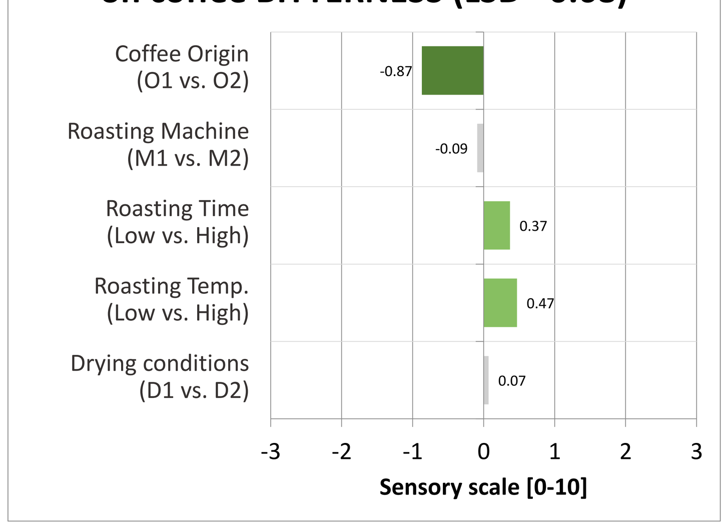
Example of interpretation for a given DoE parameter:

- **Coffee origin:** Impact on foam properties, in particular Foam1 and Foam4 (significantly higher with Origin 2). Overall flavor and odour intensity also higher, but not enough to be detected by the panel if the origin is the only change in the product. However, attributes 4 and 5 higher for Origin 2 ⇨ very significant change in the quality of the odour.
- **Roaster type:** globally much less impact than coffee type, but still a consistent trend for roaster 1 to increase foam properties and decrease aroma and flavour perception

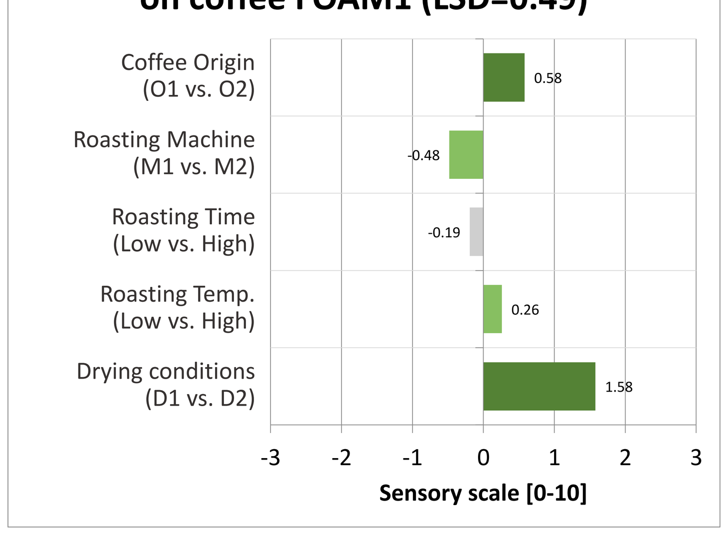
Example of interpretation for a given sensory attribute:

- **Bitterness:** mainly impacted by coffee origin, but combined roasting conditions (time and temperature) could be as impactful.
- **Foam1:** mainly impacted by drying conditions, then coffee origin. To a lower extent, roasting also impacts foam1 (e.g. area to be further investigated)

Impact of DoE parameters on coffee BITTERNESS (LSD= 0.68)



Impact of DoE parameters on coffee FOAM1 (LSD=0.49)



Conclusion

This poster presents a proposal for a simple way of analyzing sensory data in factorial DoE context. This approach allows calculations of the significance of the DoE parameters based on an error independent of the DoE and that represents production and measurement variability. The two subsequent LSD-based thresholds give the opportunity of a simple reading of the output to identify the strongest effects and the potentially impacting parameters that could be investigated further.