Food Quality and Preference 83 (2020) 103924

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Accounting for the dimensionality of the dependence in analyses of contingency tables obtained with Check-All-That-Apply and Free-Comment

Benjamin Mahieu*, Michel Visalli, Pascal Schlich

Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

ARTICLE INFO

Keywords:
Correspondence analysis
Dimensionality test
Monte-Carlo test
Confidence ellipses
Chi-square per cell

ABSTRACT

Check-All-That-Apply (CATA) and Free-Comment (FC) provide a so-called contingency table containing citation counts of words or descriptors (columns) by products (rows). This table is most often analysed using correspondence analysis (CA). CA aims at decomposing dependence between products and descriptors into axes of maximal and decreasing dependencies, which is reasonable if the dependence has been previously established by a chi-square test. However, the p-value of this test is not valid when the observations are not independent or when the contingency table contains too many low expected citation rates. In addition, rejecting independence with a chi-square test only means that at least the first CA axis captures some dependence. This paper presents a test to determine the number of axes of the CA that capture significant dependence and proposes a Monte-Carlo approach to compute valid p-values for this test. The variability in the products' coordinates in the CA space is often evaluated by means of a total bootstrap procedure. The paper proposes to rely on this test to determine the number of axes to consider for the Procrustes rotations of such a procedure. Finally, to investigate which words are cited more often for each product, the paper proposes performing Fisher's exact tests per cell on the derived contingency table obtained by reversing the CA computations on the axes capturing significant dependence. The benefits of accounting for the dimensionality of the dependence in the analyses are demonstrated on real CATA data.

1. Introduction

In recent years, new consumer-oriented methods have emerged to overcome the limitations of sensory descriptive analysis (Valentin, Chollet, Lelièvre, & Abdi, 2012; Varela & Ares, 2012), including word citation occurrence-based methods, which aim to collect product descriptions from consumers using either their own words or a mutual predefined list of descriptors. These descriptions are collected without any quantification or product comparison. The most commonly used word citation occurrence-based methods are Check-All-That-Apply (CATA) (Adams, Williams, Lancaster, & Foley, 2007) and Free-Comment (FC) as response to open-ended questions (ten Kleij & Musters, 2003). Ultra-flash profiling (UFP) (Perrin & Pagès, 2009) and labelled sorting (Abdi & Valentin, 2007) could also be seen as word citation occurrence-based methods, but the word-based descriptive data are not the main output when using these two methods.

Data collected from a CATA or FC task are stored in a so-called contingency table containing citation counts of words or descriptors (columns) by products (rows). Each cell of the contingency table contains the number of times a product was described by a word. The first step to study such a dataset is to test for overall differences between products. In the context of contingency tables collected using FC, this is usually performed using a chi-square test (Galmarini, Symoneaux, Chollet, & Zamora, 2013; Lahne, Trubek, & Pelchat, 2014; Lawrence et al., 2013; Symoneaux, Galmarini, & Mehinagic, 2012). However, computing the p-value of the chi-square test using the chi-square distribution is valid only if the following conditions are met: (i) the observations are independent, (ii) no expected cell count is less than five in the contingency table (Agresti, 2007) and (iii) the contingency table is not sparse (Renter, Higgins, & Sargeant, 2000). In the context of contingency tables obtained using CATA or FC, these conditions are rarely met, especially the first condition, as all subjects evaluate all the products by assessing all the words. In the context of contingency tables collected using CATA, to address the issue of the non-validity of the chi-square distribution, Meyners, Castura, and Carr (2013) proposed to test for overall differences between products using a Monte-Carlo test based on combination of Cochran's Q statistics. In both contexts, if overall difference between products is not established, pursuing further analyses is not recommended. When overall difference between products is established, then a correspondence analysis (CA) (Benzécri, 1973) can

* Corresponding author.
E-mail address: benjamin.mahieu@inrae.fr (B. Mahieu).

https://doi.org/10.1016/j.foodqual.2020.103924
Received 22 November 2019; Received in revised form 27 February 2020; Accepted 28 February 2020
Available online 29 February 2020

be performed to visualise the association between products and words on a factorial map that decomposes the dependence between products and words into axes of maximal and decreasing dependencies. Furthermore, it is common to represent the variability in the products' coordinates in the CA space using confidence ellipses on the CA map. Confidence ellipses can be constructed in two ways: parametric bootstrap using a multinomial distribution (Antúnez, Ares, Giménez, & Jaeger, 2016; Oppermann, de Graaf, Scholten, Stieger, & Piqueras-Fiszman, 2017; Ringrose, 2012) or total bootstrap based on resampled subjects (Alcaire et al., 2017; Cadoret & Husson, 2013; Vidal, Ares, Hedderley, Meyners, & Jaeger, 2018). There are, to the best of our knowledge, two approaches to interpret relations between products and words in an objective manner. The first approach consists of computing the chi-square per cell on the contingency table (Symoneaux et al., 2012) to list words significantly more or less cited for each product; these words contribute the most to the global chi-square statistic. The second approach consists of performing the Multidimensional Alignment (MDA) on CA coordinates to interpret the cosine of the angle between product vectors and word vectors in the full CA space (Carr, Dzuroska, Taylor, Lanza, & Pansini, 2009; Meyners et al., 2013).

When overall difference between products is established, it only means that at least the first axis of the CA captures a significant dependence. From that result, there is a need to know how many other axes capture a significant dependence. Moreover, all computations performed with the analyses presented above are performed without considering how many axes capture a sufficient dependence to be considered significant. Thus, these methods do not take into account the dimensionality of the dependence and potentially add noise or miss important information needed for the interpretation.

The present paper proposes an approach that considers the dimensionality of the dependence when analysing CATA or FC data. The first section introduces a test of dimensionality based on chi-square statistic and on a Monte-Carlo approach to compute valid p-values. Chi-square statistic was chosen over the alternative Monte-Carlo test proposed by Meyners et al. (2013) because this latter is based on combination of Cochran's Q statistics that are not related to CA. The paper then explains how to take into account the information provided by the test when investigating the variability in the products' coordinates in the CA space and the relations between products and words. In the second section of this paper, the results obtained with this new approach are compared to those provided by the traditional analyses. In the last part, the benefits and limitations of both approaches are discussed. Finally, a global conclusion is given.

2. Material and methods

2.1. Testing dependence captured by the CA axes

Because CA and chi-square statistic belong to the same rationale, they are tightly related to each other. The tight relation between CA and chi square statistic gives interesting properties that enable testing the dependence captured by the CA axes. For this reason, the subsequently proposed test relies on chi-square statistic and not the test based on combination of Cochran's Q statistics proposed by Meyners et al. (2013). Further, contrarily to the Cochran's Q test that tests for equality of citation proportions across products for a given word, the chi-square test tests for independence between products and words and thus takes into account the total numbers of citations of the products (their margins).

The chi-square statistic of a contingency table is linked to the eigenvalues of the CA performed on this contingency table by the following equation:

$$\chi^2 = N \times \sum_i \lambda_i$$

where χ^2 is the chi-square statistic of the contingency table, N is the sum of all the cells of the contingency table, and λ_i is the i-th eigenvalue of the CA.

The sum of the eigenvalues of the CA can be seen as the effect size or the absolute intensity of the dependence between rows and columns. It is equal to the chi-square statistic divided by N and is thus based only on the observed and expected probabilities of being in each cell of the contingency table. Contrary to the chi-square statistic, it is independent of the sample size. Based on the above equation, it is possible to test for the dependence of each CA axis with a stepwise procedure (Camiz & Gomes, 2013). The idea is to test, at each step, whether removing the dependence captured by the axes of all the previous steps still results in rejecting independence in the sense of the chi-square test, i.e., if there is still enough dependence to be considered significant.

Suppose that we have a contingency table X of size n × p. The rank of X is equal to the minimum of (n-1) and (p-1) or less if there is a singularity. Let us denote this rank D. Let k vary from 1 to D until independence is not rejected for an axis. The principle of the stepwise procedure is as follows:

(i) At the k-th step, compute the following statistic: $Q_k = N \times \sum_{i=k}^{D} \lambda_i$
(ii) Compare this statistic to the quantiles of a chi-square distribution with (n-k)(p-k) degrees of freedom to obtain a p-value
(iii) If this p-value is less than the predetermined α risk, then set k = k + 1.

Running this procedure until independence is not rejected provides the number of CA axes that capture some significant dependence and thus the dimensionality of the data in the sense of dependence. The statistic computed at step k = 1 is equal to the statistic of the chi-square test. At step k (1 ≤ k ≤ D), the test is conceptually equivalent to perform a chi-square test on the derived contingency table represented only by the k-th to the D-th CA axes.

In practice, as stated in introduction (Section 1), computing the p-value of the chi-square test using the chi-square distribution is not valid in the context of contingency tables collected using CATA and FC. To overcome this limitation, a Monte-Carlo approach (Adery, 1968) is proposed. In such an approach, a large number of datasets are simulated under the null hypothesis investigated and then the statistic of interest is computed for each simulated dataset. These computations enable the user to obtain an empirical distribution under the null hypothesis with no probabilistic assumption. The statistic of interest computed on the real dataset is then compared to those of the simulated distribution under the null hypothesis, and the p-value is the proportion of the simulated statistics more extreme than or equal to the observed one. Here, the null hypothesis is independence between products and words on the k-th axis and the statistic of interest is Q_k.

The simulated data under the null hypothesis must be consistent with the nature of the data. In our case, the contingency table is obtained by summing the number of citations of each word for each product across the subjects. Simulating data by considering only the information provided by the observed contingency table, using, for example, Patefield's algorithm (Patefield, 1981), omits the subjects' individual information and thus is not appropriate. To overcome this limitation, independence can be simulated by randomly reallocating each word citation to a product by subject. However, this approach is problematic because it does not take into account the semantic nature of the words, so it could lead to unrealistic individual simulated data. For example, if a subject used the words "hard" and "soft" to describe a set of products, one can hope that both of these words were not used to describe the same product, but that could happen after random reallocation. For these reasons, this approach is also not appropriate. A more appropriate alternative to simulate consistent data consists of considering whole descriptions instead of words. Here, a description refers to the set of words used by one subject to describe one product. As these descriptions are indeed observed, they are realistic from a semantic point of view.

Thus, to obtain an empirical p-value for the test of dependence of each axis of the CA, a Monte-Carlo approach following these steps is proposed:

(i) Simulate B contingency tables by permuting the product labels of descriptions at the individual level and then compute the corresponding virtual contingency table
(ii) Perform a CA on each of the simulated contingency tables
(iii) Compute all $Q_k(1 \leq k \leq D)$ statistics for each of the simulated contingency tables
(iv) Compute the p-value of each Q_k as: $\frac{1+\sum_{s=1}^{B} I(Q_{k_S} \geq Q_{k_{obs}})}{1+B}$

where I is the identity function equal to 1 when its argument is true and 0 otherwise, B is the number of simulations (set to 1000 in following examples), $Q_{k_{obs}}$ is the observed statistic at step k, Q_{k_S} is the S-th $(1 \leq S \leq B)$ statistic at step k computed from the simulations and 1 stands for the observed contingency table (Davison & Hinkley, 1997).

The permutation procedure proposed here is the same one as the one proposed by Meyners et al. (2013) and is similar to the one proposed by Meyners and Pineau (2010) and Wakeling, Raats, and MacFie (1992).

2.2. Accounting for the dimensionality of the dependence when investigating the variability in the products' coordinates in the CA space

Performing a CA on the word-by-product contingency table does not account for the subject's variability, which means that it is impossible to assess the stability of the products' coordinates in the CA space, and thus it is impossible to know if the products are significantly discriminated. Computing the products' confidence ellipses with parametric bootstrap (Ringrose, 2012) presents two major limitations. First, it does not take into account the subjects' individual source of variation. Second, it assumes observations are independent from each other, for both products and words, which is not the case for CATA and FC data as explained in Section 2.1. This approach is thus not appropriate. The total bootstrap methodology (Cadoret & Husson, 2013) is well suited to compute confidence ellipses for the products' coordinates in a CA space. This methodology consists of generating virtual panels with random resampling with replacement of the actual panel. Then, the products' configurations of the virtual panels are rotated on the products' configuration of the actual panel thanks to Procrustes rotations. The total bootstrap methodology enables to take into account the specificity of the subjects' individual data as well as the dependence between observations. The main issue when using this methodology is to determine how many axes to take into account in the Procrustes rotations. It seems that this decision is usually arbitrary and can lead to taking into account for example two axes (Alcaire et al., 2017; Vidal et al., 2018) or four axes (Antúnez, Vidal, de Saldamando, Giménez, & Ares, 2017). The more axes one takes into account when performing the Procrustes rotations, the more degrees of freedom are available to find an optimal rotation and thus, the smaller the ellipses. Then, the decision to take into account only two axes can probably be explained by the fact that this is the most conservative option and thus protects from over-interpretation. However this practice can lead to overestimating the variability in the products' coordinates and thus to underestimating products' discrimination. It is necessary to have an objective criterion for selecting the number of dimensions of the space in which the Procrustes rotations must be performed. For that purpose, applying Procrustes rotations in the subspace generated by the significant CA axes is proposed.

2.3. Accounting for the dimensionality of the dependence when investigating relations between products and words

The two approaches presented in the introduction, the chi-square per cell and the MDA, differ in how they consider the data, but none of them considers the dimensionality of the dependence. In addition, MDA is flawed by the fact that it considers the angle between a product vector and a word vector but not their norms. Indeed, the vector norm represents the strength with which a product or a word deviates from the independence, which is crucial information that must be taken into account. To account for all the information, scalar products should be used instead of MDA. Even if the scalar products are the valid way to interpret relations between the product vectors and the word vectors in the CA space, it still has two limitations. First, the values of scalar products can be negative or positive and they are not bounded, thus they are not intuitive, difficult to interpret and can only be compared relative to each other. Second, to the best of our knowledge, there is no criterion to determine if a given scalar product is large enough to consider the association significant. Thus, the other approach, chi-square per cell, was retained. Nevertheless, this approach has some limitations. The chi-square distribution is not valid for use in this context because of the reasons evoked in introduction (Section 1) and even more because chi-square distribution is not adapted for 2 × 2 contingency tables (Yates, 1984). This limitation can be overcome using the Fisher's exact test (Fisher, 1935). This test has the benefit of not relying on any distribution and then requires no specific conditions to be met. The second limitation is that chi-square per cell is performed on the raw dataset and thus on all axes of dependence, which may result in accounting for axes that are just noise and thus may lead the user to over-interpret his or her data. To overcome this limitation and determine which words are the most cited for each product, the following approach is proposed:

(i) Establish the number of significant CA axes in the sense of dependence using the procedure presented in Section 2.1
(ii) Reverse the CA computations on the significant axes to compute the derived contingency table corresponding to the significant axes
(iii) Perform Fisher's exact tests per cell on the derived contingency table accounting for the significant axes

The step of reversing the CA computation on the significant axes is detailed in the Appendix.

2.4. Case study datasets

The study took place at the Centre for Taste and Feeding Behaviour, Dijon, France. Fifty-nine regular (at least once per two weeks) consumers of red wine (16 men, 43 women, 18 to 60 years old) were recruited from a population registered in the ChemoSens Platform's PanelSens database. This database has been registered with the relevant authority (Commission Nationale Informatique et Libertés—CNIL—authorisation no. 1148039). The subjects were compensated for their participation in the study. They carried out a CATA task on four French red wines from different regions: Bordeaux (Bor), Languedoc (Lan), Gamaret wine from Beaujolais (Gam) and Val de Loire (Val). For each product, the CATA task was carried out by sensory modality: visual, olfactory and gustatory. The gustatory description was itself divided into global perception and aromas. All the CATA descriptors were selected thanks to the expertise of wine professionals. The collected data were then stored in four contingency tables, one per step, by cross tabulating the citation counts of the descriptors (columns) by the products (rows).

2.5. Analyses

All analyses and computations were performed using R 3.5.1 (R Core Team, 2018). The examples are given using contingency tables collected with CATA but it is important to remember that all the presented approaches can be used with contingency tables collected with FC.

The aim of this case study is to compare the results provided by the analytical methods proposed in Sections 2.1 to 2.3 to those from methods belonging to the chi-square rationale commonly used on contingency tables. For that purpose, the results provided by the chi-square distribution and the Monte-Carlo approach for computing the p-values of the tests of dependence were compared, as well as the difference in results after performing the Procrustes rotations in the total bootstrap procedure using either the significant axes (when more than two) or the first two axes. For the tests of dependence, any p-value less than the α risk of 5% was considered significant. Ellipses of the total bootstrap procedure were computed with an α risk of 5%. In addition, the results of the use of Fisher's exact tests per cell accounting for all the axes were compared to the results from Fisher's exact tests per cell accounting for the significant axes in the sense of dependence. The Fisher's exact tests were conducted with a one-sided greater alternative hypothesis, which means that only cells with a larger observed value than the expected value were investigated. The results of these tests are presented with two different levels of α risk, namely $\alpha = 5\%$ and $\alpha = 15\%$. The motivation for this is not to miss descriptive information concerning the products. The results presented for $\alpha = 5\%$ can be considered as significant descriptions of the products while the results presented for $\alpha = 15\%$ can be considered as tendencies in the description of the products.

It is important to highlight here that the aim of the following case study was not to conduct full interpretation ending with product comparisons, but to compare only the outputs of the proposed analyses to those of the more traditional ones in order to underline the potential differences between them.

3. Results

3.1. Dependence of CA axes

Table 1 shows similar conclusions between the results provided by the chi-square distribution and the Monte-Carlo approach for the tests of dependence of the first axes. For the gustatory global perception data, regarding the other axes, the same conclusions are also provided by the two approaches: two axes are significant in the sense of dependence. In contrast, differences exist between the results provided by the chi-square distribution and the Monte-Carlo approach concerning the tests of dependence of the second and the third axes for the olfactory data and gustatory aromas data. According to the non-valid chi-square distribution, only the first axis is significant in the sense of dependence whereas the Monte-Carlo approach reveals that there are actually two significant axes in the sense of dependence for the gustatory aromas data and three significant axes in the sense of dependence for the olfactory data.

3.2. Variability in the products' coordinates in the CA space

Fig. 1 shows information in line with the tests of dependence based on the Monte-Carlo approach.

For the visual sense CA, Fig. 1 (a) shows that ellipses confirm the results provided by the Monte-Carlo approach since the ellipses' projections on the second axis strongly overlap.

For the olfactory sense CA, Fig. 1(d) shows that the third axis indeed captures some dependence and information as it isolates the product Val from the others. If the usual relative criterion of accounting for approximately 70–80% of the inertia was used for this CA, the third dimension would not have been considered and thus some information would have been lost. Further, the comparison of Fig. 1(b) and Fig. 1(c) is a great example of possible misinterpretations and missed information resulting from arbitrarily setting the number of axes to two to perform the Procrustes rotations in the total bootstrap procedure. Indeed, looking at Fig. 1(b), the product Val seems not to be different from the products Gam and Lan whereas it is indeed on the third axis as well as on the second axis when all significant axes are considered for Procrustes rotation (Fig. 1(c)). This information is taken into account when setting the relevant number of axes to perform the Procrustes rotations. Thus looking at Fig. 1(c), we can see that Val is different from Gam and Lan. This example shows the real importance of taking into account all the significant axes in the sense of dependence to perform the Procrustes rotations in the total bootstrap procedure.

For the gustatory global perception CA, ellipses also confirm the results provided by the Monte-Carlo approach since the products Val and Lan are different from the products Bor and Gam on the second axis (Fig. 1(e)).

For the gustatory aromas CA, ellipses, computed with the most conservative option, show that the second axis captures a significant dependence as two product pairs (Val vs. Bor & Val vs. Lan) are different on the second axis, while the p-values computed using the chi-square distribution suggest that this second axis is not significant. In this example, the Monte-Carlo approach, compared to the chi-square distribution, seems to be better aligned with the information provided by the ellipses.

3.3. Relations between products and words

Fig. 2 shows that using Fisher's exact tests per cell on all the axes leads to the over-interpretation of some dependent relations that are not significant. Indeed, for the visual data, when accounting for all the axes, there are tendencies for the product Gam to be more associated with the words Black and Opaque whereas when accounting only for the first significant axis, the product Gam is definitely associated with the word Violet and not associated with the words Black and Opaque. The product Val, when accounting for all the axes, is associated with the word Violet whereas when accounting only for the first significant axis, the product Val is not associated with any words. For the gustatory aromas data, when accounting for all the axes, there are tendencies for the products Gam and Lan to be more associated with the word Red fruit whereas when accounting for the first two significant axes, the product Gam is not associated with any words, and the product Lan is definitely associated with the word Red fruit. These two examples show the need to perform Fisher's exact tests per cell using only the

Table 1
P-values of the test of dependence for each axis of each correspondence analysis performed on the four contingency tables computed by either the chi-square distribution or the Monte-Carlo approach.

Sensory modality	Computation of the p-value	Chi-square/Axis 1	Axis 2	Axis 3
Visual sense	Chi-square distribution	< 0.001	0.9882	0.9403
	Monte-Carlo approach	< 0.001	0.5134	0.3016
Olfactory sense	Chi-square distribution	< 0.001	0.0545	0.5132
	Monte-Carlo approach	< 0.001	0.0019	0.0089
Global perception from the gustatory sense	Chi-square distribution	< 0.001	0.0309	0.8652
	Monte-Carlo approach	< 0.001	< 0.001	0.1448
Aromas from the gustatory sense	Chi-square distribution	0.0032	0.3378	0.8635
	Monte-Carlo approach	< 0.001	0.0069	0.2507

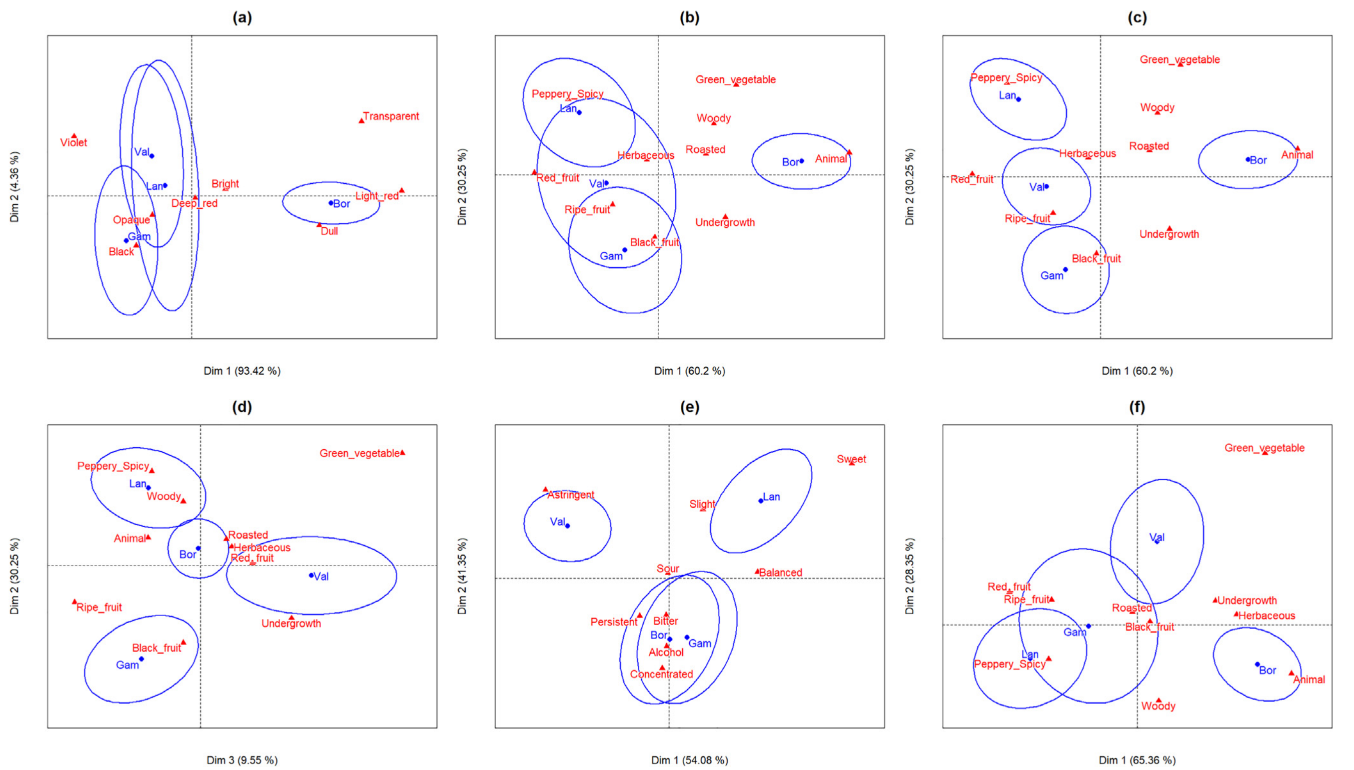

Fig. 1. Correspondence analysis of the four contingency tables with confidence ellipses computed with total bootstrap: (a) axes 1–2 of the visual sense with total bootstrap considering the first two axes, (b) axes 1–2 of the olfactory sense with total bootstrap considering the first two axes, (c) axes 1–2 of the olfactory sense with total bootstrap considering the three axes, (d) axes 3–2 of the olfactory sense with total bootstrap considering the three axes, (e) axes 1–2 of the global perception from the gustatory sense with total bootstrap considering the first two axes, (f) axes 1–2 of the aromas from the gustatory sense with total bootstrap considering the first two axes.

information provided by the significant axes in the sense of dependence. This approach prevents the user from over-interpreting some of the associations that are not sufficiently strong to be considered significant and prevents the user from missing some significant associations due to tests performed on a dataset containing noise.

The example of the global perception gustatory data shows that the differences between accounting for all the axes and accounting for the significant axes in the sense of dependence sometimes do not drastically change the conclusions. In this example, the differences are only based on some tendencies of associations.

For the olfactory data, by construction, no difference exists since all the axes are significant.

4. Discussion

The Monte-Carlo approach had a real benefit in the computation of the p-values of the chi-square test and the tests of dependence of the CA axes. Indeed, it enabled the consistent estimation of the distribution under the null hypothesis that takes into account the nature of the data. The examples presented showed that p-values computed with the Monte-Carlo approach and with the chi-square distribution do not always lead to different conclusions. However, it is common to find differences between these two approaches. As shown in the examples, the Monte-Carlo approach always provided information in line with the one provided by the confidence ellipses contrary to the chi-square distribution. This finding shows that in addition to its theoretical benefit of taking into account the nature of the data, in practice the Monte-Carlo approach also provided conclusions consistent with other information. Furthermore, in the given examples, the p-values of the Monte-Carlo approach were systematically lower than those computed using chi-square distribution. This finding suggests a higher power in dimensionality detection for the Monte-Carlo approach. Despite its benefits, the Monte-Carlo approach has a limitation: the computational time. Simulating 1000 contingency tables with the procedure explained in Section 2.1.2 takes between 10 and 20 s. If the user wants to simulate more contingency tables to better estimate the distribution under the null hypothesis, the computational time can rapidly increase.

To the best of our knowledge, testing the dependence of the CA axes has never been used in sensory and consumer research. This test is a great improvement in the analysis of contingency tables collected with CATA and FC. It enables the determination of the number of dimensions in which the dependence between products and words, if any, is large enough to be considered significant according to a statistical criterion. It prevents misinterpretations or over-interpretations and missing relevant information provided by CA axes beyond the first plan. The result of this test is also a solid basis on which further computations can rely such as the total bootstrap procedure and the investigation of associations between products and words. For the total bootstrap procedures applied on CATA and FC data, these tests are a real improvement as they provide an objective and relevant manner of determining how many axes must be taken into account for the Procrustes rotations, which prevents the user from considering two products as not being significantly discriminated when they are indeed.

Fisher's exact tests were performed with a one-sided greater alternative meaning that only observed counts that were potentially larger than the expected counts were investigated. This choice was made because of the task asked to subjects. Concerning the FC task, it is asked to subjects to describe the products in their own words. It is thus reasonable to assume that the words used to describe a product are indeed descriptive of and applicable to the product. However, assuming that because a subject does not say a given word for a product implies that this word is not applicable to the product is a very strong assumption. For the CATA task, the situation is a little different: subjects are asked to quote among a list of words, which ones apply to the products. It is thus

		Fisher's exact tests per cell on all the axes				Fisher's exact tests per cell on the significant axes			
		Bor	Gam	Lan	Val	Bor	Gam	Lan	Val
Visual sense	Violet	3	20	19	25	3	20	19	25
	Opaque	23	38	33	35	23	38	33	35
	Dull	8	3	4	3	8	3	4	3
	Light red	18	3	4	6	18	3	4	6
	Bright	33	24	30	27	33	24	30	27
	Deep red	34	34	37	35	34	34	37	35
	Black	9	19	14	15	9	19	14	15
	Transparent	9	1	4	4	9	1	4	4
		Bor	Gam	Lan	Val	Bor	Gam	Lan	Val
Olfactory sense	Black fruit	25	33	17	24	25	33	17	24
	Roasted	6	3	4	4	6	3	4	4
	Red fruit	8	19	25	26	8	19	25	26
	Green vegetable	5	0	3	4	5	0	3	4
	Peppery / Spicy	9	8	22	12	9	8	22	12
	Ripe fruit	8	13	11	7	8	13	11	7
	Animal	24	7	7	5	24	7	7	5
	Undergrowth	19	13	6	15	19	13	6	15
	Herbaceous	4	3	4	4	4	3	4	4
	Woody	21	8	16	11	21	8	16	11
		Bor	Gam	Lan	Val	Bor	Gam	Lan	Val
Global perception from the gustatory sense	Alcohol	22	26	13	17	22	26	13	17
	Slight	17	17	26	21	17	17	26	21
	Astringent	17	14	16	40	17	14	16	40
	Bitter	15	14	10	12	15	14	10	12
	Concentrated	23	21	10	14	23	21	10	14
	Balanced	24	18	26	12	24	18	26	12
	Sweet	3	7	13	3	3	7	13	3
	Persistent	29	28	17	27	29	28	17	27
	Sour	19	18	17	19	19	18	17	19
		Bor	Gam	Lan	Val	Bor	Gam	Lan	Val
Aromas from the gustatory sense	Red fruit	6	19	21	16	6	19	21	16
	Ripe fruit	6	11	12	10	6	11	12	10
	Green vegetable	5	2	1	9	5	2	1	9
	Black fruit	28	26	21	23	28	26	21	23
	Roasted	5	4	5	5	5	4	5	5
	Peppery / Spicy	13	13	25	12	13	13	25	12
	Herbaceous	7	3	3	5	7	3	3	5
	Woody	20	13	15	7	20	13	15	7
	Undergrowth	22	13	10	18	22	13	10	18
	Animal	18	6	5	7	18	6	5	7

Fig. 2. Contingency tables of the four CATA tasks. The highlighted cells show the significant results of Fisher's exact tests per cell considering all the axes and the significant results of Fisher's exact tests per cell considering the significant axes in the sense of dependence. The cells highlighted in light green are significant for $\alpha = 5\%$, and those highlighted in deep green are significant for $\alpha = 15\%$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reasonable to assume that a descriptor that was not used to describe a product was not perceived by the subject. This can be considered a more active decision than not to cite some words in a FC task, but still the guideline was to “check-all-that-apply” and not to “not check what does not apply”. Considering these points, the decision of performing one sided greater alternative tests or two-sided alternative tests is up to the discretion of the user.

As an overall limitation, it has to be mentioned that the practical results provided through the examples arose from datasets where only four products were evaluated using CATA. The relevant results of this paper need to be confirmed on other datasets with more products and with different levels of similarity between the products.

5. Conclusion

This paper introduced a complete set of statistical tools enabling to account for the dimensionality of the dependence in contingency tables obtained with CATA and FC. First, this set includes a chi-square-based test for determining the number of significant axes in CA of a contingency table. As p-values derived from chi-square distribution are not valid in the context of contingency tables based on CATA or FC data, an alternative Monte-Carlo approach was proposed. Secondly, it was shown that the Procrustes rotations in a total bootstrap procedure to derive product confidence ellipses should be done in the subspace defined by the significant axes. Finally, to investigate which words are cited more often for each product, the paper proposed to perform Fisher’s exact tests per cell on the derived contingency table obtained by reversing the CA computation on the axes capturing significant dependence. These new tools should help the users of CATA and FC to analyse their data with more precision as the methods removed noise due to non-significant dimensions in term of dependence between products and attributes or words.

CRediT authorship contribution statement

Benjamin Mahieu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Michel Visalli:** Conceptualization, Software, Validation, Resources, Data curation, Writing - review & editing. **Pascal Schlich:** Conceptualization, Validation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This paper is part of a Ph.D. financed by the Region Bourgogne-Franche-Comté and the company SensoStat.

Appendix:. Reversing the correspondence analysis computations

Let X be a contingency table. Performing a correspondence analysis on X consists of computing the standardised residual matrix R from X and then factorising R using Singular Value Decomposition (SVD). Factorising R using a SVD consists of writing R as follows:

$$R = UDV^{'}$$

The SVD of R is performed with weights for rows and columns equal to their respective marginal probabilities. The coordinates of the rows and the columns as well as the eigenvalues of the CA can directly be computed from U, D and V. For more details on this process and the computations, one can refer to Bock (2011).

Reversing the CA computations on the significant axes consists of computing R_{sig} as follows:

$$R_{sig} = U_{sig}D_{sig}V^{'}_{sig}$$

where U_{sig} is determined from the rows coordinates of only the significant axes, D_{sig} is determined from the eigenvalues of only the significant axes and $V^{'}_{sig}$ is determined from the columns coordinates of only the significant axes. Therefore, non-significant dependence is discarded. One critical aspect in the computations of U_{sig} and $V^{'}_{sig}$ is to determine if the software used to perform the CA returns principal coordinates or standard coordinates of the rows and the columns. U_{sig} and $V^{'}_{sig}$ have to be weighted back by the observed marginal probabilities before the computation of R_{sig}.

X_{sig} can then be computed from R_{sig} using the observed expected probabilities and the observed grand sum of X.

References

Abdi, H., & Valentin, D. (2007). Some new and easy ways to describe, compare, and evaluate products and assessors. *Proceedings of SPISE,* 5–18.
Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). *Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. 7th Pangborn Sensory Science Symposium. Minneapolis, USA*.
Adery, C. A. H. (1968). A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society. Series B (Methodological), 30*(3), 582–598.
Agresti, A. (2007). *An Introduction to categorical data analysis* (second ed.). New York: John Wiley & Sons.
Alcaire, F., Antunez, L., Vidal, L., Zorn, S., Gimenez, A., Castura, J. C., et al. (2017). Comparison of static and dynamic sensory product characterizations based on check-all-that-apply questions with consumers. *Food Research International, 97*, 215–222.
Antúnez, L., Ares, G., Giménez, A., & Jaeger, S. R. (2016). Do individual differences in visual attention to CATA questions affect sensory product characterization? A case study with plain crackers. *Food Quality and Preference, 48*, 185–194.
Antúnez, L., Vidal, L., de Saldamando, L., Giménez, A., & Ares, G. (2017). Comparison of consumer-based methodologies for sensory characterization: Case study with four sample sets of powdered drinks. *Food Quality and Preference, 56*, 149–163.
Benzécri, J.-P. (1973). *Analyse des données. Analyse des correspondances, vol. 2*. Paris: Dunod.
Bock, T. (2011). Improving the display of correspondence analysis using moon plots. *International Journal of Market Research, 53*(3), 307–326.
Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference, 28*(1), 106–115.
Camiz, S., & Gomes, G. C. (2013). *Joint correspondence analysis versus multiple correspondence analysis: A solution to an undetected problem. Classification and Data Mining*.
Carr, B. T., Dzuroska, J., Taylor, R. O., Lanza, K., & Pansini, C. (2009). *Multidimensional Alignment (MDA): A simple numerical tool for assessing the degree of association between products and attributes on perceptual maps. 8th Pangborn sensory science symposium. Florence, Italy*.
Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.
Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, 98*(1).
Galmarini, M. V., Symoneaux, R., Chollet, S., & Zamora, M. C. (2013). Understanding apple consumers' expectations in terms of likes and dislikes. Use of comment analysis in a cross-cultural study. *Appetite, 62*, 27–36.
Lahne, J., Trubek, A. B., & Pelchat, M. L. (2014). Consumer sensory perception of cheese depends on context: A study using comment analysis and linear mixed models. *Food Quality and Preference, 32*, 184–197.
Lawrence, G., Symoneaux, R., Maitre, I., Brossaud, F., Maestrojuan, M., & Mehinagic, E. (2013). Using the free comments method for sensory characterisation of Cabernet Franc wines: Comparison with classical profiling in a professional context. *Food Quality and Preference, 30*(2), 145–155.
Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference, 30*(2), 309–319.
Meyners, M., & Pineau, N. (2010). Statistical inference for temporal dominance of sensations data using randomization tests. *Food Quality and Preference, 21*(7), 805–814.
Oppermann, A. K. L., de Graaf, C., Scholten, E., Stieger, M., & Piqueras-Fiszman, B. (2017). Comparison of Rate-All-That-Apply (RATA) and Descriptive sensory Analysis (DA) of model double emulsions with subtle perceptual differences. *Food Quality and Preference, 56*, 55–68.
Patefield, W. M. (1981). Algorithm AS 159: An efficient method of generating random R × C tables with given row and column totals. *Applied Statistics, 30*(1).
Perrin, L., & Pagès, J. (2009). Construction of a Product Space from the Ultra-Flash

Profiling Method: Application to 10 Red Wines from the Loire Valley. *Journal of Sensory Studies, 24*(3), 372–395.

R Core Team (2018). *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing*.

Renter, D. G., Higgins, J. J., & Sargeant, J. M. (2000). *Performance of the exact and chi-square tests on sparse contingency tables. Conference on Applied Statistics in Agriculture*.

Ringrose, T. J. (2012). Bootstrap confidence regions for correspondence analysis. *Journal of Statistical Computation and Simulation, 82*(10), 1397–1413.

Symoneaux, R., Galmarini, M. V., & Mehinagic, E. (2012). Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality and Preference, 24*(1), 59–66.

ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: A complementary method to preference mapping. *Food Quality and Preference, 14*(1), 43–52.

Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International Journal of Food Science & Technology, 47*(8), 1563–1578.

Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*(2), 893–908.

Vidal, L., Ares, G., Hedderley, D. I., Meyners, M., & Jaeger, S. R. (2018). Comparison of rate-all-that-apply (RATA) and check-all-that-apply (CATA) questions across seven consumer studies. *Food Quality and Preference, 67*, 49–58.

Wakeling, I. N., Raats, M. M., & MacFie, H. J. H. (1992). A new significance test for consensus in generalized procrustes analysis. *Journal of Sensory Studies, 7*(2), 91–96.

Yates, F. (1984). Test of significance for 2 × 2 contingency tables. *Journal of the Royal Statistical Society. Series A (General), 147*(3).