





## ARTICLE

# Logistic regression in threshold detection in replicated discrimination sensory tests

 Camilla Marques Barroso\*<sup>1</sup> and  Elayne Penha Veiga<sup>1</sup>

<sup>1</sup>Department of Statistics, Institute of Exact and Technological Sciences, Federal University of Lavras, Lavras - MG, Brazil.

\*Corresponding author. Email: camilla.barroso@ufla.br

(Received: August 8, 2022; Revised: November 18, 2022; Accepted: December 20, 2022; Published: December 1, 2023)<sup>1</sup>

### Abstract

Sensory discrimination tests often consider the possibility that each assessor performs a test more than once, that is, they are performed with replications. However, these experiments are generally analyzed as if there were no repetitions or as if all judgments came from different assessors, which the usual analysis assumes that the probability of success is the same for all of them. It is emphasized the importance of considering possible differences in the perception of the assessors, which is consistent with Thurstone's psychological ideas about perception and decision processes. In this article it is suggested a mixed generalized linear model that considers the random effect of the assessor. Particularly in dose-response experiments it is showed that this model allows estimating detection thresholds for several treatments simultaneously. The results allowed to conclude that the proposed model presents good results and can be used in the analysis of replicated sensory tests. We will assume that the variation in the assessor's responses can be modeled as a random effect in a mixed generalized linear model. This model should be able to produce a better analysis, with the ability to detect differences between treatments with greater power and produce narrower confidence intervals for the estimates.

**Keywords:** Thurstonian modeling. Logistic regression. Mixed generalized linear models. Thresholds. Replicated sensory tests.

## 1. Introduction

It is possible that two samples are chemically different in formulation, but these differences are imperceptible to humans. In these situations, discrimination sensory tests can be used to determine whether there are sensory differences between two or more samples. Particularly, when we test the assessor's response to certain doses of an ingredient, we have a so-called dose-response experiments or trials.

In this situation, the higher the concentration or dose of the ingredient, the more noticeable it becomes. We can fit a model that gives us information about the effect of the ingredient concentration on the assessor's response. An interesting measure is the detection threshold which represents the lowest perceptible concentration.

The American Society for Testing and Materials (1978) provides the following definition for a detection threshold: "There is a concentration below which the taste of a substance will not be detectable under any practical circumstances, and above which individuals with a normal palate would detect promptly the presence of the substance".

Lawless & Heymann (2010) present several practical techniques for measuring thresholds such as the forced-choice limits method based on the ASTM E-679 standard method<sup>3</sup> and an alternative graphical solution that will be described next, in Section 2.

Discrimination sensory tests are often conducted with replications to the assessors in order, for example, to increase power without increasing the number of assessor (Brockhoff, 2003; Meyners & Brockhoff, 2003). These tests consider the possibility that each assessor performs a test more than once and are generally analyzed as if there were no repetitions, that is, as if all judgments were from different assessors. The number of correct answers is compared to a tabulated value. If the number of correct answers is greater than or equal to the table, it is concluded that there is a significant difference to the level of probability observed (Roessler *et al.*, 1978).

However, Thurstone (1927) concluded that the discrimination process corresponding to a given stimulus is not fixed. Perceptions are momentary because they are assumed to vary from judgment to judgment and thus can be represented as random variables. Thurstonian modeling considers this relationship by approaching two ideas: variation of perception and the concept of decision rule.

The first idea considers that variations in the neural mechanism of the assessor cause the intensity of the perception to the product not to be constant during the test. The second idea is based on the concept of a decision rule or cognitive strategy which is a rule applied by the assessor to produce an answer about the perception of the samples. For the same assessor, this rule is independent of the stimulus involved, that is, in a discrimination test, regardless of the number of repetitions, the decision rule of a specific assessor will always be the same. Decision rules among assessors may be different and due to this fact, some decision rules may be more efficient than others and thus, some assessors may perform better (Meilgaard, Carr & Civile, 2006).

This kind of situations is common in experiments involving sensory analysis, hence the need to improve data evaluation techniques presenting these characteristics. It is important to correctly deal with the possible differences in the perception of the assessors, considering the psychological ideas about the perception and decision processes (Brockhoff & Christensen, 2010; Kunert & Meyners, 1999).

This article focuses on problems of variability in detection and the challenges this poses for researchers who use thresholds as a measure of the sensitivity of individuals to a given stimulus. Several methods have been proposed for the analysis of replicated discrimination tests. The variable of interest is often the detection (yes or no) of a given ingredient in the product. The binomial distribution is commonly used in the analysis of these tests under the assumption that the assessors' choices are independent, and the probabilities of success do not vary from trial to trial. Harris & Smith (1982); Ennis & Bi (1998) address violations of the last assumption and the two-parameter-indexed Beta-Binomial model is used to explain this overdispersion (Skellam, 1948). The model assumes that the individual hit probabilities are randomly distributed according to the beta distribution and that the binomial distribution is valid considering the repetitions of a given assessor.

A variant of the beta-binomial model is proposed in Bi & Ennis (1998) in which the beta-binomial model is combined with a Thurstonian psychometric function and used to obtain estimates of sensory differences.

Hunter *et al.* (2000) as cited in Duineveld & Meyners (2008) suggest the use of a mixed generalized linear model with an overdispersion factor to model the number of correct responses. The analysis extends the usual model to binomial repeated measures considering the sensory distances of the Thurstonian approach and the random effects of individual thresholds. In Brockhoff & Muller (1997) this model is discussed in a sensorial context and they propose an approach to dose-response assays with replication based on a random effects threshold model. The focus of the work is the study of individual thresholds. In the model it is assumed that the responses observed in the test are determined by specific thresholds of the randomly distributed

subjects among others.

In Kunert (2001) a binomial mixture model was suggested for analyzing discrimination tests, assuming that the population of assessors consists of two types: non-discriminating and discriminating. Discriminators are those that have a higher probability of being right than the probability of being right by chance, considering some discrimination ability.

Comparisons between these models, beta-binomial models, mixed generalized linear model, and binomial mixture model, were presented in Brockhoff (2003), in addition to corrected versions of the beta-binomial model and the mixed generalized linear model.

Bi (2003) understands that the difficulties in the methodology of discrimination tests reside in the assumption that subjects have the same ability to respond correctly to the test, which conflicts with psychological ideas about the processes of perception and decision. Thus, the author suggests that a Bayesian approach can overcome these difficulties by treating the proportion parameter as a random variable. This was probably the first time that a Bayesian framework was used for the analysis of discrimination tests, deriving a posteriori credibility interval for the proportion, and using so-called Bayes factors to decide on alternative hypotheses. However, the author still does not explore the possible replications commonly used in discrimination tests.

The purpose of our research is to adjust a mixed model that incorporates the random effect of the assessor to estimate the global detection thresholds. To estimate these thresholds, we will use the graphic technique presented by Lawless & Heymann (2010).

Initially, the ideas of Kunert & Meyners (1999), which consider the skills of the assessor, are considered.

Several authors applied the graphical technique to estimate detection thresholds. Ziegler *et al.* (2019), for example, applied detection thresholds to assess the impact of compounds (TDN) that impart an off-flavor to gasoline in wines. However, the analyzes did not considered the random effect of the assessor.

Lima Filho *et al.* (2015) present the idea that when a sample is less preferred than the other, it does not mean that it is sensorially rejected and, therefore, they argue that acceptance tests are more indicated when one wants to investigate the point at which the product sensory rejection begins to occur. The authors propose a new methodology for determining two sensory thresholds: the compromised acceptance threshold (CAT), which represents the stimulus intensity in which the acceptance of the product becomes significant, and the hedonic rejection threshold (HRT), referring to the point of transition between sensory acceptance and rejection. The approach uses a 9-point hedonic scale to assess the samples. The results are submitted to a paired t test (using scores from the control sample and scores from the adulterated sample) and are adjusted to regression models, allowing the identification of the intensity of the stimulus from which significant changes in sensory acceptance (CAT) and rejection (HTM) of the product occur.

Although for years, there have been studies that propose models considering the possible random effect of the assessor, even today the usual analysis of detection thresholds in food science, in general, does not consider this idea.

The model proposed in this article allows the analysis of detection thresholds in replicated discrimination tests. The example presents simulated data from a replicated triangular test; however, our considerations can be extended to other discrimination tests.

Basically, we will assume that the variation in the assessor's responses can be modeled as a random effect in a mixed generalized linear model. This model should be able to produce a better analysis, with the ability to detect differences between treatments with greater power and produce narrower confidence intervals for the estimates.

Once the importance is justified, the aim of this research is to present a model applied to the analysis of detection thresholds considering the random effect of the assessor in replicated discrimination tests. We will present an application of this technique using simulated data from a replicated triangular test.

In section 2, the suggested methodology and the data that will be used to illustrate this

methodology are established. In section 3 we present the results and discussion obtained are presented and, in turn, section 4 presents the conclusion.

## 2. Materials and Methods

In this section we explain the suggested methodology and describe the data that will be used to illustrate the methodology. These are simulated data from a replicated triangular test.

### 2.1 Data

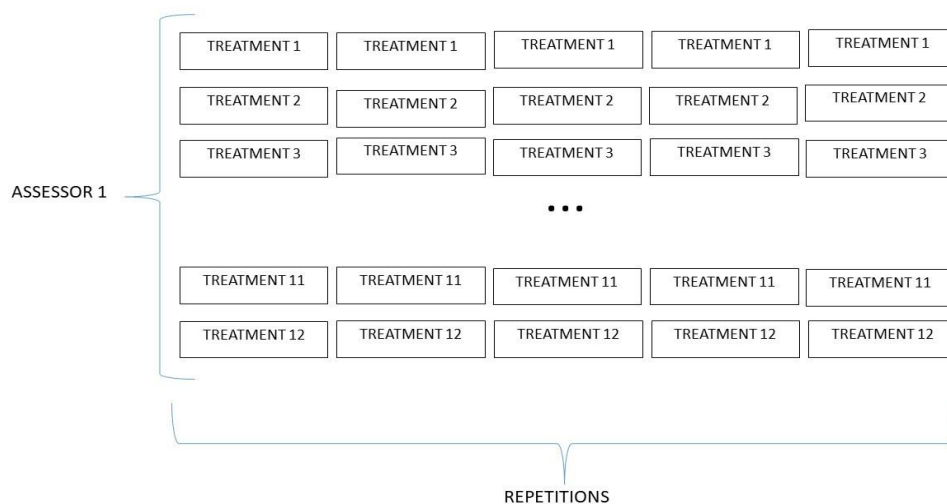
The data used in this work were simulated considering a replicated triangular discrimination test conducted in a factorial scheme to evaluate the addition of different types and concentrations of an ingredient in samples of a given product.

Triangular tests are discrimination tests in which three coded samples are presented to the assessor, two of which are the same and one different (Kemp, Hollowood & Hort, 2009). These tests are often conducted in a factorial scheme to evaluate the addition of different concentrations of ingredients to some product. The panelist performs a triangle test (or more in the case of replicate tests) for each combination of adulterated versus control (unadulterated) product.

The factors studied in the simulation were: ingredient types at two levels (A and B) and ingredient concentration at six levels (5%, 10%, 15%, 20%, 25% and 30%). We consider that the experimental design was conducted in randomized blocks, with each of the 30 assessors considered constituting a block. In the simulation, each assessor had five chances for each combination of factors (treatments) and should have identified, in each trial, the sample that he considers different from the others. In this case, the specific qualities of each sample do not matter, but the difference between them does.

In total there are 12 treatments (combination of 2 ingredients and 6 concentrations) that were compared with the control (unadulterated samples). In the test, the assessor is compelled to make the choice even though no difference is perceived.

Figure 1 presents the scheme of the experiment simulated in this work to illustrate the proposed methodology. The scheme is presented to only one assessor; however, it is extended to all other assessors.



**Figure 1.** Scheme of the simulated experiment used to illustrate the proposed methodology.

### 2.2 Detection Thresholds

In replicated discrimination tests, it is possible to calculate the average proportion of correct responses by the group for each ingredient at each concentration and plot a graph of the observed

average proportion of hits versus ingredient concentration. As concentration increases, the average hit ratio of the group should go from a level close to random hit to close to 100% correct.

In general, the higher the concentration of the ingredient, the more noticeable it becomes. These experiments are known as dose-response experiments or assays. The dose-response curve described has the mathematical properties of an accumulated continuous distribution function and exhibits a sigmoid or S-shape. The logistic distribution is frequently used in this type of analysis (Altshuler, 1981; Coleman & Marks, 1998).

From the data of the simulated experiment described in Section 2.1 we can fit a model that provides information about the effect of the concentration of the ingredient on the assessor's response. The threshold of detection is an interesting measurement commonly used and is defined as the level below which no sensation emerges from a stimulus and above which a sensation reaches the assessor's consciousness.

For each ingredient, there will be a distinct adjusted dose-response curve at which the detection threshold will be estimated, that is, the point at which the assessors change their response. Up to a certain concentration, they did not detect the presence of the ingredient, and after a certain point, they started to detect it.

In practice, it appears that there is a variability at the point which the assessors change their answers. In a test sequence, the switching point may be different from one assessor to another. This led to the establishment of common rules for defining the detection threshold, such as the level that corresponds to 50%, on average, of correct answers (Lawless & Heymann, 2010).

The dose-response curve is applied to model the relationship between ingredient concentration and the probability of correct responses. As justified in other sections, it is important to consider the random effect of the assessor. In this sense, the variation of assessors is considered and modeled as a random effect in a mixed generalized linear model.

### 2.3 Logistic regression

In this section, the structure of the generalized linear model is extended by adding the effect of the assessors as a random component. Therefore, this section considers a mixed model with fixed ingredient and concentration effects, as well as a random assessor effect.

Given a random effects vector (assessor), the number of hits,  $\mathbf{y}$ , in a repeated trial, are (conditionally) independent binomial variables,

$$y_{ij} \mid \mathbf{b} \sim \text{binomial}(\pi_{ij}, m_{ij}), \quad (1)$$

where  $y_{ij}$  the number of hits by assessor  $j$  in treatment  $i$ , where  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  e  $m_{ij}$  the number of repetitions of assessor  $j$  in treatment  $i$  being  $m = 1, \dots, M$ . In the simulation used in this article, we considered that each assessor had five chances to perform the test ( $I = 12, J = 30, M = 5$ ). However, it is common in practice for assessors not to participate in all replications of a treatment and the number of repetitions varies from one assessor to another.

Furthermore, we assume that  $\pi_{ij}$ , the probability of correctness of assessor  $j$  in treatment  $i$  is independent of the session and that  $\mathbf{b} \sim N(0, \mathbf{G})$ , where the covariance matrix  $\mathbf{G}$  may depends on an unknown vector of variance components.

The conditional mean  $\mu_{ij} = E(y_{ij} \mid \mathbf{b})$  is associated with the linear predictor

$$\eta_{ij} = x'_{ij}\boldsymbol{\beta} + z'_{ij}\mathbf{b}, \quad (2)$$

where  $x_{ij}$  and  $z_{ij}$  are known and  $\boldsymbol{\beta}$  a vector of unknown parameters (the fixed effects), through a known link function  $g(\cdot)$  such that

$$g(\pi_{ij}) = \eta_{ij}. \tag{3}$$

The model proposed by Kunert & Meyners (1999), a variation of the model proposed by Brockhoff & Schlich (1998), considers

$$\pi_{ij} = g^{-1}(\eta_{ij}) = c + (1 - c)p_{ij}, \tag{4}$$

where  $\pi_{ij}$  is the probability of correctness of assessor  $j$  in treatment  $i$ ;  $c$  is the probability of hitting at random and  $p_{ij}$  is the probability associated with the skill of the assessor, that is, the probability of the assessor  $j$  identifying the different sample in treatment  $i$  and not just guessing.

Thus, if the assessor has no ability to detect the difference between the products, we have  $p_{ij} = 0$  and  $\pi_{ij} = c$ . If he knows, or has any discernment, this corresponds to  $p_{ij} > 0$  and  $\pi_{ij} > c$ . Often,  $p_{ij}$  is represented by the logistic function

$$\pi_{ij} = c + (1 - c) \left( \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} \right). \tag{5}$$

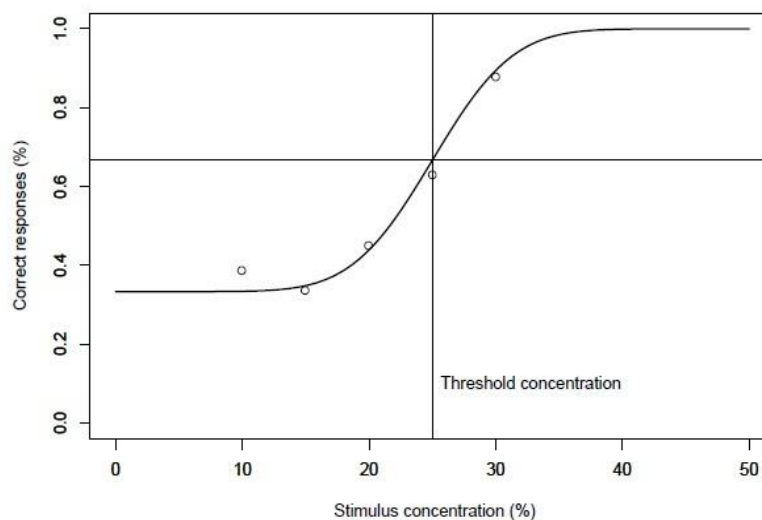
The detection threshold corresponds to  $p_i = 50\%$  of correct answers, that is, it is the concentration obtained by doing

$$\frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} = 0,5, \tag{6}$$

or even  $\eta = 0$ . In particular, for the triangular test, we have  $c = \frac{1}{3}$  and the detection threshold will be the concentration associated with 67% correct responses in the graph,

$$\pi_{ij} = \frac{1}{3} + \left(1 - \frac{1}{3}\right) 0,5 = 0,67 \tag{7}$$

At this point, we have the greatest slope of the curve, that is, where there is greater variability in the responses. Figure 2 exemplifies the estimation of the detection threshold based on the graph of the proportion of correct responses as a function of the concentration of the ingredient or stimulus.



**Figure 2.** Percentage of correct responses as a function of adulterant concentration indicating the detection threshold.

The regression models up to the fifth degree presented in equations (8-12) were adjusted for each ingredient to verify which one best explains the behavior of the response variable. The variable

$x$  represents the concentration of the ingredient.

$$\eta = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \beta_5x^5 \quad (8)$$

$$\eta = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 \quad (9)$$

$$\eta = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 \quad (10)$$

$$\eta = \beta_0 + \beta_1x + \beta_2x^2 \quad (11)$$

$$\eta = \beta_0 + \beta_1x \quad (12)$$

Although maximum likelihood and restricted maximum likelihood methods have become standard procedures in mixed linear models, likelihood-based inference in generalized linear mixed models (MLGM) is still a computational challenge. For this reason, several approaches to inference about MLGM have emerged, trying to solve, or avoid, the computational difficulties.

The likelihood function in an MLGM does not normally have a closed form expression (with, of course, the exception of the normal case). In fact, such a probability may involve high-dimensional integrals that cannot be evaluated analytically. Thus, an approximation becomes one of the natural alternatives. We will use the *glmer* function from the **lme4** package of R for inference which applies Laplace approximation and adaptive Gauss-Hermite quadrature (Bates *et al.* 2015; R Core Team, 2020). Here the percentage of correct answers is converted into a probability measure  $\pi_{ij}^*$  that follows the logistic model for implementing the algorithm.

$$\pi_{ij}^* = \frac{3}{2}\pi_{ij} - \frac{1}{2} = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}}, \quad (13)$$

where  $\pi_{ij} = C_{ij}/M_{ij}$ , where  $C_{ij}$  is the number of correct answers and  $M_{ij}$  is the number of repetitions of assessor  $j$  in treatment  $i$ . So,

$$\pi_{ij}^* = \frac{3C-M}{2M} = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}}. \quad (14)$$

Meilgaard (1991) shows that this conversion can generate probabilities  $\pi_{ij}^*$  equal to 0%, 100% or even negative. The model does not accommodate probabilities equal to 100% and 0% and certainly, no results less than 0%. For tuning in computer programs, the authors suggest arbitrarily manipulating these probabilities and replacing them with 99.5%, 0.5% and 0.1%, respectively.

## 3. Results and Discussion

### 3.1 Descriptive data analysis

In this section, an exploratory analysis of the data described in section 2.1 is presented. Table 1 presents the average proportion of hits observed for each ingredient.



**Table 1.** Average proportion of hits for each ingredient (A and B) at each concentration

Ingredient concentration	Average proportion of hits observed	
	A	B
5%	0.3733	0.3200
10%	0.3666	0.5333
15%	0.3866	0.9533
20%	0.7000	1.0000
25%	0.9533	1.0000
30%	1.0000	1.0000

Overall, the average hit ratio seems to increase as the concentration of the ingredient in the sample is also increased. Ingredient B seems to be the most easily detectable, starting at 15% almost all assessors already detect the difference.

### 3.2 Adjustment

For each ingredient, it is possible to adjust regression models in order to verify which one best explains the behavior of the response variable. The models represent the relationship between the linear predictor ( $\eta$ ) and concentrations of ingredients.

In the simulated example, the first-degree regression model presented the lowest Akaike information criterion (AIC), being, therefore, the model that best fitted the data for all ingredients.

Table 2 presents the mean estimates, 95% confidence intervals, standard error and p-value for the coefficients of the fixed (for comparison purposes) and mixed models, that is, for the models that do not consider and those that consider the effect assessor randomness, respectively.

**Table 2.** Estimates of means, 95% confidence intervals (Lower limit-LL and upper limit-UL), standard error (SE) and p-value for the coefficients of the regression model

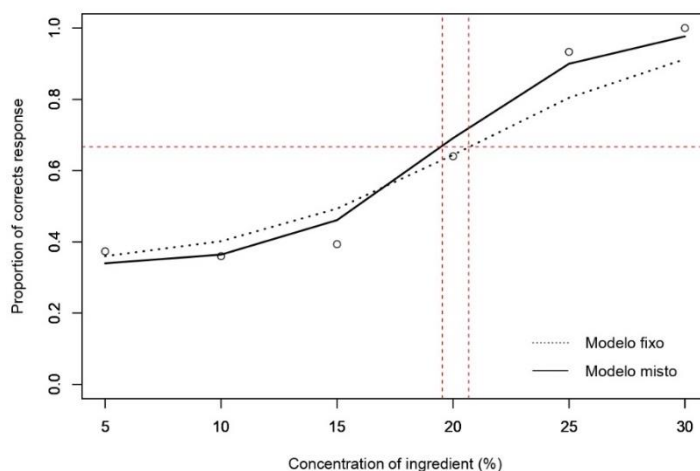
	coef	Fixed Model				Mixed Model					
		LL	mean	UL	SE	p-value	LL	mean	UL	SE	p-value
A	$\beta_0$	-6.1748	-4.1989	-2.3709	0.9658	<0.0001	-8.1553	-6.1997	-4.6204	0.8940	<0.0001
	$\beta_1$	0.1504	0.2031	0.2634	0.0286	<0.0001	0.2385	0.3173	0.4148	0.0445	<0.0001
B	$\beta_0$	-6.6294	-4.5779	-2.6944	0.9979	<0.0001	-9.3022	-6.8962	-5.0325	1.0776	<0.0001
	$\beta_1$	0.1960	0.2599	0.3355	0.0353	<0.0001	0.3157	0.4267	0.5719	0.0646	<0.0001

Modeling the assessor random effect produced a better analysis, with a lower value of the Akaike criterion and the ability to detect differences between treatments with greater power. Peltier *et al.* (2014) highlight that considering a assessor as a fixed effect can lead to the conclusion that the differences in treatments are greater than they really are.

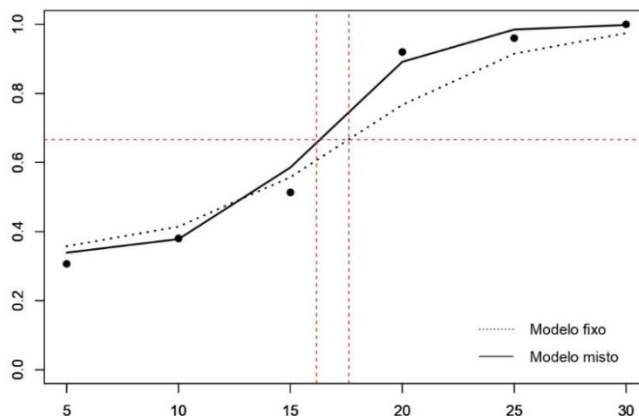
This means that not considering the random effect of the assessor may overestimate the detection threshold estimates, that is, with this model it can be wrongly concluded that ingredients are detected at certain concentrations when they will actually be detected at lower concentrations.

Figures 3 and 4 illustrate the observed mean proportions of hits (points) and the adjusted mean curves (fixed and mixed models) for the proportion of hits for ingredients A and B respectively.





**Figure 3.** Average proportions of observed hits (points) and fitted curves (fixed and mixed models) for adulterant A.



**Figure 4.** Average proportions of observed hits (points) and fitted curves (fixed and mixed models) for adulterant B.

### 3.3 Thresholds

The detection thresholds associated with ingredients A and B can be graphically interpolated through the fitted curves or estimated through the regression equations by making  $\eta = 0$  which is equivalent to  $\pi = 0,67$  as seen in equation (7) of Session 2.3,

$$\hat{\eta}_A = \hat{\beta}_0 + \hat{\beta}_1x = -6,1997 + 0,3173x, \tag{15}$$

$$\hat{\eta}_B = \hat{\beta}_0 + \hat{\beta}_1x = -6,8962 + 0,4267x. \tag{16}$$

Table 3 presents the estimates of detection thresholds (x in %) for ingredients A and B.

**Table 3.** Estimates of detection thresholds (%) for adulterants A and B

Ingredients	Threshold (%)
A	19.54
B	16.16

Overall, as the ingredient concentration increases, the easier it is to detect the difference. For both ingredients, the curves showed a good fit compared to the observed data, with ingredient B being more easily detectable than ingredient A, with a detection threshold around 16% and as the ingredient concentration approaches 30%, the hit probability for both ingredients approach 100%.

The curve associated with ingredient A showed a slightly smoother behavior, with this ingredient being more difficult to detect, with a detection threshold around 20%.

In this work we analyzed simulated data from a replicated triangular test and it was applied the graphical approach (or regression model fitting) to estimate the group detection thresholds for each ingredient. This technique has been frequently used, however, in general, works do not fit models that consider the random effect of the assessor.

The incorporation of the random effect was considered by Linander *et al.* (2019) when fitting a model to analyze data from a paired comparison test. However, the analysis focused on interpreting estimates of underlying sensory differences rather than detection thresholds.

The main concern on our work was to obtain information about sensory differences between products. However, when there is interest in obtaining information on preference between products, it has been noticed in the literature the use of preference tests as discussed by Prescott *et al.* (2005), which considers the analysis of rejection thresholds.

In this sense, Murray *et al.* (2019) used a graphical approach to estimate detection thresholds to determine the acceptability of new bioactive compounds added to milk-based beverages considering a 2-AFC forced choice test. In this context, Lima & Filho *et al.* (2015) also apply the graphical approach to estimate detection thresholds, suggesting, however, that acceptance tests are more indicated than preference tests, since one sample is less preferred than the other, it does not mean that it is sensorially rejected. The analysis uses a 9-point hedonic scale to evaluate the samples and the results are fitted to regression models, allowing the identification of the transition point between sensory acceptance and rejection.

These authors fit regression curves and use interpolation to obtain detection thresholds, as we suggested in this work. However, the analyzes still do not consider the random effect of the assessor. The model proposed in this article should be able to produce a better analysis, that is, more accurate estimates for the detection thresholds. In addition, the model can also be applied to estimate individual detection thresholds.

Brockhoff & Muller (1997) carried out this study, however, not taking into account the random effect of the assessor. They consider that individual thresholds are randomly distributed among individuals. The model takes into consideration the sensory distances of the Thurstonian approach and considers the random effects of individual thresholds.

Duineveld & Meyners (2008), in their study on discrimination rates, concluded that to obtain reliable conclusions about the distribution of discrimination rates, at least 5 or 6 repetitions of the triangular test are necessary. In future, we can explore the idea and analyze the number of repetitions needed to obtain reliable estimates of detection thresholds.

## 4. Conclusions

Our work substantiates the importance of considering a model that takes into account the random effect of the assessor in the analysis of detection thresholds in replicated discrimination tests. The proposed model has greater detection power, resulting estimates that are more accurate to detection thresholds in dose-response experiments. Not considering the random effect of the assessor may underestimate these estimates, that is, one can erroneously conclude that ingredients are detected at certain concentrations when in fact they should be detected at lower concentrations.

The proposed analysis still allows the use of unequal numbers of replications, which is more difficult to deal with in other approaches. This is most important when tests are not repeated on the same day but over several days or even weeks, so that some assessors are likely to miss one or several sessions.

Considering the random effect of the assessor is still challenging for researchers in the sensory area due to the difficulty associated with choosing the appropriate model, implementing and interpreting the results. Thus, justified the importance, this article presents the model and functions that can be directly applied in the estimation of detection thresholds.

## Acknowledgments

Authors would like to thank reviewers and editor for their valuable comments.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Altshuler, B. Modeling of Dose Response Relationships. *Environmental Health Perspectives*, **42**, 23-27, (1981).
2. American Society for Testing and Materials, ASTM. Standard practice for determining odor and taste thresholds by a forced-choice ascending concentration series method of limits, E-679. *Annual Book of Standards*, **15**,36-42, (2008a).
3. ASTM. Compilation of odor and taste threshold values *data*. American Society for Testing and Materials, Philadelphia, (1978).
4. Bates, D., Mächler, M., Bolker, B.M., & Walker, S.C. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1-48, (2015).
5. Bi, J. Difficulties and a way out: A bayesian approach for sensory difference and preference tests. *Journal of Sensory Studies*, **18**, 1-18, 2003.
6. Bi, J., & Ennis, D.M. A thurstonian variant of the beta-binomial model for replicated difference tests. *Journal of Sensory Studies*, **13**, 461-466, (1998).
7. Brockhoff, P.B. The statistical power of replications in difference tests. *Food Quality and Preference*, **14**, 405-417, (2003).
8. Brockhoff, P.B., & Christensen, R.H.B. Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference*, **21**, 330-338, (2010).
9. Brockhoff, P.B., & Schlich, P. Handling replications in discrimination tests. *Food Quality and Preference*, **9**, 303-312, (1998).
10. Brockhoff, P.M., & Muller, H-G. Random effect threshold models for dose-response relationships with repeated measurements. *Journal of the Royal Statistical Society: Series B*, **59**, 431-446, (1997).

11. Coleman, M., & Marks, H. Topics in dose-response modelling. *Journal of Food Protection*, **61**, 1425-1582, (1998).
12. Duineveld, K., & Meyners, M. Hierarchical Bayesian analysis of true discrimination rates in replicated triangle tests. *Food Quality and Preference*, **19**, 292-305, (2008).
13. Ennis, D.M., & Bi, J. The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies*, **13**, 389-412, (1998).
14. Harries, J.M., & Smith, G.L. The two-factor triangle test. *Food Science Technology*, **17**, 153-162, (1982).
15. Hunter, E.A., Piggot, J.R., & Lee, K.Y.M. Analysis of discrimination tests. In: Société Française de Statistique (Ed.), Actes des 6èmes Journées Européennes Agro-Industrie et Méthodes Statistiques, Pau, January, 19-21, (2000).
16. Kemp, S.E., Hollowood, T., & Hort J. *Sensory Evaluation: a practical handbook* (1st ed.). John Wiley & Sons, (2009).
17. Kunert, J. On repeated difference testing. *Food Quality and Preference*, **12**, 385-391, (2001).
18. Kunert, J., & Meyners, M. One the triangle test with replications. *Food Quality and Preference*, **10**, 477-482, (1999)
19. Lawless, H.T., & Heymann, H. *Sensory evaluation of food: principles and practices* (2nd ed. Springer, 2010).
20. Lima Filho, T., Minim, V. P. R., da Silva, R. C. S. N., Lucia, S. M. D., & Minim, L. A. Methodology for determination of two new sensory thresholds: Compromised acceptance threshold and rejection threshold. *Food Research International*, **76**, 561-566, (2015).
21. Linander, C.B., Christensen, R.H.B., Cleaver, G., & Brockhoff, P.B. Individual differences in replicated multi-product experiments with Thurstonian mixed models for binary paired comparison data. *Food Quality and Preference*, **75**, 220-229, (2019).
22. Meilgaard, M.C. Testing for Sensory Threshold of Added Substances. *American Society of Brewing Chemist*, (1991).
23. Meilgaard, M.C., Carr, B.T., & Civille, G.V. *Sensory Evaluation Techniques* (4th ed.). CRC Press, (2006).
24. Meyners, M., & Brockhoff, P.B. The design of replicated difference tests. *Journal of sensory studies*, **18**, 291-324, (2003).
25. Murray, N.M., Jacquier, J.C., O'Sullivan, M., Hallihan, A., Murphy, E., Feeney, E.L., & O'Riordan, D. Using rejection thresholds to determine acceptability of novel bioactive compounds added to milk-based beverages. *Food Quality and Preference*, **73**, 276-283, (2019).

26. Peltier, C., Brockhoff, P.B., Visalli, M., & Schlich, P. The MAM-CAP table: A new tool for monitoring panel performances. *Food Quality and Preference*, **32**, 24-27, (2014).
27. Prescott, J., Norris, L., Kunst, M., & Kim, S. Estimating a consumer rejection threshold for cork taint in white wine. *Food Quality and Preference*, **16**, 345-349, (2005).
28. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (2020).
29. Roessler, E.B., Pangborn, R.M., Sidel, J.L., & Stone, H. Expanded statistical tables for estimating significance in paired-preference, paired-difference, duo-trio and triangle tests. *Journal of Food Science*, **43**, 940-943, (1978).
30. Skellam, J.G. A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, **10**, 257-261, (1948).
31. Thurstone, L.L. A law of comparative judgment. *Psychology Review*, **34**, 273-286, (1927).
32. Ziegler, M., Gok, R., Bechtloff, P., Winterhalter, P., Schmarr, H-G., & Fischer, U. Impact of matrix variables and expertise of panelists on sensory thresholds of 1,1,6-trimethyl-1,2-dihydronaphthalene known as petrol off-flavor compound in Riesling wines. *Food Quality and Preference*, **78**, 103735, (2019)