



Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons

Linander, Christine Borgen; Christensen, Rune Haubo Bojesen; Cleaver, Graham; Brockhoff, Per Bruun

Published in:
Food Quality and Preference

Link to article, DOI:
[10.1016/j.foodqual.2019.103864](https://doi.org/10.1016/j.foodqual.2019.103864)

Publication date:
2020

Document Version
Early version, also known as pre-print

[Link back to DTU Orbit](#)

Citation (APA):
Linander, C. B., Christensen, R. H. B., Cleaver, G., & Brockhoff, P. B. (2020). Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons. *Food Quality and Preference*, 81, Article 103864. <https://doi.org/10.1016/j.foodqual.2019.103864>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal Pre-proofs

Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons

Christine Borgen Linander, Rune Haubo Bojesen Christensen, Graham Cleaver, Per Bruun Brockhoff

PII: S0950-3293(18)30812-7

DOI: <https://doi.org/10.1016/j.foodqual.2019.103864>

Reference: FQAP 103864

To appear in: *Food Quality and Preference*

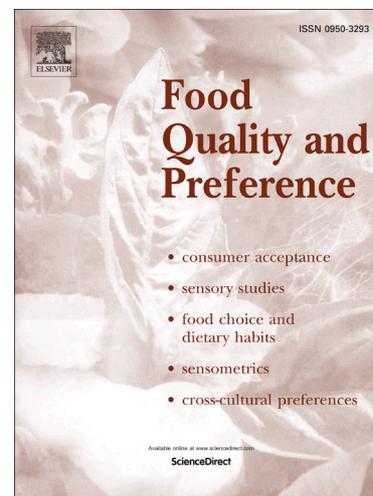
Received Date: 28 September 2018

Accepted Date: 9 December 2019

Please cite this article as: Linander, C.B., Bojesen Christensen, R.H., Cleaver, G., Brockhoff, P.B., Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons, *Food Quality and Preference* (2019), doi: <https://doi.org/10.1016/j.foodqual.2019.103864>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Ltd.



Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons

Christine Borgen Linander^{a,b,*}, Rune Haubo Bojesen Christensen^{a,c}, Graham Cleaver^d, Per Bruun Brockhoff^a

^aDTU Compute, Section of Statistics and Data Analysis, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kongens Lyngby, Denmark

^bClinical Research Center, Copenhagen University Hospital, Hvidovre, Kettegård Alle 30, DK-2650 Hvidovre, Denmark

^cChristensen Statistics, Bringetoft 7, DK-3500 Værløse, Denmark

^dUnilever Research and Development, Port Sunlight, Wirral, UK, CH63 3JW (retired)

Abstract

When considering sensory discrimination studies, multiple d-prime values are often obtained from several sensory attributes. In this paper, we introduce principal component analysis as a way of gaining information about d-prime values across sensory attributes. Specifically, we propose estimating d-prime values using a Thurstonian mixed model for binary paired comparison data and then using these estimates in a principal component analysis. Binary paired comparisons are a sensitive way to test products with only subtle differences. When analyzing data with a Thurstonian mixed model, product-specific as well as assessor-specific d-prime values are obtained. Principal component analysis of these values results in information about products and assessors across multiple sensory attributes illustrated by product and attribute maps. Furthermore, the analysis captures individual differences. Thus, by using d-prime values from a multi-attribute 2-AFC study in principal component analysis insights that are typically obtained considering quantitative descriptive analysis are obtained.

Keywords: d-prime values, discrimination testing, assessor information, multi-product setting, Principal Component Analysis

1. Introduction

One could ask the question: is it meaningful to make a multi-attribute 2-AFC study when it is possible to consider standard profiling e.g. quantitative descriptive analysis (QDA)? In short, our answer is yes. Discrimination testing, including the 2-AFC test, is sometimes preferred because it brings value in its own right or simply because specific comparisons are of interest. Furthermore, by conducting a multi-attribute 2-AFC study, it is possible to obtain the insights that are typically obtained in a QDA: product attribute maps (biplots) and individual differences. To achieve this it is necessary to use principal component analysis (PCA). PCA is a well-established type of analysis for QDA and the solid arguments for using PCA on QDA data are equivalently relevant for a multi-attribute 2-AFC study. QDA is a commonly used method in sensory science and Lawless & Heymann (1998) gives a nice introduction to this topic. If there are many evaluations, fatigue could be an issue for the 2-AFC study. However, when designing a multi-attribute 2-AFC study, this aspect can be considered by e.g. having the assessors evaluate the samples over multiple sessions or even days.

In discrimination studies, several attributes can be considered. One approach is to do an analysis for one attribute at a time (e.g. Linander et al. (2019)). For other types of sensory data, e.g. sensory profiling data, many attributes can be present. Such data are often analyzed by Principal Component Analysis (PCA) (Næs & Risvik (1996), Næs et al. (2010), Lawless & Heymann (1998)). PCA is a well-known multivariate analysis that is used in many applications e.g. in Chemometrics (Varmuza & Filzmoser (2009)) as well as sensory and consumer science.

When considering discrimination studies, an advantage of using δ , the measure for sensory differences, is that δ does

*Corresponding author. E-mail address: christine.borgen.linander@regionh.dk (C. B. Linander).

not depend on the discrimination test protocol, see e.g. Ennis (1993). Thus, it is common to obtain d-prime values, the estimates of δ , when analyzing discrimination studies. In the analysis of sensory data Luciano & Næs (2009) considered PCA using the estimates obtained from a regular two-way ANOVA.

In this paper, we will be considering principal component analysis of d-prime values obtained from sensory discrimination tests using the binary paired comparison. We will consider two approaches; one analyzing the raw d-prime values, transforming data by the inverse of the psychometric function, and one where the d-prime values are obtained from analyzing data by a model. More specifically, Linander et al. (2019) introduced a Thurstonian mixed model for binary paired comparison data. This model can be used to analyze multiple sensory attributes one at a time. From each analysis, assessor-specific as well as product-specific d-prime values are obtained, and these will be used in a principal component analysis. Thus, information about products as well as assessors are obtained across sensory attributes when considering the principal component analysis.

In a world that constantly evolves, e.g. due to health initiatives, it is beneficial for companies to be able to compare a product with versions that are subtle modifications of this product by investigating whether any of the new variants have a similar sensory profile to the known product across many attributes.

Considering discrimination tests for subtle product differences, the 2-AFC test is much simpler than the 3-AFC, and almost as sensitive. Thus, 2-AFC is the preferred test to evaluate a specific sensory attribute (Dessirier & O'Mahony, 1999; Van Hout et al., 2011; Ennis & Jesionka, 2011). For PCA to be relevant, various sensory attributes must be evaluated for a group of products. Therefore, we extend the methodology for binary paired comparisons by describing how PCA can be used to evaluate d-prime values obtained from such studies. Moreover, the 2-AFC does not use a scale and will not lead to variability of using a scale. Furthermore, the 2-AFC enables a constant reference, thus providing a rather simple way to obtain relative sensory profiles. However, there are limitations associated with the use of 2-AFC that we will also address in this paper.

The methodology presented in this paper is a step towards building a bridge between discrimination tests and descriptive analysis by considering principal component analysis using d-prime values obtained from a discrimination test. PCA makes it possible to obtain knowledge about possible correlations among sensory attributes that would not have been detected if only univariate analyses were considered. Furthermore, the biplot presenting the results from the PCA gives an easy way to gain information about similarities between products across multiple sensory attributes. Additionally, it is possible to gain knowledge about the assessors, e.g. if obvious mistakes are made it would be possible to identify these in the assessor biplot.

The paper is structured such that we introduce PCA using d-prime values in Section 2. We consider PCA using model-based product d-prime values in Section 3. In Section 4, we consider PCA using assessor specific d-prime values. Section 5 presents important aspects to consider when doing PCA using product d-prime values. We conclude the paper with a discussion in Section 6.

2. PCA using d-prime values

Principal component analysis is a well-known multivariate analysis used in many applications, including sensory science. Many books and papers exist, giving a good introduction to PCA e.g. Næs et al. (2010), Varmuza & Filzmoser (2009), Bro & Smilde (2014).

In this paper, principal component analysis is considered using d-prime values obtained from a sensory discrimination study. We consider discrimination studies using binary paired comparisons in the sense of unbounded 2-AFC tests where no correct answer exists. Furthermore, we assume that the comparisons are of the type product A vs. control, product B vs. control and so forth.

Throughout the paper, we use biplots to illustrate the results obtained from the PCA, since these show useful information in a nice way. Biplots can be used to illustrate relations among products and attributes as well as to obtain information about the individuals. Basically, a biplot shows the scores and loadings in a combined figure, under given conditions, making it possible to interpret features regarding products and attributes e.g. correlations among the attributes. The angle between two arrows can be used to get information about the correlations among the two attributes for the dimensions displayed in the biplot; an acute angle implies the attributes are positively correlated, an obtuse angle implies negatively correlated, and a right angle implies the attributes are uncorrelated. Furthermore, the length of the arrow approximates to the standard deviation of that attribute. Gabriel & Odoroff (1990) gives a detailed description of interpreting biplots, and for a thorough description of the definition of a biplot, we refer the reader to

90 Gabriel (1971).

91 Interpreting correlations is a fundamental part of understanding the phenomenon the sensory study is intended to in-
92 vestigate. Furthermore, understanding the correlations between attributes can be useful in the sense that information
93 is gained about possible combinations of a product. e.g. if two attributes are positively correlated and one would like
94 to increase one attribute while decreasing the other, a positive correlation indicates that this will be impossible.

95 One important aspect of the d-prime values is that they should be comparable such that two similar values ensure
96 similar characteristics of the corresponding products. The d-prime values we consider in this paper are comparable,
97 since the d-prime values can be both positive as well as negative, depending on the perception of the product relative
98 to the control. Thus, positive values indicate that the product has a stronger intensity of the sensory attribute than the
99 control. Similarly, negative values indicate a weaker intensity of the product.

101 2.1. *Scaling and centering*

102 In many applications, data are centered as well as scaled before doing the PCA. The scaling is usually important
103 because variables can be measured using different scales. However, when considering variables within sensory panel
104 studies, such differences in scales rarely occur. Thus, PCA is often applied without scaling the variables when consid-
105 ering experiments in sensory panel studies (Borgognone et al., 2001; Næs et al., 2010; Lawless & Heymann, 1998).
106 For the d-prime values, no scaling is used, since all of the values are on the same scale, namely the d-prime scale
107 where the d-prime values will mostly lie in the interval from -3 to 3 when negative d-prime values are allowed as in
108 the binary paired comparison. Therefore, with the same reasoning as in sensory science in general, no scaling is done
109 using d-prime values in a principal component analysis.

110 In sensory science, as well as in many other applications, data are often centered when doing a PCA. This will also
111 be the general approach in this paper. There might be situations where all (or a majority) of the d-prime values from
112 a binary paired comparison are either positive or negative. In such situations, it is important to center the data when
113 doing the principal component analysis. The impact of not centering is that all the sensory attributes will appear to be
114 positively correlated, even when this is not reflecting the correct nature of the potential correlations among the sensory
115 attributes. Thus, we recommend centering when using raw d-prime values for the PCA. With regard to the product-
116 specific d-prime values from a Thurstonian mixed model, both centered and non-centered PCA are considered. Each
117 of these makes it possible to get different types of information on the products. In Section 3.3 we will consider the
118 specific situation where an interpretation exists using non-centered product-specific d-prime values.

119 2.2. *PCA using raw product d-prime values*

120 In this section, we consider PCA using raw product d-prime values that are found by transforming the data. Let x_{ij}
121 be the number of times product i is chosen out of a total of n_{ij} answers for the j th sensory attribute. Let $p_{ij} = x_{ij}/n_{ij}$
122 be the proportion of times product i is chosen for the j th attribute. Data are transformed into d-prime values using the
123 inverse of the psychometric function for the binary paired comparison:

$$124 f_{paired}^{-1}(p_{ij}) = \Phi^{-1}(p_{ij})\sqrt{2} = d'_{ij} \quad (1)$$

125 The psychometric function in (1) is used by many authors (e.g. Linander et al. (2019) and Brockhoff & Christensen
126 (2010)) but derived in earlier work by Ennis (1993) and dating all the way back to Thurstone's pioneering work
127 (Thurstone, 1927). The d-prime values found by (1) are the so-called raw d-prime values that are used in the PCA.
128 We illustrate this approach by an example.

129 We simulate binary paired comparison data to illustrate the advantages of doing a principal component analysis using
130 raw d-prime values compared to considering univariate analyses. We impose structure among products and sensory
131 attributes in the data in order to find this structure in the results of the PCA. The comparisons considered are of the
132 type product vs. control and the simulated data are the number of times a product is chosen.

133 We consider 10 products and 6 sensory attributes for 25 assessors. We impose positive as well as negative correlations
134 among the sensory attributes. More specifically, attributes 1, 2 and 3 are positively correlated and attribute 4 and
135 attribute 5 are negatively correlated. Furthermore, attributes 4 and 6 are positively correlated.

136 We impose similarities and differences between the products. More specifically, products 1, 2 and 3 are similar and
products 7, 8 and 9 are similar but different to products 1, 2 and 3. Products 4 and 5 are completely different as

137 having low or high sensory intensity when the other product has the opposite sensory intensity. We refer the reader
138 to Appendix A for more details regarding the simulation of the data. The d-prime values for the simulated data are
139 shown in Table 1. We use these values in a principal component analysis.

141 [Table 1 about here.]

142 We consider the centered PCA which results in the biplot shown in Figure 1.

144 [Figure 1 about here.]

145 The arrows for attributes 4 and 6 are close together, implying that these two attributes are positively correlated. The
146 same holds for attributes 1 and 3. Furthermore, the arrow for attribute 2 is pointing in the same direction as the arrows
147 for attributes 1 and 3, implying that attribute 2 is also positively correlated with attributes 1 and 3. Additionally,
148 the arrow for attribute 5 is pointing towards the opposite direction of the arrows for attributes 4 and 6, implying that
149 attribute 5 is negatively correlated with attributes 4 and 6. These findings correspond to the structure we imposed on
150 the data.

151 Products 1, 2 and 3 are close together, implying similarities between these products. Similarly, products 7 and 8 are
152 also rather close, implying similarities between these two products. Furthermore, product 9 is not too far away from
153 products 7 and 8, and therefore it resembles these products. Additionally, products 4 and 5 are in opposite directions
154 with respect to the first PC, implying differences among these two products. Furthermore, products 6 and 10 are the
155 products with the strongest sensory intensity of attribute 5. However, they are rather far from one another, as well as
156 the other products, implying that they overall are different from the other products.

157 The findings regarding correlations among the sensory attributes would not have been detected if only univariate
158 analyses were considered. Furthermore, the biplot gives an easy way to gain information about similarities between
159 products across multiple sensory attributes.

160 2.3. Discrimination study

161 In this section, we describe an existing discrimination study that we use as an example in the remainder of this
162 paper. This study is the same discrimination study as is used in Linander et al. (2019). We briefly explain the structure
163 of the discrimination study and refer the reader to Linander et al. (2019) for further details.

164 The overall aim of this study was to find a new product that has some of the same sensory characteristics as an existing
165 product. In this study, the assessors were comparing different test products with the same control product. A sample
166 of a test product as well as a sample of the control product were applied to an assessor's own skin. The assessor had
167 to choose the sample with the strongest intensity of the attribute in question.

168 The study included eight test products assessed by a maximum of 25 assessors. Not all assessors evaluated all the
169 test products. The assessors evaluated eight attributes, five of these were assessed immediately after application to the
170 skin. In addition, three of these attributes were re-assessed 5 minutes after application. Each assessor evaluated each
171 test product twice by making one comparison in two consecutive sessions. Thus, the maximum number of assessments
172 for each test product is 50. The number of evaluations for the test products range from 40 to 46.

173 The desired sensory characteristics for the new product is to be less sticky and at the same time not to be greasier than
174 the control product. Furthermore, the new product should be at least as silky as the control product, though preferably
175 silkier. This will be evaluated by observing the sensory profile for the products in the biplots.

176 3. PCA using product-specific d-prime values obtained from a Thurstonian mixed model

177 In Section 2.2, the PCA was applied using a set of d-prime values found by using the inverse of the psychometric
178 function. In this section, we will model the probabilities of a product being chosen. The d-prime values obtained from
179 such a model will then be analyzed by PCA. It is not necessary to fully understand the model to be able to comprehend
180 the results of the PCA of the d-prime values. Thus, for readers without interest in how the model is defined, Section
181 3.1 can be omitted.

182 3.1. Thurstonian Mixed Model

183 In this section, we describe the Thurstonian mixed model (Linander et al., 2019) that we use to find the d-prime
184 values for the PCA. The Thurstonian mixed model analyzes data for one sensory attribute at a time. Thus, to obtain
185 the results for all sensory attributes in the data, the analysis must be conducted multiple times.

186 The Thurstonian mixed model models data obtained from binary paired comparison studies with comparisons like
187 product A vs. control, product B vs. control and so forth. Each observation is binomially distributed:

$$Y_{lmk} \sim \text{binomial}(p_{lm}, 1)$$

188 where $l = 1, \dots, L$ represents the products, $m = 1, \dots, n_l$ represents the assessors for the l th product and $k =$
189 $1, \dots, K$ represents the sessions ($K = 2$ and $L = 8$ for the discrimination study described in Section 2.3). We
190 assume that p_{lm} , the probability of the m th assessor choosing the l th product, is independent of the sessions:

$$p_{lm} = P(Y_{lmk} = 1)$$

191 It is possible to impose a linear structure of p_{lm} which explains the variables that are affecting these probabilities. We
192 consider a model where the probabilities are explained by products as well as assessors:

$$p_{lm} = f_{\text{paired}}(\mu + \alpha_l + b_m) \quad (2)$$

193 where f_{paired} is the psychometric function with its inverse given in (1). Additionally, μ is the overall average differ-
194 ence between the products and the control and α_l is the difference for the l th product to the average product-difference
195 μ . Thus, the sensory difference for the l th product to the control is

$$\delta_l = \mu + \alpha_l$$

196 Furthermore, b_m is the random effect of the m th assessor where $b_m \sim N(0, \sigma_m^2)$ which are independent for all m .
197 b_m is the difference for the m th assessor to the average product-difference μ on the d-prime scale. Thus, the sensory
198 difference, on the d-prime scale, between the products and the control for the m th assessor is $\tilde{b}_m = \mu + b_m$. We refer
199 the reader to Linander et al. (2019) for further details regarding the Thurstonian mixed model.

200 The results from the Thurstonian mixed model that we use in this paper are product-specific as well as assessor-specific
201 d-prime values for all of the sensory attributes.

202 3.2. Centering

203 3.2.1. Background

204 An important aspect of PCA is whether or not to center the data before doing PCA. For the product-specific d-
205 prime values, both situations will be considered, since each of these contributes valuable information regarding the
206 products. As we will show below, when centering the product-specific d-prime values, the information regarding the
207 control is removed. However, when the d-prime values are used without centering, the information about the control
208 is maintained in the PCA.

209 To be able to distinguish the estimates of the product-specific d-prime values obtained for the different sensory at-
210 tributes, an additional sub-script will be used:

$$d'_{lj} = \hat{\mu}_j + \hat{\alpha}_{lj}, \quad l = 1, \dots, L$$

211 where $j = 1, \dots, J$ represents the sensory attribute and $\hat{\mu}_j$ and $\hat{\alpha}_{lj}$ are the estimates obtained from the analysis of the
212 j th attribute. Thus, d'_{lj} is the sensory difference for the l th product to the control for the j th sensory attribute.

213 The model in (2) is over-parameterized, thus it is assumed that for each j :

$$\sum_{l=1}^L \hat{\alpha}_{lj} = 0 \quad (3)$$

214 When centering the product-specific d-prime values, the mean value of the d'_{ij} s for each j is subtracted. $\bar{d}'_{.j}$; the mean
 215 value over i , for a given j , reads:

$$\begin{aligned}\bar{d}'_{.j} &= \frac{1}{L} \sum_{l=1}^L d'_{lj} \\ &= \frac{1}{L} \sum_{l=1}^L (\hat{\mu}_j + \hat{\alpha}_{lj}) \\ &= \frac{1}{L} L \hat{\mu}_j + \frac{1}{L} \sum_{l=1}^L \hat{\alpha}_{lj} \\ &= \hat{\mu}_j\end{aligned}$$

216 where the last equality follows from (3). Therefore, the centered d-prime values are given as:

$$\begin{aligned}d'_{ij} - \bar{d}'_{.j} &= d'_{ij} - \hat{\mu}_j \\ &= (\hat{\mu}_j + \hat{\alpha}_{lj}) - \hat{\mu}_j \\ &= \hat{\alpha}_{lj}\end{aligned}$$

217 Thus, when doing the PCA using the centered d-prime values, the $\hat{\alpha}_{lj}$ s are used. Hence, when interpreting the results
 218 of the PCA, the information regarding the control has been removed. Recall that α_{lj} merely expresses the difference
 219 from the l th product to the average product-difference μ_j for the j th attribute. Thus, when considering the PCA on
 220 the centered d-prime values, it is possible to compare the products with each other, but not with the control.

221 3.2.2. Example

222 In this section, we consider the centered PCA using the data from the discrimination study described in Section
 223 2.3. More specifically, we consider the product-specific d-prime values which are obtained from model (2) and listed
 224 in Table 2. Note that the d-prime value for product H for `Silky` (0 minutes) equals $-\infty$. We have chosen to re-
 225 analyze data with a 0 changed to 1 for one assessor to obtain a finite value. We believe this is a sensible choice
 226 for handling the extreme value, since the lowest number of evaluations for `Silky` is 40. Thus, the imputed value
 227 expresses a rather large difference between H and the control. This approach results in an imputed value of -3.32 for
 228 product H for `Silky` (0 minutes). Several approaches exist for handling extreme values, and we will address some of
 229 these in Section 5.1.

230 [Table 2 about here.]

231 The PCA using the centered product-specific d-prime values results in the biplot shown in Figure 2.

232 [Figure 2 about here.]

233 The first principal component, PC1, is explained by `Sticky` and `Silky` initially, as well as after five minutes, for
 234 both attributes. Test products H and D are placed in opposite directions with respect to PC1; H is stickier and less
 235 silky than D. Test products A and G are positioned close to each other with respect to PC1. Thus, A and G have similar
 236 sensory properties regarding stickiness and silkiness. The second principal component, PC2, is primarily explained
 237 by `Thickness`. Test products A and G are the thickest products whereas H is the least thick product. Furthermore,
 238 test products B, F, E, C and D are similar with respect to `Thickness`. In conclusion, test products A and G are
 239 similar with respect to the sensory attributes `Thickness`, `Silky` and `Sticky` initially and after 5 minutes for the
 240 two latter attributes. Furthermore, test product H is very different from all of the other test products. Additionally, test
 241 product D is by far the most silky test product.

242 3.3. No centering

243 3.3.1. Background

244 When the d-prime values are not centered, the information about the control is maintained. The d-prime values
 245 that are used are the $\hat{\mu}_j + \hat{\alpha}_{lj}$ s. The difference $\hat{\mu}_j + \hat{\alpha}_{lj}$ is the estimated sensory difference between the l th product and
 246 the control for the j th sensory attribute. Therefore, all the d-prime values used for the PCA are differences between the
 247 products and the control. Thus, the origin in the biplot corresponds to the control. Hence, when differences between
 248 products and the control are of interest, the non-centered PCA will provide insight about these.

249 Before doing a non-centered PCA, it is important to consider whether the set of d-prime values is suitable for this
 250 type of analysis. As written in Section 2.1, a non-centered PCA only works if the d-prime values are positive as well
 251 as negative. Otherwise, this analysis risks introducing correlations that are not necessarily there. It can be difficult
 252 to determine whether the d-prime values are scattered such that enough positive and negative values occur. For now,
 253 we do not have any specific recommendations, only that one must look at the signs of the d-prime values to decide
 254 whether it seems reasonable to do the non-centered PCA. Thus, future research could investigate whether it is possible
 255 to determine more specific guidelines regarding the distribution of the d-prime values.

256 3.3.2. Example

257 In this section, we consider the non-centered PCA using the data from the discrimination study described in
 258 Section 2.3. As in Section 3.2.2 we use the imputed value of -3.32 instead of the extreme value. Considering the
 259 d-prime values in Table 2, a number of positive as well as negative values are occurring. Thus, we believe that it is
 260 sensible to consider the non-centered PCA. The biplot for the PCA using the non-centered product-specific d-prime
 261 values is shown in Figure 3.

262 [Figure 3 about here.]

263 The arrows for *Thickness* and *Absorption* are short. That, in combination with many small d-prime values,
 264 makes it extremely difficult to conclude anything regarding *Thickness* and *Absorption* from Figure 3 (the sign
 265 of the values will easily change, since other attributes will affect the values of the scores more heavily). Considering
 266 *Greasy* (initially and after 5 minutes), all of the test products are less greasy than the control, since they are placed
 267 opposite the direction of the arrows for *Greasy*.

268 Generally, when interpreting the biplot from a non-centered PCA, we use the perpendicular line for each attribute; the
 269 straight line going through the origin that is orthogonal to the arrow for the attribute in question. For simplicity, we
 270 have only included the perpendicular line for one attribute (*Silky*) as an illustration. However, it is possible to obtain
 271 the perpendicular lines for all attributes using the dot product. The perpendicular line is used to determine the relation
 272 between the product and the control for that attribute. Thus, for a product to have similar intensity as the control, for a
 273 given sensory attribute, the product should be positioned closely to the imaginary/perpendicular line. For a product to
 274 have a stronger sensory intensity of the attribute in question, it should be closer to the tip of the arrow than the control.
 275 Similarly, for a product to have a weaker sensory intensity than the control, it should be further away from the tip of
 276 the arrow than the control.

277 The distance to the perpendicular line for the product to have similar intensity depends on the specific objective of the
 278 study. Thus, it is up to each sensory scientist to evaluate whether a product is close enough to the perpendicular line
 279 for that product and the control to have similar sensory intensities.

280 With respect to silkiness, the majority of the test products are less silky than the control. Furthermore, it appears that
 281 test product C is just about as silky (initially) as the control, and test product D is the only test product which clearly
 282 is silkier than the control. Considering stickiness, test products G and H are stickier (initially) than the control, and A
 283 is more or less as sticky (initially) as the control. It appears that the remaining test products are less sticky (initially)
 284 than the control. Furthermore, it appears that the only test product which is stickier than the control after 5 minutes is
 285 test product H.

286 It is possible, from Figure 3, to identify the most interesting test products based on the most important sensory prop-
 287 erties. These properties are pre-specified by the sensory scientist and they are not chosen based on the results from the
 288 PCA. Recall that a test product is interesting when it is as silky as the control as well as being less greasy and sticky.
 289 Thus, the most interesting test products are C and D with D being the more promising of the two since it is silkier
 290 than C. It should be noted that the desired characteristics are decided by the sensory scientist (e.g. when designing the

study). Thus, for other studies other sensory attributes might be the most important.

4. PCA using assessor-specific d-prime values from a Thurstonian mixed model

Individual differences are important and relevant to investigate in relation to sensory studies, and for QDA data many methods exist enabling this information. In this section, we show how information about assessors can be obtained considering PCA using assessor-specific d-prime values from a multi-attribute 2-AFC study.

To gain knowledge about which assessors are proportionally choosing products similarly on average across products, the $b_{m,j}$ s are considered for each attribute. To be able to distinguish the estimates obtained for the different attributes, an additional sub-script will be used. Thus, $b_{m,j}$ is the difference for the m th assessor to the average product-difference μ_j , on the d-prime scale, for the j th attribute. The centered and non-centered PCA will give similar results, since $E(b_{m,j}) \approx 0$, thus the assessor-specific d-prime values $b_{m,j}$ s are almost centered. Considering the $b_{m,j}$ s, it is possible to investigate how the assessors proportionally choose products compared to the average.

4.1. Example

In this section, we consider PCA using the data from the discrimination study described in Section 2.3. More specifically, we consider the assessor-specific d-prime values obtained from model (2). When testing skin care products, it is important to capture individual subject differences, since people have different skin types which respond differently to such products.

[Figure 4 about here.]

The biplot for the assessor-specific d-prime values is shown in Figure 4. The first principal component is explained by *Silky*, mostly evaluated after five minutes, but also initially. The assessors 7 and 8 are proportionally choosing products the most with respect to silkiness. Assessor 19 is the assessor proportionally choosing the control the most regarding *Silky*. The second principal component is mostly explained by *Greasy* (initially and after five minutes). Assessor 23 is the assessor proportionally choosing products the most with respect to *Greasy*. There are assessors who are proportionally choosing products similarly on average across products. The assessors 12 and 15 are close in the biplot. Assessors 22, 18 and 11 are also close together. Furthermore, it appears that the assessors in the lower left quadrant have a tendency to proportionally choose products more than the average assessor across the attributes, since all of the arrows points toward left and/or down.

In our opinion, Figure 4 can be used to look for patterns of the assessors' way of selecting products which might be missed otherwise. However, this is not the same as being able to interpret the quality of an assessor. We believe that to be able to interpret the quality of the assessors, prior knowledge about the 'correct' product differences must be available.

When the arrow is short, it was easier for the assessors to evaluate that attribute compared to the other attributes. However, it is not possible, from the biplot, to explain whether this is a consequence of the attribute being easier to assess or whether the products were more differentiable with respect to this attribute.

The correlations among the attributes can be used to look for obvious misunderstandings among the assessors in the evaluation of the attributes, e.g. if two attributes that are positively correlated are shown to be negatively correlated, this could indicate that more training of the panel is needed. Furthermore, if some correlations appear to be counter-intuitive, this could be a result of the assessors misunderstanding how to evaluate some of the attributes or simply because they find evaluating these attributes difficult.

The center of the biplot corresponds to an average assessor. The further away from the center an assessor is positioned, the more different assessments from that assessor are compared to an average assessor. If all the assessors are close to the center, the assessors are assessing similarly. The more the assessors are spread out, the more diverse the assessors are assessing across products. Thus, when interpreting the biplot of the assessor-specific d-prime values, information is obtained about how variable the assessors were in the study. Therefore, insights are achieved into how the product averages are obtained. Often, the aim of a panel is to include a diverse group of assessors, and in such situations, it is to be expected that the assessors will be variable. However, if the panel consists of similar assessors, e.g. assessors with the same skin type, then it is likely that the assessors will be less variable. Thus, whether the variability of

338 assessors could be an issue depends on the given objective of a study.
 339 The interpretation of Figure 4 gives the idea of interpreting such a biplot. However, in this specific discrimination
 340 study we have a limitation of not all assessors evaluating all products. Thus, some of the differences seen in Figure 4
 341 might be due to this design issue. Thus, when considering the biplot for the assessors, preferably all assessors should
 342 evaluate all products.

343 5. Considerations regarding PCA using product d-prime values

344 In this section, we describe important aspects to consider when analyzing product d-prime values with PCA.
 345

346 5.1. Handling extreme values

347 As was occurring in Section 3.2.2, there is a likelihood of observing d-prime values equal to $\pm\infty$. In this section,
 348 we suggest approaches for handling such extreme values. We consider PCA using the so-called raw d-prime values
 349 as well as model-based d-prime values. Some approaches handling the extreme values are applicable for both types
 350 of d-prime values. Other solutions only work for one or the other type. An approach which only works for the raw
 351 d-prime values, is to use the proportion $1/n$ rather than $0/n$ (or $(n-1)/n$ rather than n/n) when finding the d-prime
 352 values. Changing the proportion ensures finite d-prime values. However, this raises the question of how large n should
 353 be before this change of proportion is meaningful. For small n , this change will imply a completely different result,
 354 e.g. if $n = 2$ then $x = 0$ results in a d-prime value of minus infinity, whereas $x = 1$ results in a d-prime value of 0.
 355 For $n = 50$, the proportion $1/n$ corresponds to a d-prime value of -2.9 , which is still expressing a large difference
 356 between the products. For this approach to be reasonable, a rather large n is needed and a topic for future research
 357 is to investigate when n is 'large enough'. An equivalent approach for the model-based d-prime values is to change
 358 one data point from 0 to 1 and then re-analyze data by the Thurstonian mixed model to find the value to replace the
 359 extreme value. Using this approach, it is important to only extract the imputed value from the new analysis.
 360 Another approach is to use a different proportion than $1/n$ when imputing a value for the raw d-prime values. It might
 361 be possible to weight the data in a better way compared to using $1/n$ considering the raw d-prime values. Thus, an
 362 issue for future research is to investigate whether it is possible to claim that, in general, one weight is better than
 363 another.

364 Other approaches valid for both types of d-prime values, are to either remove the sensory attribute(s) with extreme
 365 values or to remove the product(s) with extreme values. Obviously, these approaches only work if not too many
 366 different attributes or products have extreme values. However, if there are many extreme values, one could argue
 367 whether the PCA is meaningful to consider at all.

368 Yet another approach is to choose an arbitrary value that expresses a large difference between the products, and use that
 369 instead of the extreme value. Our preliminary findings, considering the data from the discrimination study described
 370 in Section 2.3, indicate that with only one extreme value the results of the PCA seem to be affected as expected when
 371 changing the size of the imputed value; the more extreme the imputed value, the more extreme the allocation of that
 372 product for the attribute the value is imputed for becomes, without affecting the interpretation regarding the remaining
 373 products as well as attributes.

374 It is important to carefully consider how to handle the extreme values and future research could investigate in more
 375 detail how different approaches affect the results of the PCA.

376 5.2. Product Selection

377 As described in Section 2.1, it is important that the d-prime values used for PCA are positive as well as negative.
 378 When designing a discrimination study, it is possible to minimize the risk of getting clustered d-prime values in the
 379 way the products are selected. In the case of an extreme control, the likelihood of getting d-prime values with the
 380 same sign increases compared to having a non-extreme control.

381 Having d-prime values with the same sign particularly becomes an issue when considering the non-centered PCA in
 382 Section 3.3. This approach is only valid when a few of the attributes have d-prime values with the same sign.

383 Thus, considering binary paired comparisons to be used in PCA, preferably a non-extreme control should be used.
 384 Currently, we do not have a definition of when a control is extreme vs. non-extreme. However, it would be relevant

385 to investigate whether it is possible to determine when a control is non-extreme. Furthermore, it would be interesting
386 to investigate how the results from the centered PCA are affected by an increasing number of attributes with values
387 of the same sign. Additionally, it would be interesting to investigate whether it is possible to find a cut-off value for
388 the percentage of attributes with an extreme control where the results are affected such that they no longer reflect the
389 correct nature of data.

390 6. Summary and discussion

391 In this paper we have presented a first step towards building a bridge between descriptive analysis and discrim-
392 ination testing. More specifically, we have suggested a way to gain knowledge about products and assessors across
393 sensory attributes by considering d-prime values obtained from a binary paired comparison, and we have used these
394 in a principal component analysis. As written in Section 1, insights about products and individual differences are
395 obtained when considering a multi-attribute 2-AFC study. Furthermore, the binary paired comparison is preferable
396 when considering subtle product differences due to its simplicity as well as the lack of a scale in this discrimination
397 test. Additionally, the 2-AFC test provides a constant reference giving relative sensory profiles. However, to the best
398 of our knowledge, it has not been documented in the literature whether increased sensitivity will be obtained in a
399 multi-attribute 2-AFC study. It is beyond the scope of this paper to investigate this, but it would be interesting to
400 compare the sensitivity for QDA with the sensitivity for a multi-attribute 2-AFC study as a part of future research on
401 this topic.

402 When considering PCA using d-prime values, various aspects are important to consider carefully in order to ensure a
403 valid interpretation of the results. We have addressed the importance of centering the d-prime values to ensure that we
404 do not force incorrect correlations between the sensory attributes. There is one exception, where it is valid to consider
405 the non-centered PCA; using the product-specific d-prime values from a Thurstonian mixed model given that both
406 positive and negative d-prime values are present in the data.

407 An interesting continuation is to investigate whether it is possible to determine guidelines for a distribution of positive
408 and negative d-prime values.

409 Another important aspect to consider is how to handle extreme values. We have suggested different ways of dealing
410 with extreme values, although more work is needed to figure out in more detail how different approaches affect the
411 results of the PCA.

412 Additionally, we have dealt with how to choose the product used as the control in the evaluations; it is important to
413 choose a non-extreme control. Currently, we do not have a definition of when a control is non-extreme, and it would
414 be relevant to investigate the possibility of making such a definition. An interesting continuation of the work presented
415 in this paper is to consider other types of paired comparisons such as product A vs product B, product B vs product C
416 and so forth. Furthermore, it would be interesting to consider other ways of obtaining the d-prime values than consid-
417 ering binary paired comparisons. As mentioned in Section 2, it is important that the d-prime values are comparable.
418 Thus, it is important to ensure that this is fulfilled when investigating whether it is possible to use d-prime values that
419 are not obtained from a binary paired comparison. However, further research is needed to comprehend the impact of
420 how the d-prime values are obtained.

421 In order to facilitate easy choices regarding similarities between products and control, it would be relevant to investi-
422 gate whether a d-prime interpretation of distances exists in the biplot.

423 Future research regarding the biplot for the assessor d-prime values could be to investigate patterns if additional in-
424 formation is available for the sensory scientist, e.g. whether the assessors are clustered according to their skin type.
425 Furthermore, it would be relevant to consider whether it is generally possible to identify good and bad assessors.

426 As this paper is aimed at being a first step in building a bridge between descriptive analysis and discrimination testing,
427 it would be relevant to continue building this bridge in the future. One way to continue this work is by comparing the
428 performance of principal component analysis of data from descriptive analysis studies with the performance of PCA
429 using d-prime values from the binary paired comparison.

430 Acknowledgments

431 The research that led to this paper is funded by the Technical University of Denmark and Unilever U.K. Central
432 Resources Limited. Unilever also provided the data used as an example in sections 3 and 4. Furthermore, the authors

433 would like to thank the editor as well as two anonymous reviewers for comments that led to a substantial improvement
434 of this paper. Additionally, the first author would like to thank Thomas Kalleose for many rewarding discussions.

435 **Appendix A. Details regarding the example in Section 2.2**

436 In this appendix, we describe in more detail how we simulated the data used in Section 2.2.
437 The comparisons considered are all of the type product vs. control in a binary paired comparison. Thus, the simulated
438 data values are the number of times a product is chosen for each comparison of a product with the control.
439 The number of times product i is chosen for attribute j is binomially distributed:

$$Y_{ij} \sim \text{binomial}(n_{ij}, p_{ij})$$

440 where n_{ij} is the total number of evaluations and p_{ij} is the probability that the i th product is chosen for the j th
441 sensory attribute. Since we are simulating binomial data, we impose the structure among products and attributes on
442 the probabilities for a product being chosen. We consider 10 products and 6 sensory attributes for 25 assessors. Thus,
443 $i = 10$ and $j = 6$.

444 The attributes 1, 2 and 3 are made positively correlated by having the same probabilities for all products:

$$p_{i1} = p_{i2} = p_{i3} \quad i = 1, \dots, 10$$

445 Attribute 4 and attribute 5 are negatively correlated, one having high probabilities where the other has low probabili-
ties:

$$p_{i5} = 1 - p_{i4} \quad i = 1, \dots, 10$$

446 Additionally, attributes 4 and 6 are positively correlated:

$$p_{i4} = p_{i6} \quad i = 1, \dots, 10$$

447 Products 1, 2 and 3 are made similar as having the same probabilities:

$$p_{1j} = p_{2j} = p_{3j} \quad j = 1, \dots, 6$$

448 Similarly, products 7, 8 and 9 are similar but with a different probability than products 1, 2 and 3:

$$p_{7j} = p_{8j} = p_{9j} \neq p_{1j} \quad j = 1, \dots, 6$$

449 Products 4 and 5 are made opposite in the sense where one has high probabilities the other has low probabilities:

$$p_{5j} = 1 - p_{4j} \quad j = 1, \dots, 6$$

450 The probabilities that we used to simulate the data for the example in Section 2.2 are shown in Table 3.
451

452 [Table 3 about here.]

453 The simulated data, shown in Table 4, are transformed into d-prime values using:

$$f_{paired}^{-1}(p_{ij}) = \Phi^{-1}(p_{ij})\sqrt{2} = d'_{ij}$$

454 where the p_{ij} s are the proportions of the data corresponding to each entry of Table 4 being divided by 25.
455

[Table 4 about here.]

456 **References**

- 457 Borgognone, M. G., Bussi, J., & Hough, G. (2001). Principal component analysis in sensory analysis: covariance or correlation matrix? *Food*
458 *Quality and Preference*, *12*, 323–326.
- 459 Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*, 2812–2831.
- 460 Brockhoff, P. B., & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality*
461 *and Preference*, *21*, 330–338.
- 462 Dessirier, J.-M., & O'Mahony, M. (1999). Comparison of d' values for the 2-afc (paired comparison) and 3-afc discrimination methods: Thurstonian
463 models, sequential sensitivity analysis and power. *Food Quality and Preference*, *10*, 51 – 58.
- 464 Ennis, D. M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, *8*, 353–370.
- 465 Ennis, J. M., & Jesionka, V. (2011). The power of sensory discrimination methods revisited. *Journal of Sensory Studies*, *26*, 371–382.
- 466 Gabriel, K. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*, 453–467.
- 467 Gabriel, K., & Odoroff, C. L. (1990). Biplots in biomedical research. *Statistics in medicine*, *9*, 469–485.
- 468 Lawless, H. T., & Heymann, H. (1998). *Sensory Evaluation of Food*. Springer.
- 469 Linander, C. B., Christensen, R. H. B., Cleaver, G., & Brockhoff, P. B. (2019). Individual differences in replicated multi-product experiments with
470 thurstonian mixed models for binary paired comparison data. *Food Quality and Preference*, *75*, 220–229.
- 471 Luciano, G., & Næs, T. (2009). Interpreting sensory data by combining principal component analysis and analysis of variance. *Food Quality and*
472 *Preference*, *20*, 167 – 175.
- 473 Næs, T., Brockhoff, P. B., & Tomic, O. (2010). *Statistics for Sensory and Consumer Science*. Wiley.
- 474 Næs, T., & Risvik, E. (1996). *Multivariate analysis of data in sensory science*. Elsevier.
- 475 Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.
- 476 Van Hout, D., Hautus, M. J., & Lee, H.-S. (2011). Investigation of test performance over repeated sessions using signal detection theory: comparison
477 of three nonattribute-specified difference tests 2-afcr, a-not a and 2-afc. *Journal of Sensory Studies*, *26*, 311 – 321.
- 478 Varmuza, K., & Filzmoser, P. (2009). *Multivariate Statistical Analysis in Chemometrics*. CRC Press. Taylor & Francis Group.

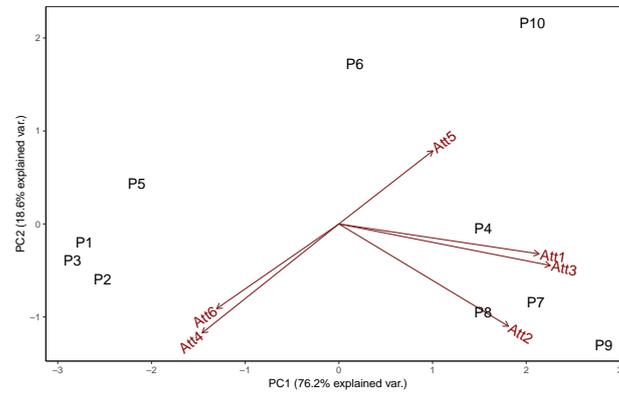


Figure 1: The biplot for the centered d-prime values from the simulated data.

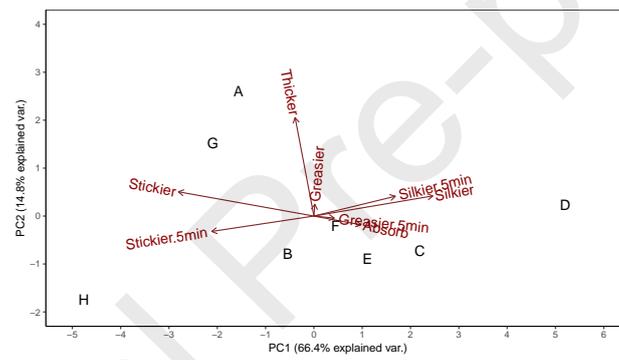


Figure 2: The biplot for the centered product specific d-prime values.

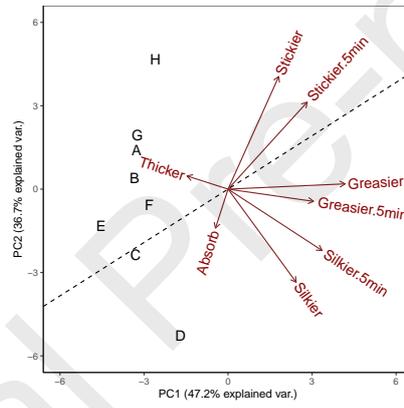


Figure 3: The biplot for the non-centered product-specific d-prime values. The dashed line is perpendicular to the arrow associated with Silkier.

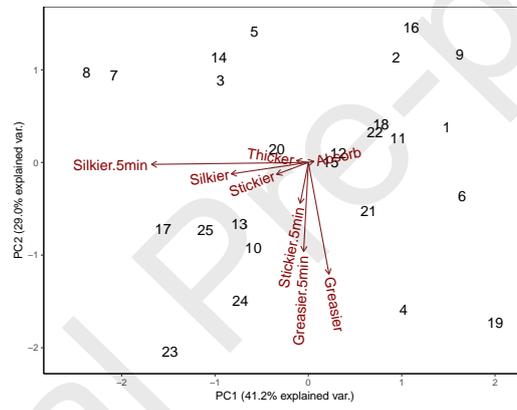


Figure 4: The biplot for the centered assessor specific d-prime values b_j .

Table 1: Simulated d-prime values for the binary paired comparison.

Product	Attribute					
	1	2	3	4	5	6
1	-1.00	-0.82	-0.82	1.41	-0.07	0.66
2	-1.00	-0.51	-0.51	0.66	-1.19	1.00
3	-1.41	-0.36	-1.00	1.19	-0.51	0.66
4	0.82	1.19	1.41	-0.82	0.51	-0.51
5	-0.51	-0.51	-1.41	0.21	-0.36	0.21
6	-0.07	-0.21	0.07	-1.19	0.51	-1.41
7	1.66	1.41	1.99	-0.36	0.07	-0.51
8	1.00	1.99	1.19	-0.21	0.07	-0.51
9	1.99	2.48	1.99	-0.21	0.51	-0.66
10	1.00	0.07	1.19	-1.99	1.66	-1.66

Table 2: Product specific d-prime values found by using the model given in (2).

Test Product	Sticky		Greasy		Silky		Thickness	Absorption
	0 min	5 min	0 min	5 min	0 min	5 min	0 min	0 min
A	0.64	-1.36	-1.19	-1.76	-1.15	-1.57	2.94	-1.35
B	-1.44	0.08	-1.98	-0.51	-1.47	-2.39	0.58	1.50
C	-2.60	-2.15	-1.79	-1.22	-0.03	-0.89	-0.13	0.46
D	-3.42	-3.47	-0.67	-0.65	2.28	0.83	-0.01	0.88
E	-1.93	-2.22	-2.74	-1.80	-1.02	-1.47	-0.04	0.76
F	-0.40	-1.88	-2.21	-1.19	-0.64	-0.37	-0.09	-0.47
G	0.70	-0.01	-1.81	-1.21	-1.89	-1.77	2.43	0.99
H	2.06	0.88	-1.04	-1.82	-Inf	-3.03	-1.09	-1.58

Table 3: Probabilities used when simulating the binary paired comparison data.

Product	Attribute					
	1	2	3	4	5	6
1	0.30	0.30	0.30	0.70	0.30	0.70
2	0.30	0.30	0.30	0.70	0.30	0.70
3	0.30	0.30	0.30	0.70	0.30	0.70
4	0.80	0.80	0.80	0.30	0.70	0.30
5	0.20	0.20	0.20	0.70	0.30	0.70
6	0.50	0.50	0.50	0.20	0.80	0.20
7	0.90	0.90	0.90	0.40	0.60	0.40
8	0.90	0.90	0.90	0.40	0.60	0.40
9	0.90	0.90	0.90	0.40	0.60	0.40
10	0.70	0.70	0.70	0.10	0.90	0.10

Table 4: The simulated number of times a product was chosen.

Product	Attribute					
	1	2	3	4	5	6
1	6	7	7	21	12	17
2	6	9	9	17	5	19
3	4	10	6	20	9	17
4	18	20	21	7	16	9
5	9	9	4	14	10	14
6	12	11	13	5	16	4
7	22	21	23	10	13	9
8	19	23	20	11	13	9
9	23	24	23	11	16	8
10	19	13	20	2	22	3