Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1914

One citation, one vote! A new approach for analysing check-all-that-apply (CATA) data using L1-norm methods

C. Chaya, J. C. Castura, and M. J. Greenacre

September 2025

One citation, one vote! A new approach for analysing check-all-that-apply (CATA) data using L1-norm methods

C. Chaya[1*], J.C. Castura[2] and M.J. Greenacre[3]

[1*]Department of Agricultural Economics, Statistics and Business Management, Business Innovation and Management Group, Universidad Politécnica de Madrid, Spain.
[2]Compusense Inc., Guelph, Canada.
[3]Department of Economics and Business, Universitat Pompeu Fabra, and Barcelona School of Management, Barcelona, Spain.

*Corresponding author(s). E-mail(s): carolina.chaya@upm.es;
Contributing authors: jcastura@compusense.com;
michael.greenacre@upf.edu;

Abstract

A unified framework is provided for analysing check-all-that-apply (CATA) product data following the "one citation, one vote" principle. CATA data arise from studies where A assessors evaluate P products by describing samples by checking all of the T terms that apply. Giving every citation the same weight, regardless of the assessor, product, or term, leads to analyses based on the L1 norm where the median absolute deviation is the measure of dispersion. Five permutation tests are proposed to answer the following questions. Do any products differ? For which terms do products differ? Within each of the terms, which products differ? Which product pairs differ? On which terms does each product pair differ? Additionally, we show how products and terms can be clustered following the "one citation, one vote" principle and how principal component analysis using the L1-norm (L1-PCA) can be applied to visualize CATA results in few dimensions. Together, the permutation tests, clustering methods, and L1-PCA provide a unified approach that provides robust results

measured in citation percentages. The proposed methods are illustrated using a data set in which 100 consumers evaluated 11 products using 34 CATA terms. R code is provided to perform the analyses.

Keywords: check-all-that-apply (CATA), pick-any, permutation tests, median absolute deviation (MAD), L1 norm, L1-norm principal component analysis (L1-PCA)

1 Introduction

Sensory and consumer studies are often designed to investigate how humans perceive foods, beverages, or non-food products. These studies usually involve structured sample evaluations of the products by human assessors with data analysis by sensometric methods. Some studies obtain precise sensory characterizations of the products from panelists trained to identify and scale sensory attributes from a consensus lexicon. Other studies investigate how consumers perceive and conceptualise products. Both types of studies frequently use the check-all-that-apply (CATA) question format (Ares & Jaeger, 2023; Meyners & Castura, 2014), which provides a list of *terms* to allow *assessors* to characterise *products*. For example, a CATA question might instruct the assessor to check as many sensory or emotion terms from the list as describe the product sample under evaluation. Such a study leads to a three-way array (products-by-terms-by-assessors) of raw data.

Since checked terms, called *citations* or *elicitations*, can be analysed as count data, CATA data are often aggregated over the assessors to obtain a two-way products-by-terms table of frequencies. Thus, a CATA frequency table counts the assessors who cited each term for each product.

CATA frequency tables are often analysed as contingency tables (Dooley et al., 2010; Mahieu et al., 2021; Meyners & Castura, 2014; Meyners et al., 2013; Valentin et al., 2012), in which a normalisation or adjustment is applied that in some sense equalises the contributions of variables with different citation rates. The widely used normalisation inherent in the chi-squared test has this feature: each cell frequency is expressed relative to the reciprocal of the square root of its expectation, given the column and row marginal sums. However, a CATA frequency table is not a contingency table. In a contingency table, each member of the assessor sample contributes to only one table cell. In a CATA frequency table, each cell indicates how many assessors cited a particular term (column) for a particular product (row). For this reason, we propose to analyse raw counts without normalising terms or products. We will investigate differences between products for each term separately (univariately) and over all terms simultaneously (multivariately), using the novel approach of treating each citation equally. Our proposal is embodied in the following principle: one citation, one vote!

This proposal provides a new way of summarising and analysing CATA data that we have not seen used previously. Treating every citation equally leads us to use medians (not means) as our measure of central tendency, medians of absolute deviations from medians (not variances) as our measure of dispersion, and median absolute deviations (not Euclidean distances) to quantify differences when grouping products and grouping terms in cluster analysis. Additionally, we use L1-norm principal component analysis (L1-PCA; Ke & Kanade, 2005), not the conventional L2-norm PCA (L2-PCA), to summarise the product information visually in few dimensions.

As we will show, treating all citations equally gives many advantages. Terms are not rescaled, as in some other analyses, so infrequently cited terms having small product differences are not adjusted to carry the same weight as frequently cited terms having large differences between products. These analyses yield results that are easy to interpret and communicate. A drawback of our approach is we cannot test hypotheses in the classical manner (e.g., chi-squared tests), but we overcome this disadvantage using resampling procedures (e.g. permutation tests, bootstrapping).

Data and methods are introduced in Section 2. The data come from a study involving 100 consumers who used 34 emotion terms to evaluate 11 blackcurrant squash products (fruity beverages) available commercially in the United Kingdom (Ng et al., 2013a). There, we also describe the proposed statistical methods, all of which are aligned with the "one citation, one vote" principle. We describe how to conduct distribution-free permutation tests, as well as how to adjust for multiple testing. We describe how to perform cluster analysis of products and terms, as well as how to explore relationships between products and terms using L1-PCA. Section 3 presents the results obtained from applying these methods to the blackcurrant squash CATA data. We discuss these results and make conclusions in Section 4.

2 Materials and Methods

2.1 Data from the blackcurrant squash study

This paper uses data extracted from research by Ng et al. (2013a) in which consumers evaluated 11 commercial blackcurrant squashes, representing a range of sensory and packaging properties observed in the UK market. The original labelling of the 11 squash products (P1 to P11) in Ng et al. (2013a) is supplemented here by two additional letters (ab). Letter a indicates the economical, niche and standard market-value categories, respectively a = e, n, s. Letter b indicates products with and without added sugar, respectively b = w, o. For example, the first product P1sw is a standard product with added sugar.

A preliminary study described by Ng et al. (2013b) was conducted to select relevant conceptual terms with emotional or functional connotations (Thomson et al., 2010) for evaluating blackcurrant squashes under three conditions: blind, pack, and informed. In this manuscript, we analyse results from the

blind condition only, in which consumers evaluated each sample without seeing the package. In this condition, consumers' emotional conceptualizations of the samples are not influenced by information about brand, assumptions about the market-value category (niche vs standard vs economical), added sugar, or use of non-nutritive sweeteners, if any.

Results from the preliminary study were used to finalise a CATA questionnaire. Assessors were tasked with checking all terms that described how they felt immediately after tasting each sample. The 34 emotion terms of the blind condition can be found in Table 2.

In total, $A = 100$ assessors evaluated the $P = 11$ products according to the $T = 34$ terms. Products were presented by experimental design to balance sample order effects. Terms were presented by experimental design to balance term position effects.

The three-way raw data, depicted schematically in Fig. 1 contains indicator variables: zeros indicate terms not cited; ones indicate terms cited. From these raw data, the CATA frequency table is the $P \times T$ table formed by summing the A slices across the assessors.

2.2 A new approach, simple and robust

As described in Section 1, when considering how to analyse CATA data, we began with the following premise: each citation has exactly the same importance, regardless of the term elicited, the product evaluated, or the assessor who made the judgment. Since every assessor evaluates all products using the same terms, the frequency scale (i.e., counts) and the percentage scale (i.e., expressing each cell as a percentage of A) are equivalent. For ease of reference, the table of citation percentages will be called a CATA table.

Many previous approaches for analysing CATA data are based on classical statistics and involve squared differences, which are used to compute variances, least-squares estimates, chi-squared statistics, and so on. Even a simple statistic such as the arithmetic mean, which estimates an expected value, is based on least-squares estimation: the mean is the value that minimises the sum of squared differences to the sample values. The mean is sensitive to outliers because least-squares estimation gives greater weight to more-distant sample values and lesser weight to more-central ones. For example, the counts 10 and

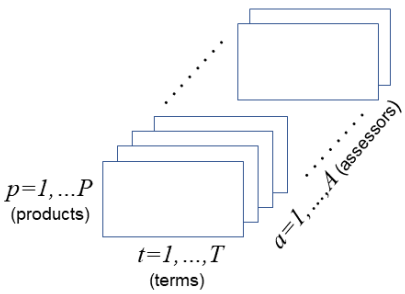

Fig. 1 Structure of raw CATA data in the form of a three-way array, where cell (p, t, a) has the value 1 if assessor a cited term t for product p; otherwise, the cell value is 0. The three-way array is shown as a series of A tables corresponding to the assessors. The summation of these $P \times T$ tables over all the A assessor slices yields the $P \times T$ CATA frequency table of counts. For CATA percentages, the counts in each cell are expressed as a percentage of A.

12 have the absolute difference 2 and the squared difference 4, whereas the counts 10 and 14 have the absolute difference 4 and the squared difference 16. Squaring amplifies larger differences more than smaller differences, which is inconsistent with the "one citation, one vote" principle.

To treat every response equally, we will measure differences by their absolute values and use the median as a measure of centrality. The median is the value that minimises the sum of absolute differences to the sample values (Schwertman et al., 1990). We measure dispersion about the median by finding the median of the absolute deviations between the sample values and the median, called the *median absolute deviation* (MAD) about the median. Since absolute differences are less sensitive to outliers, the median and MAD are more robust measures of central tendency and overall difference, and fit the "one citation, one vote" principle.

If $\mathbf{E}$ is a CATA (percentage) table, then each element e_{pt} gives the percentage of the A assessors who cited term t for product p. In this case, $\mathsf{M}(t) \equiv \mathsf{M}\{e_{pt}\,;\ p = 1, \ldots, P\}$ denotes the median citation percentage for term t across the P products, i.e., the median of column t of matrix $\mathbf{E}$. The MAD for term t is defined as the median of absolute deviations from the median:

$$\mathrm{MAD}(t) \equiv \mathsf{M}\{|e_{pt} - \mathsf{M}(t)|\,;\ p = 1, \ldots, P\} \tag{1}$$

For example, in the CATA table that we will analyse soon, the following percentage values are observed for the set of 11 products for the two terms *Happy* and *Sickly*:

Happy (%)	54	31	15	22	40	25	55	28	41	15	48
Sickly (%)	21	25	19	15	12	30	12	35	17	14	12

Happy has the median percentage 31% across the products, while *Sickly* has the median percentage 17%. Within each set, we calculate the MAD dispersion values using the sample values and their respective medians. The dispersions are 10% for the term *Happy* and 5% for the term *Sickly*.

We draw attention to three points. First, as just illustrated, the measures of central tendency and dispersion usually increase in tandem. When least-squares methods are applied in multivariate analyses, terms having unequal variances are often rescaled to equalise their contributions to the analysis; for example, in correspondence analysis (Benzécri, 1973; Greenacre, 2016), the chi-square normalisation rescales variables and samples so all are on the same footing. In our case, following the "one citation, one vote" principle, we do not attempt to balance the influence of any terms, products, or assessors. Larger differences between products based on citations of the term *Happy* are considered to discriminate the products better than the smaller differences between products based on citations of the term *Sickly*. Second, the summary values used (median, MAD) are very easy to understand and interpret because they are measured in percentages, which is exactly the same scale used in the CATA table. Third, the median and the MAD are more robust to outliers than

measures based on squared differences, such as the mean and variance, which are influenced more strongly by outlying sample values.

When it comes to hypothesis testing, instead of using classical statistics with known, usually asymptotic theoretical distributions based on squared differences and least-squares methods, we will take a distribution-free approach, described next.

subsectionPermutation testing

To test differences between products on the selected terms, we make use of the simple, non-parametric approach called permutation testing; for a complete treatment, see Good (2004), and for previous applications in the analysis of CATA tables, see Meyners et al. (2013) and Mahieu et al. (2021). As we will show, this approach can be used to test various hypotheses, such as whether individual cells in the CATA table are larger or smaller than expected or whether a pair of products differs on any of the terms. The approach naturally generalises to multivariate tests of whether products differ globally, accounting for all terms simultaneously.

The null hypothesis assumes that there is no difference between products. If the null hypothesis is true, then any product differences have occurred by chance. We can simulate a potential outcome under this null hypothesis by permuting product labels within each assessor, leading to a new CATA table. The permutation procedure requires the raw data and is repeated many times, each time producing another random CATA table. In this study, we have generated $B = 9\,999$ random CATA tables, which generate a null distribution of the chosen test statistic. These CATA tables represent a range of potential outcomes that could have occurred by chance assuming the products in fact elicit the same set of terms. The chosen test statistic is computed for the observed CATA table and for each of the B simulated CATA tables, giving a null distribution comprising $B+1$ values in total. Suppose the value of the test statistic from the observed table is so extreme that we almost never see values as or more extreme. The low probability of observing such an extreme outcome, indicated by a very small p-value, leads us to conclude that the null hypothesis is improbable. Based on this result, we would conclude that products differ with a high level of confidence, as indicated by the small p-value estimated from the null distribution.

Now, we describe the procedure in more specific detail. Operationally, each simulation takes the form of random permutations of the product labels within each assessor. Fig. 1 shows the set of citations in the $P \times T$ matrix "slice" of the array corresponding to a particular assessor. Permuting the P rows of this matrix assigns rows to different product labels but preserves dependencies in elicitations across the multivariate terms. For example, a particular assessor who tends to endorse the terms *Pleased* and *At ease* together or not at all has data dependencies that are kept intact by permuting the rows (products) keeping the columns (terms) intact. This procedure is applied to data from all A assessors. Since rows are assigned to product labels independently per assessor, any two assessors will usually not have the same product rows assigned to the

same product labels. After permuting product rows within assessors, assessor matrix slices are aggregated into a simulated $P \times T$ CATA table. Repeating this procedure B times yields B random CATA tables, which together with the original CATA table, gives $B + 1$ CATA tables in total.

Suppose we want to evaluate the evidence against a null hypothesis that products have the same citation percentage for a particular term t. The test statistic chosen to quantify the observed product differences is the MAD statistic given by Eq. (1), that is, the median absolute deviation from the P products to their median value. The null distribution for term t is comprised of the MAD statistic from the observed CATA table and B MAD values from the random CATA tables. The p-value is the proportion of these $B + 1$ MAD values that are greater than or equal to the observed MAD statistic. The MAD statistic and corresponding p-value can be obtained for each of the T terms in the same manner.

Since the observed MAD statistic is part of the null distribution, at least one MAD value from the null distribution always meets this criterion. For this reason, if a null distribution contains $B+1 = 10\,000$ MAD values, the smallest possible p-value (0.0001) is obtained if the observed MAD statistic is larger than all other MAD values in the null distribution.

The beauty of permutation testing is that the researcher can select any suitable test statistic, even one with an unknown statistical distribution. Since we have generated $B = 9\,999$ random CATA tables, a p-value can be estimated to four decimal places. Performing more permutations would give even more precise p-values. In the present context, we propose the permutation tests described in Table 1. The testing procedure is straightforward and, in many cases, permutation results obtained for one test can be re-used in another test.

Test #	Type	Question addressed	Max. tests
1	Multivariate	Do any product citation percentages differ across the terms?	1
2	Univariate	Do any product citation percentages differ in term t?	T
3	Elementwise	Does the product p citation percentage differ from the median citation percentage in term t?	PT
4	Pairwise multivariate	Does this pair of products differ in citation percentages across the terms?	$\frac{1}{2}P(P-1)$
5	Pairwise univariate	Does this pair of products differ in citation percentages of term t?	$\frac{1}{2}P(P-1)T$

Table 1 Permutation tests proposed. The column 'Max. tests' gives the maximum number of tests for a data set with P products and T terms.

Tests in Table 1 are interrelated. The MAD statistic for each elementwise test (Test 3) is the absolute difference of the percentage of product p on term t to the median percentage of term t. The MAD statistic for each univariate test (Test 2) is the median of the P elementwise MAD statistics for term t in Test 3. The MAD statistic for the multivariate test (Test 1) is the median of the T univariate MAD statistics in Test 2. Additionally, the MAD statistic for each univariate paired product comparison (Test 5) is the absolute difference between the two products on term t. The MAD statistic for the multivariate paired comparison of these two products (Test 4) is the median of the T MAD statistics from these univariate paired comparisons of these two products in Test 5.

The multivariate test (Test 1) evaluates whether there are global product differences, considering all terms. A significant test result encourages further investigations. When making these additional investigations, multiple hypotheses are tested, introducing statistical considerations that will be discussed in the next section. For this reason, we will discuss the issue of multiple testing in Section 2.3, then further describe Tests 2–5.

2.3 Multiple testing and false discovery

In a null hypothesis test, a true null hypothesis is rejected with Type I error rate α set by the researcher. If multiple independent hypotheses are each tested at level α, the probability of making one or more Type I errors can be higher, or even substantially higher, than α. For this reason, many researchers adjust their protocol when conducting multiple tests. In this section, we discuss how multiple testing adjustments will be applied in the analysis of the CATA data.

Suppose we investigate each term to determine if citation percentages differ significantly from the median on each of the $T = 34$ terms (Test 2). If the products are, in fact, the same with respect to all terms, the probability of making at least one Type I error is $1 - (1 - 0.05)^{34} = 0.825$, far larger than the nominal rate 0.05. Had there been 100 terms, we would make at least one Type I error with probability 0.994—a near certainty.

Since controlling the Type I error rate across all tests reduces statistical power severely, researchers often elect to control the false discovery rate (FDR) instead. A rejected null hypothesis, called a "discovery", can be either true or false. The discovery is true (i.e., correct) if the rejected null hypothesis is, in fact, false. The discovery is false (i.e., incorrect) if the rejected null hypothesis is, in fact, true. When FDR=0.05, the proportion of false discoveries among all discoveries is controlled at level 0.05.

We are not conducting independent null hypothesis tests of independent predictor variables; rather, our tests are dependent hypothesis tests of correlated response variables. For this reason, it could be argued there is no reason to control the FDR. In this case, our rationale for controlling the FDR is simply to restrict communication of results to terms that merit the most attention.

Controlling the FDR for M tests using the Benjamini-Hochberg (BH) step-up procedure (Benjamini & Hochberg, 1995) involves sorting the p-values in ascending order alongside the arithmetic series of M BH values, $\{\frac{1}{M}0.05,\ \frac{2}{M}0.05, \ldots,\ \frac{M-1}{M}0.05,\ 0.05\} = (\{1, 2, \ldots, M\}/M) \cdot 0.05$, starting at $\frac{1}{M}0.05$ (corresponding to the conservative Bonferroni test), and increasing by increments of $\frac{1}{M}0.05$ to 0.05 (corresponding to the Type I error rate). The BH critical value is the largest p-value that is smaller than its corresponding BH value. All p-values less than or equal to this critical value are declared significant. These results can be visualised in a heatmap colour-coded to emphasise the significant p-values. This BH procedure can be applied to Tests 2–5 in Table 1. In each case, we use FDR=0.05, but each BH series differs according to the maximum number of hypotheses tested, given in column 'Max. Tests' of Table 1.

Apart from Tests 1–3 on the CATA table values themselves, it is also of interest to test their pairwise differences, univariately or multivariately. The multivariate product paired comparison Test 4 evaluates which of the $\frac{1}{2}P(P-1)$ product pairs differ significantly across the terms. The test statistic for a particular product pair is the median of the T absolute differences in citation percentages between the two products in the CATA table. Using the permutation procedure, we obtain the corresponding MAD values between the two products computed on the B random CATA tables. These B simulated MAD values, together with the test statistic, constitute the null distribution.

As before, the p-values are obtained from the null distribution and evaluated using the BH procedure, then visualised in a products-by-products ($P \times P$) heatmap.

Doing the pairwise test for each term t (Test 5), the null distribution is comprised of the test statistic, which is the observed signed difference, along with the similar differences in the B random CATA tables. The p-value is again the proportion of absolute outcomes in the null distribution as large or larger than the absolute test statistic. It is possible to test only product pairs determined to differ significantly based on the multivariate test (Test 4). Here, all product pairs on all terms are tested using the BH procedure, and the results are visualised for each term in a $P \times P$ heatmap.

2.4 Hierarchical cluster analysis

A researcher might want to cluster products into groups where products in the same group are conceptualised similarly and products in different groups are conceptualised differently. We perform a hierarchical clustering of the products using complete-linkage clustering, which tends to yield relatively tight, widely separated clusters since the distance between the furthest member of each group is considered whenever clusters are merged. For consistency with the "one citation, one vote" principle, we measure the dissimilarity of any pair of products using the median of the pairwise MAD values from the T terms, which is the test statistic from Test 4 in Table 1.

A researcher might also want to cluster terms into groups where terms in the same group are cited in a similar manner and terms in different groups are cited differently. To do so, we centre the CATA table by subtracting the column medians, then apply complete-linkage clusters based on the paired comparisons of the terms. The distance between any pair of terms is the median of the pairwise MAD values from the P products in this centred table.

2.5 L1-norm principal component analysis (L1-PCA)

Without loss of generality, suppose matrix $\mathbf{X}$ has median-centred columns. L1-PCA decomposes $\mathbf{X}$ into two matrices: a score matrix $\mathbf{T}$ and loading matrix $\mathbf{P}$. Dimension reduction is possible since retaining only K components minimises the sum of absolute residuals

$$\|\mathbf{X} - \mathbf{T}_K \mathbf{P}_K^{\mathsf{T}}\|_1 \tag{2}$$

where $^{\mathsf{T}}$ denotes matrix transposition and $\|\cdot\|_1$ denotes the L1 norm, that is, the sum of absolute elements. Several L1-PCA algorithms have been proposed for approximating $\mathbf{X}$ with $\mathbf{T}_K \mathbf{P}_K^{\mathsf{T}}$. We use the algorithm proposed by Ke & Kanade (2005).

Additionally, we show the uncertainty of the products in the L1-PCA visualization. To do so, we conduct bootstrapping at the assessor level. Each assessor has PT values in his or her set of responses (a slice of the array shown in Fig. 1.) Assessors are randomly chosen with replacement, until another three-way array of A slices, as shown in Fig. 1, is constructed, where some assessors are included more than once, some not at all. From this bootstrapped array another CATA table is formed, and its rows are projected onto the original solution already established using linear programming to minimize the sum of absolute deviances, subject to constraints (Ke & Kanade, 2005). The L1-PCA algorithm we used (Ke & Kanade, 2005) is robust to initialization values and reaches more precise solutions by allowing more iterations towards a smaller convergence tolerance. A solution could be considered converged if changing the initialization values, allowing more iterations, and specifying a smaller convergence tolerance does not provide meaningful improvement. After repeating this bootstrap procedure 1000 times, each time projecting the rows to the solution, the points are overlaid with 95% covering ellipses, which estimate the confidence regions for the products. If confidence ellipses on two product means do not overlap, this is an indication, but not conclusive proof, of a significant difference between the products. In the present L1 norm framework, we prefer to judge this pairwise product difference using the permutation test previously proposed.

The total dispersion is the L1 norm of $\mathbf{X}$. The L1-PCA with K components extracts the largest possible sum of absolute deviances from $\mathbf{X}$. The proportion

of the L1 norm extracted in these K components,

$$\text{prop}_K = \frac{\|\mathbf{T}_K \mathbf{P}_K^{\mathsf{T}}\|_1}{\|\mathbf{X}\|_1} \tag{3}$$

quantifies the success of this L1-PCA in approximating $\mathbf{X}$. The solutions in L1-PCA are not nested as in regular PCA, so percentages of explanation cannot be ascribed to individual dimensions, only to specific K-dimensional solutions. To obtain a scree plot similar to that usually produced in PCA, the L1-PCA is performed separately for solutions of dimension $K = 1, 2, \ldots, \min\{P, T\}$, and the gains in explanation are then computed, going from the solutions of dimension $K - 1$ to K (where, for $K = 0$, the explanation is 0). Different researchers will make different choices for how many components to retain. Since we were interested in product separation, we kept only the first K components that separated products, i.e., where one or more product confidence ellipse excluded the origin.

2.6 Software

Data analyses were conducted in R 4.2.2 (R Core Team, 2024). Functions `madperm()` and `mad.dist()` in the R package **cata** (Castura, 2025) were used to conduct permutation tests and to calculate MAD-based distances for cluster analyses. R code in Supplementary Material shows how to reproduce the permutation tests presented here. Running these permutation tests took approximately about four minutes on a laptop running 64-bit Windows 10 Pro 22H2 with an Intel® Core™ i5-1235U, 1300 Mhz, 10 Core(s), 12 Logical Processor(s) and 16 GB RAM. Functions `l1pca()` and `l1projection()` in the R package **pcaL1** (Jot et al., 2023) were used to conduct L1-PCA and to project bootstrapped sample points onto the L1-PCA biplots. Function `CIplot_biv()` in the R package **easyCODA** (Greenacre, 2018) used these points to calculate and draw confidence ellipses. Heatmaps and dendrograms were drawn using the function `pheatmap()` in the R package **pheatmap** (Kolde, 2019).

3 Results

3.1 Tests and heatmaps

In the blackcurrant squash data set, counts and percentages are identical since $A = 100$ consumers evaluated the products.

Test 1 conducts a global multivariate test to determine whether product differences exist on any term. The test statistic is the median of the MAD values from the 34 terms, which was 5% (at the extreme right). Fig. 2 shows the null distribution for this test statistic.

A high proportion (0.96) of the null distribution has the median value (2.0%). There are relatively few values at 1.0%, 1.5%, 2.5% and 3.0%. The value of 5%, which is the observed test statistic, is alone at the extreme right.

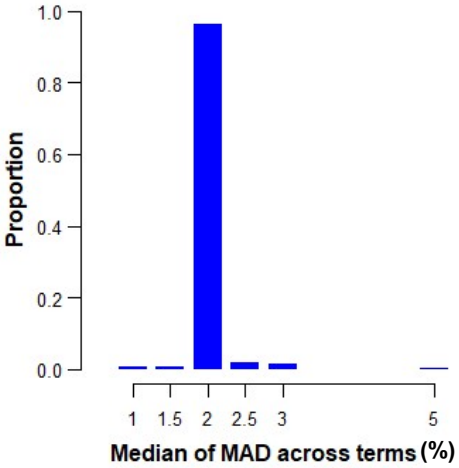

Fig. 2 Null distribution values for the overall test statistic for product differences (x axis) vs proportion of occurrence (y axis). The median absolute deviance (MAD) is calculated for each term; the observed test statistic is the median of the MAD values across the 34 terms, which was 5% (at the extreme right).

Since none of the randomised values equal or exceed the observed value, the global multivariate test indicates significant global differences ($p = 0.0001$) and justifies further analyses.

Next, we investigate which of the T terms discriminate the products. For the term *Happy*, citation percentages range from 15% to 55%. The MAD value for *Happy* is 10%, which is the median "distance" of the 11 citation percentages from the median of 31%, on the percentage scale. Fig. 3 shows graphically both the median (as a wide vertical bar) and MAD (as a narrow bar on top) for each term. This plot shows the median and MAD values are correlated. We emphasise the MAD bars of terms that significantly discriminate the products, as determined by the univariate test results (Test 2), which are described next.

To evaluate univariate test results, we conducted Test 2 permutation tests (Section 2.2), from which we obtained 34 p-values. These results are presented in Table 2, alongside the BH series used to control the FDR at 0.05 (Section 2.3). The 15 terms achieving the smallest possible p-value ($p = 0.0001$) are presented in no particular order. The critical value (0.0125) was obtained from term *Sickly*, which had the largest p-value that was smaller than its BH step-up value. This critical value controls the FDR at level 0.05. The 22 terms on which the product differed significantly had MAD values of 5% and larger, whereas the 12 other terms had MAD values of 4% and smaller (Table 2).

Fig. 4 (top row) shows a heatmap visualisation of the univariate (Test 2) p-values in Table 2. Terms that significantly discriminate the products are shown in green, with deeper colours indicating a more significant test result. Terms above the BH critical value, which are non-significant across the products, are shown in light grey.

Cluster analysis of the terms, as described in Section 2.4, organises the terms into a two-cluster solution, shown at the top of Fig. 4. Positively valenced terms (cluster 1) are contrasted against negatively valenced terms (cluster 2). Term cluster 1 consists of 15 positively valenced terms, 11 of

which discriminated the products: *Comforted*, *Guilty pleasure*, *Trust*, *Approval*, *Pleasant surprise*, *Interested*, *At ease*, *Pleased*, *Satisfaction*, *Happy* and *Good*. The lexicon developers did not classify the term *Guilty pleasure* as either positively or negatively valenced (Ng et al., 2013a), but these term clusters suggest that consumers conceptualised *Guilty pleasure* as positively valenced. The remaining four terms in this cluster (*Desire*, *Warm*, *Attentive*, *Reminiscence*) did not discriminate the products. Term cluster 2 is comprised of 19 negatively valenced terms, of which 11 discriminated the products significantly (*Disappointment*, *Displeasure*, *Annoyed*, *Shocked*, *Discontent*, *Uncomfortable*, *Disgust*, *Unhappy*, *Unpleasant surprise*, *Sickly*, *Regret*) and eight did not (*Bored*, *Curious*, *Angry*, *Worried*, Cautious, *Confused*, *Sceptical*, *Resentment*).

Fig. 4 (lower part) also shows a $P \times T$ heatmap visualization of the elementwise (Test 3) p-values obtained by analysing, for each term, which of the P products differs from the median value. We decided a priori to conduct all $P \times T = 11 \times 34 = 374$ elementwise tests (Section 2.2) with the FDR controlled at level 0.05 (Section 2.3). In 374 elementwise tests, 89 p-values were less than or equal to the BH critical value of 0.0118.

Significant elementwise p-values were colour-shaded, where the cell colour was determined by the sign of the difference between the observed and expected count: a product significantly below the term median was coloured red; a product significantly above the term median was coloured blue. Products that were not significantly different from the median are coloured light grey.

At least one product differed significantly from the median for all but four terms: *Attentive*, *Worried*, *Cautious*, and *Confused*. These four terms are among the 12 non-significant terms (coloured light grey in the top row of Fig. 4). Inspection of the columns for the other eight non-significant terms show at least one significant elementwise test. This result shows that individual

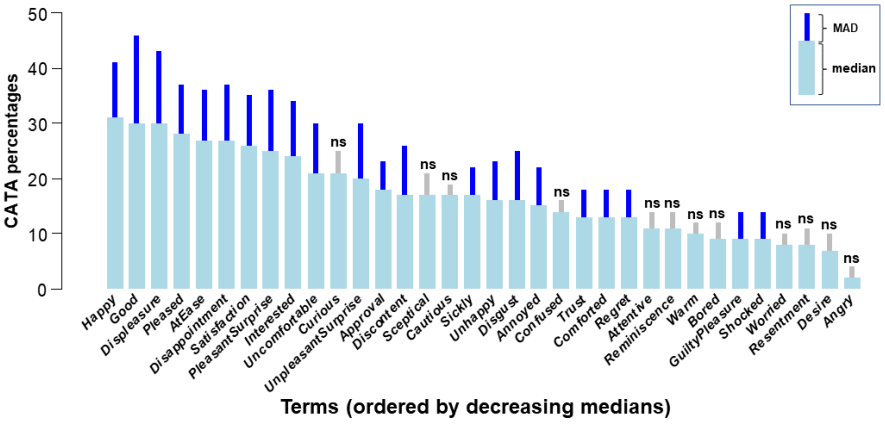

Fig. 3 For each term, the median citation percentage of the 11 products is shown (wide vertical bars) with the median absolute deviation (MAD) of each product represented by a thin vertical bar on top. The MAD line is shown in dark blue for the 22 significant terms and in light grey (denoted 'ns') for the 12 non-significant terms.

Term	MAD (%)	Permutation p-value	BH step-up value
Happy	10	0.0001	0.0015
Unhappy	7	0.0001	0.0029
Uncomfortable	9	0.0001	0.0044
At ease	9	0.0001	0.0059
Pleasant surprise	11	0.0001	0.0074
Unpleasant surprise	10	0.0001	0.0088
Disappointment	10	0.0001	0.0103
Satisfaction	9	0.0001	0.0118
Discontent	9	0.0001	0.0132
Disgust	9	0.0001	0.0147
Interested	10	0.0001	0.0162
Good	16	0.0001	0.0176
Displeasure	13	0.0001	0.0191
Annoyed	7	0.0001	0.0206
Pleased	9	0.0001	0.0221
Shocked	5	0.0002	0.0235
Guilty pleasure	5	0.0003	0.0250
Regret	5	0.0011	0.0265
Trust	5	0.0013	0.0279
Comforted	5	0.0060	0.0294
Approval	5	0.0094	0.0309
Sickly	5	0.0125	0.0324
Resentment	3	0.0704	0.0338
Sceptical	4	0.0719	0.0353
Desire	3	0.0793	0.0368
Curious	4	0.1059	0.0382
Attentive	3	0.1121	0.0397
Angry	2	0.1573	0.0412
Reminiscence	3	0.1891	0.0426
Bored	3	0.1982	0.0441
Worried	2	0.4821	0.0456
Warm	2	0.6587	0.0471
Confused	2	0.7350	0.0485
Cautious	2	0.7961	0.0500

Table 2 Results for the univariate test (Test 2) controlling the false discovery rate at 0.05 show sorted p-values alongside the Benjamini-Hochberg (BH) series. The 22 terms having a p-value below the critical value (0.0125) were considered significant. The non-significant terms below the double line did not discriminate the products.

products can differ significantly from the median of a term, even if the product citation percentages do not differ significantly overall on that term.

Fig. 4 shows the dendrogram from the complete-linkage cluster analysis of products to the left of the elementwise heatmap. The rows (products) are sorted to accommodate this dendrogram which shows a crisp two-cluster solution. Product cluster 1 is composed of five standard products with added sugar (P5sw, P7sw, P11sw, P1sw, P9sw). Product cluster 2 is composed of six more

varied products: four products without added sugar (P10eo, P8so, P2so and P6so) and two niche products with added sugar (P3nw and P4nw).

The elementwise heatmap reveals approximate characterisations of products in the two product clusters. Members of product cluster 1 elicited significantly higher citation percentages than the median product in positively valenced terms in term cluster 1 and significantly lower citation percentages than the median product in negatively valenced terms in term cluster 2. By contrast, members of product cluster 2 elicited significantly higher citation percentages than the median product in negatively valenced terms in term cluster 2 and significantly lower citation percentages than the median product in positively valenced terms in term cluster 1.

We draw attention to selected product results (rows) in the elementwise heatmap related to product cluster 2. Like other members of product cluster 2, the economical product P10eo had higher-than-median citation percentages for negatively valenced terms in term cluster 2 and lower-than-median citation percentages for positively valenced terms in term cluster 1. Additionally, P10eo elicited significantly higher-than-median citation percentages for *Bored*, *Sceptical*, and *Resentment* and a significantly lower-than-median citation percentage for *Curious*. Unlike the other 10 products, P10eo had relatively high intensities of catty, confectionery blackcurrant and artificial sweetness (Ng et al., 2012), which might explain why this product elicited these negatively valenced terms.

Three standard products without added sugar (P2so, P6so and P8so) were also allocated to cluster 2. Besides of the general trend of product cluster 2, P6so and P8so elicited higher-than-median citation percentages for *Sickly* in

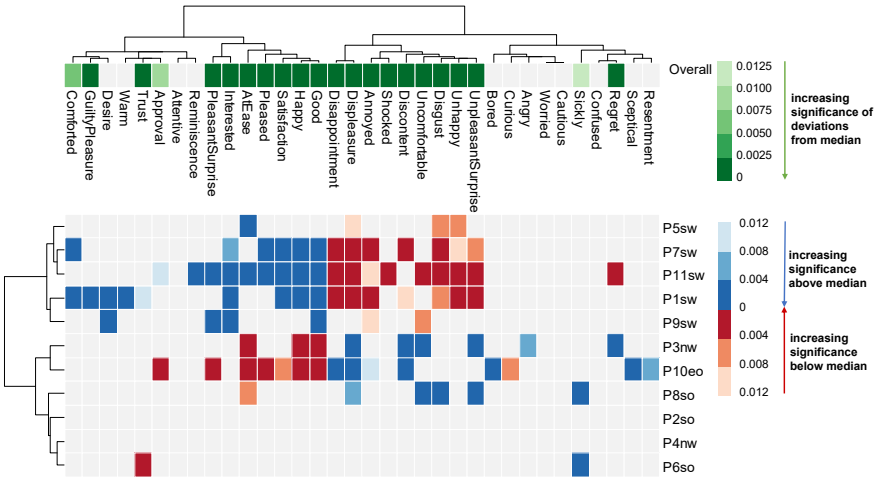

Fig. 4 Heatmap of p-values from univariate tests (Test 2 results shown in top row interpreted with uppermost legend) and elementwise tests (Test 3 results shown in lower matrix interpreted with the two lower legends). Complete-linkage cluster analysis results are shown for clustering of terms (top) and clustering of products (left). In Test 2 and in Test 3, the false discovery rate was 0.05. [Products are uniquely numbered (P1 to P11). The first letter indicates the market-value category: economical (e), niche (n), standard (s). The second letter indicates with (w) or without (o) added sugar.]

	P5sw	P7sw	P11sw	P1sw	P9sw	P3nw	P10eo	P8so	P2so	P4nw
P7sw	3.0									
P11sw	3.0	2.0								
P1sw	5.0	2.5	4.0							
P9sw	4.0	4.0	4.0	4.0						
P3nw	**12.5**	**14.5**	**14.5**	**15.5**	**12.0**					
P10eo	**11.0**	**14.0**	**14.5**	**16.5**	**13.0**	4.0				
P8so	**10.0**	**12.5**	**12.5**	**12.0**	**8.5**	5.0	**5.5**			
P2so	**8.0**	**10.0**	**10.5**	**10.0**	**7.0**	**6.0**	**8.0**	4.5		
P4nw	**8.5**	**10.5**	**12.5**	**11.0**	**9.0**	4.0	**5.5**	4.5	4.0	
P6so	**10.0**	**13.0**	**14.0**	**13.5**	**10.5**	4.0	4.5	4.0	4.0	2.0

Table 3 Median absolute difference in citation percentages across all terms from each of the 55 multivariate product paired comparisons. The false discovery rate was 0.05. The 34 significantly different pairs are shown in boldface.

term cluster 2. In a previous study, P6so and P8so were observed to have the highest intensities of artificial sweetness and natural processed blackcurrant (Ng et al., 2012), suggesting the relatively high *Sickly* citation rates were due to cloying oversweetness and processed flavour. These same sensory properties could be the reason for a lower-than-median citation percentage of *Trust* for P6so.

Compared to other products, P3nw was more intense in green/leafy and earthy flavours, more bitter, and more astringent (Ng et al., 2012). If consumers in the blind sensory evaluation did not expect these sensory properties, it could explain why this niche product elicited negatively valenced terms more often than other products. Individual tests (Tests 3 and 5) can be significant even when an overall test (Test 2) is not. Some researchers might enforce greater consistency by conducting Test 3 only for terms that were significant based on Test 2, but conducting all tests helps in the interpretation of the consumers' response elicited by P3nw, adding some cues for the identification of specific differences. Although products did not differ significantly in citation percentages for *Angry* based on the univariate test results (Test 2; Table 2, Fig. 3), P3nw elicited higher-than-median citation percentages for *Angry* based on elementwise tests (Test 3; Fig. 4) and univariate pairwise tests (Test 5; results not shown). Again, these results might be explained by the sensory profile of P3nw mentioned above.

Test 4 investigates which product pairs differ considering all T terms simultaneously. For each of the 55 paired comparisons, the test statistic is the median absolute paired difference across the T terms. Permutation testing (Section 2.2) yields a p-value for each product pair. The critical value that controlled FDR at 0.05 (Section 2.3) was 0.0281. Based on this critical value, 34 paired comparisons were significantly different. Table 3 shows the multivariate paired comparison test statistics, emphasising these significant differences. The rows and columns of this table are ordered as in Fig. 4.

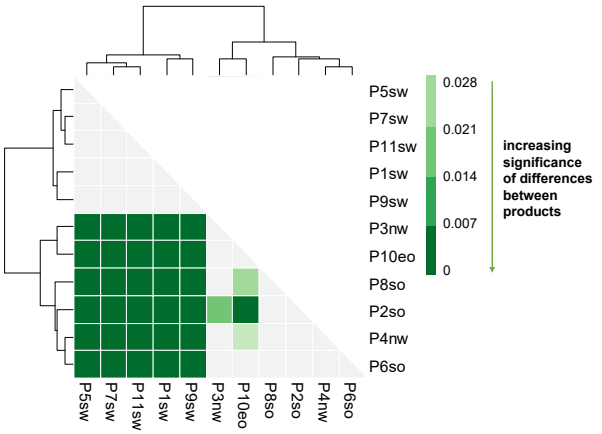

Fig. 5 Heatmap of p-values for the 55 multivariate product paired comparisons in Table 3 (Test 4). With the false discovery rate controlled at 0.05, 34 of the 55 pairs were significantly different (shown in green). The diagonal and cells for non-significant tests results are shown in light grey.

Fig. 5 shows the dendrogram from the complete-linkage cluster analysis of products on horizontal and vertical axes of this heatmap, the same dendrogram that appeared in Fig. 4 on the left side. The heatmap combined with the dendrogram facilitates the interpretation. It shows standard products with added sugar (product cluster 1) differ significantly from products without added sugar and niche products (product cluster 2). There was no evidence of significant paired differences within product cluster 1, but various paired differences were found within product cluster 2. Among products without added sugar, the economical product P10eo had a different emotional profile than the standard products P8so and P2so. The two niche products, P3nw and P4nw, elicited a different emotional response than products P2so and P10eo, respectively, from the same product cluster 2. Although we can venture reasonable guesses why each of these product pairs differed by referring to the element-wise test results (Fig. 4), a better understanding is obtained from univariate product paired comparisons, which will be evaluated next.

Each univariate paired comparison test (Test 5) investigates whether a term discriminates the product pair. Since this study is exploratory but also illustrative, we decided a priori to investigate all product pairs across all terms. Each test statistic is the signed difference in citation percentages between the two products. Using the permutation test procedure (Section 2.2), we obtained $TP(P-1)/2 = 1870$ p-values, which were placed alongside the BH arithmetic series to obtain the BH critical value 0.0130. Using this critical value to control the FDR at level 0.05, we obtained 487 significant test results based on Test 5.

To illustrate the univariate paired comparison test, we return to *Happy* and *Sickly*, two terms introduced in Section 2.2. Both these terms significantly discriminated the products (Table 2). As expected, *Happy* ($p = 0.0001$) discriminated the products better than *Sickly* ($p = 0.0125$), and discriminated more product pairs based on the Test 5. We found 24 product pairs differ significantly in *Happy* citation percentages and 13 product pairs differ significantly in *Sickly* citation percentages.

The p-values for *Happy* and *Sickly* are visualised in heatmaps in Fig. 6. In these heatmaps, the direction of the difference in citation percentages is obtained by subtracting the column product from the row product (i.e. rows minus columns), where positive and negative values are colour-coded blue and red, respectively, for product pairs that differ significantly. The differences above and below each main diagonal have the opposite sign but the same absolute magnitude, hence the blue–red symmetry across the diagonal. Products are ordered and shown together with the dendrograms as in Fig. 5, which show the results from the complete-linkage cluster analysis of products.

Products did not differ significantly in *Happy* citation percentages; however, every product in product cluster 1 had a higher observed *Happy* citation percentage than the products in product cluster 2. Within product cluster 1, the *Happy* citation percentage was highest for P7sw and lowest for P5sw. Due to these differences, more (6) significant paired differences were detected between P7sw and the products in product cluster 2 than were detected (3) between P5sw and products in product cluster 2. P3nw and P10eo had the lowest Happy citation percentages (both 15%); both these products differed significantly from all five products in product cluster 1.

Paired comparison results for *Sickly* are driven entirely by two standard products without sugar from product cluster 2. P6so and P8so had *Sickly* citation percentages of 35% and 30%, respectively, whereas citation percentages for other products lie between 12% and 25%. P6so had a significantly higher citation percentage than five other products, three in product cluster 1 and two in product cluster 2. P8so had a significantly higher citation percentage than eight other products, including all members of product cluster 1 and three products in product cluster 2.

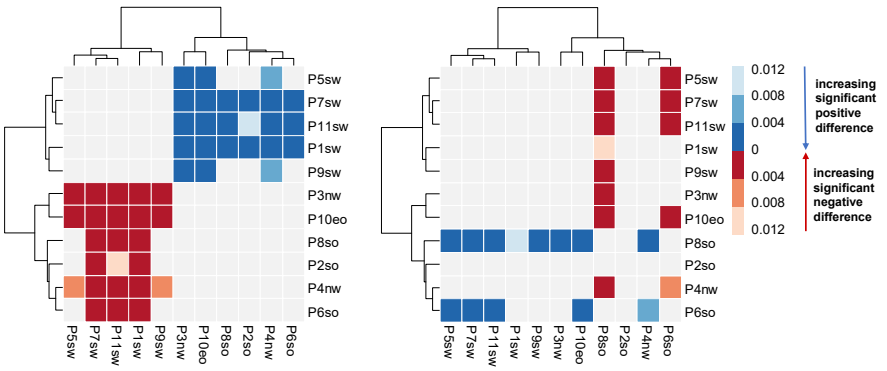

Fig. 6 Results for univariate paired comparisons of products on *Happy* and *Sickly* with the false discovery rate controlled at 0.05 across all such tests. The diagonal and cells for non-significant tests results are shown in light grey. Blue and red cells indicate significant differences; see legend for details.

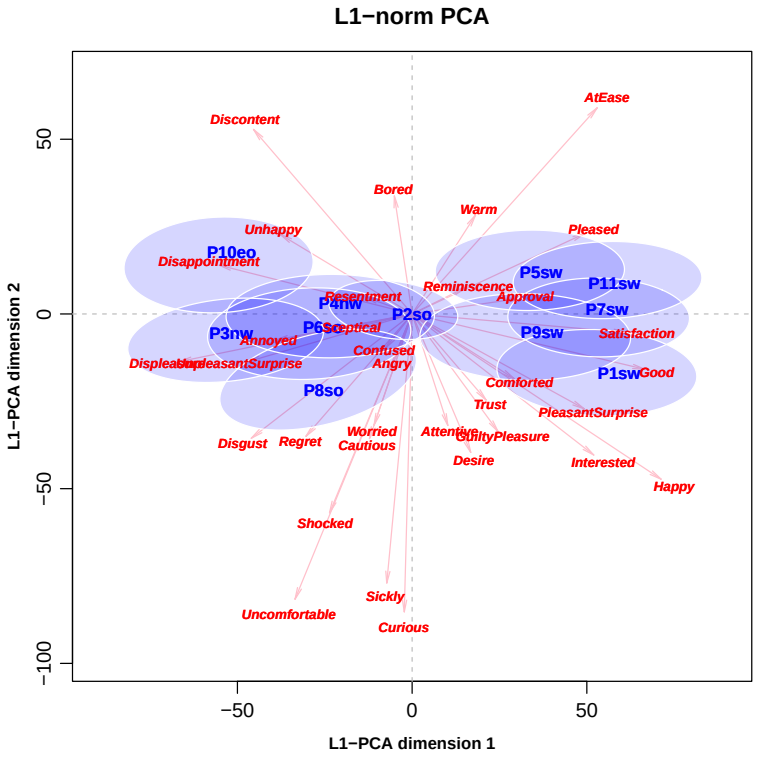

Fig. 7 L1-PCA biplot of the CATA table, with 95% confidence ellipses for the product medians.

3.2 Dimension reduction using L1-PCA

The L1-PCA biplot (Fig. 7) gives a compact summary of the data structure and shows several results that have been observed in the permutation tests. The scale of these two components is percentage units. The biplot shows scores (products, as points) visualised together with the loading vectors, which are rescaled for legibility. The origin coincides with product P2so, which lies at the median.

Based on Eq. (3), the proportion of the CATA table explained in two components is 0.657. The proportion of the CATA table explained by one L1-PCA component is 0.586, so the gain in the two-dimensional solution is 0.071. Fig. 8A shows the equivalent of a scree plot for the L1-PCA, that is, the gains in explanation of the CATA table for increasing dimensions. To show the quality of the L1-PCA two-dimensional solution, the concordance between the CATA table percentages $\mathbf{E}$ and the estimated values obtained from the two-dimensional L1-PCA solution is shown in the scatterplot of Fig. 8B. Notice in Eq. (2) that $\mathbf{X}$ is median-centered, hence the medians have to be added to $\mathbf{T}_K\mathbf{P}_K^{\mathsf{T}}$ to obtain the PCA estimates for the solution in $K = 2$ dimensions.

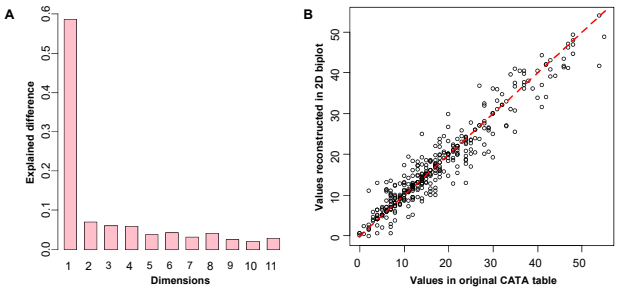

Fig. 8 A. Scree plot of the L1-PCA, showing the gains in explaining the CATA table for solutions of increasing dimension (K). B. Scatterplot of CATA table percentages (horizontal axis) and the values reconstructed from the two-dimensional L1-PCA biplot of Fig. 7.

The five standard products with added sugar (i.e., with codes ending -sw) are aligned with attributes such as *At ease*, *Good*, *Happy*, *Interested*, *Pleasant surprise*, and *Satisfaction*. The attributes *Disappointment*, *Displeased*, *Unpleasant surprise* and *Unhappy* are aligned with products P10eo and P3nw, attributes that also tend to characterise the products P4nw and P6so, but since their 95% confidence ellipses overlap the origin, it cannot be concluded that their sensory profiles differ from the median sensory profile, at least based on the results in this plane.

In most cases, the significant multivariate paired product differences (Fig. 5) have non-overlapping ellipses in the L1-PCA plane shown (Fig. 7). For example, both Test 4 and the L1-PCA indicate particular product pairs differ: P10eo and P2so, P3nw and P2so, and P10eo and P8so. In Fig. 7, ellipses for P2so and P8so do not overlap with the ellipse for P10eo (which had catty notes), even though all of these products were in product cluster 2.

The five standard products with added sugar are well separated from the other six products, with the exception of slightly overlapping ellipses for P9sw and P2so, and P9sw and P4nw, which are better separated in higher dimensions. Table 3, however, confirms the strict separation of these two product clusters statistically.

In some cases, interpretations from the two analyses differ. For example, the multivariate pairwise test (Fig. 5) did not detect a significant difference between P1sw and either P5sw or P11sw, even though their 95% confidence ellipses are separated in the first L1-PCA plane (Fig. 7). The discrepancy probably occurs because the multivariate pairwise test considers all attributes simultaneously, whereas the L1-PCA results in this plane considers only the linear combinations of attributes in the two components shown.

4 Discussion

4.1 Constraining the number of tests

In many CATA studies, all assessors evaluate each product once using the same CATA terms. A CATA frequency table from such studies has the unique property that each cell is the result of a separate counting of the same sample

of respondents. This property suggests a wide variety of hypothesis tests. We have listed five such tests from which the researcher can choose.

Researchers can constrain which tests are conducted so only relevant aspects are investigated. For example, three researchers might apply Test 5 in different ways depending on the study objectives. One researcher might apply Test 5 only to terms that differ significantly based on Test 2 results. A second researcher might apply Test 5 only to product pairs that differ significantly based on Test 4 results. A third researcher might apply Test 5 only to the significant terms based on Test 2 results for only the significant product pairs based on Test 4 results.

4.2 Comparison with conventional tests

Among the proposed permutation tests in Table 1, the univariate test (Test 2) and the univariate pairwise test (Test 5) are respectively analogous to the Cochran's Q test (Cochran, 1950; Meyners et al., 2013) and McNemar's test (McNemar, 1949; Meyners et al., 2013). Results from these two tests are provided in Supplementary Material to facilitate comparisons with permutation testing results in Section 3.

Under a true null hypothesis that P products elicit the same citations proportion, Cochran's Q test statistic follows the asymptotic chi-squared distribution with $P - 1$ degrees of freedom. Cochran's Q test detected product differences in 30 terms, more than were detected using the permutation testing approach (22). Although controlling FDR will tend to reduce the number of significant test results, in this data set, the number of significant Cochran's Q test results was the same with and without FDR control, and the number of significant results from Test 2 was also the same with and without FDR control (Supplementary Material, Table S1).

We also conducted McNemar's test to determine whether two products elicit the same citation proportion for each of the 55 product pairs on each of the 34 attributes. In each of these 1870 tests, differentiating citations (i.e. in which the term is endorsed for only one of the two products) will be binomially distributed. With FDR controlled at level 0.05, McNemar's test detected 434 significant paired differences, fewer than were reported based on the permutation testing approach (487). When FDR was uncontrolled, McNemar's test detected 639 significant paired differences, fewer than were detected using the permutation testing approach (660). These results show that the permutation testing approach is at least sometimes more powerful than the exact version of McNemar's test. An advantage of McNemar's test, however, is that it returns consistent output rapidly from an exact distribution that does not require a permutation testing procedure (Supplementary Material, Table S2).

Previously, Meyners et al. (2013) also proposed a global test similar to Test 1 based on the randomisation testing procedure, which was approximated by the sum of Cochran's Q test statistics. This test is often highly significant in typical consumers tests, as in this data set (Fig. 2).

4.3 Future studies and limitations

The proposed permutation tests might be applied to other types of data also. For example, quantitative sensory data collected on rating scales could be converted to percentages, then analysed using method proposed in Table 1, where the L1-PCA can provide dimension reduction and summary. Other approaches, such as taxicab correspondence analysis (Choulakian, 2006), might also be investigated further.

The method for constructing confidence ellipses in the L1-PCA could be investigated further to determine whether these ellipses have approximately frequentist properties and whether other confidence-ellipse constructing procedures are superior (Castura et al., 2022, 2023).

Since multivariate pairwise tests (Test 4) are based on median MAD values, it seems possible that real differences could be missed if the products are strongly separated on a few terms, but a large number of non-differentiating or irrelevant terms are included in the study. Further study would be needed to evaluate the power of the proposed permutation tests compared to conventional tests for analysing CATA data.

A seed should be set before conducting permutation tests to ensure reproducibility. Software referenced in this manuscript can facilitate such matters. Permutation testing requires enough permutations to achieve stable results. Our results based on 9 999 permutations were usually close (within ± 0.01) to results based on 99 999 permutations. Researchers are advised to determine whether results are stable by conducting additional permutations or confirming that the conclusions are the same if a different seed is used.

Neither term clusters nor product clusters seemed to be influenced by marginal citation percentages of the terms and products, respectively; however, a potential drawback with the proposed method of clustering terms is that ties might increase when there are relatively few products. Usually, there are more terms than products, but if the number of terms is small, ties might arise when clustering products also.

5 Conclusion

In this study, we tested specific hypotheses given in Table 1 using permutation tests that make no particular assumptions about the distribution of the data under the null distribution. Due to the large number of tests, we opted to control the false discovery rate to restrict communication systematically to the significant test results in which we were most confident. This approach provided a data-driven way to determine which results to report that we found superior to simply choosing a smaller Type I error rate.

Our new framework for analysing CATA data is simple and robust. This framework emerges from the principle that every single citation recorded in a CATA product-by-term table counts the same as any other citation, no matter the assessor, product, or term. This “one citation, one vote” principle implies a different set of statistical measures based on the L1 norm, contrary to other

methods based on the L2 norm. When using the L1 norm, the measure of dispersion is based on the absolute differences in the CATA values, which are reported in percentage units, whereas methods based on the L2 norm are based on squared differences, often with some type of normalization. In the present approach, data are analysed "as is", meaning there are no potentially controversial data transformations. Thus, even though there are differences in the frequencies of checking between terms and between assessors, we avoid any attempt to normalize the data, both term-wise and assessor-wise, since such normalizations would imply that citations get varying numbers of "votes", depending on the terms and/or the assessor.

Thanks to the absence of any transformation on the CATA table, the proposed analyses are often easier to understand than analyses based on the L2 norm. Our approach using the L1 norm introduces a considerable simplification into the analysis of CATA tables for the food researcher.

The methods described in this paper are being introduced into the R package **cata** (Castura, 2025), which will also include this paper's data on blackcurrant squashes.

Author contributions

CC: Data curation, Formal analysis, Investigation, Resources, Writing – review & editing.

JCC: Conceptualization, Formal analysis, Software, Validation, Writing – original draft, Writing – review & editing.

MJG: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

Supplementary material

Supplementary material is provided in the form of two files.

1. A file with five sections: (S1) R code for replicating results, (S2) Cochran's Q test results, (S3) McNemar's test results, (S4) principal component analysis and correspondence analysis, and (S5) a description of the 3D video animation of the L1-PCA results in Suppl. Video V1.
2. Suppl. Video V1. Score plot for L1-PCA of the CATA table with three components showing 95% confidence ellipoids for the product coordinates. The video shows rotation around the vertical second axis, so the dispersion of the product ellipsoids on the third dimension can be seen.

References

Ares G, Jaeger SR (2023). Check-all-that-apply (CATA) questions with consumers in practice: Experimental considerations and impact on outcome. In: Delarue J, Lawlor JB (eds) Rapid Sensory Profiling Techniques, 2nd edn. Elsevier, p 257–280

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benzécri JP (1973). L'Analyse des Données. Tôme II: L'Analyse des Correspondances. Dunod, Paris

Castura JC (2025). cata: Analysis of check-all-that-apply (CATA) data. URL https://CRAN.R-project.org/package=cata, R package version 0.1.0.27

Castura JC, Rutledge DN, Ross CF, Næs T (2022). Discriminability and uncertainty in principal component analysis (PCA) of temporal check-all-that-apply (TCATA) data. Food Quality and Preference 96:104,370. https://doi.org/10.1016/j.foodqual.2021.104370

Castura JC, Varela P, Næs T (2023). Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis. Food Quality and Preference 107:104,843. https://doi.org/10.1016/j.foodqual.2023.104843

Choulakian V (2006). Taxicab correspondence analysis. Psychometrika 71:333–345. https://doi.org/10.1007/s11336-004-1231-4

Cochran WG (1950). The comparison of percentages in matched samples. Biometrika 37(3/4):256–266. https://doi.org/10.2307/2332378

Dooley L, Lee Y, Meullenet JF (2010). The application of check-all-that-apply (cata) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. Food Quality and Preference 21(4):394–401. https://doi.org/10.1016/j.foodqual.2009.10.002

Good P (2004). Permutation, Parametric and Bootstrap Tests of Hypotheses. Third Edition. Springer

Greenacre M (2016). Correspondence Analysis in Practice, 3rd edn. Chapman & Hall / CRC Press, Boca Raton, Florida

Greenacre M (2018). Compositional Data Analysis in Practice. Chapman & Hall / CRC Press, Boca Raton, Florida

Jot S, Brooks JP, Visentin A, Park YW (2023). pcaL1: L1-Norm PCA Methods. R package version 1.5.7

Ke Q, Kanade T (2005). Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition

(CVPR'05), pp 739–746 vol. 1, https://doi.org/10.1109/CVPR.2005.309

Kolde R (2019). pheatmap: Pretty Heatmaps. URL https://CRAN.R-project.org/package=pheatmap, R package version 1.0.12

Mahieu B, Schlich P, Visalli M, Cardot H (2021). A multiple-response chi-square framework for the analysis of free-comment and check-all-that-apply data. Food Quality and Preference 93:104,256. https://doi.org/10.1016/j.foodqual.2021.104256

McNemar Q (1949). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12:153–157. https://doi.org/10.1007/BF02295996

Meyners M, Castura JC (2014). Check-all-that-apply questions. In: Ares G, Varela P (eds) Novel Techniques in Sensory Characterization and Consumer Profiling, vol 1. CRC Press, p 271–306

Meyners M, Castura JC, Carr BT (2013). Existing and new approaches for the analysis of cata data. Food Quality and Preference 30:309–319. https://doi.org/10.1016/j.foodqual.2013.06.010

Ng M, Lawlor JB, Chandra S, Chaya C, Hewson L, Hort J (2012). Using quantitative descriptive analysis and temporal dominance of sensations analysis as complementary methods for profiling commercial blackcurrant squashes. Food Quality and Preference 25:121–134. https://doi.org/10.1016/j.foodqual.2012.02.004

Ng M, Chaya C, Hort J (2013a). Beyond liking: Comparing the measurement of emotional response using EsSense Profile and consumer defined check-all-that-apply methodologies. Food Quality and Preference 28:193–205. https://doi.org/10.1016/j.foodqual.2012.08.012

Ng M, Chaya C, Hort J (2013b). The influence of sensory and packaging cues on both liking and emotional, abstract and functional conceptualisations. Food Quality and Preference 29(2):146–156. https://doi.org/10.1016/j.foodqual.2013.03.006

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/

Schwertman NC, Gilks AJ, Cameron J (1990). A simple noncalculus proof that the median minimizes the sum of the absolute deviations. The American Statistician 44(1):38–39. https://doi.org/10.1080/00031305.1990.10475690

Thomson DM, Crocker C, Marketo CG (2010). Linking sensory characteristics to emotions: An example using dark chocolate. Food Quality and Preference 21(8):1117–1125. https://doi.org/10.1016/j.foodqual.2010.04.011

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. https://doi.org/10.1111/j.1365-2621.2012.03022.x