



HAL
open science

A multiple-response chi-square framework for the analysis of Free-Comment and Check-All-That-Apply data

Benjamin Mahieu, Pascal Schlich, Michel Visalli, Hervé Cardot

► **To cite this version:**

Benjamin Mahieu, Pascal Schlich, Michel Visalli, Hervé Cardot. A multiple-response chi-square framework for the analysis of Free-Comment and Check-All-That-Apply data. *Food Quality and Preference*, 2021, 93, pp.104256. 10.1016/j.foodqual.2021.104256 . hal-03355886

HAL Id: hal-03355886

<https://hal.inrae.fr/hal-03355886v1>

Submitted on 24 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Title

2 A multiple-response chi-square framework for the analysis of Free-Comment and
3 Check-All-That-Apply data

4 Authors

5 Benjamin Mahieu^a, Pascal Schlich^a, Michel Visalli^a, Hervé Cardot^b

6 ^aCentre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE,
7 Université Bourgogne Franche-Comté, F-21000 Dijon, France.

8 ^bInstitut de Mathématiques de Bourgogne, CNRS, Univ. Bourgogne Franche-Comté,
9 Dijon, France.

10 Corresponding author:

11 Benjamin Mahieu; Pascal Schlich

12 benjamin.mahieu@inrae.fr; pascal.schlich@inrae.fr

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

15 Highlights

- 16 - The usual chi-square framework is not well suited to analyze FC and CATA
17 data
- 18 - A modified multiple-response chi-square framework is introduced
- 19 - This new framework takes into account the specificities of FC and CATA data
- 20 - R package (MultiResponseR) available upon request to the authors and on
21 GitHub

22 Abstract

23 Free-Comment (FC) and Check-All-That-Apply (CATA) provide a contingency table
24 containing citation counts of descriptors by products. The analyses performed on this
25 table are most often related to the chi-square statistic. However, such practices are

26 not well suited because they consider experimental units as being the citations (one
27 descriptor for one product by one subject) while the evaluations (vector of citations
28 for one product by one subject) should be considered instead. This results in
29 incorrect expected frequencies under the null hypothesis of independence between
30 products and descriptors and thus in an incorrect chi-square statistic. Thus, analyses
31 related to this incorrect chi-square statistic, which include Correspondence Analysis,
32 can lead to wrong interpretations. This paper presents a modified chi-square square
33 framework dedicated to the analysis of multiple-response data in which experimental
34 units are the evaluations and which is, therefore, better suited to FC and CATA data.
35 This new framework includes a multiple-response dimensionality test of dependence,
36 a multiple-response Correspondence Analysis, and a multiple-response test per cell
37 to investigate which descriptors are significantly associated with which product. The
38 benefits of the multiple-response chi-square framework over the usual chi-square
39 framework are exhibited on real CATA data. An R package called “MultiResponseR”
40 is available upon request to the authors [and on GitHub](#) to perform the multiple-
41 response chi-square analyses.

42 **Keywords**

- 43 - Chi-square statistic
- 44 - Multiple-response Correspondence Analysis (MR-CA)
- 45 - Multiple-response dimensionality test of dependence
- 46 - Multiple-response tests per cell
- 47 - Analysis of multiple-response data

48 **1. Introduction**

49 Free-Comment (FC) ([ten Kleij & Musters, 2003](#)) and Check-All-That-Apply (CATA)
50 ([Adams, Williams, Lancaster, & Foley, 2007](#)) are word citation occurrence-based
51 methods that aim at collecting product descriptions from consumers using either their
52 own words or a mutual predefined list of descriptors. These descriptions are collected
53 without any quantification or product comparison. At the panel level, the collected
54 data constitute count data that are usually stored in a contingency table that contains

55 the number of times each descriptor (in columns) was cited for each product (in
56 rows).

57 The analysis of these data starts by testing whether overall differences exist between
58 the products. Two approaches can be distinguished to do so. The first one consists of
59 performing a chi-square test while the second one is based on a combination of
60 Cochran's Q statistics (Meyners, Castura, & Carr, 2013). Pursuing the analyses
61 further is only recommended if the existence of overall differences between products
62 is established. In this case, these differences can be visualized using
63 Correspondence Analysis (CA). CA enables to represent the structure of the
64 dependence between products and descriptors on a factorial map that decomposes
65 the whole dependence into axes of maximal and decreasing dependence. As a final
66 step of the analysis, it is important to determine which descriptors are significantly
67 associated with which product. Again, two approaches can be distinguished to do so.
68 The first one is multidimensional alignment (Meyners et al., 2013) that consists of
69 considering a descriptor significantly positively (resp. negatively) associated to a
70 product when their vectors in the sensory space depicted by the CA form an angle
71 lower than or equal to 45° (resp. higher than or equal to 135°). The second approach
72 consists of testing each cell of the contingency table against the null hypothesis of
73 independence using a chi-square test or a Fisher's exact test (Mahieu, Visalli, &
74 Schlich, 2020a; Symoneaux, Galmarini, & Mehinagic, 2012).

75 All of these approaches but the combination of Cochran's Q statistics are based on
76 the chi-square statistic. The chi-square statistic can be directly used to test for overall
77 differences between the products before performing the CA. The total inertia of CA is
78 the chi-square statistic divided by the grand sum of the contingency table, also called
79 phi-square index. Since multidimensional alignment relies on the CA, it depends also
80 on the chi-square statistic. Finally, the tests per cell approach directly rely on the chi-
81 square statistic since Fisher's exact test can, roughly speaking, be seen as an exact
82 chi-square test.

83 These common practices assume that all citations are independent experimental
84 units within an evaluation, which is not the case since citations of descriptors by a
85 given subject for a given product are not independent. Instead, one evaluation, i.e.
86 the entire set of descriptors cited by one subject for one product, should be
87 considered as an experimental unit (Loughin & Scherer, 1998). Indeed, considering

88 citations as experimental units implies computing incorrect expected values under the
89 null hypothesis of independence between products and descriptors (Loughin &
90 Scherer, 1998), resulting in an incorrect chi-square statistic. Subsequent analyses of
91 FC and CATA data based on this chi-square statistic are thus also incorrect and can
92 sometimes lead to wrong interpretations.

93 The present paper aims to overcome the previous limitations by introducing the
94 multiple-response chi-square framework based on the multiple-response chi-square
95 statistic of Loughin and Scherer (1998). This new framework considers experimental
96 units as being the evaluations rather than the citations. First, some notations are
97 introduced and the multiple-response chi-square test of Loughin and Scherer (1998)
98 is presented and adapted to the context of FC and CATA data. Second, the multiple-
99 response Correspondence Analysis (MR-CA) is introduced. Third, the transposition of
100 the methodologies presented in Mahieu et al. (2020a) to the multiple-response chi-
101 square framework is established. Fourth, examples of the benefits of the new
102 framework are given on real CATA data. Finally, an overall discussion and a
103 conclusion are given.

104 2. Material and methods

105 2.1. Notations and multiple-response chi-square test of homogeneity

106 Let us consider an FC or a CATA experiment where S subjects evaluated P products
107 on D descriptors. Each product $p \in \{1, \dots, P\}$ has been evaluated E_p times and the
108 total number of evaluations is equal to $E = \sum_{p=1}^P E_p$. Note that in the particular case of
109 balanced experimental design, i.e. when all subjects evaluated all products, then $E =$
110 $S \times P$. Let us denote by n_{pd} the number of citations of descriptor $d \in \{1, \dots, D\}$ for
111 product p during the E_p evaluations and by C_d the number of citations of descriptor d
112 during all the E evaluations.

113 Let us denote by π_d^p the probability of descriptor d to be cited for product p . What is
114 under investigation is whether π_d^p differs from one product to another. Using the
115 above notations, the following hypotheses are considered:

$$116 H_0: \pi_d^1 = \dots = \pi_d^P = \pi_d, \quad \forall d \in \{1, \dots, D\}$$

117 H_A : It exists $d \in \{1, \dots, D\}$ and $p, p' \in \{1, \dots, P\}$ with $p \neq p'$ such as $\pi_d^p \neq \pi_d^{p'}$

118 Note that this does not correspond to a classical test of homogeneity since, for each
119 product p , multiple descriptors can be selected. Under the null hypothesis, the
120 expected number of citations of descriptors d for product p , denoted by $E(n_{pd})$, is
121 equal to $E_p \times \pi_d$ and can be estimated by $E_p \times C_d/E$. The following test statistic,
122 called multiple-response chi-square statistic, is thus introduced:

$$123 \quad \chi_{mr}^2 = \sum_{p=1}^P \sum_{d=1}^D \frac{(n_{pd} - E_p \times C_d/E)^2}{E_p \times C_d/E}$$

124 As $E_p \times C_d/E = E \times (E_p/E \times C_d/E)$, χ_{mr}^2 can also be expressed as:

$$125 \quad \chi_{mr}^2 = \sum_{p=1}^P \sum_{d=1}^D \frac{(n_{pd} - E \times (E_p/E \times C_d/E))^2}{E \times (E_p/E \times C_d/E)}$$

126 As in [Loughin and Scherer \(1998\)](#), it can be shown that the asymptotic distribution of
127 this test statistic under the null hypothesis is complicated because descriptors might
128 not be selected independently. A reasonable option for estimating the distribution of
129 χ_{mr}^2 under the null hypothesis is to consider a Monte-Carlo approach (see Section
130 2.3.1.2).

131 2.2. The multiple-response Correspondence Analysis

132 2.2.1. Conceptual difference with the usual Correspondence Analysis for Free-

133 Comment and Check-All-That-Apply data

134 In usual CA, the products are compared to each other according to their profile. The
135 profile of each product is defined as the proportion of citations of each descriptor for
136 this product relatively to the total number of citations (all descriptors combined)
137 elicited by this same product. Thus, in the context of FC and CATA data, when
138 products elicit different average citation rates (all descriptors combined) then
139 absolute differences in descriptors' citation rates between products are distorted due
140 to this "citation rescaling". The degree of distortion depends on the degree of

141 differences in citation rates between products. For more details on the usual CA, one
142 can refer e.g. to Greenacre (2007). The previous assertions are also applicable to
143 Hellinger-distance-based CA (Rao, 1995; Vidal, Tárrega, Antúnez, Ares, & Jaeger,
144 2015) because this latter is also based on the products' profiles.

145 MR-CA overcomes the above limitation by scaling products according to their number
146 of evaluations instead of their number of received citations. It results in comparing
147 products based on their average proportions of citations for each descriptor. This
148 "evaluation scaling" only has importance in the case of unbalanced design. Indeed,
149 products that are more evaluated are likely to elicit more citations of all descriptors
150 and it is necessary to put products on an equal footing before comparing them. To
151 summarize, the propensity of some products to elicit more citations than others does
152 not affect MR-CA while it does with usual CA.

153 When applied to FC and CATA data, MR-CA can be seen as standing at the frontier
154 between the usual CA of the descriptor by product contingency table and the PCA of
155 the products' average profiles depicted by the descriptors' proportions of citations.
156 MR-CA performs the PCA of the products' average proportions of citations but
157 weighting the descriptors proportionally to their citation rate as in usual CA.

158 2.2.2. Definition

159 Similarly, to the usual CA based on the singular value decomposition of the matrix of
160 standardized residuals defined by the usual chi-square statistic, the MR-CA is based
161 on the singular value decomposition of the matrix of standardized residuals defined
162 by the multiple-response chi-square statistic. Using the notations defined in the
163 previous section, let us consider:

- 164 - \mathbf{r} a column matrix of size $P \times 1$ whose elements equals $E_p/E, p \in \{1, \dots, P\}$
- 165 - \mathbf{c} a column matrix of size $D \times 1$ whose elements equals $C_d/E, d \in \{1, \dots, D\}$
- 166 - \mathbf{D}_r a diagonal matrix of size $P \times P$ whose diagonal elements equal $E_p/E, p \in$
167 $\{1, \dots, P\}$
- 168 - \mathbf{D}_c a diagonal matrix of size $D \times D$ whose diagonal elements equal $C_d/E, d \in$
169 $\{1, \dots, D\}$
- 170 - \mathbf{X} a matrix of size $P \times D$ whose general term equal $n_{pd}/E, p \in \{1, \dots, P\}, d \in$

171 $\{1, \dots, D\}$

172 Using these notations, the MR-CA is based on the singular value decomposition of
173 the matrix \mathbf{S} defined as:

174
$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}}(\mathbf{X} - \mathbf{r}\mathbf{c}^t)\mathbf{D}_c^{-\frac{1}{2}}$$

175 Let us denote by \mathbf{U} the matrix of left singular vectors of \mathbf{S} , $\mathbf{\Gamma}$ the diagonal matrix of
176 singular values of \mathbf{S} and \mathbf{V} the matrix of right singular vectors of \mathbf{S} such that $\mathbf{S} =$
177 $\mathbf{U}\mathbf{\Gamma}\mathbf{V}^t$. Similarly to the usual CA, the principal coordinates of the products are defined
178 as $\mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}\mathbf{\Gamma}$ and the so-called contribution coordinates (Greenacre, 2013) of the
179 descriptors are defined as \mathbf{V} . Note that since this system of coordinates defines a
180 strict biplot as defined in (Gabriel, 1971), it is suggested to use arrows rather than
181 points to display the descriptors' coordinates. This could help practitioners to
182 remember to interpret relations between products and descriptors as scalar products
183 (orthogonal projection) and not "proximities". Different systems of coordinates could
184 be used for displaying results of MR-CA similarly to usual CA (Greenacre, 2006).
185 However, the one proposed here has two benefits: it enables interpreting maps
186 similarly to Principal Component Analysis (PCA) biplots and the coordinates of the
187 columns (descriptors) reflect their respective contribution to the inertia and to the
188 distances between rows (products) (Greenacre, 2006).

189 Equivalently, the MR-CA can be defined as the PCA of the matrix $\mathbf{D}_r^{-1}\mathbf{X}\mathbf{D}_c^{-\frac{1}{2}}$. This
190 latter definition of MR-CA better highlights that the distance between two products
191 $p \neq p' \in \{1, \dots, P\}$ in the sensory space depicted by MR-CA called multiple-response
192 chi-square distance is equal to:

193
$$d_{\chi_{mr}^2}(p, p') = \sqrt{\sum_{d=1}^D \frac{E}{C_d} \left(\frac{n_{pd}}{E_p} - \frac{n_{p'd}}{E_{p'}} \right)^2}$$

194 From the definition of the multiple-response chi-square distance, one can see that the
195 weight given to each product is proportional to its number of evaluations rather than
196 its number of received citations as it is in usual CA. Finally, it should be noted that the
197 number of axes obtained by MR-CA is equal to the minimum between $P - 1$ and D ,
198 as in a PCA in which descriptors act as variables and products as individuals, while in
199 usual CA it is equal to the minimum between $P - 1$ and $D - 1$. This difference in the

200 number of axes is because usual CA centers both rows (products) and columns
201 (descriptors) while MR-CA centers only rows.

202 2.3. Statistical inference for multiple-response chi-square framework

203 This section transposes the methodologies from Mahieu et al. (2020a) to the multiple
204 chi-square framework.

205 2.3.1. The dimensionality test of the dependence

206 2.3.1.1. Conceptual aims for Free-Comment and Check-All-That-Apply data

207 The aim of this test is twofold. First, it investigates if at least one axis of the MR-CA is
208 significant, that is if some overall differences exist between the products. If no axis is
209 significant, interpreting subsequent analyses including the outputs from MR-CA might
210 lead to over-interpretations. If at least one axis is significant, the second aim of the
211 test is to determine the number of axes that can be considered significant and thus
212 interpreted. Because drawing sensory conclusions based on more than three or four
213 axes can be difficult visually, the number of significant axes is taken into account in
214 subsequent proposed analyses, which are simpler to interpret from a sensory point of
215 view.

216 2.3.1.2. Technical aspects

217 It is possible to test if the dependence of each MR-CA axis is significant with a
218 stepwise procedure similarly as for the usual CA (Mahieu et al., 2020a). The idea is
219 to test, at each step k ($k > 1$), whether the hypothesis of independence between
220 products and descriptors is still rejected while the dependence captured by the axes
221 1 to $k - 1$ was removed. In other words, it is tested if the strength of the dependence
222 is still large enough to be considered significant.

223 As seen in the previous section, the total number of MR-CA axes, denoted K , is
224 equal to the minimum between $P - 1$ and D . Let us consider \mathbf{U}_k the matrix of the $K -$
225 $k + 1$ last left singular vectors of \mathbf{S} , $\mathbf{\Gamma}_k$ the diagonal matrix of the $K - k + 1$ last
226 singular values of \mathbf{S} and \mathbf{V}_k the matrix of the $K - k + 1$ last right singular vectors of \mathbf{S}
227 such that $\mathbf{S}_k = \mathbf{U}_k \mathbf{\Gamma}_k \mathbf{V}_k^t$. Let us denote by $\chi_{mr_k}^2$ the multiple-response chi-square
228 statistic of the derived contingency table corresponding to the $K - k + 1$ last axes of
229 the MR-CA denoted \mathbf{Y}_k and defined following the *reconstitution formula* as:

230
$$Y_k = \left(D_r^{\frac{1}{2}} S_k D_c^{\frac{1}{2}} + r c^t \right) \times E$$

231 The multiple-response chi-square test associated with the test statistic $\chi_{mr_k}^2$ enables
 232 testing if the k-th axis of the MR-CA captures a significant dependence between
 233 products and descriptors. Note that if $k = 1$ then this test corresponds to the multiple-
 234 response chi-square test defined in section 2.1.

235 The multiple-response chi-square statistic of the products by descriptors contingency
 236 table is related to the eigenvalues of the MR-CA by the following equation:

237
$$\chi_{mr}^2 = E \times \sum_{i=1}^K \lambda_i$$

238 where χ_{mr}^2 is the multiple-response chi-square statistic of the contingency table, E is
 239 the total number of evaluations and λ_i is the i-th eigenvalue of the MR-CA. This
 240 relation enables to compute each $\chi_{mr_k}^2$ as:

241
$$\forall k, \quad \chi_{mr_k}^2 = E \times \sum_{i=k}^K \lambda_i$$

242 To estimate the distribution of each $\chi_{mr_k}^2$ under the null hypothesis, it is proposed to
 243 randomly permute the response vectors along products within each subject (Mahieu
 244 et al., 2020a; Meyners et al., 2013; Meyners & Pineau, 2010; Wakeling, Raats, &
 245 MacFie, 1992; Winkler, Webster, Vidaurre, Nichols, & Smith, 2015), a response
 246 vector referring to all citations given for one product by one subject.

247 To summarize, the dependence between products and descriptors captured by each
 248 MR-CA axis can be tested following these steps:

- 249 (i) Simulate a large number of contingency tables by randomly permuting the
 250 response vectors along products within each subject
- 251 (ii) Perform MR-CA on each of the simulated contingency tables
- 252 (iii) Compute all $\chi_{mr_k}^{2(*)}$ statistics, $k = 1, \dots, K$, as $\chi_{mr_k}^{2(*)} = E \times \sum_{i=k}^K \lambda_i^{(*)}$ for each
 253 of the simulated contingency tables

254 (iv) Compute the p-value of each $\chi_{mr_k}^2$ as the proportion of $\chi_{mr_k}^{2(*)}$ under
255 permutation having an equal or a larger value than the observed $\chi_{mr_k}^2$.

256 2.3.2. Confidence ellipses and discrimination of the products

257 In MR-CA, as well as in every multivariate analysis providing a product map,
258 superimposing confidence ellipses on product coordinates is crucial to estimate if
259 products are well discriminated. A total bootstrap procedure (Cadoret & Husson,
260 2013) is proposed to achieve this objective. This procedure consists of generating
261 virtual panels by randomly resampling with replacement the subjects of the actual
262 panel. Then, the product configurations of the virtual panels are rotated on the
263 product configuration of the actual panel thanks to Procrustes rotations. A confidence
264 ellipse is then constructed for each product based on the coordinates of its rotated
265 bootstrap replicates. It is proposed to rely on the significant axes, indicated by the
266 test of dependence presented in section 2.3.1, to determine the number of axes to
267 account for the Procrustes rotations in the total bootstrap procedure.

268 For each pair of products, to determine if the two products are significantly different, it
269 is proposed to rely on the total bootstrap test (Mahieu, Visalli, Thomas, & Schlich,
270 2020b) considering the null hypothesis that the two products are not different. For
271 each pair of products, a canonical discriminant analysis based on the rotated
272 bootstrap replicates of the two products is performed. The rotated bootstrap
273 replicates of the two products are then projected on the axis resulting from the
274 canonical discriminant analysis. The distribution of the paired differences of the
275 projected bootstrap replicates is estimated. Finally, the probability of zero to belong to
276 this distribution is estimated and used as a p-value of the test. It is proposed to
277 perform the total bootstrap tests on the significant axes.

278 2.3.3. Determination of the significant associations between products and 279 descriptors: multiple-response tests per cell

280 2.3.3.1. Conceptual aims for Free-Comment and Check-All-That-Apply data

281 These tests aim to investigate the relations between descriptors and products. In
282 particular, they investigate for a given descriptor and a given product if this descriptor
283 is cited for this product in a proportion that significantly differs from the overall

284 average citation proportion of this descriptor all products combined. The tests can be
285 one-sided (positive differences) or two-sided (both positive and negative differences):
286 this choice is up to the discretion of the practitioner. A discussion is given about this
287 choice in Mahieu et al. (2020a).

288 *2.3.3.2. Technical aspects*

289 It is proposed to define a multiple-response test per cell to test the following
290 hypotheses for a given $p \in \{1, \dots, P\}$ and a given $d \in \{1, \dots, D\}$:

$$291 \quad H_0: \pi_d^p = \pi_d$$

$$292 \quad H_A: \pi_d^p \neq \pi_d$$

293 The multiple-response test per cell is based on a Monte-Carlo procedure. In this
294 procedure, for each product $p \in \{1, \dots, P\}$, E_p evaluations are randomly drawn among
295 the subjects having evaluated p and only one evaluation is randomly drawn among
296 each of these subjects. This enables constructing a virtual contingency table under
297 the null hypothesis accounting for both the subject structure of the data and the non-
298 independence of the citations. Indeed, one evaluation is randomly drawn from each
299 subject having evaluated p and one randomly drawn evaluation (that respect the joint
300 distributions of citations of the descriptors) contributes to several cells in the virtual
301 contingency table.

302 A large number of virtual contingency tables under the null hypothesis can be
303 generated by repeating this procedure. Then, for each cell, the proportion of $n_{pd}^{(*)}$
304 under the null hypothesis having an equal or a more extreme value than the
305 observed n_{pd} constitute a p-value of the test. The multiple-response tests per cell can
306 be performed with a two-sided alternative hypothesis or a one-sided greater
307 alternative hypothesis.

308 Finally, it is proposed to perform the multiple-response tests per cell on the derived
309 contingency table corresponding to the significant axes (Mahieu et al., 2020a),
310 denoted Y_{sig} , and defined following the *reconstitution formula* as:

$$311 \quad Y_{sig} = \left(\mathbf{D}_r^{\frac{1}{2}} \mathbf{S}_{sig} \mathbf{D}_c^{\frac{1}{2}} + \mathbf{rc}^t \right) \times E$$

312 Where $\mathbf{S}_{sig} = \mathbf{U}_{sig}\mathbf{\Gamma}_{sig}\mathbf{V}_{sig}^t$ with \mathbf{U}_{sig} the matrix of left singular vectors of \mathbf{S}
313 corresponding to the significant axes, $\mathbf{\Gamma}_{sig}$ the diagonal matrix of singular values of \mathbf{S}
314 corresponding to the significant axes and \mathbf{V}_{sig} the matrix of right singular vectors of \mathbf{S}
315 corresponding to the significant axes.

316 To perform the multiple-response tests per cell on \mathbf{Y}_{sig} rather than on the observed
317 contingency table results in a gain of power without any inflation of the type I error as
318 suggested by the simulation results presented in the [Appendix](#). [The simulation results](#)
319 [also suggest that the smaller the number of significant axes and the intensity of the](#)
320 [dependence between products and descriptors, the higher the gain of power.](#)

321 2.4. Examples

322 These examples from two CATA datasets aim to compare outputs obtained from
323 analyses belonging to the usual chi-square framework to those obtained from
324 analyses belonging to the multiple-response chi-square framework. Although these
325 examples deal with CATA datasets, note that the multiple-response chi-square
326 framework is also appropriate to analyze FC data.

327 2.4.1. Datasets

328 The datasets are the same from [Mahieu, Visalli, Thomas, and Schlich \(2021\)](#).

329 The study took place at the Barry Callebaut© Company, Belgium. Seventy regular
330 consumers of milk chocolates (at least once every two weeks) were recruited among
331 the employees of the Barry Callebaut© Company (not implied in sensory and
332 consumer research). They performed a CATA task on four milk chocolates having
333 different recipes: a standard Belgian milk chocolate, a Swiss milk chocolate, a milk
334 compound chocolate, and a protein base milk chocolate. The four products were
335 presented according to a Williams Latin square design. For each product, the CATA
336 task was carried out according to two sensory modalities: texture in the mouth
337 followed by flavor in the mouth. All the CATA descriptors were selected thanks to the
338 expertise of sensory experts from the Barry Callebaut© Company. The collected data
339 were then stored in two contingency tables, one per sensory modality, by cross
340 tabulating the citation counts of the descriptors (columns) by the products (rows).

341 Since sensory interpretation is out of the scope of this paper dedicated to the
342 comparison of the two chi-square frameworks, the descriptors were renamed *D1*, *D2*,
343 etc. and the products were renamed *P1*, *P2*, *P3*, and *P4*. Finally, for the texture
344 dataset, an additional product called *P5* was artificially created. This product is
345 exactly *P4* except that for *P5* the number of received citations for every descriptor
346 has been divided by two as compared to *P4*. This was made to illustrate the
347 differences between the multiple-response chi-square framework and the usual chi-
348 square framework.

349 2.4.2. Analyses

350 All analyses were performed using R 4.0.2 (R Core Team, 2020). The analyses
351 belonging to the multiple-response chi-square framework were performed using the R
352 package “MultiResponseR” developed for this purpose by the authors.

353 The two contingency tables were analyzed using the following procedure. An alpha
354 risk (Type I error) of 10% was considered as the significance level.

355 The dimensionality of the dependence between products and descriptors was
356 determined within each chi-square framework using the dimensionality test (2000
357 simulations) presented in Mahieu et al. (2020a) for the usual chi-square framework
358 and using the dimensionality test (2000 simulations) presented in section 2.3.1 for the
359 multiple-response chi-square framework.

360 When at least one axis was significant, the corresponding CA (usual or multiple-
361 response) was performed on the contingency table. Outputs of each CA were
362 displayed using a standard biplot (Greenacre, 2013). For each CA, confidence
363 ellipses for the products' coordinates in the sensory space were computed with a
364 total bootstrap procedure using 2000 bootstrap samples. The Procrustes rotations
365 were performed on the significant axes. For each pair of products, a total bootstrap
366 test was performed on the significant axes for assessing the significance of product
367 difference.

368 For each pair of product and descriptor (cell), a Fisher's exact test was performed for
369 the usual chi-square framework and a multiple-response test per cell as described in
370 section 2.3.3 (2000 simulations) was performed for the multiple-response chi-square
371 framework. All tests per cell were performed with a one-sided greater alternative

372 hypothesis and conducted on the derived contingency table corresponding to the
 373 significant axes.

374 3. Results

Sensory modality	Chi-square framework	Axis 1	Axis 2	Axis 3	Axis 4
Texture	<i>Usual</i>	0.447 (<0.001)	0.162 (<0.001)	0.001 (0.9970)	0 (1)
	<i>Multiple-response</i>	0.907 (<0.001)	0.323 (<0.001)	0.079 (<0.001)	0.002 (0.6146)
Flavor	<i>Usual</i>	0.243 (<0.001)	0.012 (0.0154)	0.003 (0.0914)	/
	<i>Multiple-response</i>	0.557 (<0.001)	0.089 (<0.001)	0.013 (0.0054)	/

375 **Table 1:** Eigenvalues of Correspondence Analysis and corresponding p-values (in
 376 brackets) for testing the number of significant axes in the usual and multiple-
 377 response frameworks for the two datasets

378 Table 1 shows that whatever the sensory modality and the axis considered, the
 379 eigenvalues of the CA are higher in the multiple-response framework than in the
 380 usual one. This suggests that the usual framework underestimates the dependence
 381 between products and descriptors. This line of reasoning is reinforced by the
 382 example treated by Loughin and Scherer (1998) as they obtained a lower p-value
 383 (which is partly a function of the effect size) for their chi-square test in the multiple-
 384 response framework than in the usual one. On the dimensionality of the dependence,
 385 Table 1 shows that similar conclusions are provided between products and
 386 descriptors by the two chi-square frameworks concerning the flavor dataset: three
 387 axes capture significant dependence. However, the dependence on the third axis
 388 appears more certain ($p=0.0054$) in the multiple-response chi-square framework than
 389 in the usual one ($p=0.0914$). Concerning the texture dataset, only two axes capture
 390 significant dependence within the usual chi-square framework while three axes
 391 capture significant dependence within the multiple-response chi-square framework.

392 Fig. 1 shows that for the texture dataset, the maps depicted by the two first axes of
 393 the usual CA (Fig. 1(a)) and the MR-CA (Fig. 1(b)) are very similar: all the products
 394 except *P5* and all the descriptors have the same position on the two maps. The only
 395 difference between these maps is the location of *P5* being different from *P4* and
 396 closer to the origin in MR-CA (Fig. 1(b)) as compared to usual CA (Fig. 1(a)). The
 397 reason for this difference lies in the fact that *P4* and *P5* have the same profile
 398 (repartition of citations) in the usual CA. On the contrary, the MR-CA captures that *P5*

399 received fewer citations than *P4* for all the descriptors but still following the same
400 pattern of association with the descriptors. This explains the position of *P5* relative to
401 *P4*: *P5* deviates from independence in the same direction that *P4* (same pattern of
402 association with the descriptors) but *P5* is closer to the origin of the coordinates
403 system than *P4* (received fewer citations). Concerning the third significant axis
404 obtained with the multiple-response chi-square framework on the texture dataset
405 (Fig. 1(c)), it mainly traduces that *P5* received fewer citations than *P4* for all
406 descriptors, which is logical. Note that the usual CA is unable to capture this
407 difference between *P4* and *P5*, which explains the non-significance of the third axis
408 for this CA.

409 For the flavor dataset, Fig. 2 shows that the spaces provided by the usual CA and the
410 MR-CA exhibit different configurations for both products and descriptors. For every
411 descriptor, there is at least one other product that received more citations than *P3*.
412 Thus, in MR-CA, it is associated with no descriptor, which explains its position: *P3*
413 lies at the opposite of every descriptor loadings (Fig. 2(c) & Fig. 2(d)). On the
414 contrary, in usual CA, *P3* seems to be associated with *D1*, *D5*, and *D6* and slightly
415 with *D2* (Fig. 2(a) & Fig. 2(b)). Indeed, in usual CA, the number of citations received
416 by *P3* for every descriptor is rescaled according to its total number of received
417 citations. Thus, the fact that for every descriptor there is at least one other product
418 that received more citations than *P3* is erased in the usual CA. These features of *P3*
419 are the principal explanation of the differences between the spaces provided by MR-
420 CA and usual CA, both applied on the flavor dataset. Another notable difference is
421 between the maps depicted by the two first axes of the usual CA and of the MR-CA
422 (Fig. 2(a) & Fig. 2(c)). On these maps, *P1* and *P2* appear to be more associated with
423 *D1*, *D4*, and *D6* in MR-CA as compared to the usual CA. This difference is due to the
424 opposite phenomenon that occurred with *P3*: *P1* and *P2* received much more
425 citations than *P3* and *P4* for these descriptors and the usual CA distorts this
426 difference while the MR-CA does not.

427 Concerning the total bootstrap tests, whatever the considered sensory modality and
428 whatever the considered chi-square framework, the conclusions they provided were
429 the same except when considering the pair *P4/P5* and the texture dataset. In the
430 usual chi-square framework, *P4* is for sure not different from *P5* ($p = 1$) while *P4* and

431 *P5* are significantly different in the multiple-response chi-square framework ($p <$
432 0.001). Of course, this is perfectly in line with [Fig. 1\(a\)](#), [Fig. 1\(b\)](#), and [Fig. 1\(c\)](#).

433 For texture, [Fig. 3](#) shows that differences in the significant associations concern the
434 pairs: *P2-D2*, *P5-D3*, and *P5-D8*. The pair *P2-D2* is significant in the multiple-
435 response and not in the usual chi-square framework because *P2* received more
436 citations of this descriptor than the other products [except P3](#). Concerning the product
437 *P5*, it is noticeable that in the usual framework, it is significantly associated with the
438 same descriptors as *P4* (*D3*, *D5*, and *D8*), which was expected since *P5* has the
439 same profile that *P4* in this framework. The pair *P5-D8* being significant in the usual
440 framework with a percentage of citations (25.71%) lower than the one of product *P3*
441 (28.57%) which is not significant nicely illustrates the issue of the “citation rescaling”
442 due to considering the citations as experimental units. Since *P3* and *P5* were
443 evaluated the same number of times, it is counterintuitive to have the one with the
444 lowest proportion of citations significant and not the other. However, in the multiple-
445 response framework, both *P3* and *P5* are not significantly associated with *D8*, which
446 is consistent. Regarding the pair *P5-D3*, the association is not significant in the
447 multiple-response chi-square framework while it is in the usual chi-square framework.
448 This difference is due to the “citation rescaling” that occurs in the usual chi-square
449 framework and not in the multiple-response one.

450 Concerning the flavor dataset, several differences are shown in [Fig. 3](#) between the
451 conclusions provided by the two chi-square frameworks on descriptor by product
452 significant associations. As was suggested by [Fig. 2](#), *P1* and *P2* are significantly
453 associated with *D1*, *D4*, and *D6* in the multiple-response chi-square framework while
454 only *P2* is significantly associated with only *D4* and *D6* in the usual chi-square
455 framework. This difference is because *P1* and *P2* received much more citations than
456 *P3* and *P4* for these descriptors. On the contrary, without the “citation rescaling”,
457 since *P3* and *P4* received fewer citations, they got less significance in the multiple
458 response framework; precisely, *P3-D5* and *P4-D3* are no longer significant in this
459 framework. Finally, it is noticeable that the counterintuitive conclusion in the usual
460 chi-square framework on the significant association of *D3* with *P4* and not with *P1*
461 and *P2* while these received a higher percentage of *D3* citations than *P4*, no longer
462 holds in the multiple-response chi-square framework.

463 4. Discussion

464 To the best of our knowledge, it is the first time that a chi-square framework properly
465 taking into account multiple-response data is introduced. The proposed analyses
466 including the test of dimensionality, the product confidence ellipses, the pairwise
467 product comparisons, and the product by descriptor association tests, the three of
468 them being conducted on the significant axes, are all originals. [This multiple-](#)
469 [response chi-square framework fits perfectly to FC and CATA data.](#) However, this
470 multiple-response chi-square framework is not restricted to be used only in sensory
471 and consumer science and can be used to analyze any multiple-response data
472 whatever the field they come from.

473 The examples presented in this paper showed that the multiple-response chi-square
474 framework is better suited than the usual chi-square framework to analyze FC and
475 CATA data. A major benefit of using the multiple-response chi-square framework is
476 that when the experimental design is balanced, every product is equally weighted.
477 This is more appropriate and leads to logical outputs as opposed to the usual chi-
478 square framework that can lead to counterintuitive outputs. Indeed, it sounds more
479 logical to weight the products equally and not rescale them according to their number
480 of received citations when they have been evaluated the same number of times. Note
481 that an equivalent weighting of the products using the usual chi-square framework is
482 almost impossible since products are very unlikely to receive the same number of
483 citations at the panel level. The multiple-response tests per cell introduced in this
484 paper take into account all the specific aspects of FC and CATA data, especially the
485 non-independence of citations between descriptors.

486 The conclusions provided by the two chi-square frameworks are not always
487 necessarily different. For example, they would have been almost the same on the
488 texture dataset if *P5* had not been artificially added to the dataset. [The more different](#)
489 [the citation rates \(all descriptors combined\) between products are, the more the](#)
490 [conclusions drawn from the usual chi-square framework will differ from the multiple-](#)
491 [response one.](#) The products likely receive different numbers of citations when some
492 products have few sensory characteristics while some others have a lot or when
493 some products present obvious characteristics while the characteristics of the other

494 products are more subtle; these kinds of situations are likely to occur in sensory
495 evaluation.

496 Since the multiple-response chi-square framework relies heavily on Monte-Carlo and
497 bootstrap simulations, the results of the proposed analyses are not instantaneous.
498 For the datasets used as examples, it took around 30 seconds by dataset to obtain
499 the results of all analyses. However, this computation time increases with the number
500 of evaluations and thus with the number of subjects and products. For large datasets
501 (e.g. 3000 evaluations), it takes around 5 minutes to obtain the results using the
502 settings of this paper.

503 5. Conclusion

504 For the analysis of Free-Comment and Check-All-That-Apply data, the paper
505 proposes to replace the usual chi-square framework with a new multiple-response
506 chi-square framework taking into account dependence among citations within an
507 evaluation. It is thus statistically valid while the former was not. The new framework
508 includes a test of dimensionality, a Correspondence Analysis with confidence
509 ellipses, a test for pairwise product comparison, and a test of significance of product
510 by descriptor associations. Note that ellipses, tests of product comparisons, and tests
511 of association with descriptors are the three of them computed on the significant axes
512 of dependence. The basic difference introduced by this new framework is not to
513 longer consider citations (one descriptor for one product by one subject) as
514 experimental units, but to rely on evaluations (vector of citations for one product by
515 one subject) as being the experimental units. Simulations showed that testing the
516 significance of product by descriptor associations on the significant axes of
517 dependence increased power in detecting product by descriptor associations without
518 any inflation of the type I error. [The new approaches are supported by an R package
519 called "MultiResponseR" and available upon request to the authors and on GitHub.](#)

520 Appendix: Simulations

521 To investigate the benefits and/or the downsides of performing the multiple-response
522 tests per cell on the derived contingency table corresponding to the significant axes,
523 simulations of sensory data were performed. For every simulation, 60 subjects, 5

524 products, and 10 descriptors were considered. The 5 products were considered as
525 being evaluated by the 60 consumers, as it is common in sensory evaluation. The
526 descriptors marginal probabilities were randomly chosen and were the following:
527 0.20, 0.56, 0.26, 0.23, 0.21, 0.30, 0.20, 0.42, 0.52, 0.75. From these marginal
528 probabilities, the matrix of expected probabilities under the null hypothesis of
529 independence between products and descriptors was computed. This matrix
530 contained 50 cells (5 products \times 10 descriptors).

531 Some deviation from independence was then added iteratively to these expected
532 probabilities such that at each step, one axis of dependence was added orthogonally
533 to the previous axes. On the first added axis, two products *were differentiated* on six
534 descriptors. On the second added axis, two products *were differentiated* on four
535 descriptors. On the third added axis, two products *were differentiated* on two
536 descriptors. On the fourth added axis, four products *were differentiated* on four
537 descriptors. This enabled to control the true dimensionality of the dependence
538 between products and descriptors. The cells that deviated from the null hypothesis
539 did with equal intensity but with opposite direction to keep the marginal probabilities
540 fixed. Two levels of deviation intensity were considered: 0.1 and 0.2. 8 matrices (4
541 levels of dimensionality \times 2 levels of deviation intensity) of probabilities were thus
542 generated. Each of the 8 matrices contains 50 cells (5 products \times 10 descriptors).

543 For each of these 8 matrices, 1000 datasets were simulated. Each of these datasets
544 was generated by adding 60 individual data (the subjects). Each individual data was
545 generated by performing a random Bernoulli draw for each of the 50 cells according
546 to the specified probability given in the matrix.

547 For each of the 8000 datasets (8 matrices of probabilities \times 1000 generated
548 datasets), the number of significant axes was considered unknown and was
549 determined using the dimensionality test presented in section 2.3.1. The multiple-
550 response tests per cell were then performed on either the observed table or the
551 derived contingency table corresponding to the significant axes returned by the test.
552 The p-values of the multiple-response tests per cell were stored.

553 For each combination of the factors deviation intensity (0.1 or 0.2), dimensionality
554 (one axis, two axes, etc.), and table (observed or derived) and for each of the 50
555 cells, the proportion of test (among the 1000 datasets) rejecting the null hypothesis

556 was computed at the following nominal alpha risks: 5%, 7.5%, and 10%. Then, the
557 results from a given cell were assigned either to the group H0 if its probability was not
558 modified or to the group H1 otherwise. Finally, the average proportion of rejection of
559 the null hypothesis was computed within each group (H0 or H1), number of
560 dimensions, and deviation intensity. The results are presented in [Table 3](#).

<i>Deviation intensity</i>	Dimensionality	Nominal alpha risk = 5%				Nominal alpha risk = 7.5%				Nominal alpha risk = 10%			
		H0 derived table	H0 observed table	H1 derived table	H1 observed table	H0 derived table	H0 observed table	H1 derived table	H1 observed table	H0 derived table	H0 observed table	H1 derived table	H1 observed table
0.1	1	0.020	0.034	0.521	0.434	0.030	0.052	0.592	0.507	0.040	0.071	0.644	0.562
	2	0.029	0.034	0.461	0.444	0.044	0.051	0.537	0.514	0.061	0.069	0.595	0.569
	3	0.032	0.032	0.451	0.450	0.049	0.049	0.523	0.519	0.069	0.069	0.582	0.577
	4	0.034	0.032	0.532	0.536	0.052	0.049	0.594	0.599	0.070	0.068	0.643	0.646
0.2	1	0.018	0.032	0.987	0.955	0.027	0.050	0.991	0.969	0.037	0.069	0.994	0.978
	2	0.028	0.033	0.973	0.960	0.041	0.050	0.982	0.973	0.058	0.068	0.988	0.980
	3	0.030	0.032	0.966	0.962	0.046	0.048	0.977	0.973	0.064	0.066	0.984	0.981
	4	0.029	0.030	0.973	0.974	0.046	0.046	0.982	0.982	0.066	0.066	0.987	0.987

562 **Table 3:** Average proportion of rejection of the null hypothesis among the 1000 simulations depending on the deviation intensity,
563 the dimensionality, the nominal alpha risk, the table considered, and the deviation from the null hypothesis or not.

564 Table 3 shows that the empirical type I error never exceeded the nominal alpha risk in
565 group H0 for both approaches, which suggests that both approaches are valid. It can
566 be seen that the empirical type I error in the H0 group was even slightly lower when
567 considering the derived table which is a nice feature.

568 The percentage of rejections in group H1 (estimating test power) was higher when
569 considering the derived table as compared to the observed table whatever the
570 combination of factors considered except with a dimensionality of 4. Therefore,
571 performing the multiple-response tests per cell on the derived contingency table
572 corresponding to the significant axes enables gaining power without increasing type I
573 error. It should also be noted that the smaller the dimensionality of the dependence,
574 the higher the gain of power. It is logical because a low dimensionality maximizes the
575 difference between the derived table and the observed one. Finally, it should also be
576 noted that the gain in power is higher with the lower independence deviation (0.1 vs
577 0.2), that is with the more complex/subtle situation. This is a nice feature arguing in
578 favor of performing the multiple-response tests per cell on the derived contingency
579 table corresponding to the significant axes.

580 Acknowledgments

581 This study is part of a Ph.D. financed by the Region Bourgogne-Franche-Comté and
582 the SensoStat Company.

583 The authors would like to thank Barry Callebaut© for providing the chocolate
584 samples.

585 Calculations in the Appendix were performed using HPC resources from PSIUN
586 CCUB (Centre de Calcul de l'Université de Bourgogne, France). The Institut de
587 Mathématiques de Bourgogne is supported by the EIPHI Graduate School (contract
588 ANR-17-EURE-0002).

589 References

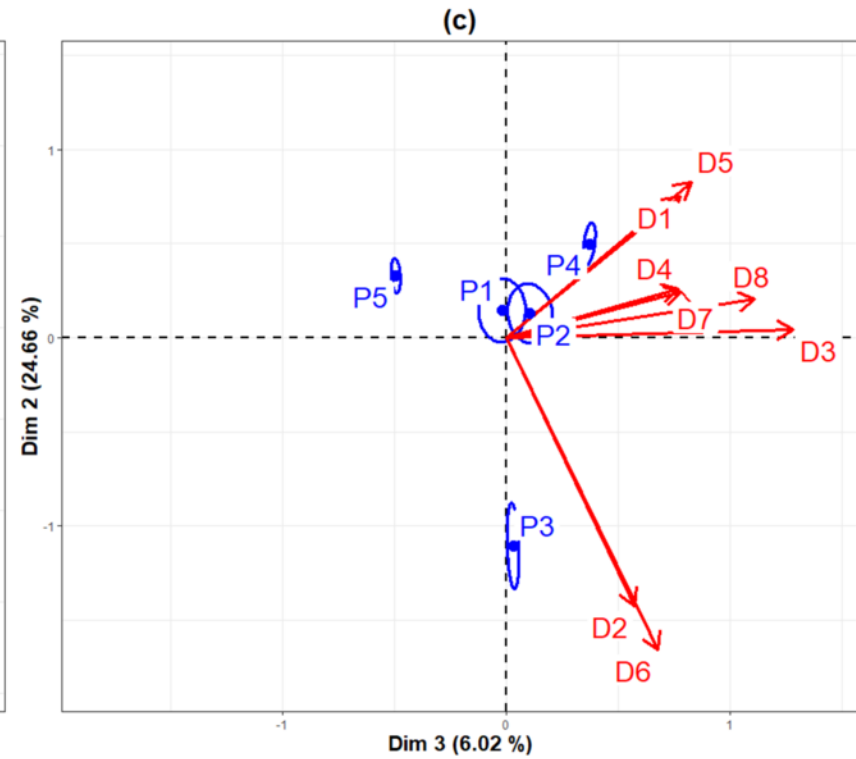
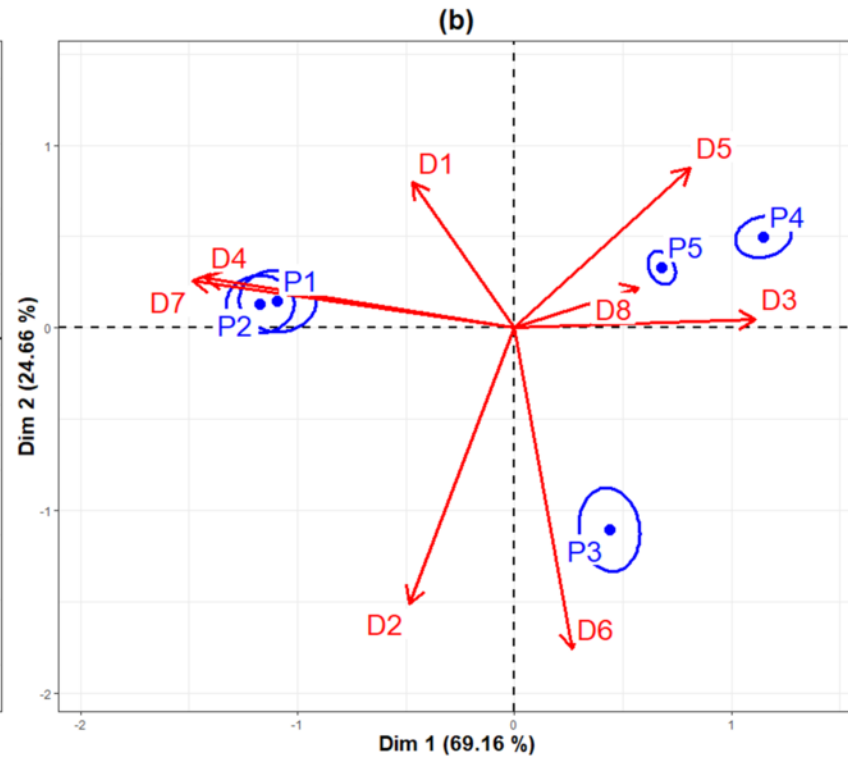
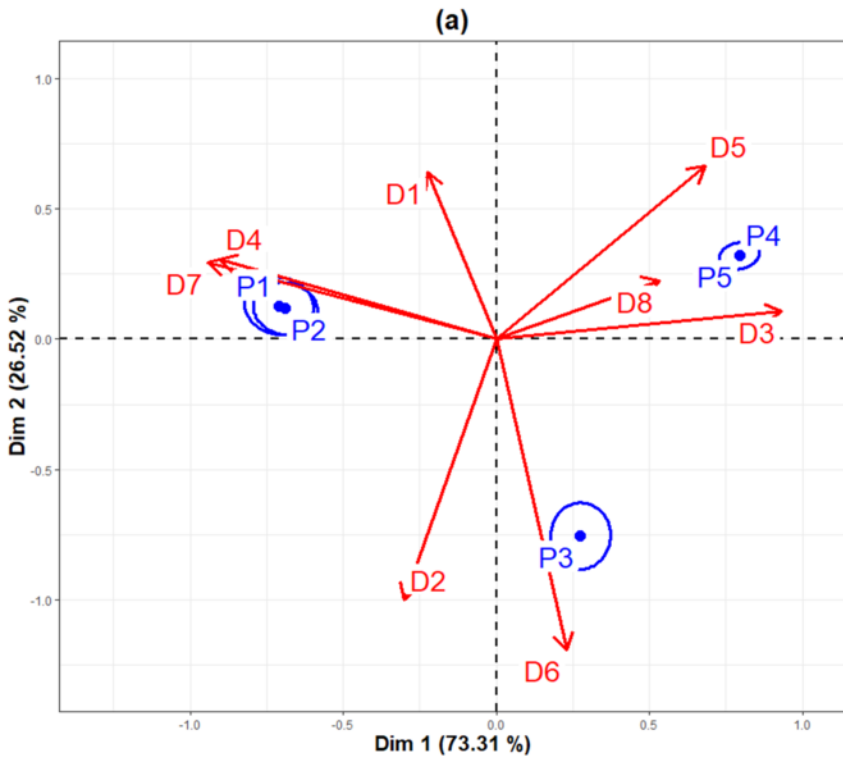
590 Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of
591 check-all-that-apply response compared to traditional scaling of attributes for
592 salty snacks. In, *7th Pangborn Sensory Science Symposium*. Minneapolis,
593 USA.

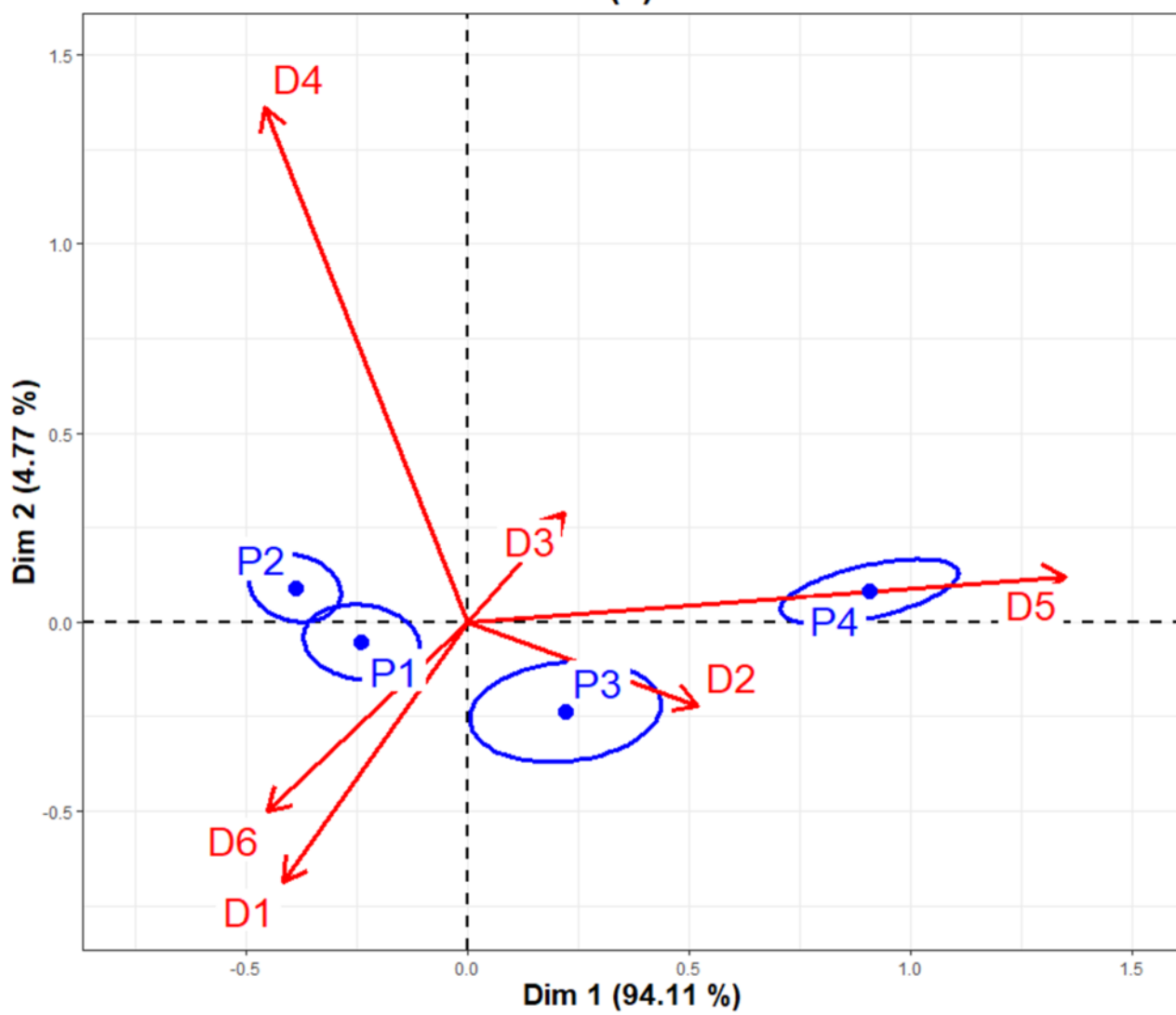
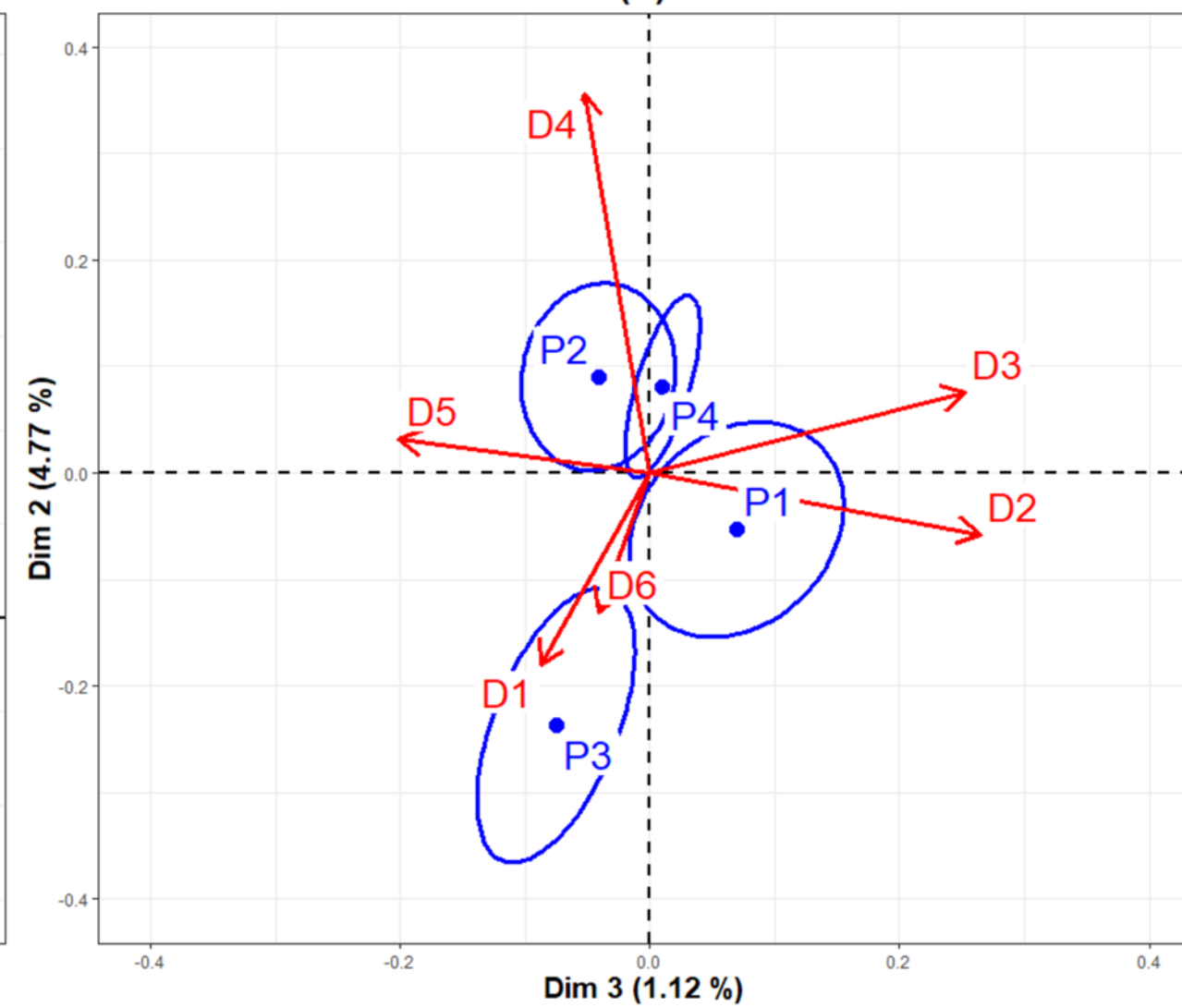
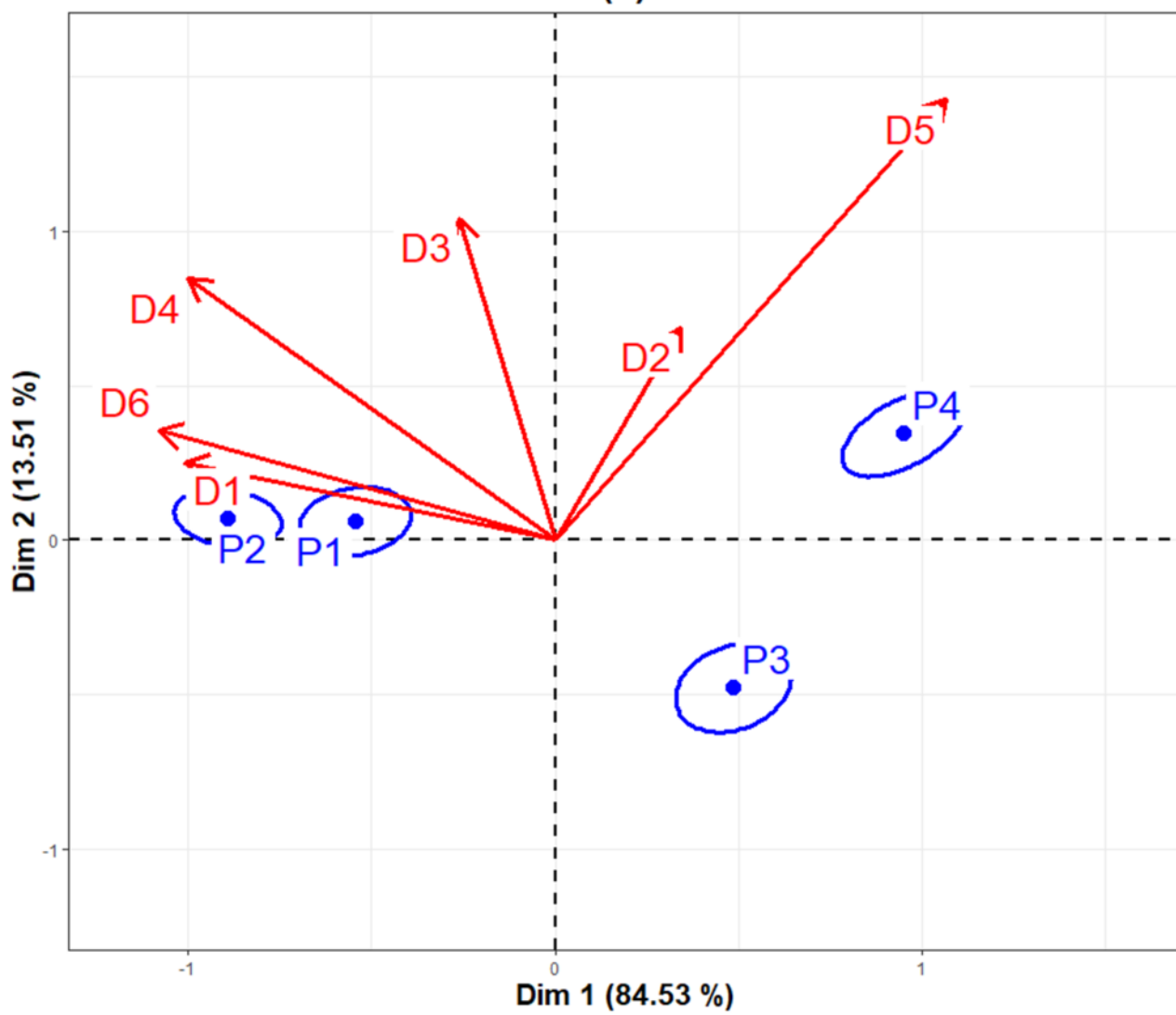
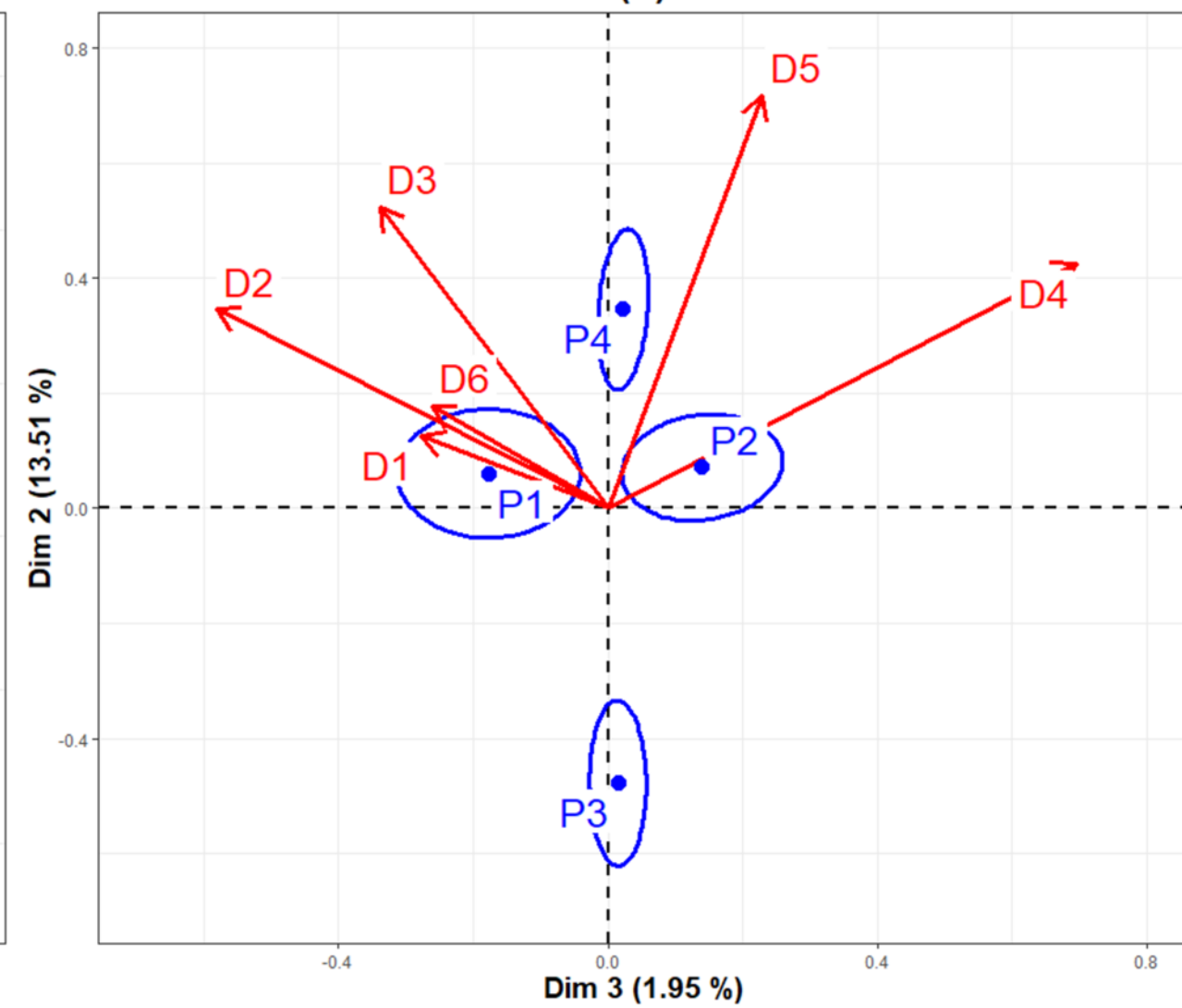
- 594 Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses
595 applied at sensory data. *Food Quality and Preference*, 28(1), 106-115.
- 596 Gabriel, K. R. (1971). The biplot graphic display of matrices with application to
597 principal component analysis. *Biometrika*, 58(3), 453-467.
- 598 Greenacre, M. (2006). Tying up the loose ends in simple, multiple, joint
599 correspondence analysis. In, *Compstat*. Heidelberg: Physica-Verlag HD.
- 600 Greenacre, M. (2007). *Correspondence Analysis in Practice, second edition*.
- 601 Greenacre, M. (2013). Contribution Biplots. *Journal of Computational and Graphical*
602 *Statistics*, 22(1), 107-122.
- 603 Loughin, T. M., & Scherer, P. N. (1998). Testing for Association in Contingency
604 Tables with Multiple Column Responses. *Biometrics*, 54(2), 630-637.
- 605 Mahieu, B., Visalli, M., & Schlich, P. (2020a). Accounting for the dimensionality of the
606 dependence in analyses of contingency tables obtained with Check-All-That-
607 Apply and Free-Comment. *Food Quality and Preference*, 83.
- 608 Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020b). Free-comment
609 outperformed check-all-that-apply in the sensory characterisation of wines with
610 consumers at home. *Food Quality and Preference*, 84.
- 611 Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2021). An investigation of the
612 stability of Free-Comment and Check-All-That-Apply in two consumer studies
613 on red wines and milk chocolates. *Food Quality and Preference*, 90.
- 614 Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for
615 the analysis of CATA data. *Food Quality and Preference*, 30(2), 309-319.
- 616 Meyners, M., & Pineau, N. (2010). Statistical inference for temporal dominance of
617 sensations data using randomization tests. *Food Quality and Preference*,
618 21(7), 805-814.
- 619 R Core Team. (2020). R: A language and environment for statistical computing. In.
620 Vienna, Austria: R Foundation for Statistical Computing.
- 621 Rao, C. R. (1995). A Review of Canonical Coordinates and an Alternative to
622 Correspondence Analysis Using Hellinger Distance. *Questiio*, 19, 23-63.
- 623 Symoneaux, R., Galmarini, M. V., & Mehinagic, E. (2012). Comment analysis of
624 consumer's likes and dislikes as an alternative tool to preference mapping. A
625 case study on apples. *Food Quality and Preference*, 24(1), 59-66.
- 626 ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey
627 responses: a complementary method to preference mapping. *Food Quality*
628 *and Preference*, 14(1), 43-52.
- 629 Vidal, L., Tárrega, A., Antúnez, L., Ares, G., & Jaeger, S. R. (2015). Comparison of
630 Correspondence Analysis based on Hellinger and chi-square distances to
631 obtain sensory spaces from check-all-that-apply (CATA) questions. *Food*
632 *Quality and Preference*, 43, 106-112.
- 633 Wakeling, I. N., Raats, M. M., & MacFie, H. J. H. (1992). A New Significance Test for
634 Consensus in Generalized Procrustes Analysis. *Journal of Sensory Studies*,
635 7(2), 91-96.
- 636 Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E., & Smith, S. M. (2015).
637 Multi-level block permutation. *NeuroImage*, 123, 253-268.

Fig. 1: Biplot from Correspondence Analysis of the texture dataset: (a) usual CA (axes 1-2), (b) MR-CA (axes 1-2) and (c) MR-CA (axes 3-2).

Fig. 2: Biplot from Correspondence Analysis of the flavor dataset: (a) usual CA (axes 1-2), (b) usual CA (axes 3-2), (c) MR-CA (axes 1-2) and (d) MR-CA (axes 3-2).

Fig. 3: Descriptors by product percentages of citations across the panel. Highlighted cells denote a significant ($\alpha = 10\%$) Fisher exact tests per cell in the usual chi-square framework or a significant ($\alpha = 10\%$) multiple-response test per cell (2000 simulations) in the MR chi-square framework



(a)**(b)****(c)****(d)**

	Usual					Multiple-response					
Texture		P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
	D1	24.29	28.57	0	18.57	9.29	24.29	28.57	0	18.57	9.29
	D2	24.29	27.14	44.29	1.43	0.71	24.29	27.14	44.29	1.43	0.71
	D3	11.43	10	50	81.43	40.71	11.43	10	50	81.43	40.71
	D4	61.43	62.86	10	4.29	2.14	61.43	62.86	10	4.29	2.14
	D5	2.86	1.43	4.29	34.29	17.14	2.86	1.43	4.29	34.29	17.14
	D6	11.43	12.86	57.14	15.71	7.86	11.43	12.86	57.14	15.71	7.86
	D7	61.43	65.71	10	2.86	1.43	61.43	65.71	10	2.86	1.43
	D8	12.86	20	28.57	51.43	25.71	12.86	20	28.57	51.43	25.71
Flavor	D1	68.57	74.29	27.14	18.57		68.57	74.29	27.14	18.57	
	D2	7.14	1.43	4.29	12.86		7.14	1.43	4.29	12.86	
	D3	37.14	34.29	12.86	31.43		37.14	34.29	12.86	31.43	
	D4	32.86	51.43	4.29	10		32.86	51.43	4.29	10	
	D5	7.14	2.86	15.71	50		7.14	2.86	15.71	50	
	D6	74.29	81.43	27.14	20		74.29	81.43	27.14	20	