Biblid: 1821-4487 (2022) 26; 3-4; p 90 - 94
UDK: 366.2
DOI: 10.5937/jpea26-41654

Original Scientific Paper
Originalni naučni rad

PENALTY ANALYSIS IS AN EFFICIENT TOOL FOR FOOD PRODUCT DEVELOPMENT. CASE STUDY WITH LEMONADES

ANALIZA KAZNI JE EFIKASAN ALAT ZA RAZVOJ PREHRAMBENIH PROIZVODA. STUDIJA SLUČAJA SA LIMUNADOM

Attila GERE [*]
Hungarian University of Agriculture and Life Sciences, Institute of Food Science and Technology, Department of Postharvest, Supply Chain, Commerce and Sensory Science, H-1118, Budapest, Villányi út, 29-43.
**e-mail:gereattilaphd@gmail.com; gere.attila@uni-mate.hu*

ABSTRACT

Although penalty analysis has been used for product development for many years, its application is less widespread. Penalty analysis provides a quick and easy way to identify those sensory attributes that significantly influence consumers' preferences. The information provided helps food product developers to focus on key sensory attributes and with their improvement the overall product preference increases. The present paper focuses on the introduction of penalty analysis to the readers through an example of lemonades. Three prototypes of lemonades were tested with 65 consumers to identify the optimal level of sugar and lemon juice content that meets the need of the majority of the consumer group. Results suggest that with only three prototypes, the proper composition is easy to define.

Keywords: *sensory, consumer, nonparametric, product development.*

APSTRAKT

Iako se analiza kazni koristi za razvoj proizvoda dugi niz godina, njena primena je manje rasprostranjena. Analiza kazne pruža brz i lak način da se identifikuju oni senzorni atributi koji značajno utiču na preferencije potrošača. Dostavljene informacije pomažu proizvođačima prehrambenih proizvoda da se fokusiraju na ključne senzorne atribute i sa njihovim poboljšanjem sveukupna preferencija proizvoda se povećava. Ovaj rad se fokusira na upoznavanje čitalaca sa analizom kazni kroz primer limunade. Tri prototipa limunade su testirana sa 65 potrošača kako bi se identifikovao optimalan nivo šećera i limunovog soka koji zadovoljava potrebe većine potrošačke grupe. Rezultati sugerišu da je sa samo tri prototipa lako definisati odgovarajući sastav.

Ključne reči: *čula, potrošač, neparametrijska, razvoj proizvoda.*

INTRODUCTION

Sensory analysis or sensory evaluation is defined by the Institute of Food Technologists (IFT) as “a scientific discipline used to evoke, measure, analyze, and interpret human reactions to meat sensory characteristics as perceived by sight, smell, taste, touch, and hearing” (Miller, 2017). During a sensory evaluation, two major types of assessors are differentiated who play an active role in food product development: trained and consumer assessors. Trained sensory assessors work in sensory panels consisting of about 8-12 individuals who received specific sensory training (ISO 8586:2012, 2012). During the sensory training, their vision, smell, and taste abilities are tested and mapped. In case there are any issues, the sensory abilities can be improved through specific sensory training. A special feature of trained sensory panels is that their performances are constantly monitored to provide valid data (Næs et al., 2010). Because of the excessive training and expertise, trained individuals are used in tests where analytical questions are used (e.g., rating the intensities of certain product attributes) and objective answers are required (e.g., subjective opinions should be excluded).

On the contrary, the consumer sensory panel consists of untrained individuals, who consume at least occasionally the tested products. They did not receive any training; therefore, the sensory researcher does not have any information on the sensory performance of consumers. However, there is no need for such information because the consumer sensory panel provides subjective opinions on the product attributes. Such subjective opinion might be for example product preference, purchase intention, or any related question (Porretta et al., 2021). Since subjective opinions are used, the number of assessors in a consumer panel should be higher compared to trained panels. It is suggested to involve about 100-300 consumers per study but not go below 60 as most multivariate methods require more participants for proper data analysis (Næs et al., 2010). Numerous different techniques and methods are used in consumer sensory testing; however, the current paper intends to place the focus on the evaluation of just-about-right data using penalty analysis. For further information on consumer sensory methods, see Valera and Ares (2014).

Penalty analysis requires consumers to rate their answers on two different types of scales: just-about-right scales and hedonic scales. Just-about-right (JAR) scales are used to identify the strengths/weaknesses of the products under investigation and to determine which sensory product attribute intensity should be increased or decreased during product development. JAR scales are usually used with untrained consumer assessors. When used with hedonic scales, the impact of JAR variables on liking can be understood. JAR scales are bipolar scales with three distinct points: too little, just-about-right, and too much. The number of categories can be increased by adding more levels between the midpoint and the endpoints; therefore, it can be increased to 5, 7, or 9 categories. Independently of the number of categories, the midpoint is always the just-about-right, the leftmost point is the too little, while the rightmost category is the too much point. Product attributes are then rated on these scales to determine if the level of the attribute is just-about-right or not. The scale is easy-to-use as it is similar to the action of setting the right temperature of bath water. However, due to its bipolar structure, participants should be noted to read the scale labels.

Hedonic responses are usually recorded on categorical scales consisting of an even or odd number of categories. The major difference between even (e.g., 10) and odd (e.g., 9) categories is that an odd number of categories always require the consumers to decide if they like the product or not (at least a minimal level).

As it has been earlier introduced, penalty analysis requires JAR and hedonic scales as well to identify the product attributes that increase or decrease product preference. Penalty analysis has been used to develop vanilla yogurts (Narayanan et al., 2014), wholegrain buckwheat enriched pasta (Škrobot et al., 2022), poppy seed flavored white chocolates (Zay and Gere, 2019), sweet potato varieties (Nakitto et al., 2022), tigernut milk (Clemente-Villalba et al., 2021) and lobsters (English et al., 2020), just to name a few.

Penalty analysis itself consists of four main steps (Pagès et al., 2014):

1. A consumer sensory panel rates a set of products using multiple JAR variables and express their overall liking on a categorical scale.
2. JAR categories are merged into three main levels. Independently of the number of categories used, the midpoint is kept as JAR level, categories lower than the midpoint go to the "too little" level, while the right side of the scale becomes the "too much" level. In the case of a 9-category scale, categories 1, 2,3, and 4 are merged to the "too little" level, category 5 becomes the JAR level, while categories 6, 7, 8, and 9 are merged to the "too much" level (Figure 1.).

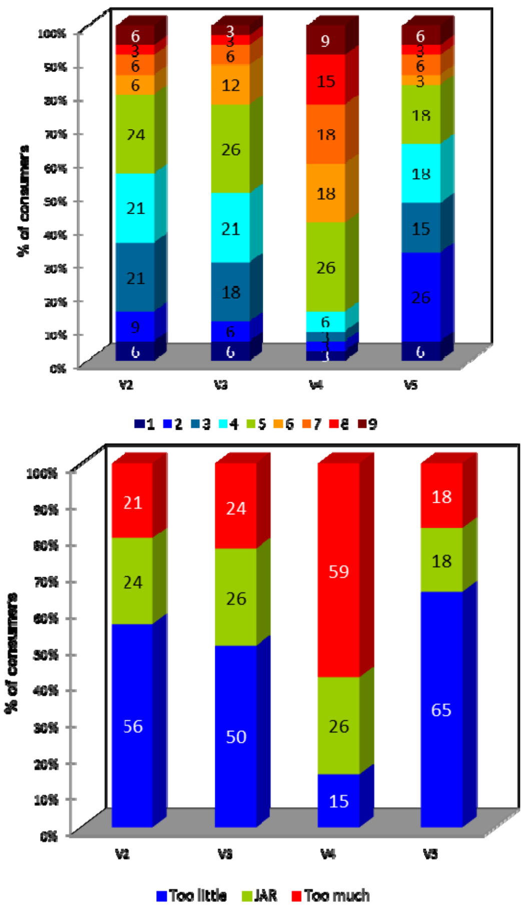

Fig. 1. Example of bar charts of the frequencies of the rated categories (left) collapsed bar chart (right). Categories 1, 2, 3, and 4 are merged to the "too little" level, category 5 becomes the JAR level, while categories 6, 7, 8, and 9 are merged to the "too much" level.

3. In the next step, the mean overall liking scores are calculated for all three groups, e.g., the mean overall liking of consumers belonging to the too weak, JAR, and too much levels are calculated.
4. The so-called mean drop values are calculated as the mean overall liking of the two endpoints is subtracted from the mean overall liking of the JAR group. Using a t-test, the overall liking scores of the non-JAR levels are compared to the overall liking scores of the JAR level, therefore a significant difference can be determined. To assess the effect of the JAR variable on overall liking, the overall liking scores of the two non-JAR levels are merged and compared to the overall liking scores of the JAR level, creating the Penalty of a given JAR variable.
5. The mean drop values are then plotted with the percentage of consumers who rated the non-JAR endpoints in a mean drop plot (Figure 2).

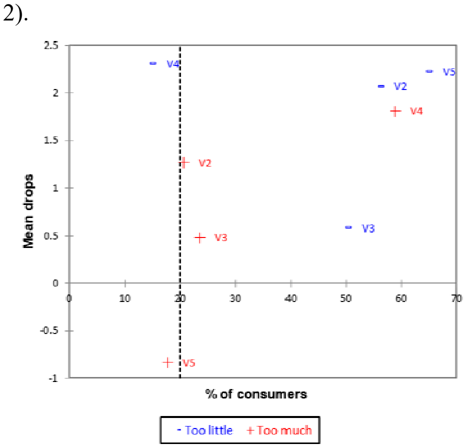

Fig. 2. Example of mean drop plot. Mean drops are calculated as the difference between the mean liking scores of the participants who rated the attribute endpoint and the mean liking scores of those who rated the product attribute as JAR.

The mean drop values are plotted against the percentage of consumers who rated the attribute as not JAR (*e.g.,* one of the endpoints). High mean drops values mean that overall liking drops highly when the attribute is not optimal (JAR), while the problem is serious when a high percentage of consumers rate the attribute endpoint. Therefore, attribute endpoints located in the upper right corner are the ones that should be addressed to improve product liking.

The present paper aims to draw attention to penalty analysis, as a quick, easy, and powerful tool suitable for conducting consumer-based food product development.

MATERIALS AND METHODS

Four lemonades were prepared using water, lemon juice, and sugar. The recipes differed in the amounts of lemon juice and sugar (Table 1). Samples were prepared in 1.5L jars and stored at room temperature until the evaluation. Ingredients were purchased at a local supermarket.

Table 1. Samples and their ingredients

Sample name	Water (ml)	Lemon juice (ml)	Sugar (g)
274	825	175	175
143	925	75	75
284	825	175	75
361	875	125	125

Samples were poured into transparent, 200 mL plastic glasses. The recommendation of Kilcast was followed as the same person prepared all samples (150 mL/person) to achieve sample homogeneity (Kilcast, 2010). Samples were coded according to ISO 6658 standard using 3-digit random numbers (see Table 1 for sample codes). Between the evaluations of the samples, consumers were instructed to use neutral non-carbonated mineral water as a taste neutralizer. Evaluations were performed under artificial daylight-type illumination, temperature control (between 22 and 24 °C), and air circulation. The consumer sensory panel consisted of 65 students (40% males, 60% females, aged between 22 and 25 years) of Hungarian university students of the Hungarian University of Agriculture and Life Sciences. Proper instructions were given to the consumers before the evaluation regarding the bipolar JAR scale and the linear hedonic scale. Consumers evaluated four JAR attributes (color, odor, sweet taste, and sour taste intensity) on a 9-point JAR scale. Overall liking was rated on a 9-point hedonic scale (1 = "dislike extremely", 9 = "like extremely"). Evaluations were conducted in the same place (Sensory Laboratory of Hungarian University of Agriculture and Life Sciences, which meets the requirements of ISO 8589:2007), between 10-12 am. Data was recorded on printed sensory ballots. Analysis of variance (ANOVA) was used with Tukey HSD *post hoc* test, where ANOVA indicated significant differences at α=95%. Penalty analysis used a 20 % consumer threshold limit. Data analysis was done using XL-Stat software (ver. 2022.4.1, Addinsoft, Paris, France).

RESULTS AND DISCUSSION

Analysis of variance of the overall liking (OAL) variables indicated a significant difference between the samples ($F(3,256) = 8.114$, $p < 0.001$). Tukey post hoc test indicated that the most liked product was 284 (OAL=5.84±2.22^{c}), followed by 361 (OAL=5.56±2.17^{bc}), 143 (OAL=4.64±2.21^{ab}) and 274 (OAL=4.20±2.19^{a}).

The numerical results of the penalty analysis are listed in Table 2. Attributes with a significant non-JAR end are listed. The first three columns give the sample code, JAR attribute name, and attribute endpoints. The fourth column gives the frequencies of the three points, *e.g.,* how many of the participants rated the attribute too weak, JAR, or too intense. This is followed by the same information expressed in percentages. Percentages are crucial here as mean drop analysis should be done only on non-JAR endpoints that meet a certain limit. This limit is usually 20 %, meaning that non-JAR attributes rated by more than 20% of consumers should be taken into account. This limit can be changed based on the size of the consumer panel. Naturally, when working with a higher number of consumers, it is advised to reduce the limit. This limit enables researchers to focus on problems that are rated by the majority of the consumer panel and to eliminate issues causing problems for a smaller number of consumers. Therefore, *t*-tests were not run for those non-JAR levels rated by less than 20%. The next column lists the results of *t*-tests, where a p-value lower than 0.05 indicates that those rated the non-JAR level, gave significantly lower overall liking scores compared to those who rated the attribute as JAR. This indicates that these attribute endpoints should be addressed. The last two columns show the Penalties and their significance tests. Penalties are calculated as the difference in the mean overall liking of those who rated the attribute as JAR and non-JAR. Although mean drops give meaningful information about the effect of the non-JAR levels on overall liking, they should be evaluated along with the number of consumers who rated them. This can be done easily using the mean drop plots (Figure 2) of the products. In Figure 2, mean drop plots of the four products are presented.

Table 2. Results of penalty analysis for all four products.

Samples	Variable	Level	Freq.	%	Mean(OAL)	Mean drops	p-value	Penalties	p-value
274		Too weak	6	9.23%	3.667	5.000			
	SweetTaste	JAR	3	4.62%	8.667			4.683	0.000
		Too intense	56	86.15%	4.018	4.649	0.000		
		Too weak	48	73.85%	3.958	1.819	0.023		
	SourTaste	JAR	9	13.85%	5.778			1.831	0.019
		Too intense	8	12.31%	3.875	1.903			
143		Too weak	57	87.69%	4.439	2.361	0.023		
	Odor	JAR	5	7.69%	6.800			2.333	0.022
		Too intense	3	4.62%	5.000	1.800			
		Too weak	29	44.62%	3.931	1.640	0.024		
	SweetTaste	JAR	21	32.31%	5.571			1.367	0.019
		Too intense	15	23.08%	4.733	0.838	0.477		
284		Too weak	37	56.92%	5.189	1.584	0.008		
	SweetTaste	JAR	22	33.85%	6.773			1.377	0.017
		Too intense	6	9.23%	6.667	0.106			
		Too weak	10	15.38%	6.100	1.011			
	SourTaste	JAR	18	27.69%	7.111			1.728	0.004
		Too intense	37	56.92%	5.189	1.922	0.003		
361		Too light	31	47.69%	4.871	1.496	0.006		
	Color	JAR	30	46.15%	6.367			1.481	0.005
		Too dark	4	6.15%	5.000	1.367			
		Too weak	7	10.77%	5.286	1.581			
	SweetTaste	JAR	15	23.08%	6.867			1.687	0.007
		Too intense	43	66.15%	5.163	1.704	0.007		
		Too weak	43	66.15%	5.070	1.597	0.014		
	SourTaste	JAR	15	23.08%	6.667			1.427	0.024
		Too intense	7	10.77%	6.286	0.381			

Each plot presents only the non-JAR attributes and uses blue and minus signs for the too weak level, while red font color and plus signs are used for the too intense levels. All mean drop plots can be divided into four distinct quadrants based on the horizontal axis representing the percentage of consumers who rated the non-JAR attributes and the vertical dashed line representing the 20% threshold discussed above. Attribute levels placed into the upper right corner have high mean drop values and were rated by a high number of consumers, therefore these are the ones that should be addressed during product development. For product 274, the sweet taste is the only one located here, meaning that reduction of sweet taste (*e.g.* using less sugar) might increase product liking. Too weak odor and sour taste are located here in the case of product 143, indicating that this product had too weak attributes for consumers. Product 284 has a too intense sour taste and too weak sweet taste in the upper right corner, indicating that less lemon juice should be used. The opposite can be seen for product 361 as too intense sweet and too weak sour tastes are located here.

A comparison of overall liking scores indicates that products 284 and 361 were the most liked ones, while penalty analysis tells us that the sour taste was too intense and the sweet taste was too weak for product 284, while the opposite was seen for product 361. As lemon juice is responsible for the sour taste, it is clear now that 175 ml is too much, while 125 ml is too weak for the consumers. Regarding sweet taste, 75 g is not enough, while 125 g is too much. A drawback of penalty analysis is that no exact amounts are defined, *e.g.* it is impossible to define the perfect amounts of sugar and lemon juice. However, a possible composition could be to change the lemon juice to a level of ~150 ml, while the sugar content could be set to ~100 g.

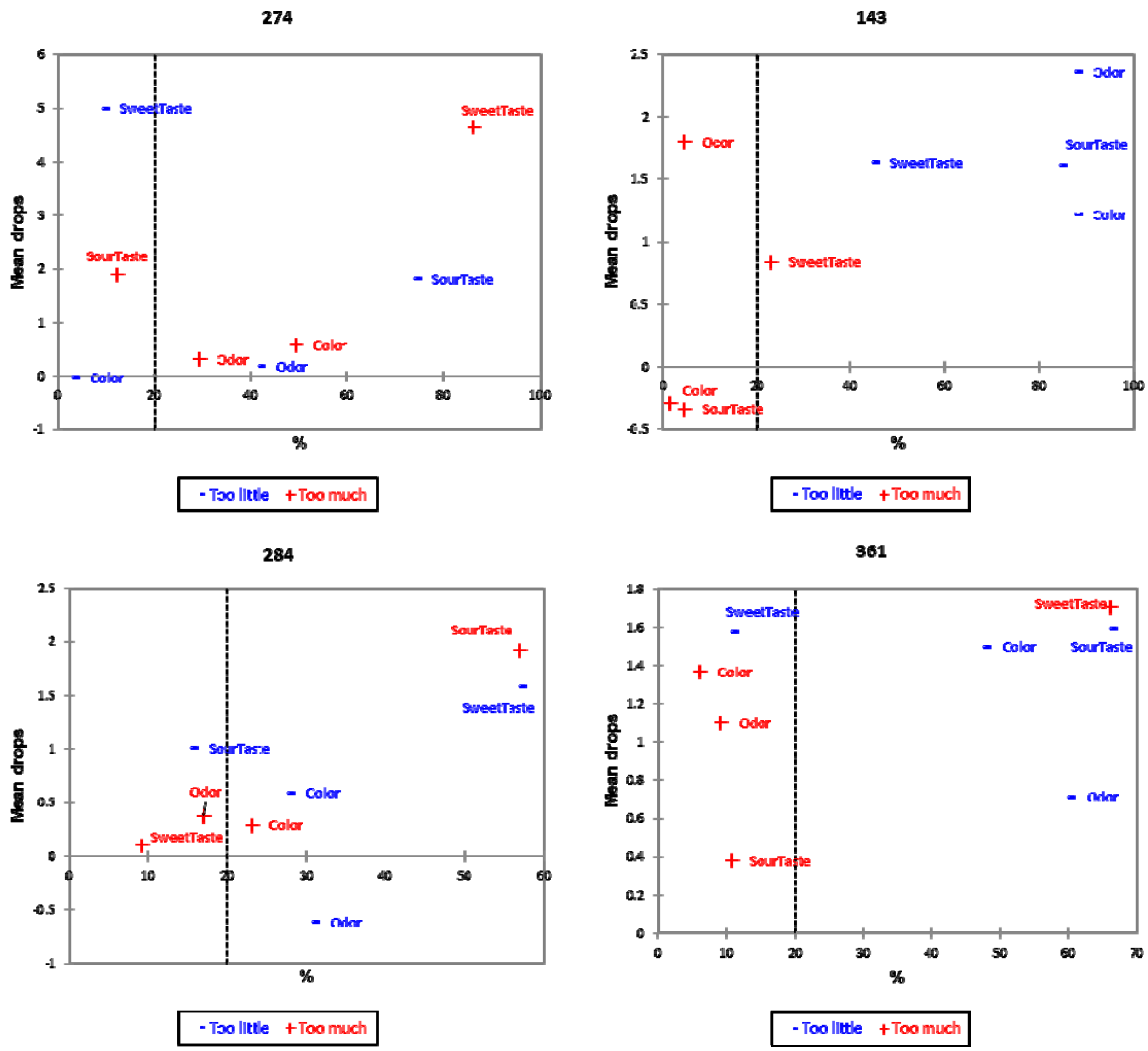

Fig. 2. Mean drop plots of the four products.

CONCLUSIONS

Just-about-right scales and penalty analysis are easy-to-use tools able to help product formulations quickly and cheaply. Naturally, the complexity of the analyses grows with the complexity of the products being tested. Although rarely done but a repeated consumer test with the same consumer panel can validate the results of changes (Gere et al., 2017).

ACKNOWLEDGEMENT: AG thanks the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the National Research, Development, and Innovation Office of Hungary (OTKA, contracts No FK 137577 and K 134260). Supported by the ÚNKP-22-5-MATE/2 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development, and Innovation Fund.

REFERENCES

Clemente-Villalba J, Cano-Lamadrid M, Issa-Issa H, et al. (2021) Comparison on sensory profile, volatile composition and consumer's acceptance for PDO or non-PDO tigernut (Cyperus esculentus L.) milk. LWT 140: 110606. DOI: 10.1016/j.lwt.2020.110606.

English MM, Scrosati PM, Aquino AJ, et al. (2020) Comparison on sensory profile, volatile composition and consumer's acceptance for PDO or non-PDO tigernut (Cyperus esculentus L.) milk. Food Research International 137: 109697. DOI: 10.1016/j.foodres.2020.109697.

Gere A, Szabó Z, Pásztor-Huszár K, et al. (2017) Use of JAR-Based Analysis for Improvement of Product Acceptance: A Case Study on Flavored Kefirs. Journal of Food Science 82(5): 1200–1207. DOI: 10.1111/1750-3841.13701.

ISO 8586:2012 (2012) ISO 8586:2012, Sensory analysis -- General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors. ISO.

Kilcast D (2010) Sensory Analysis for Food and Beverage Quality Control - A Practical Guide. Cambridge: Woodhead Publishing Limited.

Miller RK (2017) Chapter 15 - The Eating Quality of Meat: V—Sensory Evaluation of Meat. In: Toldra´ F (ed.) Lawrie´s Meat Science. Eighth Edi. Woodhead Publishing Series in Food Science, Technology and Nutrition. Woodhead Publishing, pp. 461–499. DOI: https://doi.org/10.1016/B978-0-08-100694-8.00015-7.

Næs T, Brockhoff PB and Tomic O (2010) Statistics for Sensory and Consumer Science. 1st Ed. Chichester, West Sussex, United Kingdom: John Wiley & Sons, Ltd. DOI: 10.1002/9780470669181.

Nakitto M, Johanningsmeier SD, Moyo M, et al. (2022) Sensory guided selection criteria for breeding consumer-preferred sweetpotatoes in Uganda. Food Quality and Preference 101: 104628. DOI: 10.1016/j.foodqual.2022.104628.

Narayanan P, Chinnasamy B, Jin L, et al. (2014) Use of just-about-right scales and penalty analysis to determine appropriate concentrations of stevia sweeteners for vanilla yogurt. Journal of Dairy Science 97(6): 3262–3272. DOI: 10.3168/jds.2013-7365.

Pagès J, Berthelo S, Brossier M, et al. (2014) Statistical penalty analysis. Food Quality and Preference 32. Elsevier Ltd: 16–23. DOI: 10.1016/j.foodqual.2013.07.008.

Porretta S, Moskowitz H and Gere A (eds) (2021) Consumer-Based New Product Development for the Food Industry. The Royal Society of Chemistry. DOI: 10.1039/9781839163333.

Škrobot D, Pezo L, Tomić J, et al. (2022) Insights into sensory and hedonic perception of wholegrain buckwheat enriched pasta. LWT 153: 112528. DOI: 10.1016/j.lwt.2021.112528.

Valera P and Ares G (2014) Novel Techniques in Sensory Characterization and Consumer Profiling (V Paula and A Gastóneds). 1st ed. CRC Press.

Zay K and Gere A (2019) Sensory acceptance of poppy seed-flavored white chocolates using just-about-right method. LWT 103: 162–168. DOI: 10.1016/j.lwt.2018.12.069.

Received:08.12.2022. Accepted:23.01.0222.