Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis

J. C. Castura[1*], P. Varela[2], T. Næs[2]

[1]Compusense Inc., 255 Speedvale Av. W., Guelph, Ontario, N1H 1C5, Canada

[2]Nofima AS, Osloveien 1, P.O. Box 210, N-1431 Ås, Norway

*Contact (email): jcastura@compusense.com (J.C. Castura)

Highlights

- Paired comparisons investigated after PCA
- Visual comparison of 95% confidence ellipsoids and 95% density contours
- P values and ellipsoid volumes investigate uncertainty numerically
- P values allow for rapid screening of paired difference results
- Coverage properties of confidence ellipsoids investigated via simulations

Abstract

We propose and evaluate numerical and visual methods for investigating paired comparisons after principal component analysis (PCA). PCA results can be visualized to facilitate an understanding of the relationships between the products and the sensory attributes. But identifying and visualizing significant product differences in multiple PCs simultaneously is not straightforward. A benefit of the proposed methods is that they provide a screening tool for evaluating PCA results rapidly. We begin with a real data set which is analyzed and submitted to the truncated total bootstrap (TTB) procedure. This TTB procedure simulates and analyzes results from virtual panels. The TTB-derived results form clouds of uncertainty around each product and paired comparison. Although these clouds can be visualized directly or by plotting the smallest contours that enclose 95% of their kernel-estimated densities, we propose that plotting TTB-derived 95% confidence ellipsoids provide a less cumbersome approach. We show that it is also possible to calculate P values that evaluate whether pairs of products are discriminated in the PCA subspace. The interpretation of these P values coincides with the visual interpretation of the confidence ellipsoids. The volumes of these confidence ellipsoids, which quantify uncertainty, are calculated easily. The confidence ellipsoids, the P values, and the volumes provide a simple and consistent approach for investigating paired comparisons after PCA. We illustrate the

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

methods with two real data sets, one a sensory quantitative-descriptive sensory data set from a trained panel, the other a consumer check-all-that-apply (CATA) data set. We also conduct a simulation study based on each of these data sets. The results from these simulation studies show that under repetition, the 95% confidence ellipsoids often have coverage of approximately 95%, but in some cases, coverage can be substantially lower. This indicates that the proposed ellipsoids have an approximately frequentist interpretation, but coverage varies. The complementary numerical and visual approaches can be applied to a wide range of data sets from sensory evaluation and to data from other domains.

Keywords: principal component analysis (PCA); paired comparison; Procrustes rotation; sensory evaluation; sensory profiling; bootstrap.

1. Introduction

Sensory evaluation methods are used to describe and quantify perceptions that arise when foods and beverages are consumed (Lawless & Heymann, 2010). But sensory evaluation results are sometimes challenging to interpret because the perceptions are measured on multiple correlated sensory attributes. Multivariate analyses can reveal relationships between products and attributes that are often missed if the data are only analyzed univariately. In sensory evaluation, one multivariate analysis that is often applied is principal component analysis (PCA; Mardia, Bibby & Kent, 1979).

PCA transforms data from the original, correlated variables (sensory attributes) onto mutually orthogonal principal components (PCs), which are the main directions of variability. The first PC optimally extracts variance from the results matrix; every subsequent PC optimally extracts variance from the residual variance. Coefficients in the loading matrix give the linear combinations of original variables that define each PC. The score matrix gives the product coordinates in the PCs. Dimension reduction is usually possible because, in many cases, most of the variance from the original data set is compressed into the first PCs. In this case, the retained PCs can be interpreted as “signal” and truncated PCs as “noise”. Retained PCs are often visualized two at a time, in score plots and loading plots, or together in biplots. These PCA plots facilitate understanding and communication because they often visually summarize the most important relationships between the products and sensory attributes (Lawless & Heymann, 2010).

Bootstrap-based procedures are often used to investigate uncertainty in PCA solutions (Babamoradi, van den Berg & Rinnan, 2013; Lebart, 2007; Kiers & Groenen, 2006; Husson, Lê & Pagès, 2005). These procedures are based on repeatedly simulating so-called virtual panels from the real data set, then combining the repeated results to construct the confidence regions. Since these confidence regions are based on bootstrap-derived results, they are approximations, but useful.

PCA plots are often used for comparing products within the sensory space determined by the whole set of products. But it is not straightforward to visualize significant product differences in multiple PCs simultaneously. The reason is that each product’s confidence region is constructed using results from the same virtual panels, so these mutual dependencies need to be taken into account when investigating paired comparisons. Castura, Rutledge, Ross and Næs (2022a) investigated paired

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

comparisons using a special type of bootstrap called the truncated total bootstrap (TTB; Cadoret and Husson, 2013; Courcoux et al., 2012), but only in one PC at a time. Lê and Husson (2008) submitted TTB results to a two-product Hotelling T^2 test (Mardia et al., 1979; Hotelling, 1931) to evaluate product differences numerically in a plane of two PCs, but without accounting for mutual dependencies in the data. Castura, Varela and Næs (2023) make theoretical contributions to demonstrate that both products and their paired comparisons are optimally investigated in the same PCs. They also show that calculating paired differences between TTB-derived scores gives the same result as superimposing virtual-panel paired differences on real-panel paired differences. They visualize uncertainty regions by showing contours of the kernel-estimated densities from the TTB-derived paired difference scores.

The objective of this paper is to propose and evaluate numerical and visual methods for investigating paired comparisons after PCA by building on these previous approaches. First, we will propose a way to construct confidence ellipsoids using the TTB method as an alternative to the more cumbersome kernel approach. Next, we will investigate whether these confidence ellipsoids are appropriate for assessing uncertainty of paired difference scores as compared to the nonparametric approach. Finally, we will use simulations to evaluate whether we can be approximately 95% confident that the true score values will be covered by a TTB-derived 95% confidence ellipsoid. Results from the simulation studies will inform how we interpret these confidence ellipsoids.

A researcher who interprets their results by visually inspecting the confidence regions needs to have sufficient expertise and enough time to review all of the plots. In this paper, we propose to aid and accelerate such a review process by providing numerical results that include a P value for each paired comparison. The interpretations based on these P values will coincide precisely with interpretations from visual inspection of the confidence ellipsoids. It is faster and easier to review numerical P values than it is to review many plots visually. P values, even if they are approximate, can be used to screen results to draw attention to paired differences that might otherwise go unnoticed.

An outline of the rest of the manuscript follows. Section 2 contains information on the two data sets that are explored in this paper. Section 3 contains theory and calculations. In Section 3.1, we give an overview of PCA and the TTB procedure, then describe the method for constructing confidence ellipsoids, how we calculate their volumes, and how we evaluate paired comparisons. In Section 3.2, we describe the simulation studies that are used to investigate the statistical properties of the confidence ellipsoids. In Section 4, results from the two data sets and from the simulation studies are presented and interpreted. Results are discussed in Section 5 and conclusions follow.

2. Material and methods

This section describes two data sets that will be analyzed in this manuscript. The first data set is from a trained panel that evaluated seven beverages using intensity scales. The second data set is from a consumer panel that evaluated six strawberry cultivars using a check-all-that-apply (CATA; Ares & Jaeger, 2015; Meyners & Castura, 2014) question and a list of 16 sensory attributes.

2.1. Beverages data set – At Nofima AS, a panel of 10 assessors were trained to evaluate the intensities of 23 sensory attributes using 10-cm line scales. This panel then evaluated seven beverage products

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

(Cow's Milk and six plant-based beverages, each identified by its main ingredient: Almond, Coconut, Oat, Peas, Rice, Soya) in duplicate according to an experimental design that balanced presentation order and carryover effects. Since we chose to center and variance-standardize variables before PCA, we dropped variables where product differences might have arisen by chance so that these variables did not influence the solution (Næs, Tomic, Endrizzi & Varela, 2021). Two-way analysis of variance with factors product and assessor was conducted for each of the 23 attributes. The panel failed to discriminate products with 95% confidence for three attributes (*salty*, *acidic*, *fullness*), which were dropped. Names of the 20 retained sensory attributes follow [with abbreviations used in figures]. There were eight odour/aroma attributes (*sourness* [oS], *sweetness* [oSw], *milky* [oM], *vanilla* [oV], *grainy* [oG], *coconut* [oC], *oxidized* [oO], *rancid* [oR]), five taste/mouthfeel attributes (*sour* [tS], *sweet* [tSw], *bitter* [tB], *astringent* [mA], *smoothness* [mS]), and seven flavour attributes (*milky* [fM], *vanilla* [fV], *grainy* [fG], *coconut* [fC], *cloying* [fCl], *oxidized* [fO], *rancid* [fR]). For further details on this data set, see Castura et al. (2023).

2.2. Strawberry data set – The samples from six strawberry cultivars—Festival (FE), Yvahé (YV), Yurí (YR), Guenoa (GU), L20.1 (L2), and K31.5 (K3)—were presented to 114 consumers in sequential monadic presentation format according to an experimental design that balanced presentation order and carryover effects. Consumers described each sample by responding to a CATA question with 16 sensory attributes [with abbreviations used in figures]: *sweet* [Sw]*, sour* [S]*, strawberry flavor* [fStr]*, strawberry odor* [oStr]*, flavorsome* [Flv]*, tasteless* [Tl]*, red color* [RC]*, irregular shape* [Is]*, regular shape* [Rs]*, small* [sS]*, big* [sB]*, firm* [txF]*, hard* [txH]*, soft* [txS]*, juicy* [J]*, dry* [D]. No variables were dropped before PCA because, for CATA citation rates, variables are centered but not variance-standardized before PCA (Meyners, Castura & Carr, 2013). For further details on this data set, see Ares and Jaeger (2013). This data set can be found in the R package `ClustBlock` (Llobell, Vigneau, Cariou & Qannari, 2020). It has been analyzed previously (Meyners & Hasted, 2023; Meyners & Hasted, 2022; Meyners & Hasted, 2021; Bi & Kuesten, 2023; Bi & Kuesten, 2021; Llobell et al., 2019; Meyners & Castura, 2014). The results presented herein focus on a cluster of 54 consumers (called *g2*) obtained from b-cluster analysis (Castura, Meyners, Varela & Næs, 2022b), which can be conducted using functions in the R package `cata` (Castura, 2022). Results from the *g2* consumers were also analyzed by Castura et al. (2023). For ease of reference, we will call these *g2* consumers "the consumers" and will call their data the "strawberry data".

3. Theory and calculations

In Section 3.1, we describe how we explore the real-panel data set, including how we investigate uncertainty and paired comparisons using the TTB method. In Section 3.2, we describe simulation studies that are used to understand the properties of the proposed confidence ellipsoids. In Section 3.3, we state the statistical software used to perform these methods.

3.1. Analysis of data from a real panel and investigation by the TTB method

Section 3.1.1 provides an overview of PCA of a results matrix (**X**), such as sensory results organized into a products-by-attributes matrix. Section 3.1.2 discusses paired comparisons after PCA. Section 3.1.3 gives an overview of the TTB procedure, a more detailed treatment of which is found in Castura et al. (2023).

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

The proposed methods for investigating uncertainty and paired comparisons are described in the subsections that follow (Sections 3.1.4 to 3.1.7).

3.1.1. Overview of PCA

Singular value decomposition (SVD; Mardia et al., 1979) decomposes **X**, a column-centered (*J*×*M*) matrix with rank *R*, into three matrices:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{P}^{\mathrm{T}} \quad (1)$$

Columns of the (*J*×*R*) matrix **U** contain the left singular vectors, where $\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}_R$. The singular values are on the main diagonal of the diagonal matrix **D**; their squares are the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Columns of the (*M*×R) matrix **P** contain the right singular vectors, which are identical to the eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, where $\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_R$.

Multiplication of **U** and **D** gives the score matrix **T** where

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \quad (2)$$

Dimension reduction is usually possible. In this paper, we retain PCs that collectively extract at least 80% of the variance from **X**. If the solution is truncated to retain only the first *A* PCs, then (2) can be written

$$\mathbf{X} = \mathbf{T}_A\mathbf{P}_A^{\mathrm{T}} + \mathbf{E} \quad (3)$$

where the (*J*×*A*) matrix $\mathbf{T}_A$ and the (*M*×*A*) matrix $\mathbf{P}_A$ are the truncated score and loading matrices, respectively, and **E** is the (*J*×*M*) residual matrix.

3.1.2. Paired comparisons in PCA

The same PCs optimally extract variance from both **X** and from a matrix of all its paired comparisons (Castura et al., 2023). So, it is not necessary to conduct a PCA of all paired comparisons, since we would obtain the same PCs. The difference between the scores in one row ($\mathbf{t}_1$) and another row ($\mathbf{t}_2$) in *A* PCs can be obtained by multiplying the difference between the first row ($\mathbf{x}_1$) and the second row ($\mathbf{x}_2$) in the original variables with $\mathbf{P}_A$, as shown

$$(\mathbf{t}_1 - \mathbf{t}_2)^{\mathrm{T}} = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}\mathbf{P}_A \quad (4)$$

where a vector after transposition is considered to be a row vector.

3.1.3. Using the truncated total bootstrap (TTB) procedure to investigate uncertainty after PCA

3.1.3.1. Overview of the TTB procedure

The truncated total bootstrap procedure involves multiple steps. A bootstrap procedure produces a large number of virtual panels using the real-panel results as its sampling distribution. Real-panel assessors are sampled with replacement. If assessors evaluate the products multiple times, then each time that an assessor is selected, that assessor's replicates are also sampled with replacement (Castura, et al., 2022a). Each virtual panel has the same number of assessors, replicates, products, and attributes as the real panel. Each virtual panel's results are aggregated in exactly the same manner as for the real

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

panel, then submitted to PCA. Each virtual panel's PCA solution is truncated to have the same number of PCs as the real panel. Each virtual panel's truncated score matrix is then superimposed onto the real panel's truncated score matrix by Procrustes rotation (Schönemann, 1966). This produces "clouds" of TTB-derived results which indicate the uncertainty of the real-panel results after PCA. A detailed account of this procedure is given elsewhere (Castura et al., 2023; Castura et al., 2022a; Cadoret & Husson, 2013).

3.1.3.2. Uncertainty of products and their paired comparisons

The TTB procedure resamples from the raw data to produce many possible outcomes. In the next sections, we will describe how to use the TTB results to investigate uncertainty and to make inferences about paired comparisons.

To justify 95% confidence ellipses obtained from the TTB procedure, Cadoret and Husson (2013) permute product labels in real sensory data sets, then use the TTB procedure to obtain 95% confidence ellipses for these unstructured products. They observe that these ellipses are large and tend to overlap the origin, which is consistent with an unstructured data set. But this does not demonstrate that a 95% confidence region contains 95% of the points, nor that 95% of such confidence regions would cover the true value under repetition. Later, in Section 3.2, we describe simulation studies that are used to understand the coverage properties of the confidence ellipsoids.

3.1.4. Constructing a 95% confidence ellipsoid

In the following, we construct a 95% confidence ellipsoid based on a cloud of TTB-derived difference scores. The same method can also be applied to construct an ellipsoid from a cloud of TTB-derived scores.

We base our approach on the multivariate normal distribution. In this case, a 100(1-α)% confidence ellipsoid encloses points for which

$$\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} < \chi^2_{1-\alpha,A} \tag{5}$$

(Mardia et al., 1979). For each cloud, **S** is the (*A*×*A*) covariance matrix for its points, **d** is an *A*-length vector indicating the paired difference between any point and the cloud center, and $\chi^2_{1-\alpha,A}$ is the (1-α)$^{\text{th}}$ quantile of the χ^2-distribution with *A* degrees of freedom. The left-hand side of (5) is the squared statistical (Mahalanobis) distance (Mardia et al., 1979; Mahalanobis, 1936) from a point to the cloud center. This formula will be taken as a point of departure for our proposal.

The critical value in (5) is based on a theoretical distribution which assumes that cloud points are distributed multivariate normally. But Castura et al. (2023) reported that clouds of TTB-derived scores and clouds of TTB-derived difference scores sometimes exhibit excess skewness and excess kurtosis. This means that a cloud of points may have many outlying points but also a large proportion of points concentrated near the cloud center. In such cases, more than 5% of points may lie outside the 95% confidence ellipsoids based on (5). Rather than use a more complicated statistical model for the distribution of points, we adapt (5) to better account for the observed distributions of points in the TTB-derived clouds.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

As an approximation, the confidence region of each TTB-derived cloud of difference scores is considered to be ellipsoidal and symmetric. As before, the ellipsoid shape is defined by its (*A*×*A*) covariance matrix **S**, which is calculated from the points to the cloud center. Points contained by our *A*-dimensional 100(1-α)% confidence ellipsoid satisfy

$$\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} < Q_{1-\alpha} \quad (6)$$

The only difference with (5) is that $\chi^2_{1-\alpha,A}$ is replaced by $Q_{1-\alpha}$, which is the (1-α)th quantile of the empirical distribution of the cloud points. In the results that we will report in subsequent sections, the critical values that we obtained for $Q_{1-\alpha}$ are different for every cloud and were usually slightly larger than the $\chi^2_{1-\alpha,A}$ values. So using (6) instead of (5) has a relatively small effect on our interpretations. Since $Q_{1-\alpha}$ is based on the cloud points, every ellipsoid defined by (6) contains exactly 95% of the cloud's points. Increasing the number of virtual panels yields more cloud points which provides a more precise estimate of **S**. We call these confidence ellipsoids because the cloud points are simulated study replications obtained from the TTB procedure. Later, in Section 3.2, we will consider whether this terminology is appropriate.

3.1.5. Visualizing a 95% confidence ellipsoid

By setting α=0.05, we obtain 95% confidence ellipsoids from (6), but we remind the reader that these bootstrap-derived confidence ellipsoids provide only approximate confidence levels. If *A*=1, then its shape is an interval; if *A*=2, then its shape is elliptical; if *A*≥4, then its shape is hyperellipsoidal. A two-dimensional confidence ellipse is visualized naturally in a plane. To visualize an ellipsoid in higher dimensions, we project its coordinates onto each plane of two PCs. Each projection is obtained by applying a standard matrix where the projection (or shadow) of the ellipsoid on the plane has the shape of an ellipse.

Since the confidence region of both (5) and (6) are ellipsoidal and symmetric, it was important to determine whether the TTB-derived clouds have approximately these shape characteristics. For this reason, we plotted the projections of each confidence ellipsoid based on (6) together with the cloud of points and the smallest region that contains 95% of the kernel-estimated density, which is the 95% density contour. Densities were estimated on a 500×500 grid using bivariate normal kernels, where each kernel was the product of two univariate axis-aligned normal kernels with smoothing controlled by a bandwidth determined heuristically (Venables & Ripley, 2002). This approach was used previously by Castura et al. (2023) to visualize the TTB-derived clouds nonparametrically. To evaluate the ellipse approximations in each plane, we visualized the clouds of points, the 95% density contours, and the 95% confidence ellipsoids, which each has an elliptical shape after projection.

Fig. 1 illustrates such a plot based on the Almond beverage and the Almond vs Coconut paired comparison from the beverages data set and the Festival cultivar and the Festival vs Yvahé paired comparison from the strawberry data set. The TTB-derived clouds shown from the beverages data set (Fig. 1, top row) have irregular shapes and exhibit excess skewness and excess kurtosis. Each irregular-shaped 95% density contour is almost completely contained by its 95% confidence ellipsoid, which

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

contains 95% of the cloud points due to (6). Constructing the ellipsoids based on (6) instead of (5) increased the lengths of these ellipsoid axes by about 3% in this case.

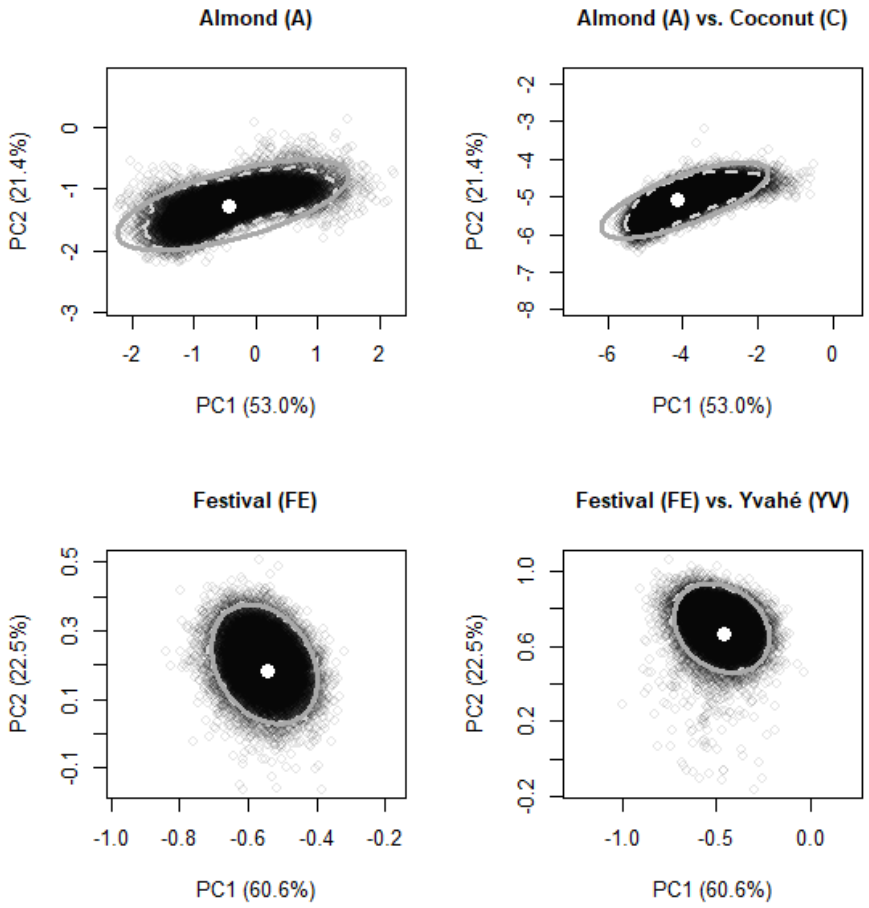


Fig. 1. TTB-derived clouds in the PC1 vs PC2 plane for the Almond (A) beverage (top left), the Almond (A) vs Coconut (C) paired comparison (top right), the Festival (FE) strawberry cultivar (bottom left) and the Festival (FE) vs Yvahé (YV) paired comparison (bottom right). The cloud centers (white points) and clouds of TTB-derived points (dark points) are overlaid with their 95% density contours (dashed light grey lines). In the top row, projections of the 95% confidence ellipsoid are shown on the plane (solid dark grey lines). In the bottom row, the 95% confidence ellipses (solid dark grey lines) almost completely obscure the 95% density contours.

In the bottom row of Fig. 1, the 95% confidence ellipses coincide closely with the contours that enclose 95% of their kernel-estimated densities. The reason that the ellipses and density contours coincide is that these clouds are approximately elliptically and symmetrically distributed. For these ellipses, the

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

axes lengths based on (5) and (6) are nearly identical. In Section 3, we will present results from all products and their paired comparisons in all planes of PCs from both of these data sets.

3.1.6. Visualization and quantification of confidence ellipsoids

To quantify the uncertainty of a difference, we calculated the volume of its *A*-dimensional confidence ellipsoid using

$$V = \frac{(\prod_{a=1}^{A} r_a)\pi^{\left(\frac{A}{2}\right)}}{\Gamma\left(\frac{A}{2}+1\right)} \quad (7)$$

(Mathai, 1999) where Γ denotes the gamma function, π is the mathematical constant, *A* is the number of dimensions, and $\prod_{a=1}^{A} r_a$ is the product of the radii, where r_a is the radius of an ellipsoid axis in dimension *a*, i.e. its semi-axis. If *A*=1 then *V* is the distance $2r_a$. If *A*=2 then *V* is an area; if *A*≥4 then *V* is a hypervolume. Although the average cloud-point distance to the centroid provides another measure of uncertainty, we use (7) for consistency since it gives the volume of the confidence ellipsoid constructed using (6).

3.1.7. Inferences for products and their paired comparisons

When using any inferential statistical method, it would be possible for two products to be very different yet not be discriminated, which can occur if sensory profiles are highly uncertain, for example due to assessor disagreement. It is also possible for two products to be discriminated, yet for the sensory differences to be slight. These same outcomes can occur when evaluating the difference between a product and particular coordinates. In all of these cases, the size of the sensory difference can be obtained from the left-hand side of (6) and the uncertainty of the difference can be obtained from (7). But this does not indicate whether the two products are different, based on the evidence in the experiment.

In a latent space obtained from correlated sensory variables, two products are discriminated if the confidence ellipsoid based on TTB-derived difference scores excludes the origin (Castura, Baker & Ross, 2016). Now, we obtain a result numerically that will indicate whether the confidence ellipsoid excludes the origin. First, we calculate the squared Mahalanobis distance between each of the TTB-derived difference scores and the cloud center to obtain the distribution *Q*. Next, we calculate the squared Mahalanobis distance between the cloud center and the origin. The larger and more extreme this latter squared Mahalanobis distance is compared with the results in *Q*, the less probable it would be to observe such an extreme result if the products are identical. In Appendix A.1, we show that following this procedure gives an approximate way of testing the null hypothesis (H_0) of no difference between products vs the alternative that the products differ (H_1). Next, we quantify the significance related to this test procedure.

The probability of observing a paired comparison that is as or more extreme than the one we actually observed under a true null hypothesis is

$$\mathrm{P} = \Pr\left(\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} \geq Q \,\middle|\, \mathrm{H}_0\right) \quad (8)$$

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

where **d** is the difference between the cloud center and the origin and *Q* is the empirical distribution of squared Mahalanobis distances for the cloud points. Observing P≤0.05 leads to a rejection of H_0 in favour of H_1 and having at least 95% confidence in this decision. Interpretations based on visual inspection of the confidence ellipsoid (6) and the numerical P value (8) coincide.

In this paper, we also use (8) to evaluate whether a product overlaps the origin. In this case, **d** is the difference between a product's cloud center and the origin. The test is relevant because a product that is located exactly at the origin of a PCA subspace is not represented in that space.

3.2. Simulation studies

We conducted two simulation studies to investigate the statistical properties of the confidence ellipsoids. Every 100(1-α)% confidence ellipsoid encloses 100(1-α)% of the points in its TTB-derived cloud due to (6). But this does not tell us if under repeated sampling 95% of ellipsoids that are constructed with confidence level 95% will cover its true value, which is the frequentist definition of a 95% confidence region (Cox & Hinkley, 1974/2000). Previously, in discussing coverage probabilities of the bootstrap procedure, Efron and Tibshirani (1994) noted that the nominal rate (e.g. 95%) is only approximate, and for a particular data set the observed and nominal coverage probabilities may differ substantially. Coverage results from our simulation studies were used to evaluate whether the confidence ellipsoids from (6) have this frequentist interpretation.

Even if it was practical to run replication studies with human assessors, doing so would not provide the necessary information because sensory profiles are never fully known, only estimated with uncertainty. Simulation studies are needed because the true sensory parameters can be specified and sampled. Our simulation studies are based on real sensory data sets. Since the true data generating functions for sensory evaluation data are unknown, our results are useful only as approximations.

Before describing our simulation studies in detail, we provide an overview of the steps for each study (Algorithm 1). The details of how this algorithm is implemented will be described in the subsections that follow. The results will be presented in Section 4.3.

Algorithm 1. Overview of simulation study steps.

1. using the real data set, estimate parameters, which are then treated as "true parameters".
2. use these true parameters to simulate a data set for a replication study.
3. get confidence ellipsoids for this replication study using the TTB procedure.
4. superimpose true parameters into the replication study solution.
5. check whether the confidence ellipsoids cover true parameters.
6. repeat Steps 2 to 5 to determine coverage probabilities for the confidence ellipsoids.

3.2.1. Simulation study 1, based on sensory quantitative descriptive analysis data set

Simulation study 1 was based on the beverages data set (Section 2.1).

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

The purpose of Step 1 was to obtain reasonable "true parameters" for a simulation study. In this study, a linear mixed effects model was fitted to the real-panel data for each attribute. The attribute intensity response depended on the fixed beverage effect, a random assessor effect, a random replicate effect, and residual error. The beverage effects and the variance estimates for the random effects were then used as the true parameters for the remainder of this simulation study.

In Step 2, simulated data for each attribute were obtained using the parameter estimates obtained from the linear mixed effects models by adding random effects sampled from the normal distribution to the fixed beverage effects. After this step, the replication study data set had same numbers of assessors, replicates, products, and attributes as the real beverages data set. In Step 3, we analyzed the replication study data set as if it was the real panel and constructed confidence ellipsoids using the TTB method (Section 3.1).

In Step 4, we submitted the true parameter values to PCA, then used Procrustes rotation to superimpose the true beverage scores onto the replication-study beverage scores. After this step, the true parameters had coordinates in the PCs of the replication study. Then, in Step 5, we checked whether each of the confidence ellipsoids covered its true value.

In Step 6, we repeated Steps 2 to 5 many times, then calculated the empirical coverage rates across all simulation studies.

3.2.2. Simulation study 2, based on consumer CATA data set

Simulation study 2 was based on the strawberry data (Section 2.2). Although this data set consists only of dichotomous data (coded `0` for not checked, `1` for checked), its data can be modelled approximately by a conventional analysis of variance (ANOVA; see Meyners & Hasted, 2023 and references therein). For this reason, the simulation study also follows the steps in Algorithm 1.

In Step 1, a mixed linear effects model with a fixed strawberry cultivar effect and a random assessor effect was fitted to the real-panel data for each attribute. These parameter estimates were treated as "true parameters" in this simulation study.

In Step 2, simulated data for each attribute were obtained by sampling the Bernoulli distribution, with citation probabilities obtained by adding the fixed strawberry cultivar estimate to an assessor effect sampled from the normal distribution. After this step, the replication study had a CATA data set with the same numbers of assessors, strawberry cultivars, and attributes as the real data set.

Steps 3 through 6 were then conducted following the same approach as described in Section 3.2.1.

3.3. Statistical software

All analyses were conducted using R version 4.2.2 (R Core Team, 2022). The TTB method was conducted with 25,000 virtual panels. To investigate the shapes of the clouds of TTB-derived points using the kernel-based approach, we followed the procedure used by Castura et al. (2023). Specifically, we calculated two-dimensional kernel-estimated densities for the cloud points using `kde2d` function in the R package `MASS` (Venables & Ripley, 2002) based on a 500×500 grid in the range of the projected cloud

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

points with the bandwidth determined by the default rule-of-thumb heuristic that is used by this function. Code from `HPDregionplot` in the R package `emdbook` (Bolker, 2020) was adapted to find the smallest region that contains 95% of the kernel-estimated density, which was enclosed by the 95% density contour. The 68% density contour was obtained by the same approach. The levels 68% and 95% were chosen by analogy to the univariate normal distribution, which has 68% and 95% of the density within one and two standard deviations of its mean, respectively. These density contours were then plotted using the R function `contour` (Becker, Chambers & Wilks, 1988). In the simulation studies, linear mixed effects models were fitted to simulation data using the R package `lme4` (Bates, Maechler, Bolker & Walker, 2015). We used the functions `rnorm` and `rbinom` in the simulation studies to sample from the normal and Bernoulli distributions, respectively.

4. Results

4.1. Beverages

4.1.1. PCA results and TTB-derived confidence regions for beverages

We obtained a products-by-attributes matrix. Its columns were variance-standardized, then submitted to PCA as described in Section 3.1. The first four PCs extract 53.0%, 21.4%, 16.4%, and 5.3% of the variance from **X**. We retained the first three PCs, which extract most (90.9%) of the variance.

Since the TTB results were to be used to investigate uncertainty in the real-panel PCA results, we confirmed that the real panel and the virtual panels extracted similar proportions of variance from their respective PCA solutions (Castura et al., 2022a). The percentage of variance extracted by the virtual panels at the 2.5th, 50th, and 97.5th percentiles were (46.5%, 51.8%, 55.2%) in PC1, (18.3%, 21.7%, 24.5%) in PC2, and (15.0%, 16.3%, 17.6%) in PC3. Since the real-panel results were well centered in these intervals, we proceeded to construct TTB-derived 95% confidence regions for the beverages and their paired comparisons.

Suppl. Fig. S1 (eComponent) shows the TTB-derived clouds for each beverage, along with projections of the 95% confidence ellipsoid and the 95% density contours. Selected results are presented in Fig. 2. Loading plots for the PC1 vs PC2 (Fig. 2a), PC1 vs PC3 (Fig. 2d), and PC2 vs PC3 (Fig. 2g) planes show the contributions of each of the sensory attributes to the PCs. Castura et al. (2023) discuss the relationships between the beverages and attributes. The 95% confidence ellipsoids for the beverages are projected onto the planes of PC1 vs PC2 (Fig. 2b), PC1 vs PC3 (Fig. 2e) and PC2 vs PC3 (Fig. 2h). For legibility, only selected paired comparisons are presented in the right column of Fig. 2; these results will be discussed in the next section.

Had the ellipsoids in Suppl. Fig. S1 been obtained from (5) instead of (6), then only the ellipsoid for the Coconut beverage would have had longer radii (by 1.6%). The ellipsoid for the Soya beverage would have had the same radii. Other beverages would have had smaller ellipsoids, by as much as 4.9% for Rice beverage.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

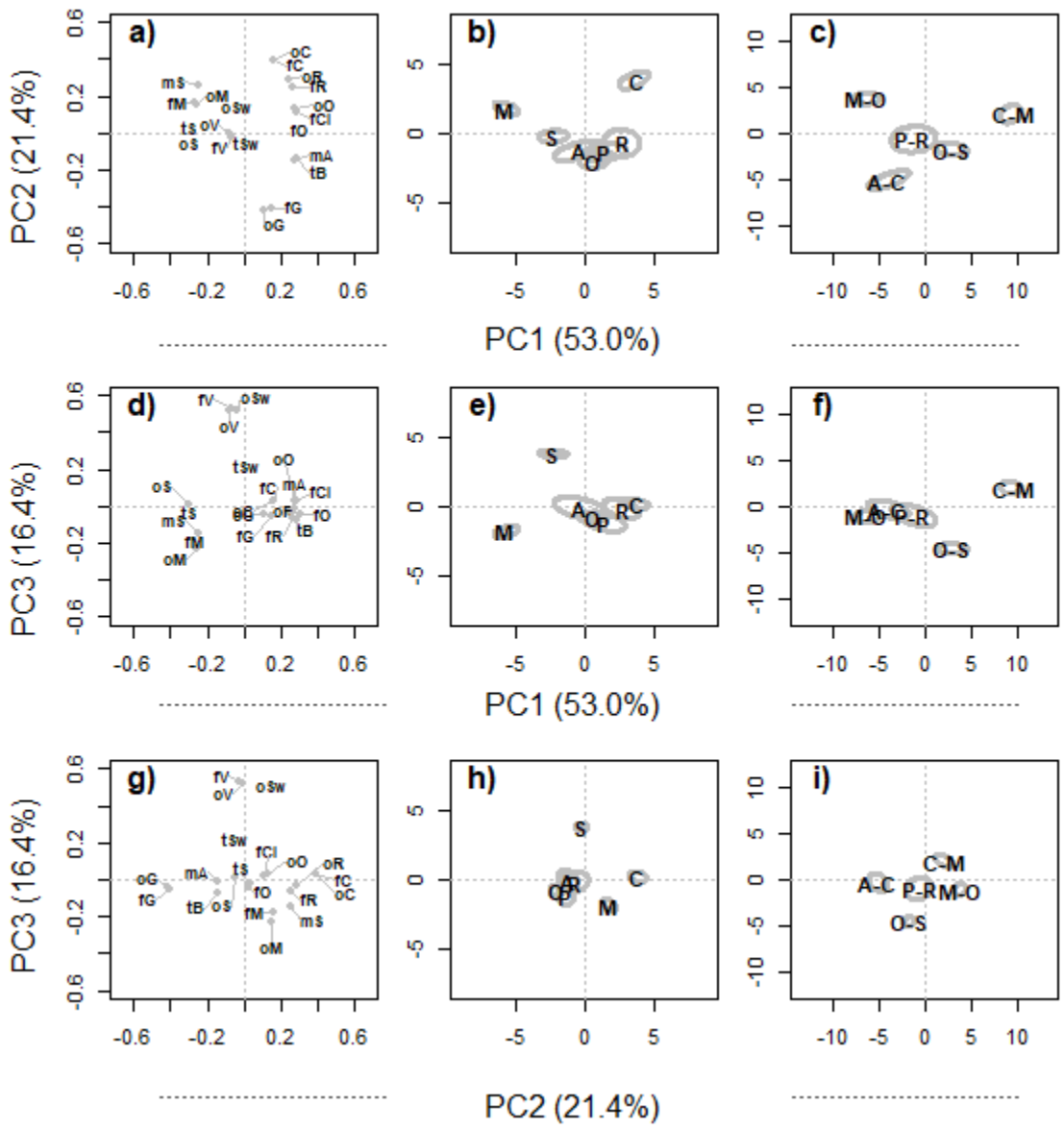


Fig. 2. Plots for the beverage data set in the planes of PC1 vs PC2 (top row; a, b, c), PC1 vs PC3 (middle row; d, e, f), and PC2 vs PC3 (bottom row; g, h, i). For each plane, three plots are shown. The left column shows a loading plot (a, d, g). The middle column shows the projection of 95% confidence ellipsoids for the beverages (b, e, h). The right column shows the projections of 95% confidence ellipsoids for selected paired comparisons (c, f, i). [Beverages: Almond (A); Coconut (C); Cow's Milk (M); Oat (O); Peas (P); Rice (R); Soya (S).]

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

4.1.2. TTB-derived confidence regions for paired differences between beverages

To investigate the paired comparisons, we projected their 95% confidence ellipsoids onto the planes of PC1 vs PC2, PC1 vs PC3, and PC2 vs PC3. Suppl. Fig. S2 (eComponent) shows the TTB-derived cloud for each paired comparison, along with projections of the 95% confidence ellipsoid and the 95% density contour. Visual inspection indicates that the best discriminated beverages are Cow's Milk, the Coconut beverage, and the Soya beverage. The right column of Fig. 2 shows results for five of the 21 paired comparisons: Almond vs Coconut (A-C); Coconut vs Cow's Milk (C-M); Cow's Milk vs Oat (M-O); Oat vs Soya (O-S); and Peas vs Rice (P-R). Axes for the paired difference score plots (Fig. 2, right column) have larger magnitudes than the axes for the score plots (Fig. 2, middle column), indicating that the variability in paired comparisons of **X** is larger than the variability in the products, as was expected (Castura et al., 2023).

Visualizing the confidence regions for the paired comparisons in all three planes gives a better understanding of their three-dimensional shapes and orientations. For example, the Almond vs Coconut confidence region excludes the origin in the PC1 vs PC2 plane (Fig. 2c), but intersects PC3 (Fig. 2f, Fig. 2i). The Almond vs Coconut confidence region is tilted away from both PC1 and PC2 (Fig. 2c) and towards attributes related to rancidity, oxidized, and cloying, suggesting that these attributes are more responsible for the variability in this pair of beverages. The Peas vs Rice confidence region gives a visual impression of being large in all three planes, which indicates that this paired comparison is associated with high uncertainty.

The five paired comparisons shown are separated from the origin in the PC1 vs PC3 plane (Fig. 2f). The panel discriminates Peas vs Rice in this plane, but not in the other two planes; only the projection of its confidence ellipsoid onto this PC1 vs PC3 plane excludes the origin.

Had the ellipsoids in Suppl. Fig. S2 been obtained from (5) instead of (6) then only the ellipsoid for the Coconut vs Cow's Milk pair would have had longer radii (by 1.2%). The ellipsoid for the Coconut vs Soya pair would have had the same radii. Other beverages would have had shorter ellipsoids, by as much as 6.3% for the Peas vs Rice pair.

4.1.3. Volume of confidence regions of beverages and their paired comparisons

Table 1 provides details on the 95% confidence ellipsoid volumes. Among the beverages, the ellipsoid volumes are largest for the Almond, Peas, and Rice beverages. These beverages have more uncertainty in their sensory profiles than the Cow's Milk, Coconut, Soya, and Oat beverages, which have smaller-volume confidence regions.

Table 1 also shows that the paired comparison ellipsoids have larger volumes than the beverage ellipsoids. These numerical results align with ellipsoid silhouettes on the three planes in Fig. 2, Suppl. Fig. S1, and Suppl. Fig. S2 (note differences in scales for the axes). Among the paired comparisons, the Cow's Milk vs Soya ellipsoid has the smallest volume and the lowest uncertainty. The Almond vs Rice ellipsoid has the largest volume and the highest uncertainty; its volume is larger than that of Cow's Milk vs Soya by an order of magnitude.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

Table 1. The volume of the 95% confidence ellipsoid for each beverage (main diagonal, underlined) and their paired comparisons (lower triangle, plain text) are shown. [Beverages: Almond (A); Coconut (C); Cow's Milk (M); Oat (O); Peas (P); Rice (R); Soya (S).]

	A	C	M	O	P	R	S
A	3.09						
C	5.30	0.69					
M	6.96	2.46	0.49				
O	8.40	2.19	2.09	0.87			
P	10.15	3.74	4.98	5.62	2.16		
R	13.40	7.32	5.45	6.04	13.14	3.23	
S	5.83	2.48	0.96	2.34	2.09	4.86	0.46

4.1.4. Inferences for beverages and their paired comparisons

We obtained P values for the beverages and their paired comparisons (Table 2) based on (8). These P values are considered to be approximations because they are based on a bootstrap procedure. If a beverage is located exactly at the origin of the three-component solution, then PCA has extracted none of its variance. P values on the main diagonal indicate that all of the beverages are significantly separated from the origin. The panel discriminates 20 of the 21 pairs (all but Oat vs Peas), often with high confidence. Cow's Milk, Coconut, and Soya beverages are discriminated from each other and from every other beverage. The Almond vs Rice paired comparison has the largest confidence ellipsoid volume and the most uncertainty (Section 4.1.3), but is discriminated. The Oat vs Peas paired comparison is measured with less uncertainty (Table 1), but is not discriminated (Table 2).

Table 2. P values are shown for the beverages (main diagonal) and their paired comparisons (lower triangle). Beverages that are discriminated from the origin and beverage pairs that are discriminated with 95% confidence are shown in bold. [Beverages: Almond (A); Coconut (C); Cow's Milk (M); Oat (O); Peas (P); Rice (R); Soya (S).]

	A	C	M	O	P	R	S
A	**0.0002**						
C	**<0.0001**	**<0.0001**					
M	**<0.0001**	**<0.0001**	**<0.0001**				
O	**0.0145**	**<0.0001**	**<0.0001**	**<0.0001**			
P	**0.0492**	**<0.0001**	**<0.0001**	0.0827	**0.0001**		
R	**0.0053**	**<0.0001**	**<0.0001**	**0.0020**	**0.0117**	**0.0007**	
S	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

4.2. Strawberry cultivars

4.2.1. PCA results and TTB-derived confidence ellipses for strawberry cultivars

The column-centered citation proportions (**X**) from the strawberry consumers were submitted to PCA as described in Section 3.1. The first four PCs extracted 60.7%, 22.5%, 9.7%, and 5.1% of the variance from **X**, respectively. We selected a two-component solution which extracts most (83.1%) of this variance. The virtual panels at the 2.5th, 50th, and 97.5th percentiles extract (48.8%, 57.4%, 65.6%) of the variance in PC1 and (16.4%, 22.6%, and 29.8%) of the variance in PC2. Since the variances extracted by the real panel are well centered in these intervals, we proceeded to obtain TTB-derived 95% confidence ellipses for both the strawberry cultivars and their paired comparisons.

Fig. 3 shows the PCA results in the plane of PC1 vs PC2. The loading plot (Fig. 3a) shows the contributions of the sensory attributes to the PCs. The score plot in Fig. 3b shows the configuration of strawberry cultivars and their 95% confidence ellipses. Castura et al. (2023) discuss the relationships between attributes and cultivars. Some of the 95% confidence ellipses partially overlap, so it is not always clear whether the cultivars are discriminated.

To evaluate each paired comparison in this plane of PCs, we plotted each cloud of points, its 95% and 68% density contours, and its 95% confidence ellipse in Suppl. Fig. S3 (eComponent). The real-panel scores from Yvahé and L20.1 are not well centered in their 68% density contours, which indicates excess skewness. But each 95% confidence ellipse coincides reasonably well with its 95% density contour.

Confidence ellipses for the strawberry cultivars in Suppl. Fig. S3 were obtained from (6). Radii based on (5) were between 2.5% smaller and 1% larger than the radii based on (6).

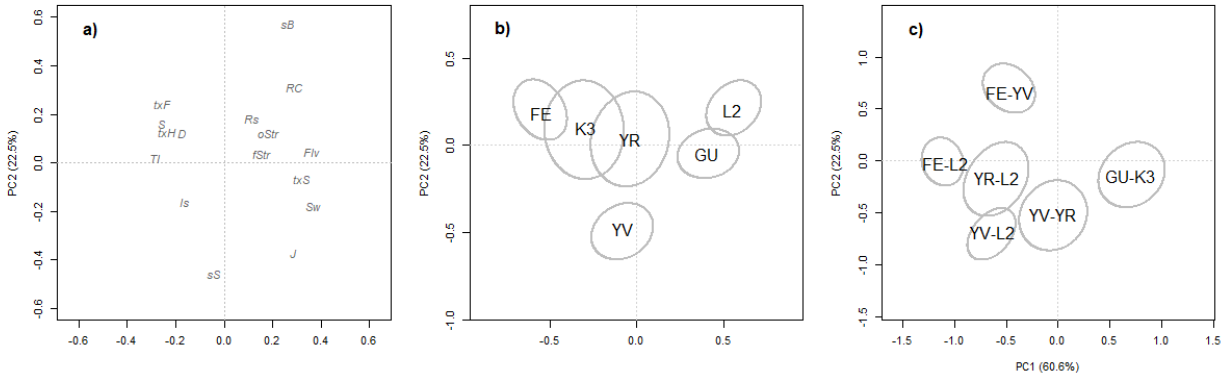


Fig. 3. PCA plots showing strawberry cultivars in the plane of PC1 vs PC2. (a) Loading plot. (b) Score plot with 95% confidence ellipses for each of the strawberry cultivars. (c) Score plot with 95% confidence ellipses for selected paired comparisons. [Strawberry cultivars: Festival (FE); Yvahé (YV); Yurí (YR); Guenoa (GU); L20.1 (L2); K31.5 (K3).]

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

4.2.2. TTB-derived confidence regions for paired differences between strawberry cultivars

Paired differences of the six strawberry cultivars were investigated by constructing their 95% confidence ellipses. For legibility, Fig. 3c shows a selection of six of the 15 paired comparisons. All of these paired comparisons are discriminated with 95% confidence because they exclude the origin.

To evaluate the 95% confidence ellipses for the paired comparisons in the plane of PCs, we plotted the cloud of points, their 95% density contours, and their 95% confidence ellipses in Suppl. Fig. S4 (eComponent). Although some paired comparisons, especially pairs that included Yvahé and L20.1, are not well centered in their 68% density contours, all 95% confidence ellipses coincided reasonably well with the 95% density contours.

Confidence ellipses for the paired comparisons in Suppl. Fig. S4, which were obtained from (6), had radii that were within 1.5% of the radii that were obtained from (5).

4.2.3. Area of confidence ellipses of strawberry cultivars and their paired comparisons

Table 3 shows confidence ellipse areas for the strawberry cultivars and their paired comparisons based on (7). Areas for the K31.5 vs Yurí cultivars were more than twice as large as the other cultivars, indicating there is less certainty associated with the sensory profiles of Yurí and K31.5 than other cultivars.

Table 3. The 95% confidence ellipse areas are shown for each strawberry cultivar (main diagonal, underlined) and their paired comparisons (lower triangle, plain text). [Strawberry cultivars: Festival (FE); Yvahé (YV); Yurí (YR); Guenoa (GU); L20.1 (L2); K31.5 (K3).]

	FE	YV	YR	GU	L2	K3
FE	0.08					
YV	0.18	0.09				
YR	0.31	0.35	0.19			
GU	0.17	0.21	0.27	0.08		
L2	0.15	0.17	0.34	0.22	0.08	
K3	0.40	0.34	0.60	0.31	0.28	0.20

The 95% confidence ellipse areas were larger for the paired comparisons than for the strawberry cultivars (Table 3), which is also apparent from inspection of the ellipses in Fig. 3, Suppl. Fig. S3, and Suppl. Fig. S4. The Yurí vs K31.5 paired comparison has the largest area and the most uncertainty. The Festival vs L20.1 paired comparison has the smallest area and the least uncertainty. Its confidence region is smaller in area than the Yurí vs K31.5 pair by a factor of 4.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

4.2.4. Inferences for the strawberry cultivars and their paired comparisons

Table 4 shows P values for the strawberry cultivars and their paired comparisons based on (8). Since these P values are based on a bootstrap procedure, they are considered to be approximations. The relatively large P value for the Yurí cultivar indicates that it is not well separated from the origin. This position near the origin suggests that its sensory profile is intermediate so it does not contribute as strongly as the other products to the PCA solution. The other cultivars are well separated from the origin and are better represented in the PCA solution.

The consumers discriminate 13 of the 15 pairs (all but Festival vs K31.5 and Yurí vs K31.5). Confidence ellipse areas for the Yvahé vs Yurí and Yurí vs L20.1 paired comparisons were relatively large, indicating higher uncertainties in their sensory profiles (Section 4.2.3). But both of these pairs were better discriminated than Guenoa vs L20.1, which in spite of having less uncertainty (Table 3) was not as well separated from the origin (Fig. 3), as indicated by its larger P value (P=0.0302). These results show that both the size of a difference and the precision of measurement are needed to evaluate how well two cultivars are discriminated.

Table 4. P values are shown for the strawberry cultivars (main diagonal) and their paired comparisons (lower triangle). Strawberry cultivars that are discriminated from the origin and cultivar pairs that are discriminated with 95% confidence are shown in bold. [Strawberry cultivars: Festival (FE); Yvahé (YV); Yurí (YR); Guenoa (GU); L20.1 (L2); K31.5 (K3).]

	FE	YV	YR	GU	L2	K3
FE	**<0.0001**					
YV	**0.0005**	**0.0004**				
YR	**0.0002**	**0.0042**	**0.8556**			
GU	**<0.0001**	**0.0004**	**0.0003**	**<0.0001**		
L2	**<0.0001**	**<0.0001**	**0.0004**	**0.0302**	**<0.0001**	
K3	0.1199	**0.0020**	0.2273	**<0.0001**	**<0.0001**	**0.0062**

4.3. Coverage proportions observed in simulated replication studies

In this section, we present results for the two simulation studies that are described in Section 3.2.

4.3.1. Results from simulation study 1 (based on the beverages data set)

We simulated 1000 replication studies based on the beverages data set, as described in Section 3.2.1. For each simulated replication study, we conducted the TTB procedure with 25,000 virtual panels, then constructed 95% confidence ellipsoids based on their results. The true values were then superimposed on each replication study's PCA solution. Table 5 shows the percentage of the 95% confidence ellipsoids that covered their true values. On average, 91.2% of the beverage ellipsoids and 92.1% of the paired comparison ellipsoids covered the true values. Coverage fell short of 95%, mostly due to lower coverage

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

rates for Cow's Milk, which was very different from the plant-based beverages. These results are considered to be approximations since they are based on a bootstrap procedure, but they are aligned approximately with a frequentist interpretation.

Table 5. Results for simulation study 1, which is based on the beverages data set. The percentage of 95% confidence ellipsoids covering its true scores are shown for each of the beverages (main diagonal, underlined) and for all paired comparisons (lower triangle). Coverage of 95% or higher are shown in bold. [Beverages: Almond (A); Coconut (C); Cow's Milk (M); Oat (O); Peas (P); Rice (R); Soya (S).]

	A	C	M	O	P	R	S
A	**96.3**						
C	93.1	86.6					
M	94.9	71.6	77.2				
O	**97.5**	**95.2**	93.0	**97.7**			
P	**96.6**	**95.0**	92.3	**96.6**	**97.0**		
R	93.7	**96.4**	75.8	**95.1**	**96.3**	92.8	
S	**95.9**	82.9	93.0	**95.5**	94.7	88.8	90.9

4.3.2. Results from simulation study 2 (based on the strawberry data)

We simulated 1000 replication studies based on the strawberry data set, as described in Section 3.2.2. For each replication study, we conducted the TTB procedure and constructed 95% confidence ellipsoids based on their results. Each entry in Table 6 shows the percentage of ellipsoids that covered its true value. On average, the 95% confidence ellipsoids covered the true values for 95.7% of the beverages and for 95.4% of the paired comparisons. These results are considered only to be approximations since they are based on a bootstrap procedure. But they indicate that the ellipses achieved approximately 95% coverage of the true values, which is consistent with a frequentist interpretation.

Table 6. Results for simulation study 2, which is based on the strawberry data set. The percentage of 95% confidence ellipsoids covering its true scores is shown for each of the strawberry cultivars (main diagonal, underlined) and for all paired comparisons (lower triangle). Coverage of 95% or higher are shown in bold. [Strawberry cultivars: Festival (FE); Yvahé (YV); Yurí (YR); Guenoa (GU); L20.1 (L2); K31.5 (K3).]

	FE	YV	YR	GU	L2	K3
FE	**96.0**					
YV	**95.8**	**95.7**				
YR	**95.4**	94.7	**95.0**			
GU	**96.1**	**96.0**	**95.8**	**96.7**		
L2	**96.8**	**96.2**	**96.5**	**96.3**	**96.6**	
K3	94.3	94.0	93.0	**95.8**	**95.1**	94.1

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

5. Discussion

5.1. Furthering the goals of interpreting sensory evaluation results

PCA is often applied to facilitate an understanding of the interrelationships between products and attributes. These results can reveal how products are characterized and how products differ. The approaches that we propose in this manuscript can help the researcher by drawing attention to results that might otherwise be missed. A low P value might draw attention to a particular paired comparison. A large ellipsoid volume might lead a research to ask why a particular product has such high uncertainty. The approaches are intended to help the researcher avoid misinterpretations. They might reveal that two products are not as different as they might appear to be (e.g. Festival vs. K31.5), or that one or both products are highly variable in how they are perceived (e.g. Almond and Rice). But the approaches we give are intended to enrich the data analysis, not to diminish it to checking whether P values are less than 0.05.

5.2. Choice of the number of components retained

We do not advise retaining every PC in which at least one paired comparison is significantly different, since it places too much emphasis on significance testing and insufficient emphasis on the objectives of data summary and dimension reduction. If we had analyzed the beverage data set and the strawberry cultivar data set using the total bootstrap based on (2) rather than the truncated total bootstrap based on (3), then the panel would have discriminated at least one paired comparison in every PC, up to the total number of PCs, as determined by the matrix rank.

In this paper, we retained as few PCs as needed to extract at least 80% of the variance from **X**. Another approach is to ensure that all variables are well summarized by the retained PCs. For example, Mardia et al. (1979) propose calculating the squared correlation between each original variable and each PC, where the sum of the squared correlations for a variable over *A* PCs indicates how well it is explained in these PCs.

In the beverages PCA solution, three PCs explain >70% of 19 of the 20 attributes, but only 18% of *sweet taste*, which was mostly (81%) explained in PC4. Rather than use a more complicated four-component PCA solution, a researcher might investigate *sweet taste* by a univariate analysis.

In the strawberry cultivars PCA solution, two PCs explain >70% of 11 of the 16 attributes. Two PCs explained >60% of *hard*, *dry*, *sour*, and *irregular shape*, but only 31% of *regular shape*. A three-component solution increases the cumulative variance extracted from 83.1% to 92.8% and explains >70% of all attributes except *dry* (61%). The added complexity of incorporating PC3 might be justified here.

Methods in Section 3.1 could be used even if only a single PC is of interest. In this case, if the uncertainties of the paired comparisons are similar, then TTB results could be pooled, as described by Castura et al. (2022a).

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

5.3. Visual evaluation of 95% confidence ellipsoids

Visual assessments of the TTB cloud shapes and the 95% density contours indicate that the clouds were often asymmetric. But the projections of the 95% confidence ellipsoids almost completely enclosed their 95% density contours, which often had a nearly ellipsoidal shape. For this reason, we find that the ellipsoids often provide a convenient approximation of the TTB-derived clouds and their irregular-shaped density contours. Ellipsoids have a limited number of parameters, are easy to calculate, and require no decisions regarding which kernel function and bandwidth to use. Using (6) instead of (5) to construct ellipsoids ensures that a 95% confidence ellipsoid contains 95% of the TTB-derived results.

5.4. Evaluation of graphical and numerical approaches

Another approach for obtaining P values could be to conduct randomization tests (Edgington & Onghena, 2007) based on the TTB-derived points. In this paper, we investigate products and their paired comparisons visually, where the conclusions link consistently with the ellipsoid volumes and P values (Section 3.1 & Appendix A.1). Reviewing results for the seven beverages and their 21 paired comparisons in three PCs would entail reviewing 28 ellipsoids in three planes. Since it is hard to read plots if so many ellipsoids are plotted together, we have presented each of these 84 ellipses in separate plots in Suppl. Fig. S1 and Suppl. Fig. S2. But reviewing so many plots can be rather tedious, particularly if it is something that needs to be done routinely. The P values proposed in Section 3.1.7 and presented in Table 2 distill certain useful information from these 84 plots into a simple numerical table. These P values can be used to screen results. They can be used to draw attention to differences that are worthy of further investigation or that might have been missed inadvertently.

Previously, Lê and Husson (2008) investigated product differences in two TTB clouds via a two-product Hotelling T^2 test (Mardia et al., 1979; Hotelling, 1931). This approach does not account for mutual dependencies because multiple results from the same virtual panel are treated as if they are independent. The test itself is based on the *F* distribution where the covariance matrix estimate has degrees of freedom related to the number of dimensions and the number of observed data. But a covariance matrix based on TTB-derived results is not estimable from observed data but rather from bootstrap-derived points, where the number of points is determined by the analyst. We find using (5) as a starting point to be natural since increasing the number of virtual panels only increases the precision of the covariance matrix estimate, not the way the ellipse is constructed. With a large number of virtual panels, the result from the *F* distribution in the Hotelling T^2 test approaches the χ^2 value, which produces ellipses that resemble (5). In our approach, TTB difference scores based on virtual panels' paired comparison results account for mutual dependencies. Confidence ellipsoids are based on the empirical distribution in (6), which are consistent with the ellipsoid volumes (7) and P values (8) that evaluate the paired comparisons numerically.

Previously, Castura et al. (2022a) proposed a noise-to-signal ratio called the reciprocal index of discriminability (R_d), where *noise* is an estimate of twice the standard error and *signal* is the distance between two products. A multidimensional R_d can be obtained by dividing the right-hand side of (6) by its left-hand side, then taking the square root. The fraction under the square root gives the critical value of the empirical distribution *Q* vs the observed squared paired distance; the square root gives the ratio

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

of the critical value and the observed distance. In both the multidimensional and unidimensional cases, R_d<1 indicates that the products are discriminated with approximately 95% confidence. The reason we used P values instead is that they are based on quantiles, so have a simpler and more familiar interpretation.

5.5. Discussion and limitations of the simulated replication studies

Due to (6), the 95% confidence ellipsoids always contain 95% of the TTB-derived points. If these 95% confidence ellipsoids were constructed by a procedure such that under repetition 95% of these ellipsoids covered the true parameter, then these confidence ellipsoids would also have a frequentist interpretation (Cox & Hinkley, 1974/2000).

In Simulation study 1 (Section 4.3.1), less than 95% of the 95% confidence ellipsoids covered the true score values. The average coverage rate was 91.2% for beverages, where three of the seven beverages had at least 95% coverage, and 92.1% for their paired comparisons, where 10 of 21 had at least 95% coverage. Most coverage rates in this study were 95 ± 3%, but coverage rates for Cow's Milk, Soya, and paired comparisons of these beverages with the Coconut and Rice beverages had coverage rates that were as low as 71.6% (Table 5). The lowest coverage rates were associated with Cow's Milk and Soya which had extreme coordinates (see score plots in Fig. 2 and Supp. Fig. 1), as did their paired comparisons with Coconut and Rice (Fig. 2; Suppl. Fig. 2). But since their true score values were extreme, it was often the case that results in 95% confidence ellipsoids would be interpreted in the same way as true score values that were not covered. Coverage rates were quite good in Simulation study 2 (Section 4.3.2), where just over 95% of the 95% confidence ellipsoids covered the true score values. The average coverage rate was 95.7% for the strawberry cultivars, where five of the six cultivars had at least 95% coverage, and 95.4% for their paired comparisons, where 12 of 15 had at least 95% coverage. In this study, all coverage rates were within 95 ± 2%.

Based on these results, we conclude that our 95% TTB-derived confidence ellipsoids often have coverage probabilities that are quite close to the nominal rate; however, in some cases, coverage differed noticeably from this rate. Our results are in line with previous reports that the coverage probability of bootstrap-derived confidence intervals deviate strongly from the nominal rate in some data sets (Efron & Tibshirani, 1994). It is worth mentioning that we would reach similar high-level conclusions had we obtained the real score values in Step 4 of Algorithm 1 by regression using (3), which is used in the partial bootstrap method (see Castura et al., 2022a) instead of by superimposition using Procrustes rotation.

The simulation study results are important because, as far as we know, the coverage of 95% confidence ellipses based on the TTB procedure has never been investigated previously. These results indicate that we must consider "95% confidence" as an approximation only, since for a particular data set we have no assurance that 95% of the 95% confidence ellipsoids constructed by this procedure will cover the true score values under repetition. Coverage rates for the 95% confidence ellipsoids cannot be worse and are probably slightly better than for the 95% density contours, which are almost completely contained inside the ellipsoids.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

The simulation studies themselves must also be considered to be approximations because the true data generating functions for sensory evaluation data are unknown. In their place, fixed and random effects that would typically be accounted for when analyzing sensory evaluation data (Section 3.2) were used to simulate raw data for the replication studies. Another way to simulate replication study data sets without assuming any particular model is to use a bootstrap procedure. But if the original data set is bootstrapped to get a replication study data set, then some assessors will tend to be selected more than once. Then, when this replication study data set is submitted to the TTB procedure, only the assessors who were part of the replication study data set will be selected for the virtual panels. Given the small number of assessors, bootstrapping twice in this way might tend to produce very dense but long-tailed TTB-derived clouds with (perhaps) artificially low coverage probabilities. For this reason, we avoided using the bootstrap procedure to simulate the replication study data sets.

Further studies might explore the coverage probabilities in simulated replication studies based on other data sets. These studies might further investigate why some coverage probabilities are lower than the nominal rate, including how column standardization or other factors affect the coverage properties. These studies might also evaluate the TTB procedure itself, to determine whether coverage is superior when using a multi-level bootstrap (bootstrapping assessors, then bootstrapping the reps of the selected assessors) instead of a single-level bootstrap (bootstrapping assessors, but leaving their reps intact) or perhaps even a single-level bootstrap in which replicates are bootstrapped and combined across assessors to obtain new, synthetic assessors.

5.6. Applications of these methods to other multivariate analyses

The methods proposed in the present manuscript could be adapted to investigate paired comparisons after other multivariate analyses. Previously, the total bootstrap procedure has been used to investigate the number of significant axes in correspondence analyses of CATA data and of free-comment data (Mahieu, Visalli & Schlich, 2020). So it is possible to expand the methods proposed in the current manuscript to investigate paired comparisons in a manner that accounts for mutual dependencies in the results from the same virtual panels. These methods might also be applied to investigate paired comparisons after multiple correspondence analysis, multiple factor analysis, and other multivariate analyses.

6. Conclusion

In this paper, we propose and evaluate methods for constructing confidence ellipsoids, for quantifying uncertainty by calculating the volumes of confidence ellipsoids, and for calculating P values which indicate numerically which paired comparisons are discriminated. These methods are used to investigate two sensory evaluation data sets visually and numerically. We present and interpret the results from these data sets. An advantage of the ellipsoidal confidence region is that its volume is easily calculated. A disadvantage of using an ellipsoid approximation is that some ellipsoids cover low-density regions where there is little evidence that the true score values are positioned. In each plane of PCs, we show that the ellipse approximations almost completely cover, and sometimes nearly coincide with, the irregular-shaped 95% density contours.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

To understand how confidence ellipsoids can be interpreted, we conducted two simulation studies based on real sensory data sets. These studies were conducted to evaluate the coverage probabilities of the proposed 95% confidence ellipsoids. Results from these simulation studies indicate that approximately 95% of ellipsoids constructed by the proposed procedure for constructing 95% confidence ellipsoids contain their true parameter values, but that coverage rates for some products and paired comparison can differ markedly from the nominal rate.

Further studies could be conducted to evaluate the advantages and disadvantages of analyzing different types of data using the proposed approaches.

Appendix A

Appendix A.1. Procedure for obtaining P values

This appendix shows how any paired comparison (P1 vs P2) can be investigated after PCA based on the first A PCs. First, we show that we obtain identical conclusions whether we investigate the P1-P2 paired difference or the P2-P1 paired difference. Then, we discuss why the approach is appropriate for testing the null hypothesis P1=P2.

The TTB procedure with B virtual panels produces a cloud of points for both P1 and P2. Coordinates of points in each cloud can be organized into a ($B \times A$) matrix. Matrix subtraction a cloud of P1-P2 paired difference scores, denoted $\mathbf{M}$. Since entries in the ($B \times A$) matrix of P2-P1 paired difference scores are identical to entries in $\mathbf{M}$ but with opposite signs, this matrix is denoted – $\mathbf{M}$.

The covariance matrices of – $\mathbf{M}$ and $\mathbf{M}$ are identical since

$$\mathbf{S} = \frac{(-\mathbf{M})^T(-\mathbf{M})}{B-1} = \frac{\mathbf{M}^T\mathbf{M}}{B-1} \quad \text{(A.1.1)}$$

If $\mathbf{d}$ is the difference between a point the center of a cloud with covariance $\mathbf{S}$, then the point and center are separated by the squared Mahalanobis distance $\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d}$. Points in $\mathbf{M}$ and – $\mathbf{M}$ with the same row index have the same squared Mahalanobis distance to their respective cloud centers since

$$\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} = (-\mathbf{d})^{\mathrm{T}}\mathbf{S}^{-1}(-\mathbf{d}) \quad \text{(A.1.2)}$$

This shows that points in $\mathbf{M}$ and $-\mathbf{M}$ have the same empirical distribution Q.

Due to (A.1.2), the squared Mahalanobis distance from the origin (zero) to the cloud center is identical to the origin (zero) is identical whether it is calculated based on the P1-P2 paired difference scores or based on the P2-P1 paired difference scores. For this reason, the P value in (8) based on P1-P2 is identical to the P value based on P2-P1.

Under a true null hypothesis, P1-P2=P2-P1=0. Such equality is never observed in practice since both P1 and P2 are observed with natural variation. Translating the P1-P2 cloud on the origin removes the product effect from the observed results, but leaves natural variation intact. Translation does not change $\mathbf{S}$ since adding a constant affects neither variances nor covariances. Since $\mathbf{S}$ is unchanged and

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

translation is applied to all points, all differences $\mathbf{d}$ are unchanged and translation does not change the distribution Q. The difference between the observed effect (P1-P2) vs no effect (zero) is measured exactly as above. Identical results are obtained from both $-\mathbf{M}$ and $\mathbf{M}$. This result shows that if the P1-P2 cloud excludes the origin, then centering the cloud on the origin will exclude the center of the P1-P2 cloud (before translation).

If $\mathbf{M}$ and $-\mathbf{M}$ are merged into a supercloud with $2B$ points, then the empirical distribution of squared Mahalanobis distances for the combined clouds would consist of B values repeated in pairs. Since the corresponding points in the two clouds are redundant, rather than added, we test the null hypothesis for this paired comparison using either $\mathbf{M}$ or $-\mathbf{M}$, not both.

Acknowledgements

We thank the anonymous reviewers providing thoughtful feedback that helped us to improve this manuscript. Authors TN and PV acknowledge financial support from the Research Council of Norway and the Norwegian Fund for Research Fees for Agricultural Products (FFL) through the project "FoodForFuture" (Project number 314318; 2021-2024).

References

Ares, G., & Jaeger, S.R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference*, *28,* 141-153. https://doi.org/10.1016/j.foodqual.2012.08.016

Ares, G., & Jaeger, S.R. (2015). Check-all-that-apply (CATA) questions with consumers in practice: Experimental considerations and impact on outcome. In: J. Delarue, J.B. Lawlor and M. Rogeaux (eds.): *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (pp. 227-245). Waltham, MA: Woodhead Publishing.

Babamoradi, H., van den Berg, F., & Rinnan, Å. (2013). Bootstrap based confidence limits in principal component analysis—A case study. *Chemometrics and Intelligent Laboratory Systems*, *120*, 97-105. https://doi.org/10.1016/j.chemolab.2012.10.007

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using `lme4`. *Journal of Statistical Software,* 67, 1-48. doi:10.18637/jss.v067.i01.

Becker, R.A., Chambers, J.M., & Wilks, A.R. (1988). *The New S Language*. Wadsworth & Brooks/Cole.

Bi, J., & Kuesten, C. (2023). Type I error, testing power, and predicting precision based on the GLM and LM models for CATA data--Further discussion with M. Meyners and A. Hasted. *Food Quality and Preference*, 104806. https://doi.org/10.1016/j.foodqual.2022.104806

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

Bi, J., & Kuesten, C. (2022). Commentary on Meyners and Hasted (2021): On the applicability of ANOVA models for CATA data, Food Quality and Preference, 92. *Food Quality and Preference*, *95*, 104340. https://doi.org/10.1016/j.foodqual.2021.104340

Bolker, B. (2020). emdbook: Ecological Models and Data in R. R package version 1.3.12. https://cran.r-project.org/package=emdbook

Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference*, *28*, 106–115. https://doi.org/10.1016/j.foodqual.2012.09.005

Castura, J.C. (2022). `cata`: Analysis of Check-All-that-Apply (CATA) data. *R Package Version 0.0.10.9*. https://CRAN.R-project.org/package=cata

Castura, J.C., Baker, A.K., & Ross, C.F. (2016). Using contrails and animated sequences to visualize uncertainty in dynamic sensory profiles obtained from temporal check-all-that-apply (TCATA) data. *Food Quality and Preference*, *54*, 90-100. https://doi.org/10.1016/j.foodqual.2016.06.011

Castura, J.C., Meyners, M., Varela, P., & Næs, T. (2022b). Clustering consumers based on product discrimination in check-all-that-apply (CATA) data. *Food Quality and Preference*, 99, 104564. https://doi.org/10.1016/j.foodqual.2022.104564

Castura, J.C., Rutledge, D.N., Ross, C.F., & Næs, T. (2022a). Discriminability and uncertainty in principal component analysis (PCA) of temporal check-all-that-apply (TCATA) data. *Food Quality and Preference*, *96*, 104370. https://doi.org/10.1016/j.foodqual.2021.104370

Castura, J.C., Varela, P., & Næs, T. (2023). Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106, 104814. https://doi.org/10.1016/j.foodqual.2023.104814

Courcoux, P., Qannari, E.M., Taylor, Y., Buck, D., & Greenhoff, K. (2012). Taxonomic free sorting. *Food Quality and Preference*, 23, 30–35. https://doi.org/10.1016/j.foodqual.2011.04.001

Cox, D.R., & Hinkley, D.V. (2000). *Theoretical Statistics*. New York: Chapman & Hall/CRC Press. (Original work published 1974).

Edgington, E., & Onghena, P. (2007). *Randomization Tests*. Boca Raton, FL: Chapman & Hall/CRC Press.

Efron, B., & Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. New York: CRC Press.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2, 360–378. https://doi.org/10.1214/aoms/1177732979.

Husson, F., Lê, S., & Pagès, J. (2005). Confidence ellipse for the sensory profiles obtained by principal component analysis. *Food Quality and Preference*, *16*, 245-250. https://doi.org/10.1016/j.foodqual.2004.04.019

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

Kiers, H.A.L., & Groenen, P.J.F. (2006). Visualizing dependence of bootstrap confidence intervals for methods yielding spatial configurations. In: S. Zani, A. Cerioli, M. Riani, & M. Vichi (eds.) *Data Analysis, Classification and the Forward Search. Studies in Classification, Data Analysis, and Knowledge Organization*. Heidelberg: Springer. https://doi.org/10.1007/3-540-35978-8_14

Lê, S., & Husson, F. (2008). SensoMineR: a package for sensory data analysis. *Journal of Sensory Studies*, *23*, 14-25. https://doi.org/10.1111/j.1745-459X.2007.00137.x

Lebart, L. (2007). Which bootstrap for principal axes methods? In P. Brito, G. Cucumel, P. Bertrand, & F. de Carvalho (Eds.): *Selected Contributions in Data Analysis and Classification* (pp. 581–588). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73560-1_55

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E.M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference 72*, 31-39. https://doi.org/10.1016/j.foodqual.2018.09.006

Llobell, F., Vigneau, E., Cariou, V., & Qannari, E.M. (2020). `ClustBlock`: Clustering of Datasets. R package version 2.3.1. https://CRAN.R-project.org/package=ClustBlock

Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Sankhya A*, 80, 1–7 (2018). https://doi.org/10.1007/s13171-019-00164-5

Mahieu, B., Visalli, M., & Schlich, P. (2020). Accounting for the dimensionality of the dependence in analyses of contingency tables obtained with Check-All-That-Apply and Free-Comment. *Food Quality and Preference*, *83*, 103924. https://doi.org/10.1016/j.foodqual.2020.103924

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate Analysis*. Toronto: Academic Press.

Mathai, A.M. (1999). *An Introduction to Geometrical Probability: Distributional Aspects with Applications.* Newark, NJ: Gordon and Breach Science Publishers.

Meyners, M., & Castura, J.C. (2014). Check-all-that-apply questions. In: P. Varela and G. Ares (eds.): *Novel Techniques in Sensory Characterization and Consumer Profiling* (pp. 271-306). Boca Raton, FL: CRC Press.

Meyners, M., Castura, J.C., & Carr, B.T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference*, *30*, 309-319. https://doi.org/10.1016/j.foodqual.2013.06.010

Meyners, M., & Hasted, A. (2023). On the choice of appropriate models for CATA data–a further reply to Bi and Kuesten. *Food Quality and Preference*, 104818. https://doi.org/10.1016/j.foodqual.2023.104818

Meyners, M., & Hasted, A. (2022). Reply to Bi and Kuesten: ANOVA outperforms logistic regression for the analysis of CATA data. *Food Quality and Preference*, *95*, 104339. https://doi.org/10.1016/j.foodqual.2021.104339

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

Meyners, M., & Hasted, A. (2021). On the applicability of ANOVA models for CATA data. *Food Quality and Preference*, *92*, 104219. https://doi.org/10.1016/j.foodqual.2021.104219

Næs, T., Tomic, O., Endrizzi, I., & Varela, P. (2021). Principal components analysis of descriptive sensory data: Reflections, challenges, and suggestions. *Journal of Sensory Studies*, *36*, e12692. https://doi.org/10.1111/joss.12692

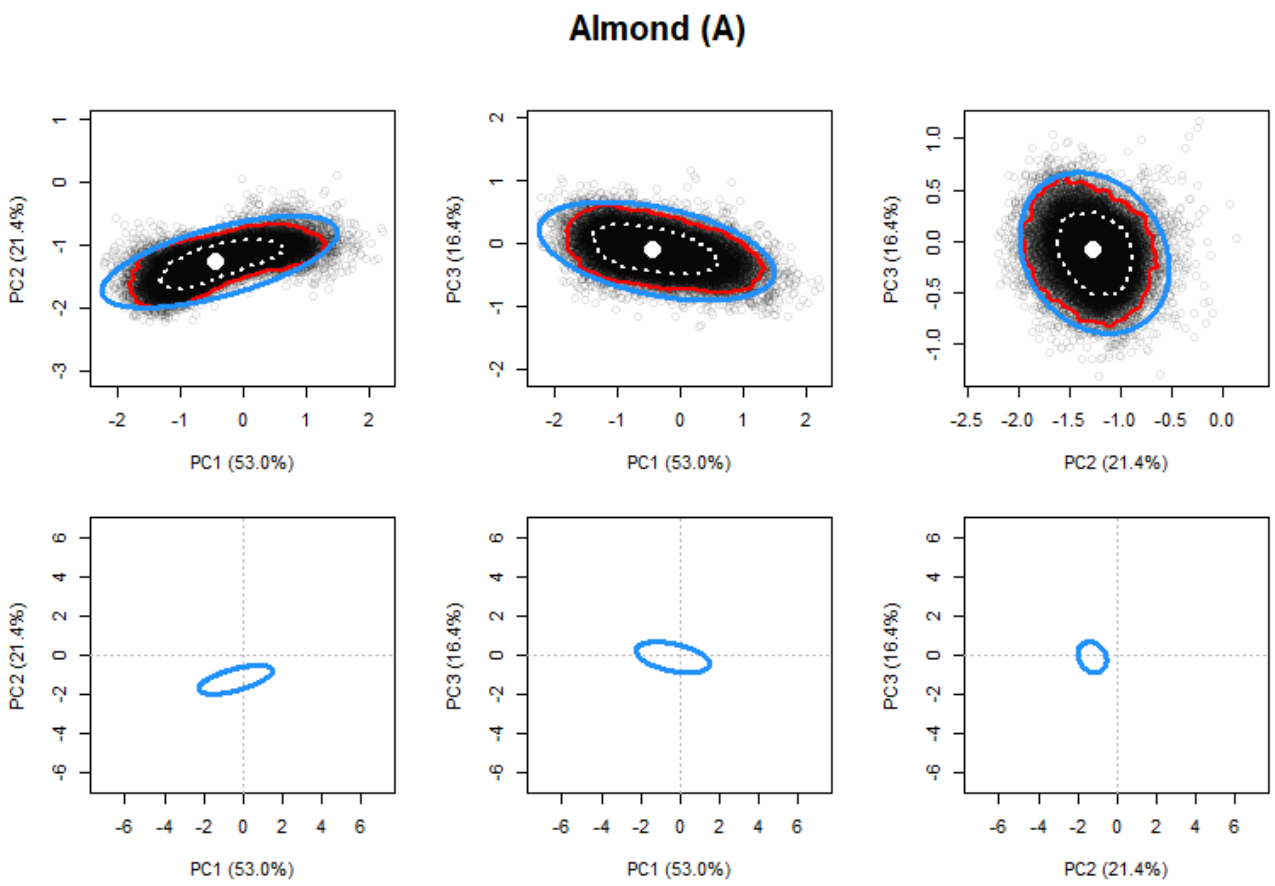
R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schönemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1–10. https://doi.org/10.1007/BF02289451

Venables, W.N., & Ripley, B.D. (2002). *Modern Applied Statistics with S.* Fourth edition. Springer.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/
Pre-print. See the published manuscript in *Food Quality and Preference* for all quotes and citations
https://doi.org/10.1016/j.foodqual.2023.104843

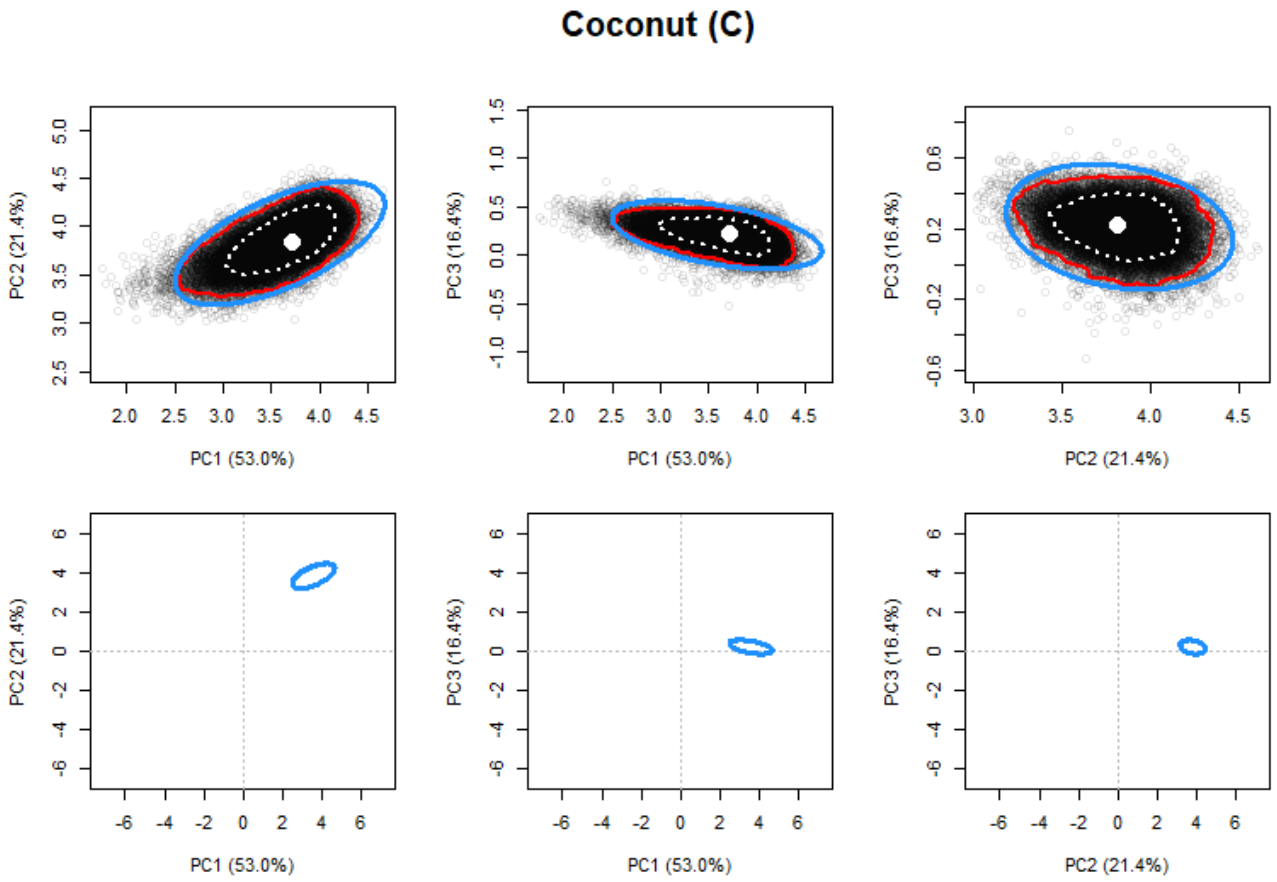
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1a. Almond (A) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

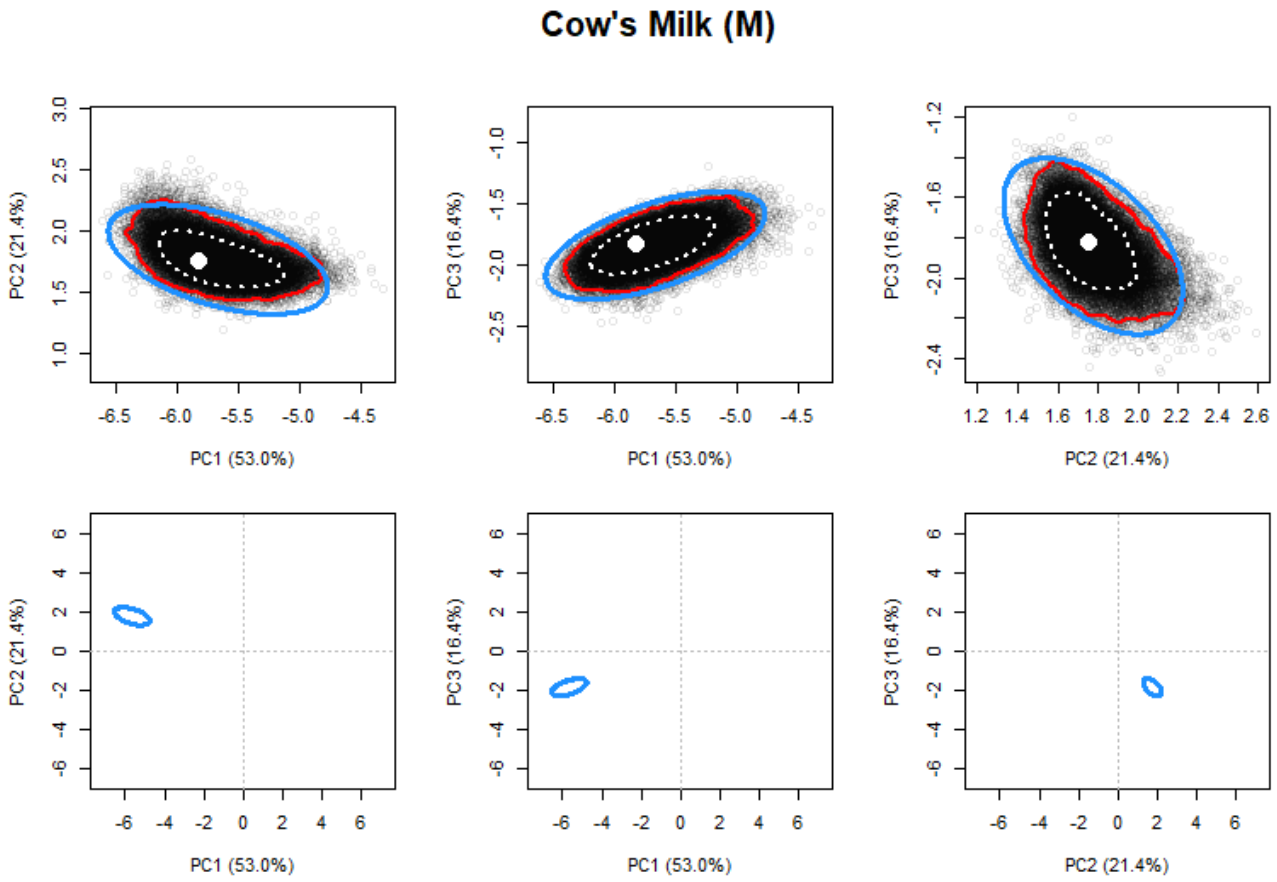
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1b. Coconut (C) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

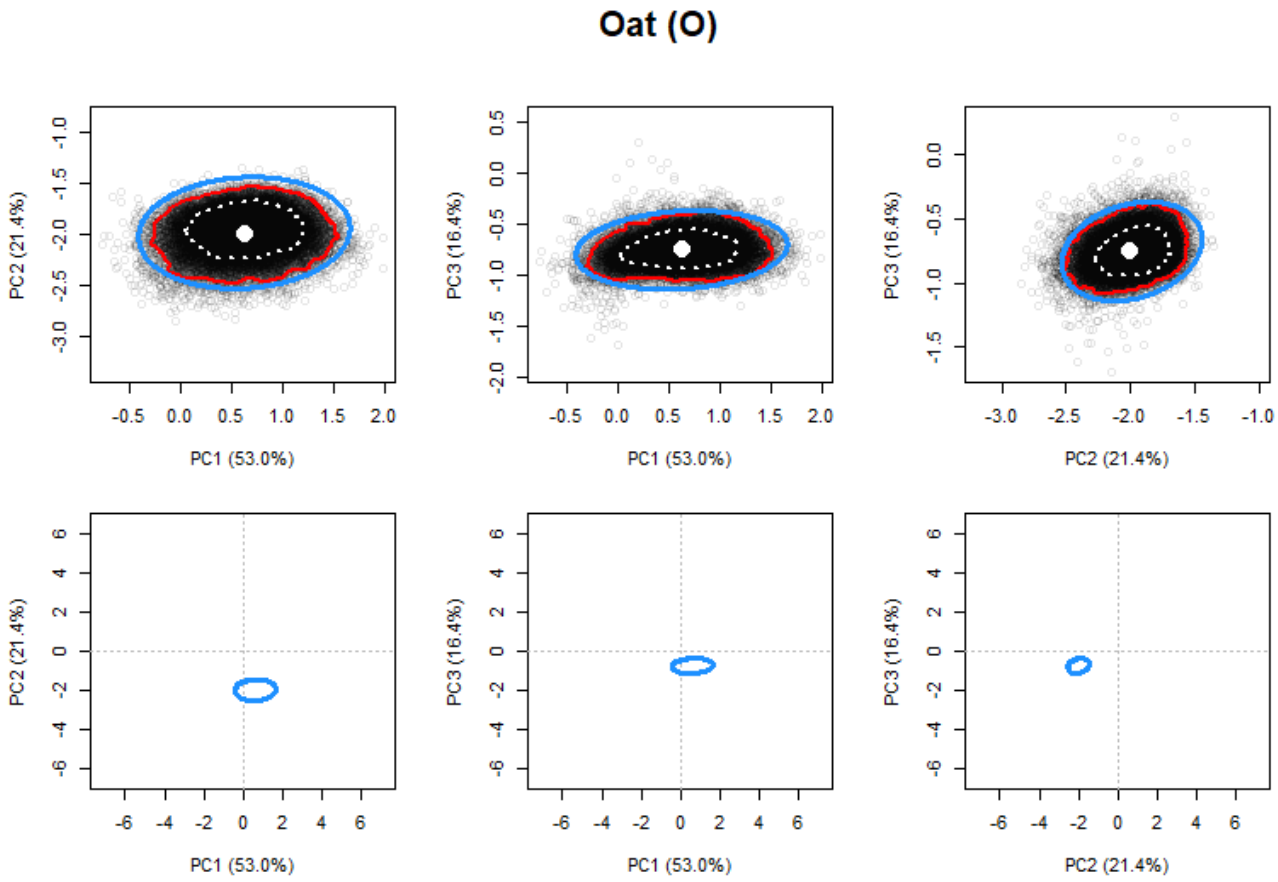
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1c. Cow's Milk (M) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

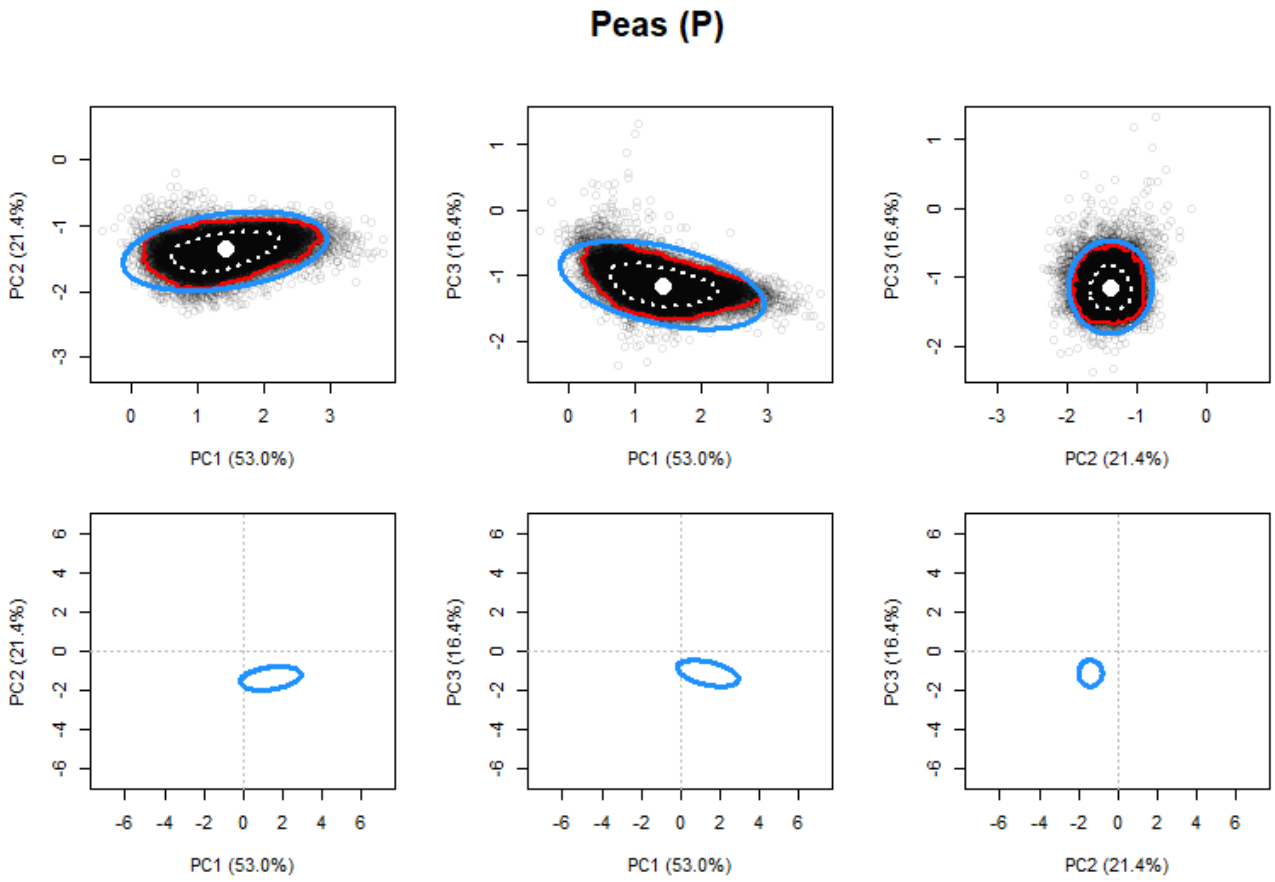
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1d. Oat (O) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

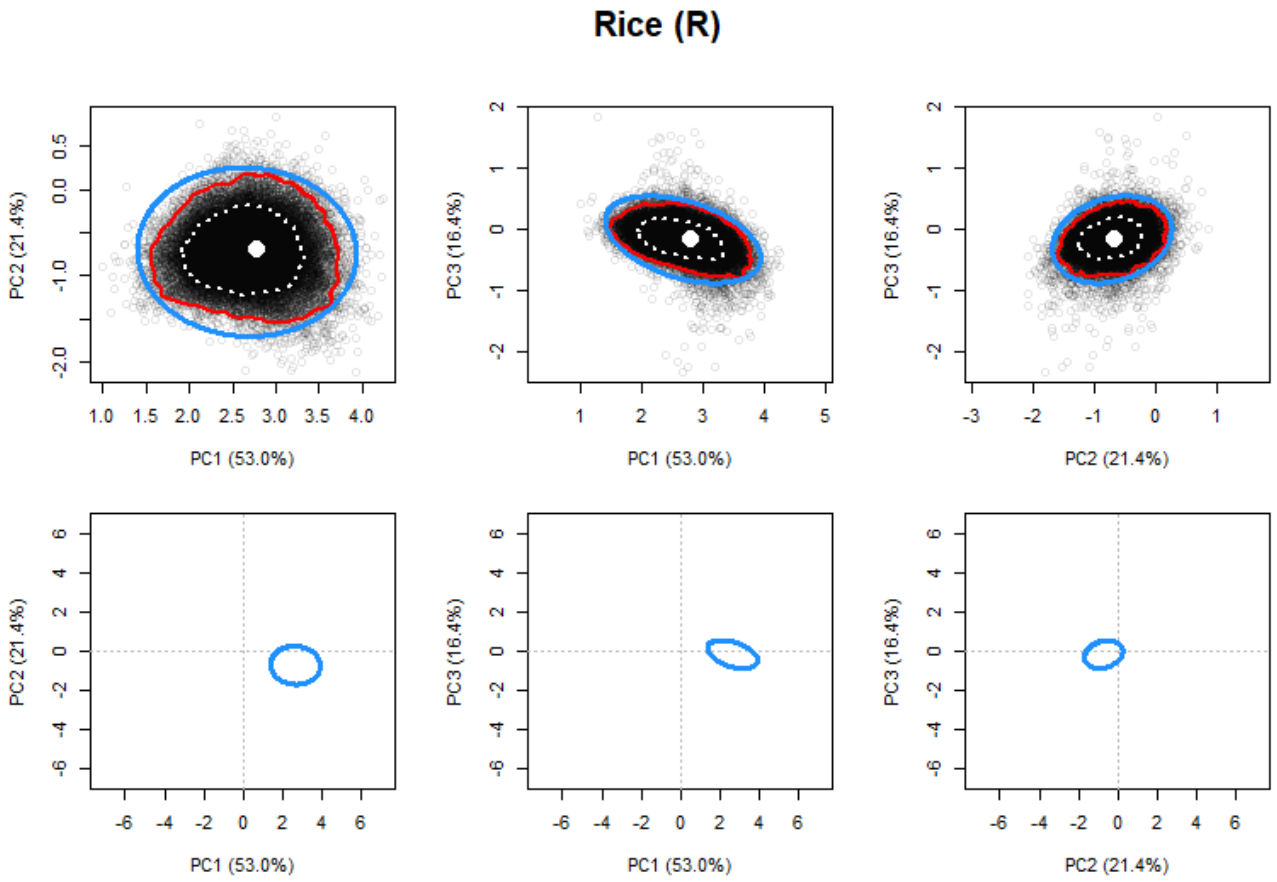
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1e. Peas (P) beverage results in PC1 vs PC2 (left column column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

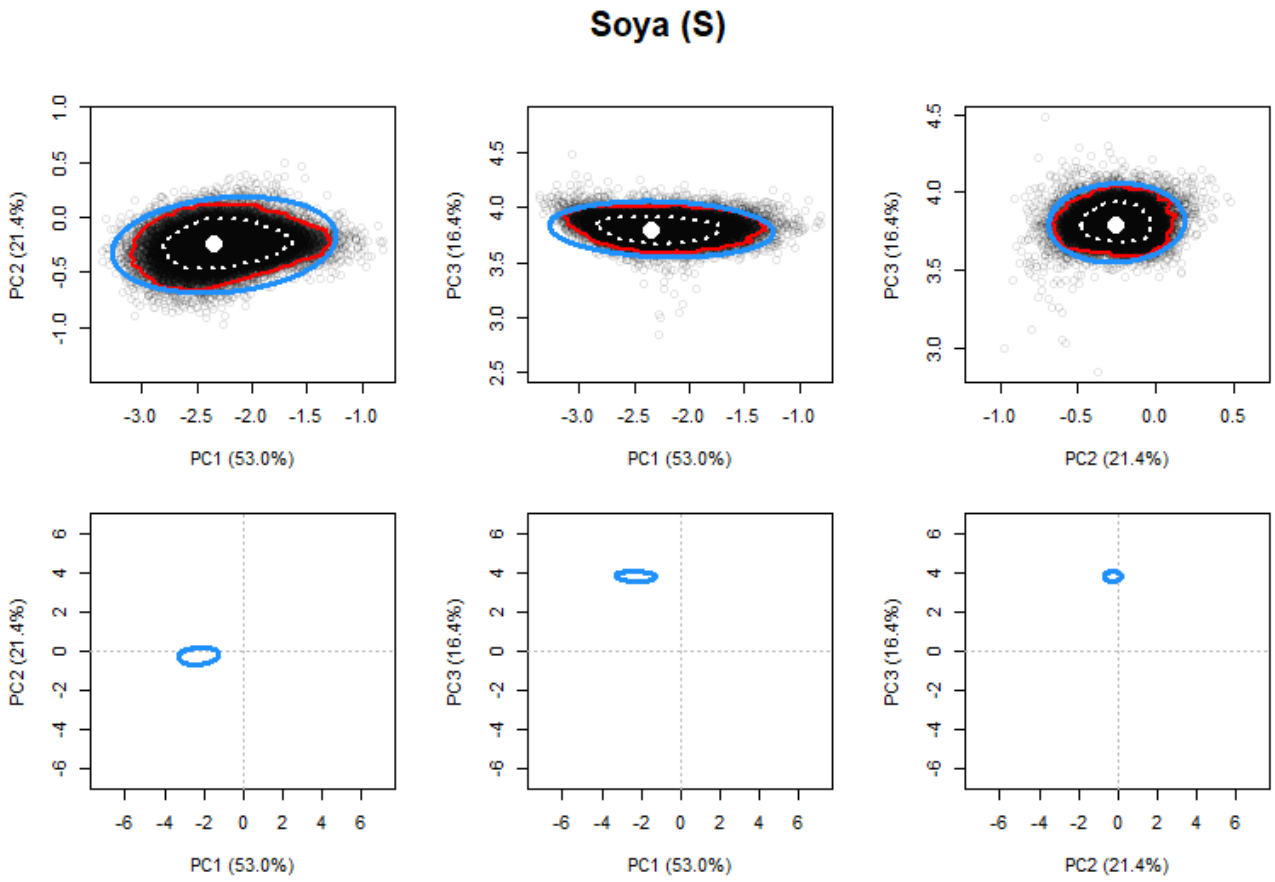
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1f. Rice (R) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

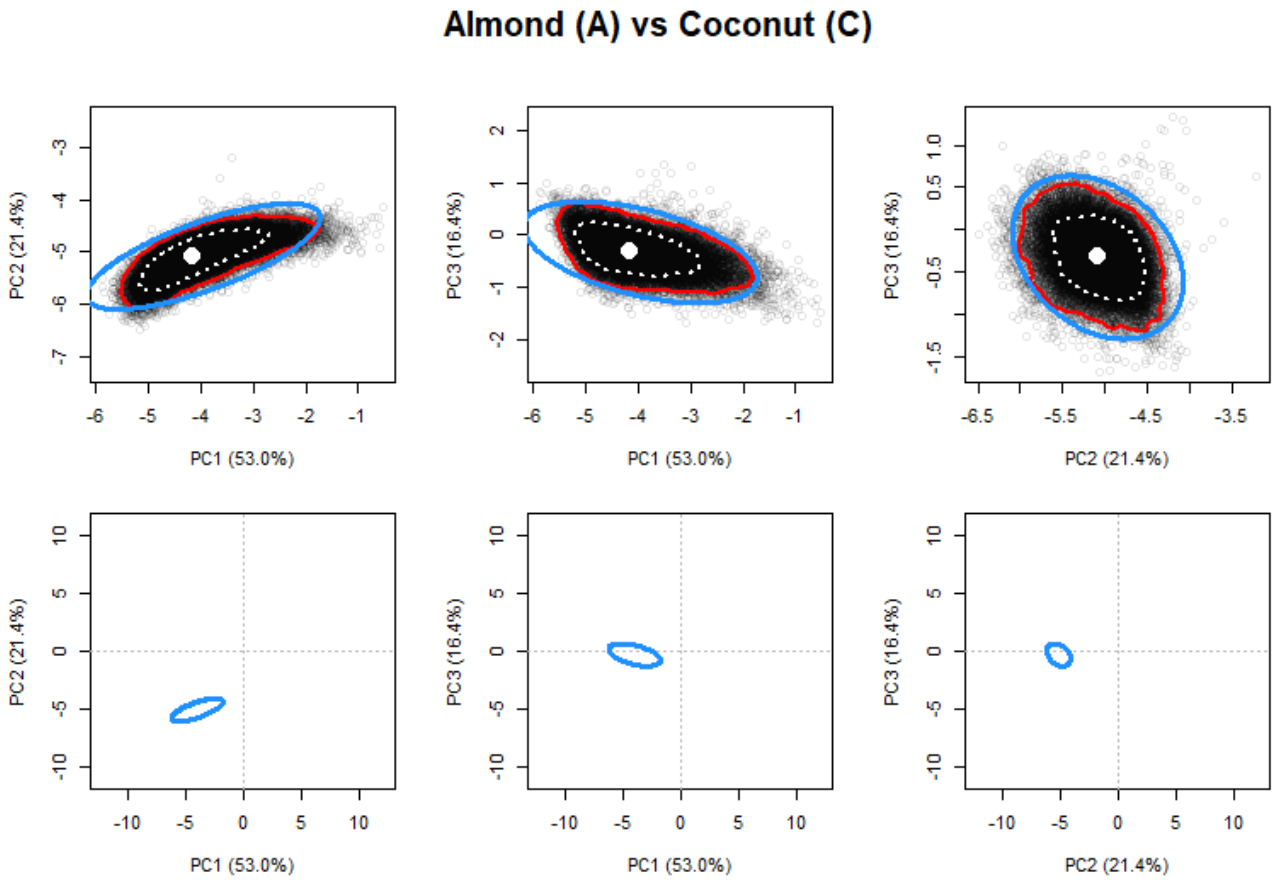
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 1) JC Castura, P Varela, T Næs



Suppl. Fig. S1g. Soya (S) beverage results in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the projection of 95% confidence ellipsoid (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). Each plot in the bottom row zooms out to show the origin and the projection of 95% confidence ellipsoid onto the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

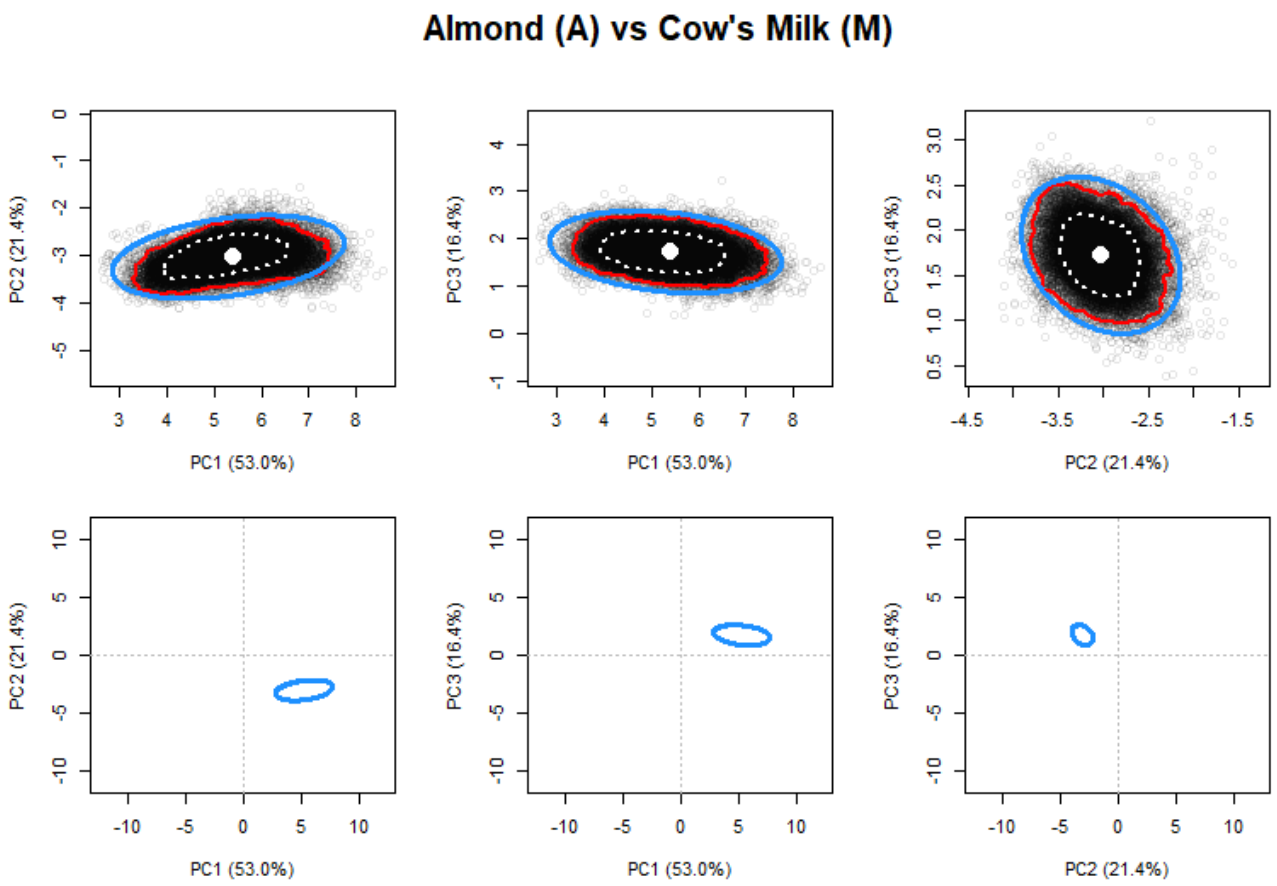
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2a. Almond (A) vs Coconut (C) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

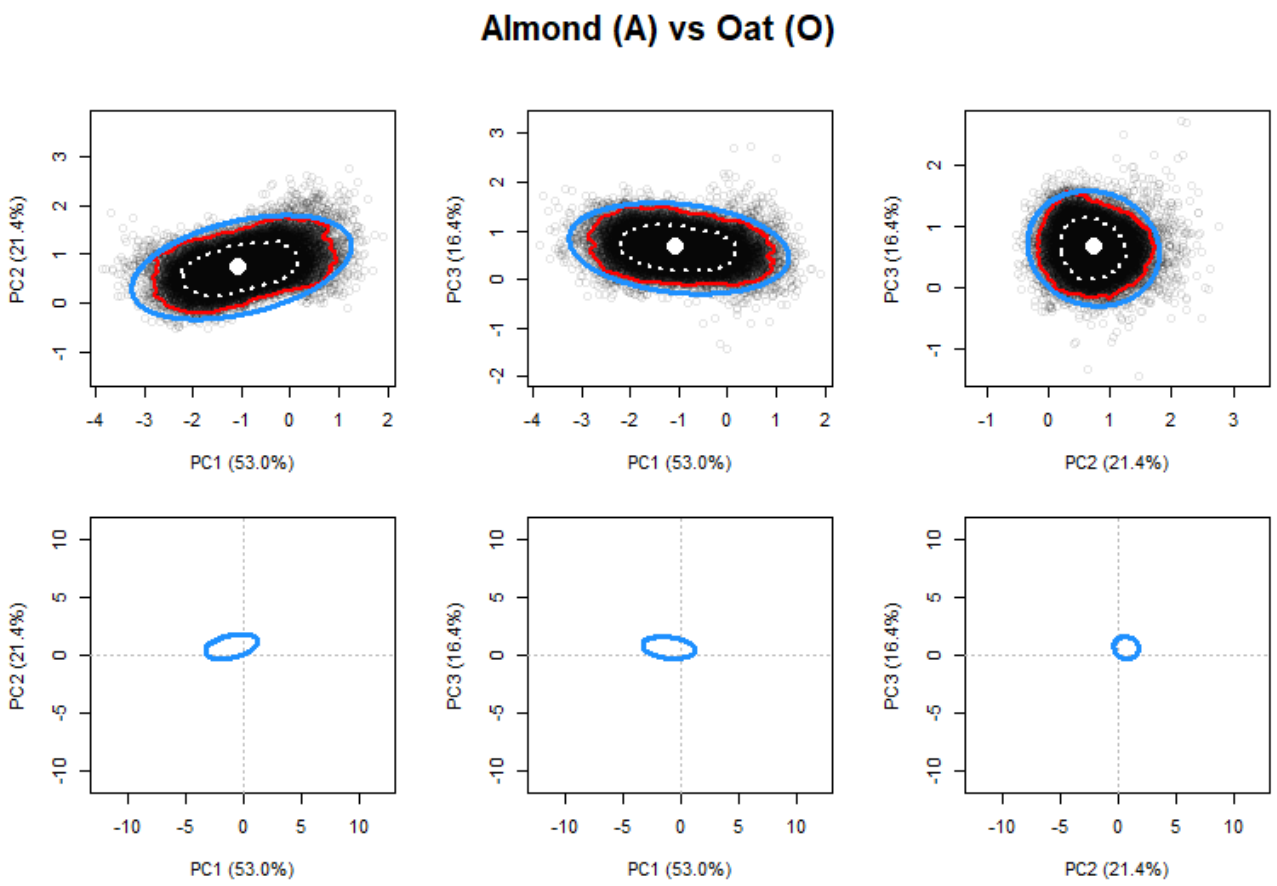
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2b. Almond (A) vs Cow's Milk (M) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

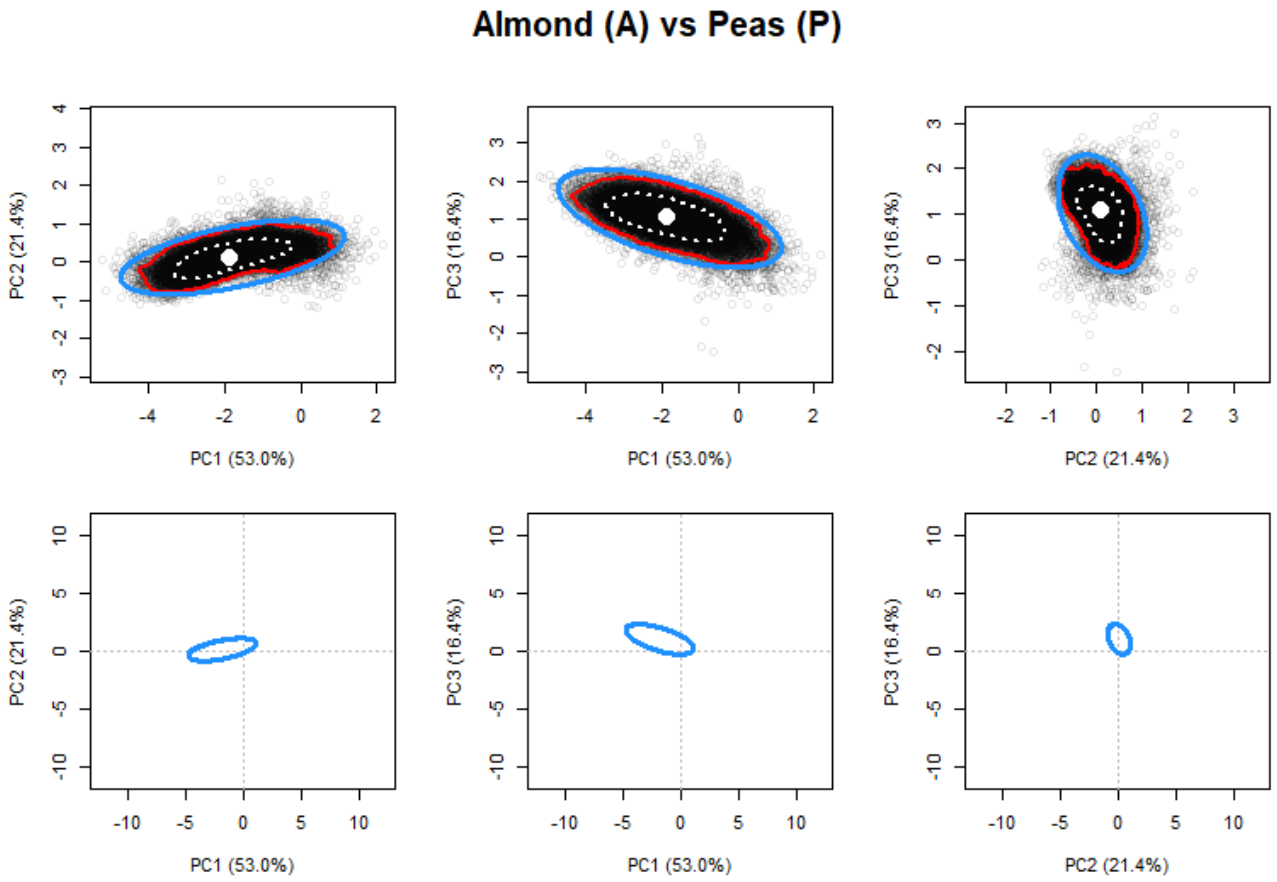
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2c. Almond (A) vs Oat (O) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

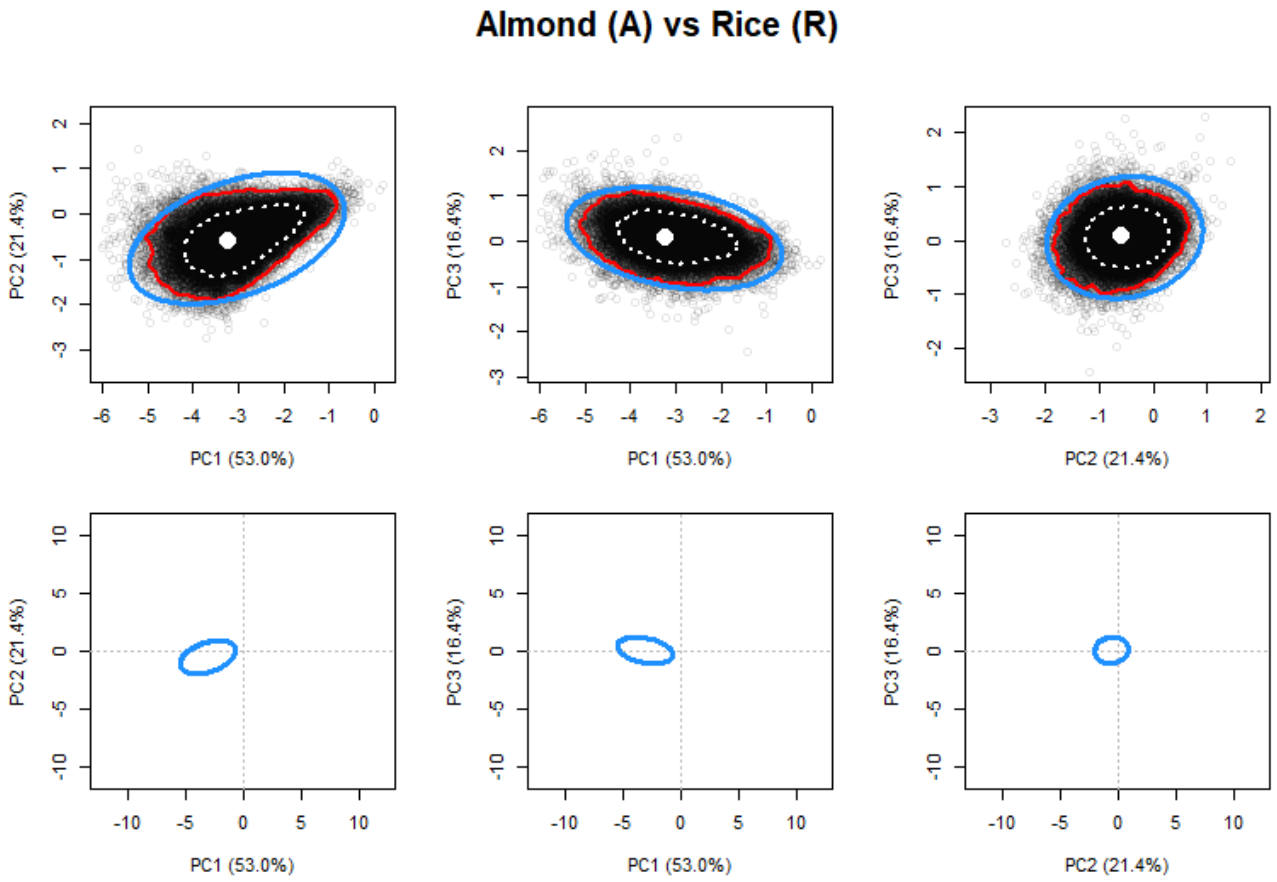
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2d. Almond (A) vs Peas (P) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

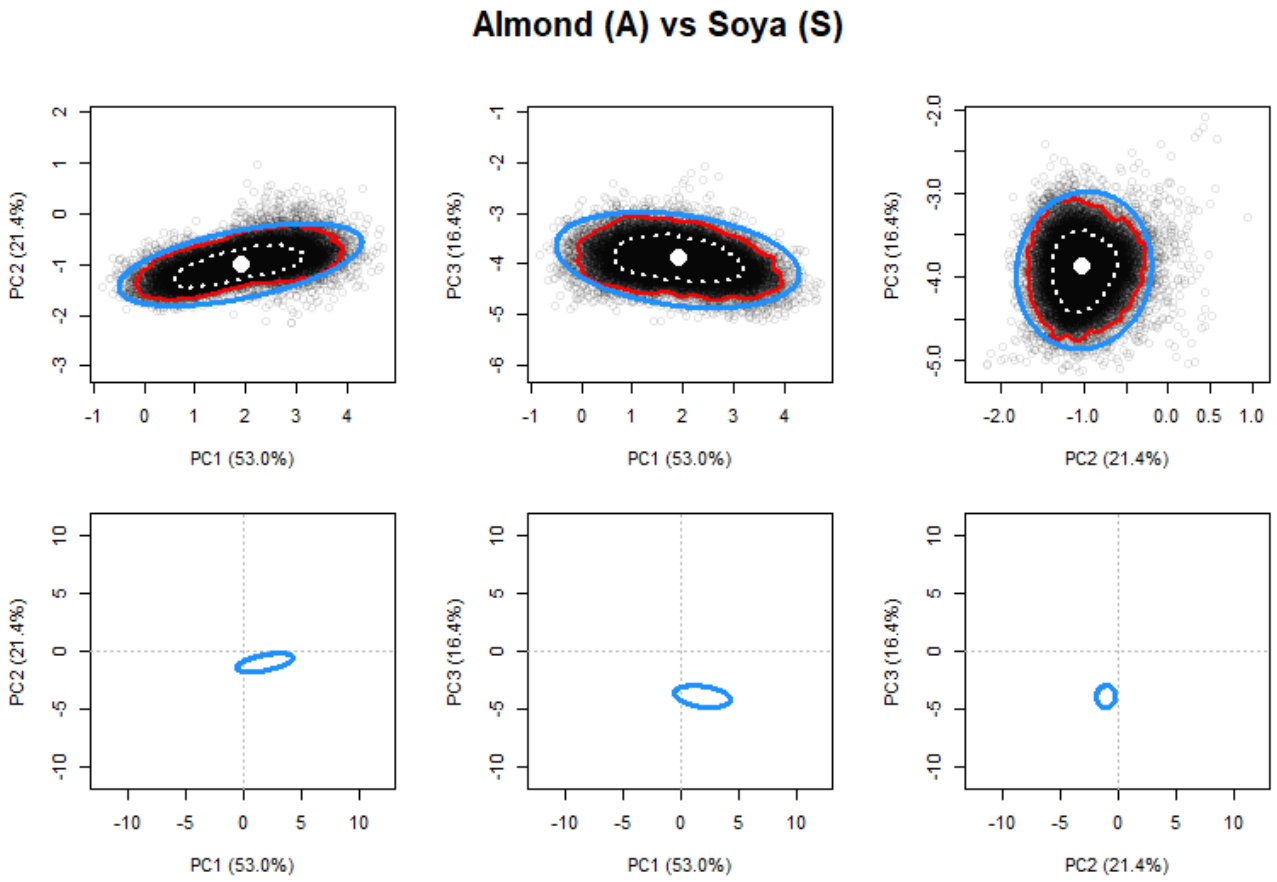
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2e. Almond (A) vs Rice (R) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

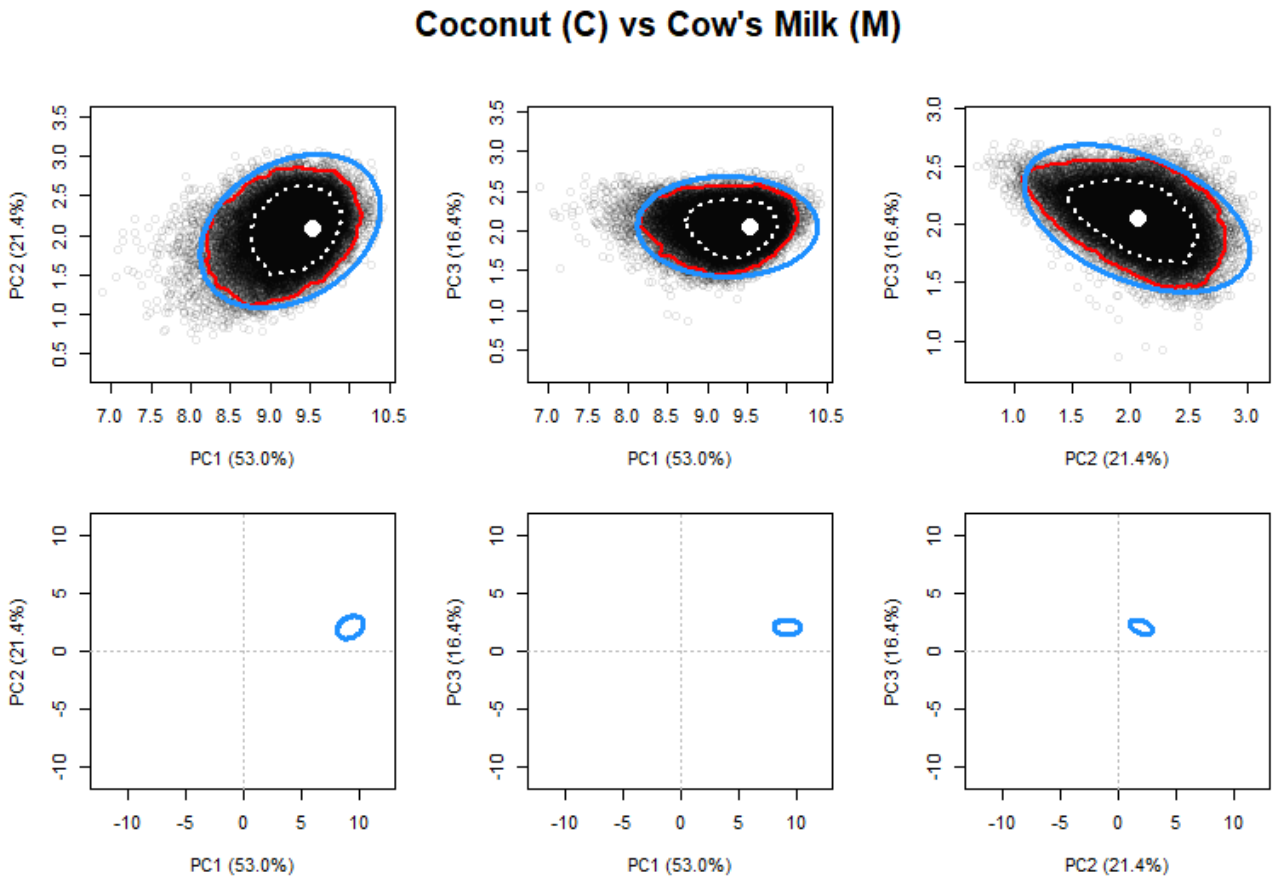
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2f. Almond (A) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

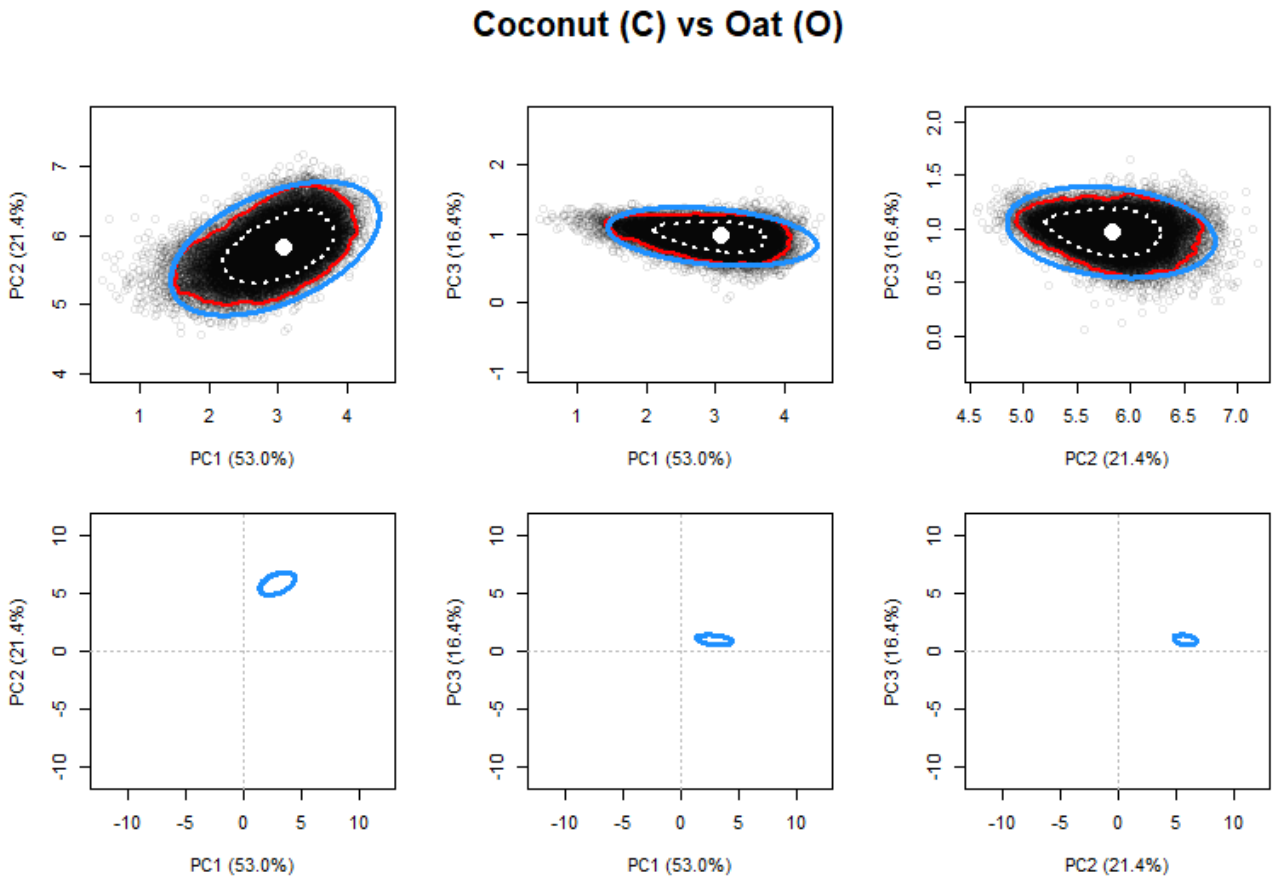
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2g. Coconut (C) vs Cow's Milk (M) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

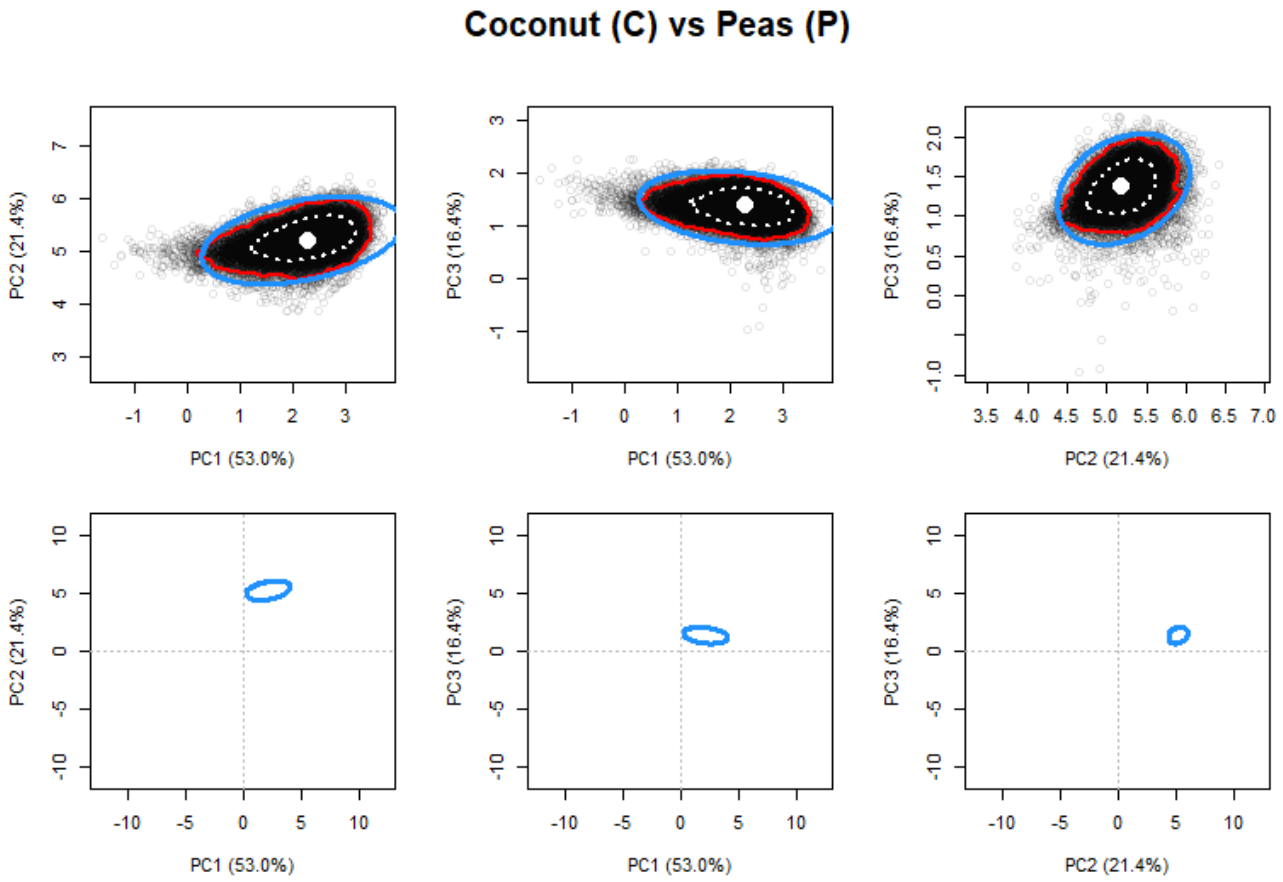
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2h. Coconut (C) vs Oat (O) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

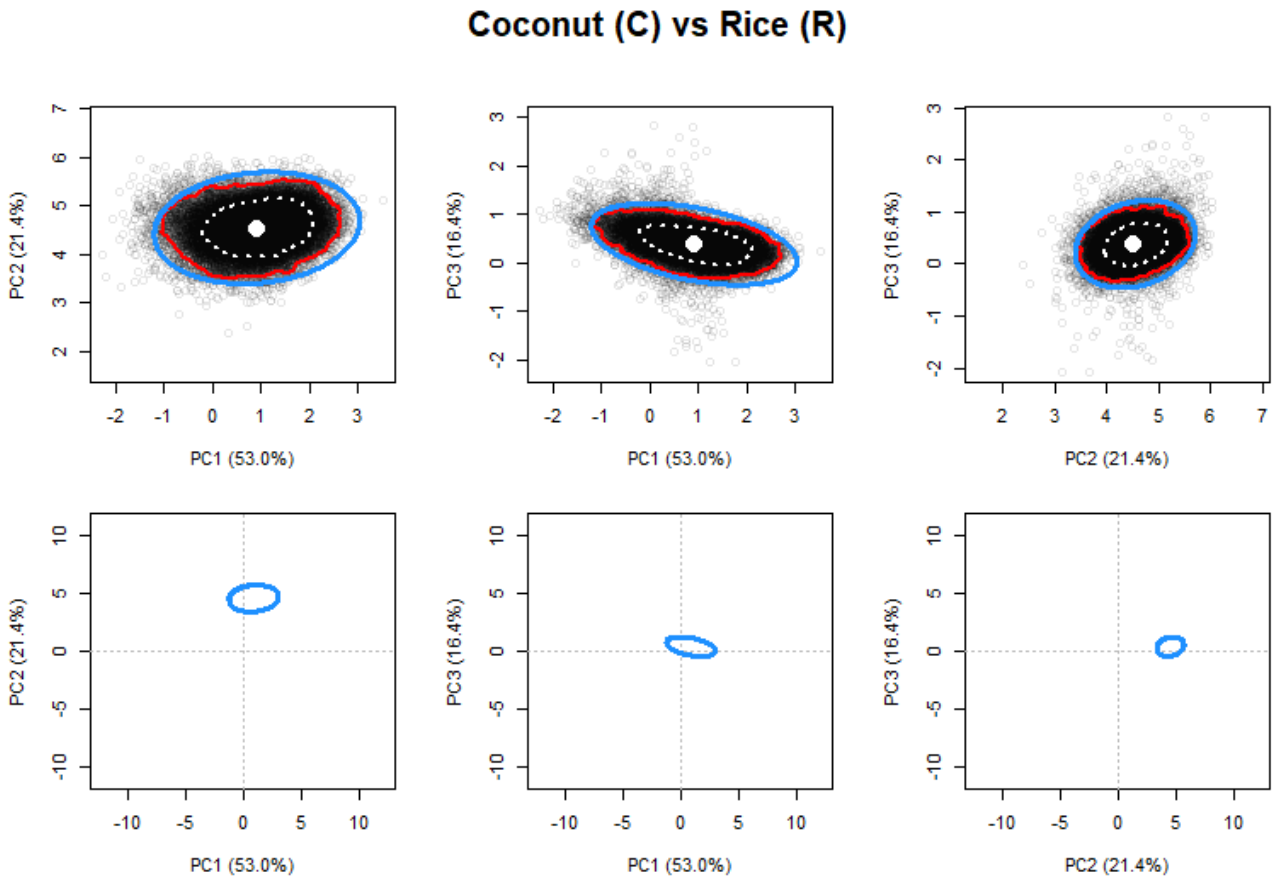
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2i. Coconut (C) vs Peas (P) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

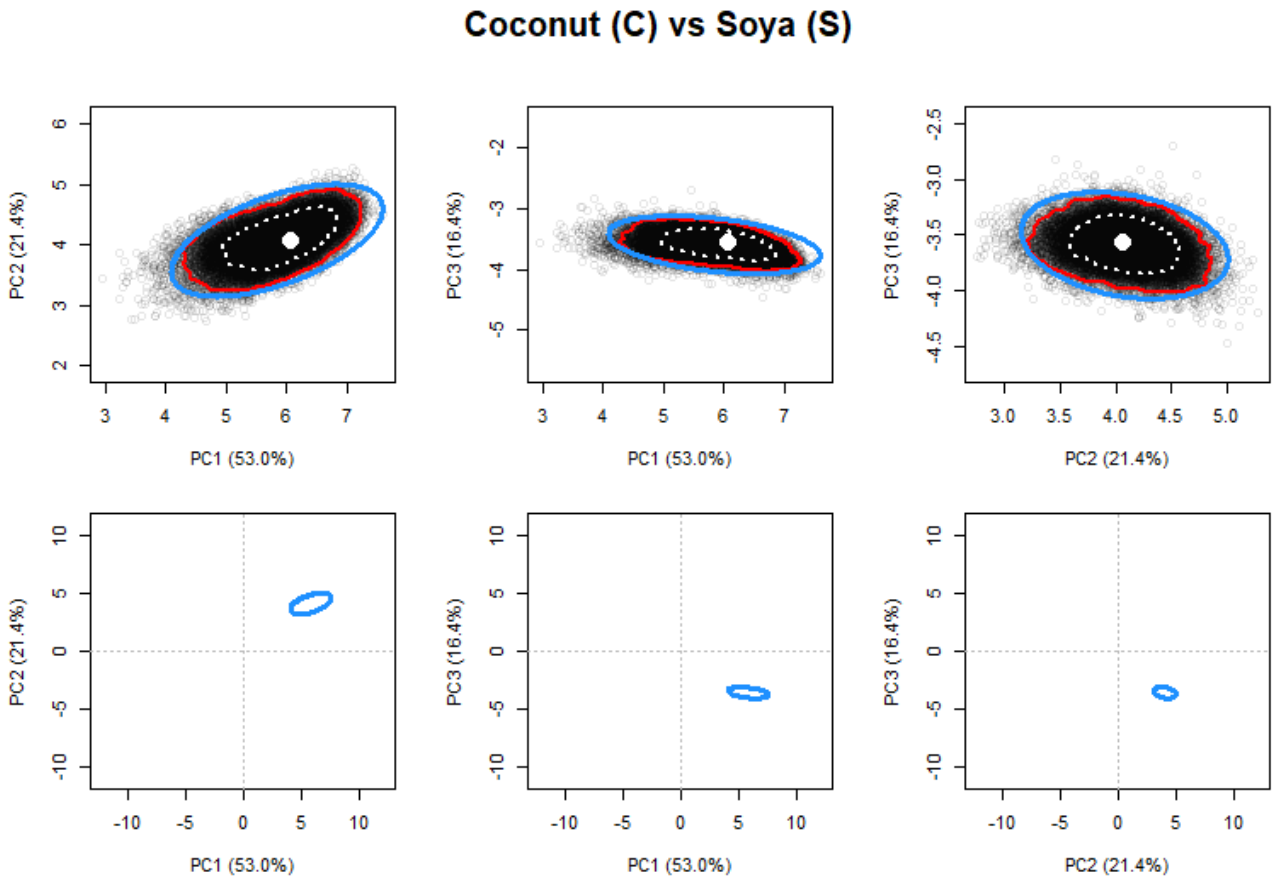
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2j. Coconut (C) vs Rice (R) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

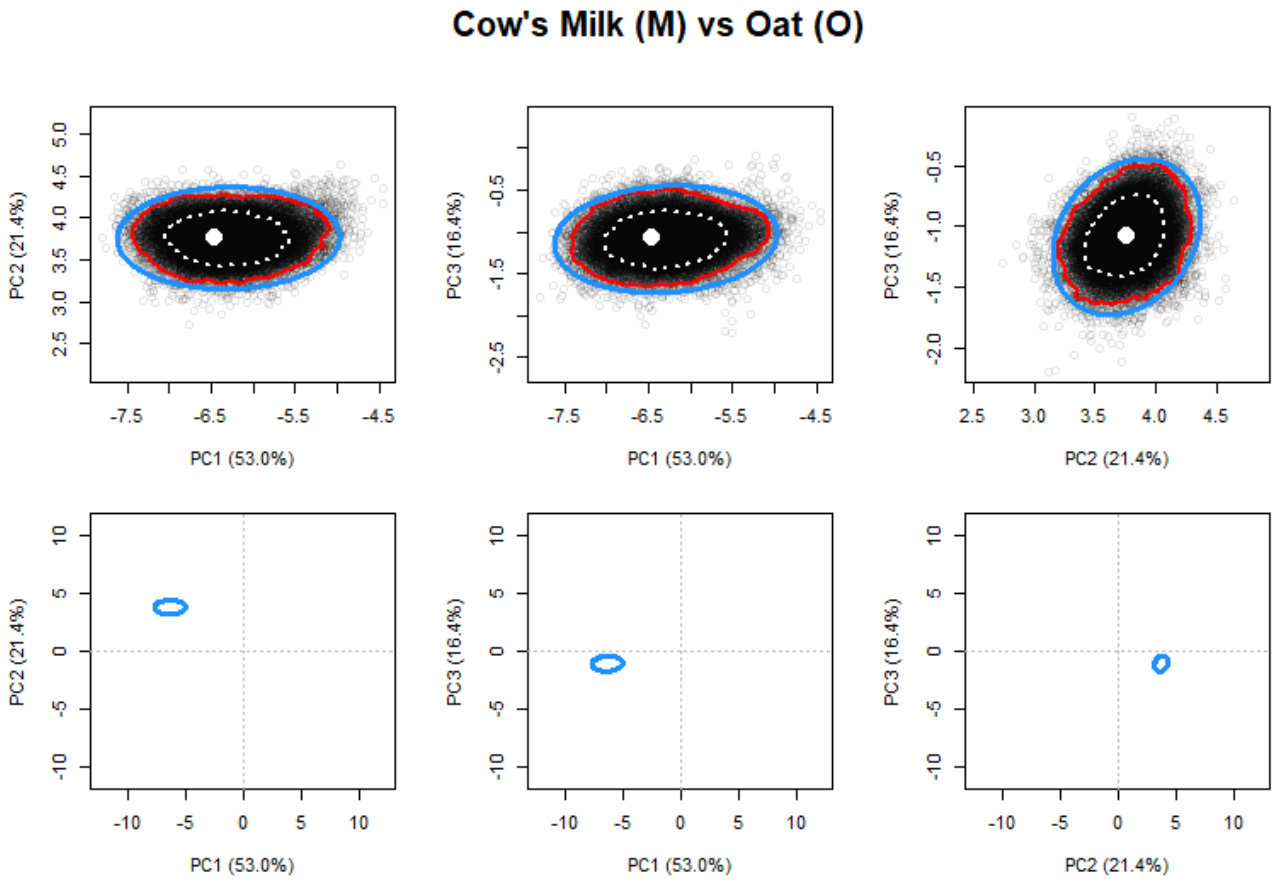
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2k. Coconut (C) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

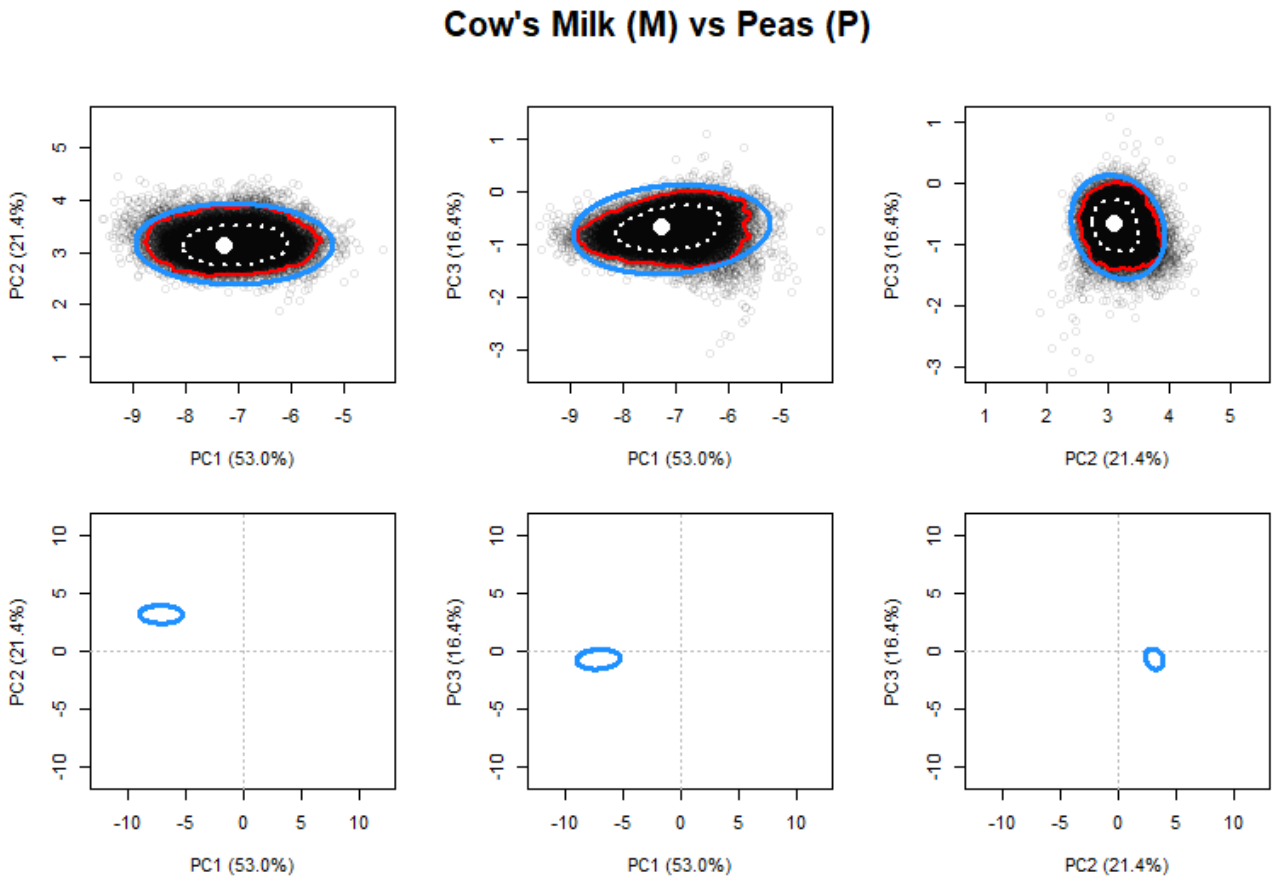
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2l. Cow's Milk (M) vs Oat (O) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

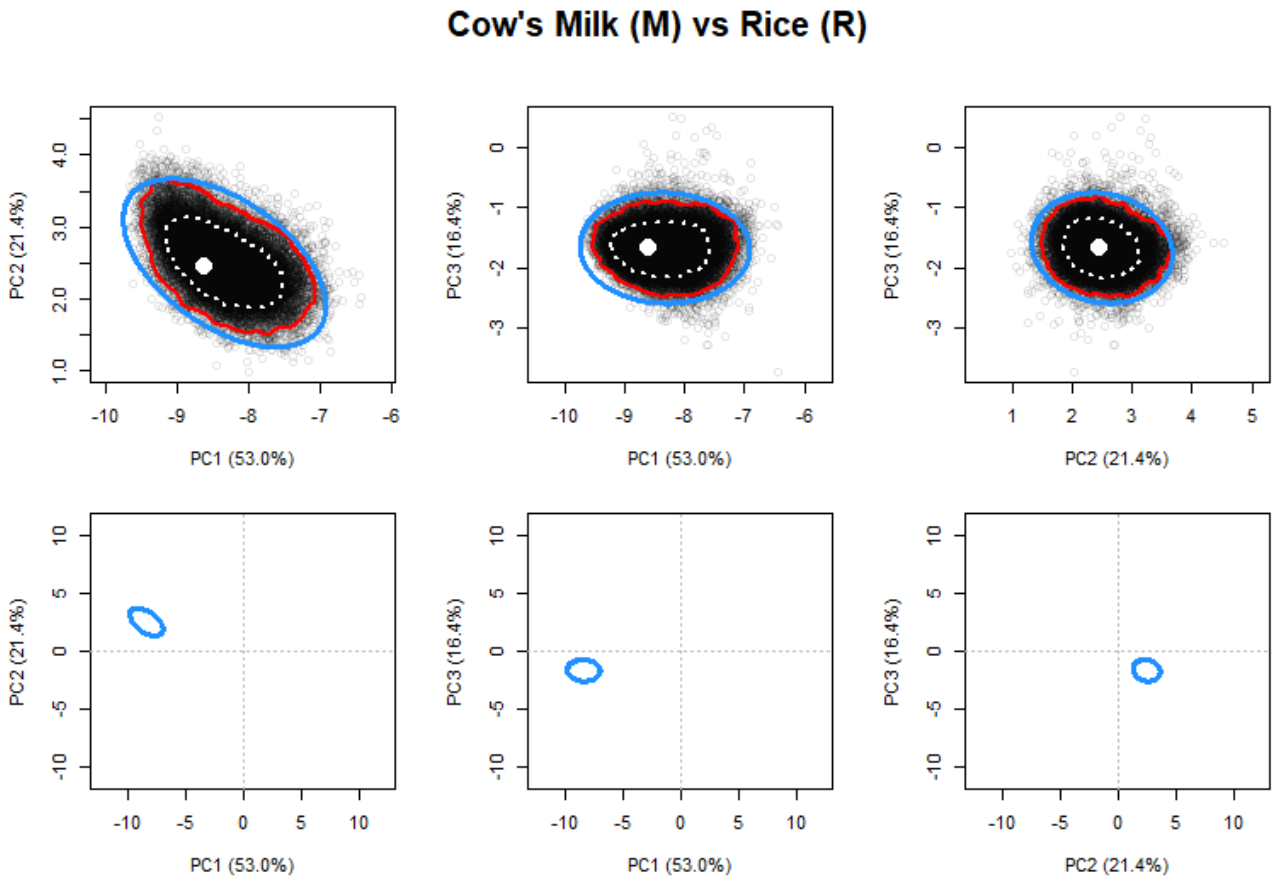
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2m. Cow's Milk (M) vs Peas (P) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

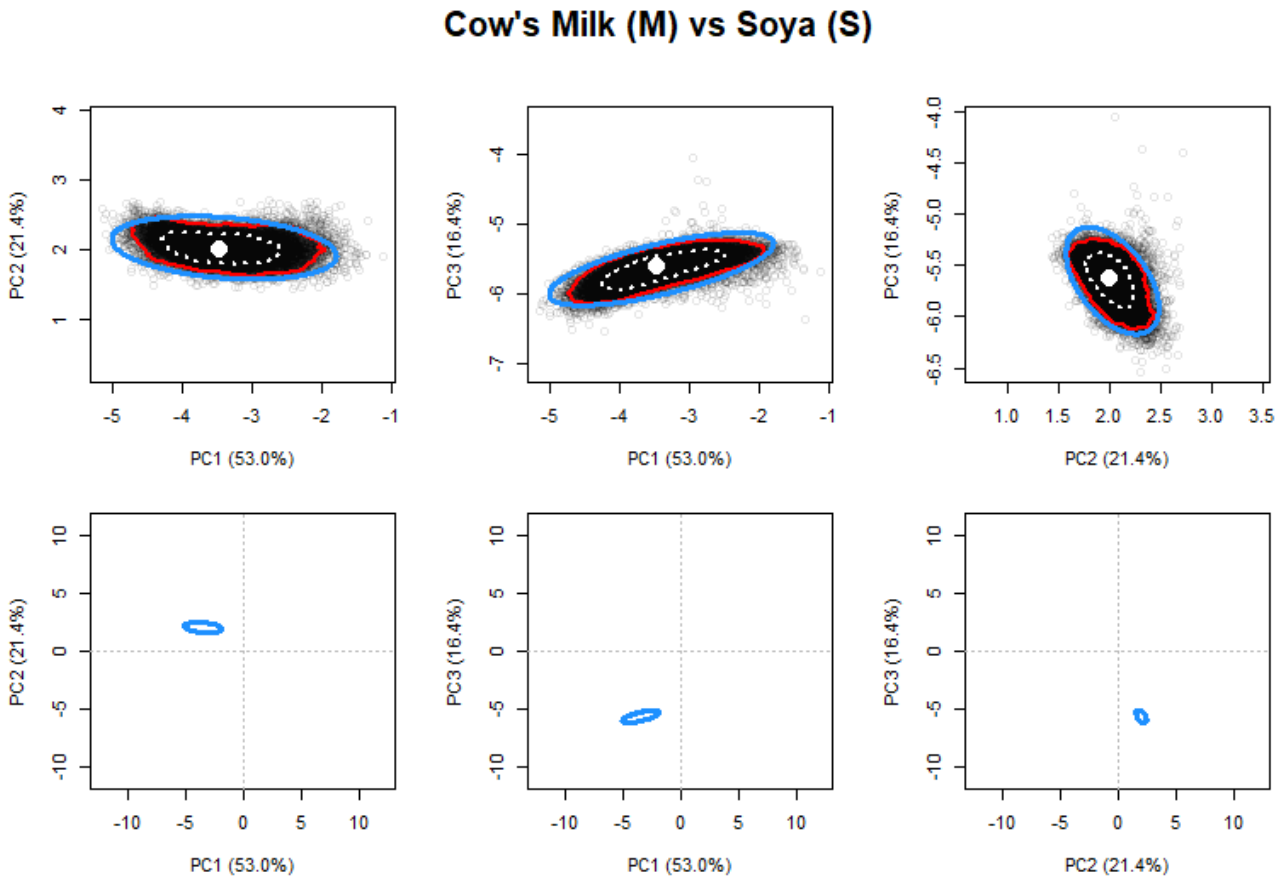
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2n. Cow's Milk (M) vs Rice (R) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

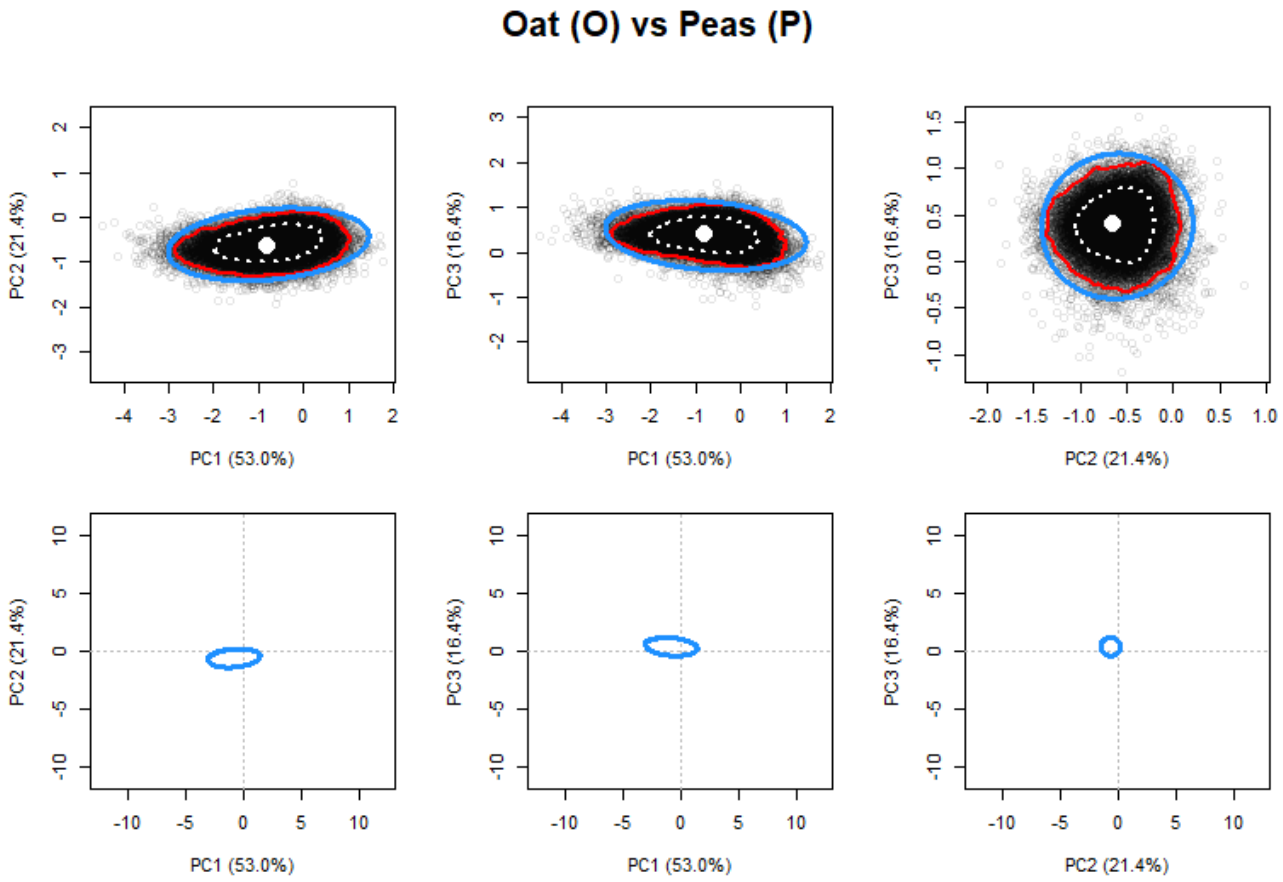
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2o. Cow's Milk (M) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

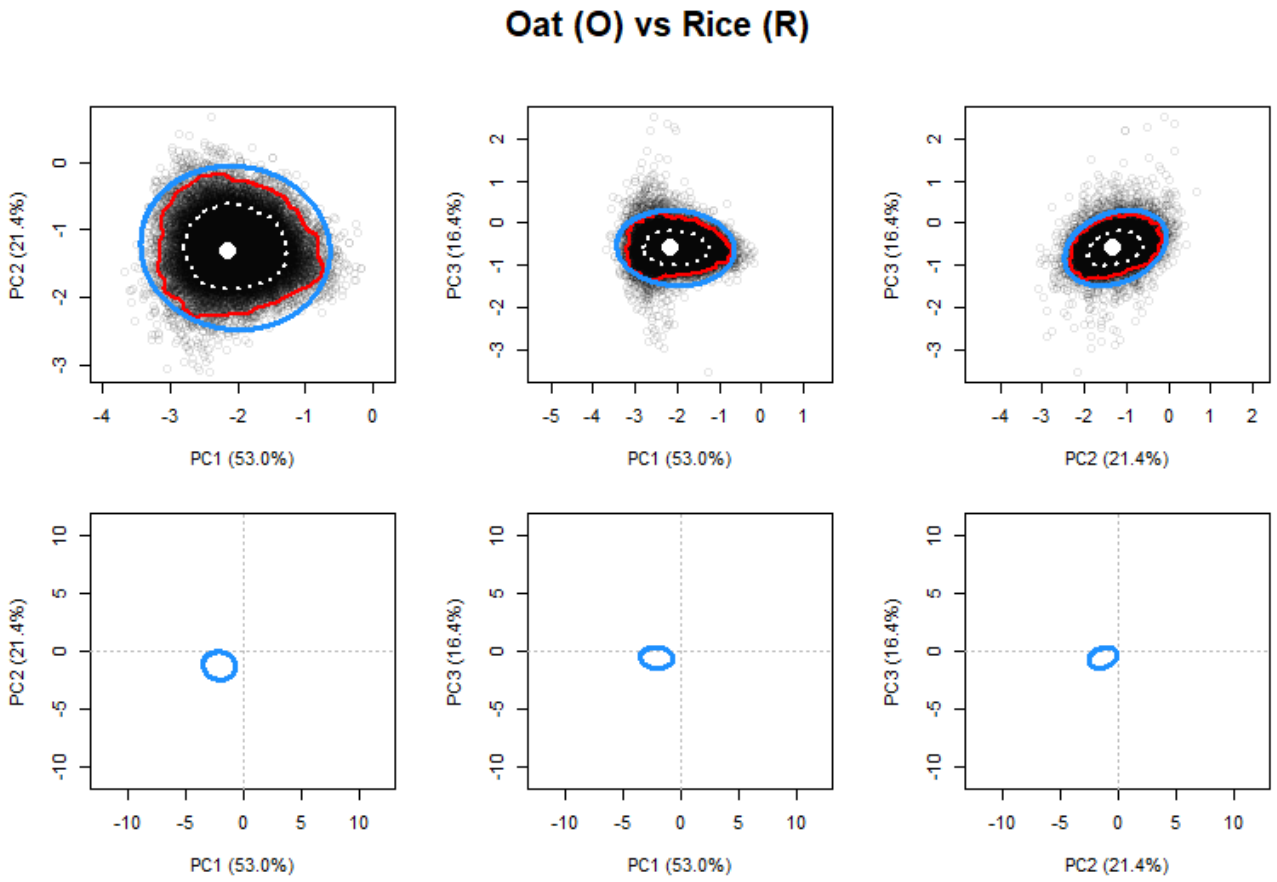
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2p. Oat (O) vs Peas (P) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

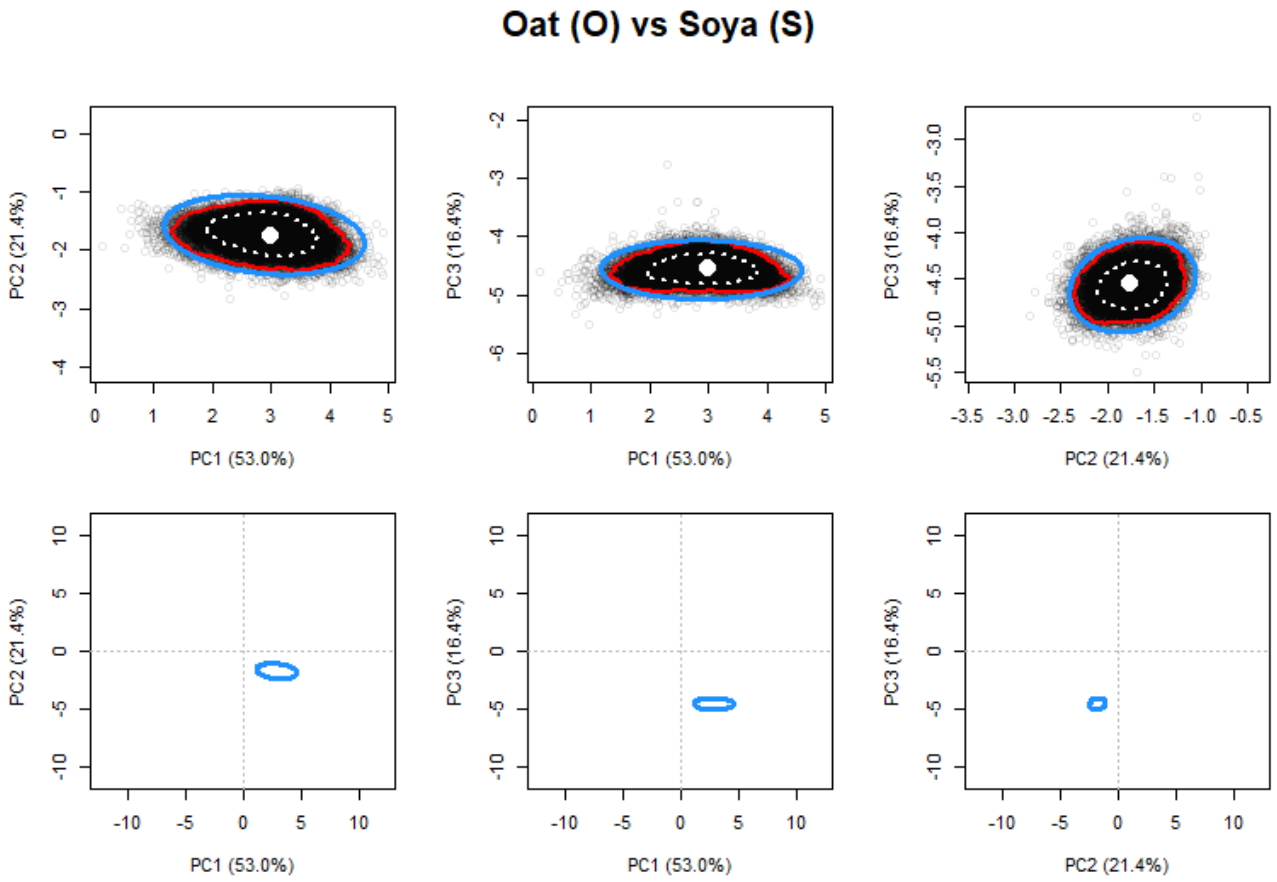
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2q. Oat (O) vs Rice (R) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

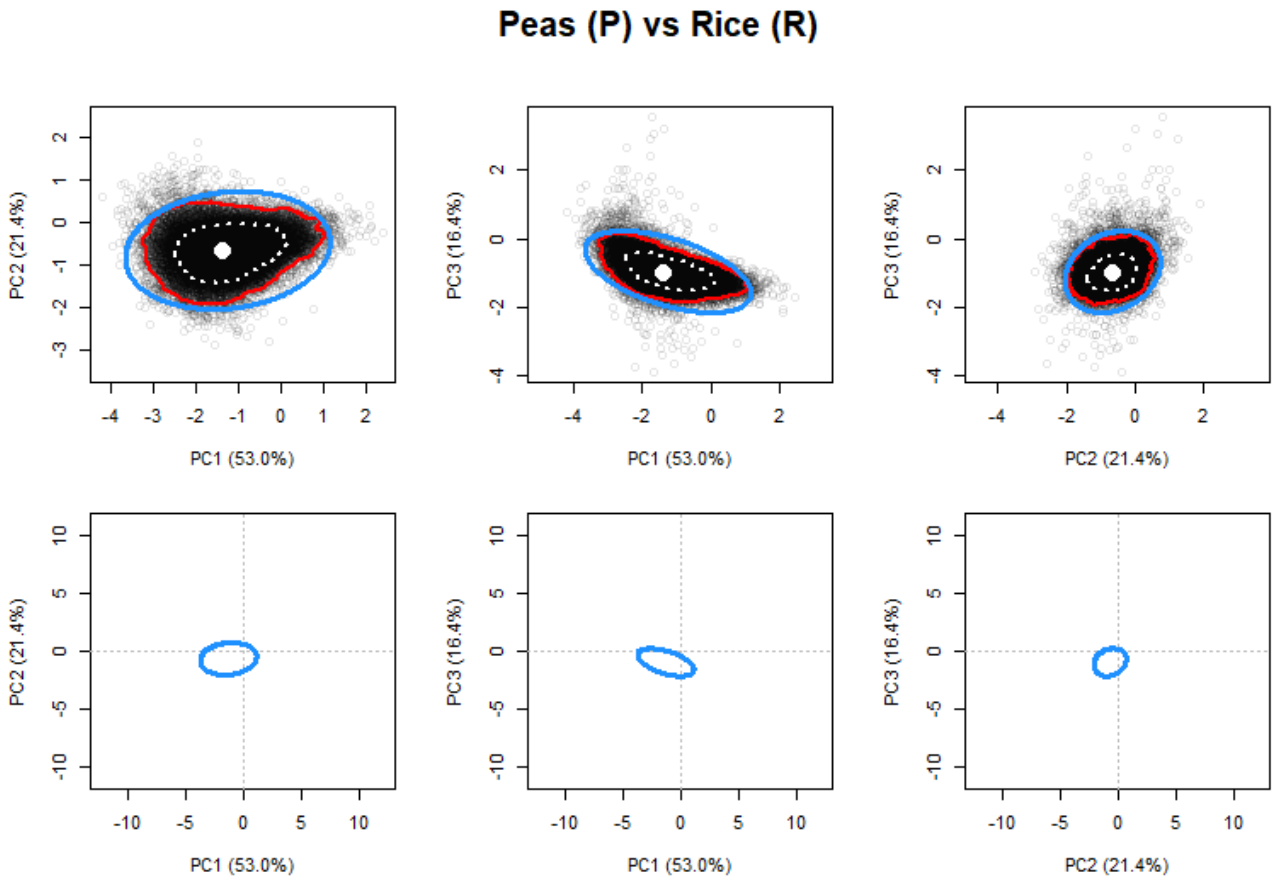
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2r. Oat (O) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

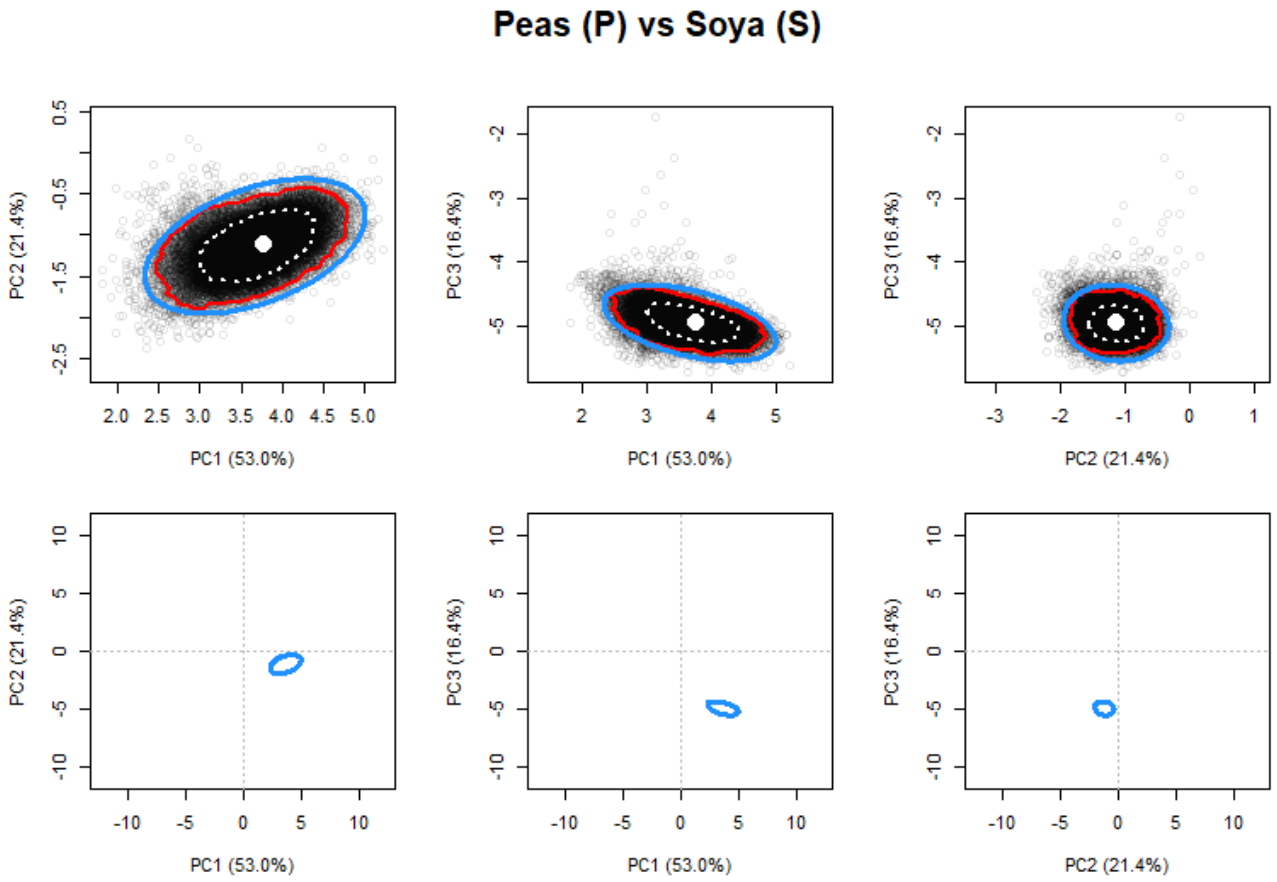
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2s. Peas (P) vs Rice (R) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

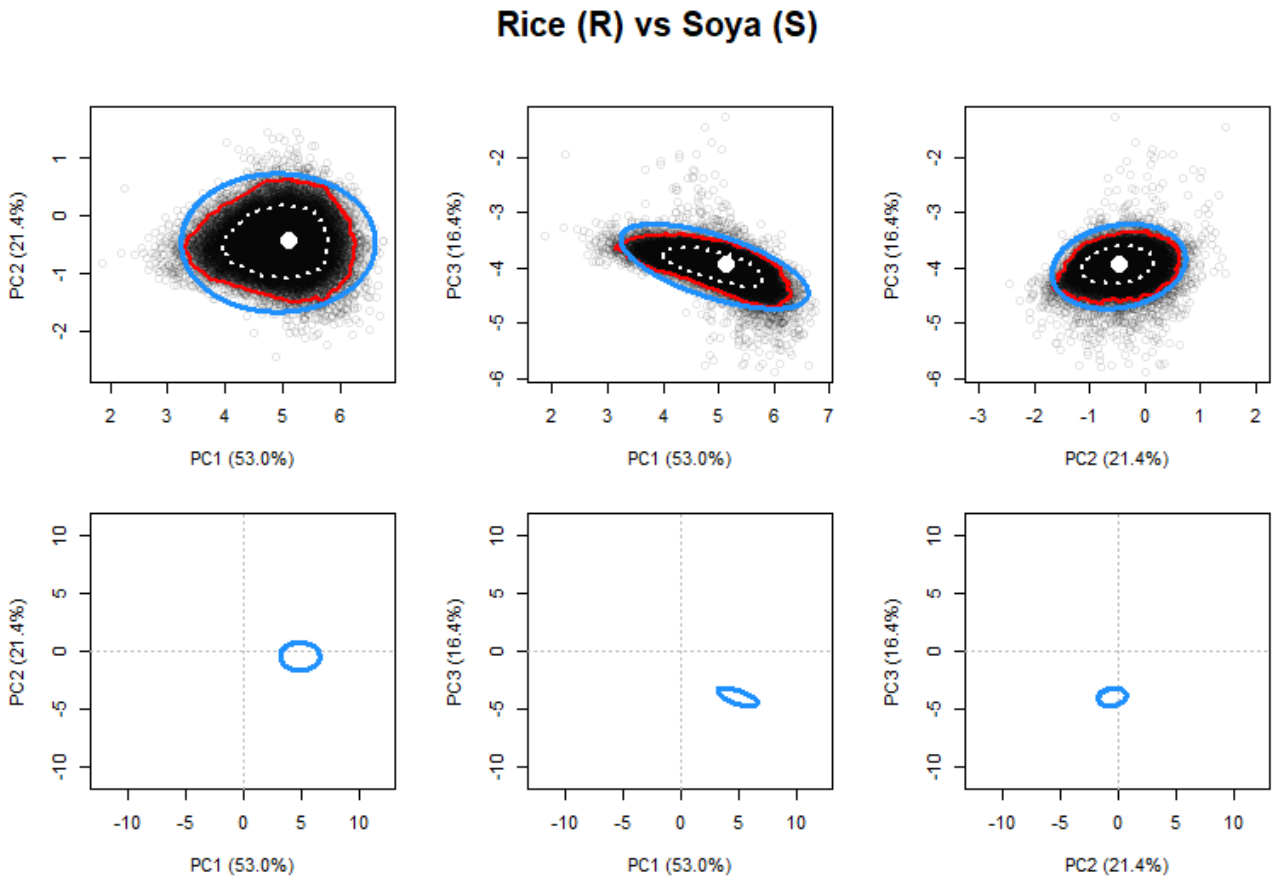
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2t. Peas (P) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

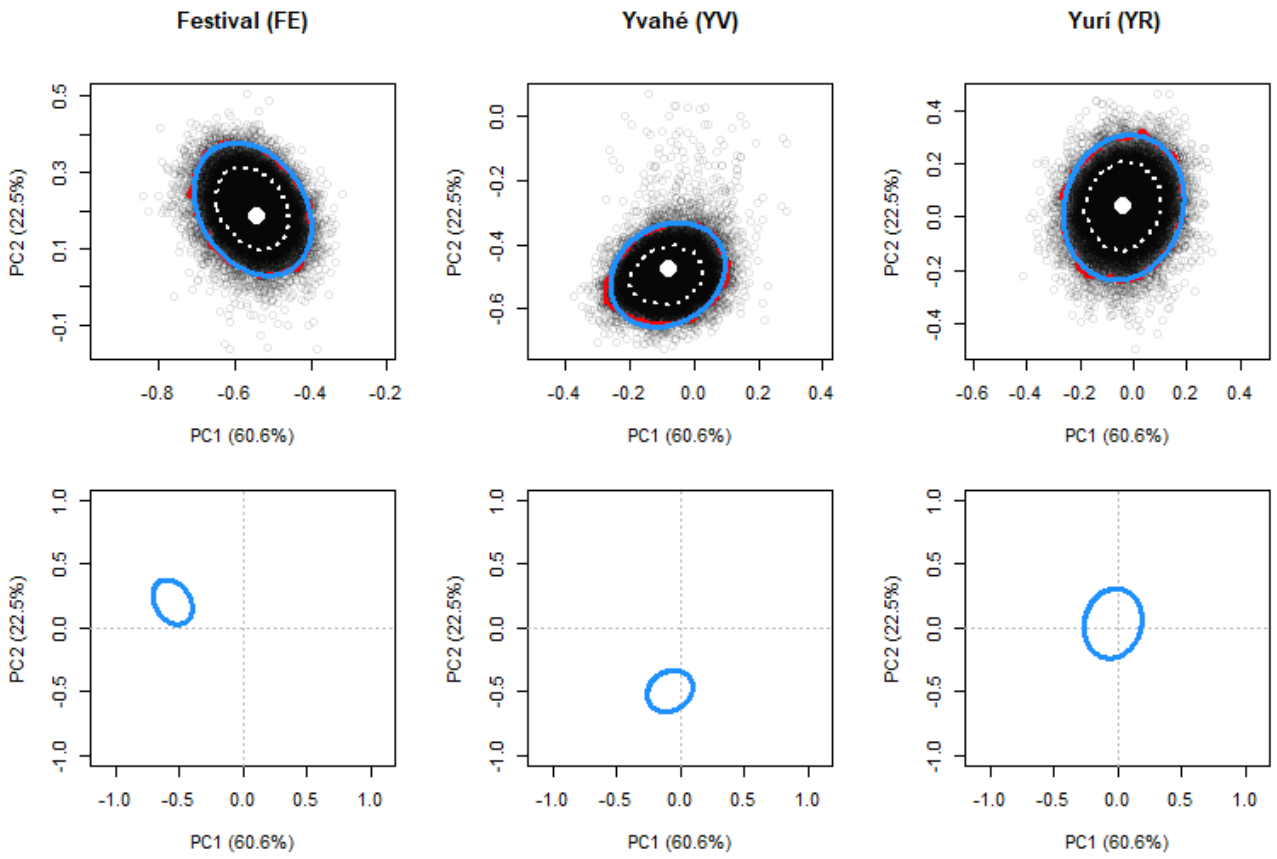
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 2) JC Castura, P Varela, T Næs



Suppl. Fig. S2u. Rice (R) vs Soya (S) beverage paired comparison in PC1 vs PC2 (left column), PC1 vs PC3 (center column), and PC2 vs PC3 (right column) planes. Each plot in the top row zooms in on the real-panel difference scores (white dot) which are overlaid by the projection of 95% confidence ellipsoid (blue line) and contours enclosing 95% (red line) and the 68% (white dotted line) of the highest kernel-estimated density regions for TTB-derived difference scores (grey points). Each plot in the bottom row zooms out to show the projection of 95% confidence ellipsoid onto the plane and the origin.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

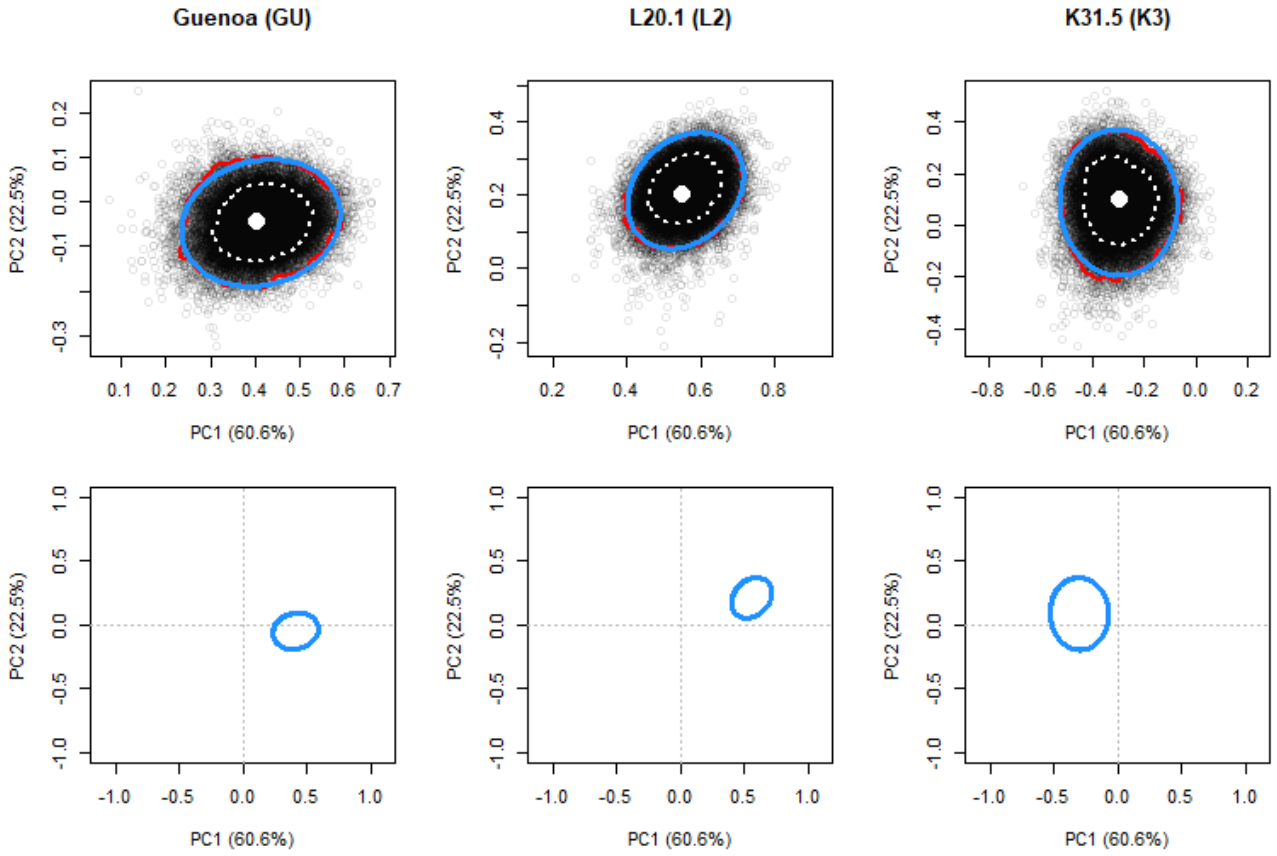
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 3) JC Castura, P Varela, T Næs



Suppl. Fig. S3a. Results for the Festival (FE; left column), Yvahé (YV; center column), and Yurí (YR; right column) strawberry cultivars in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

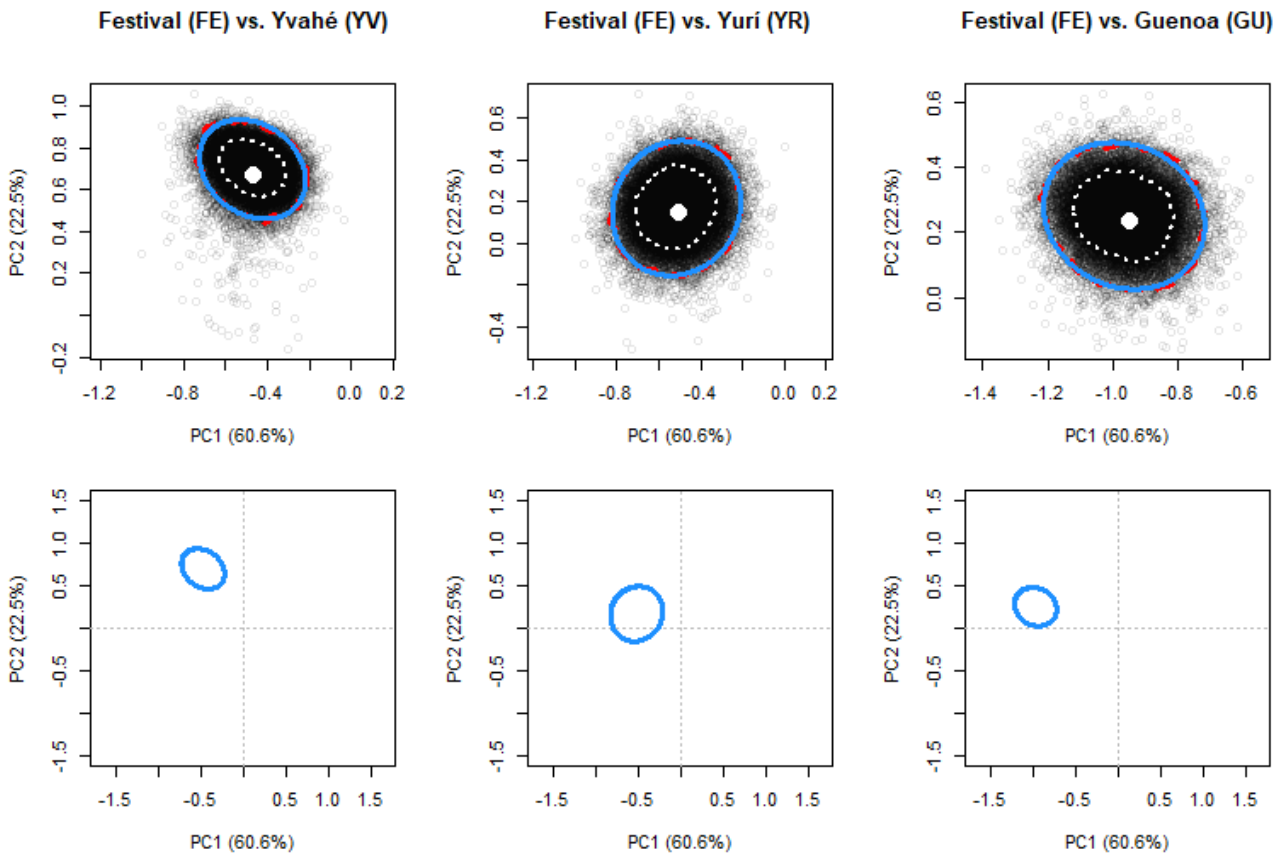
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 3) JC Castura, P Varela, T Næs



Suppl. Fig. S3b. Results for the Guenoa (GU; left column), L20.1 (L2; center column), and K31.5 (K3; right column) strawberry cultivars in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel scores (white dot) and the TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), the 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

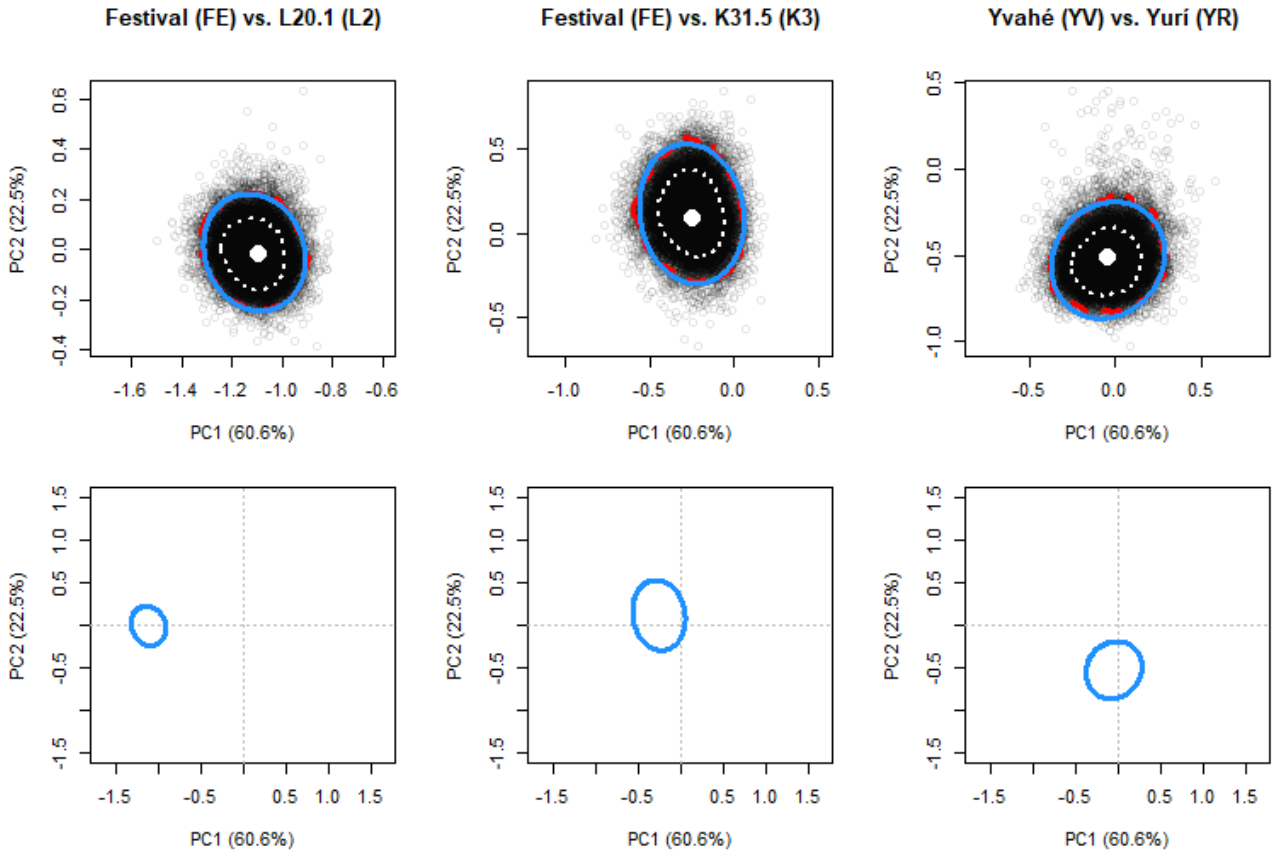
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 4) JC Castura, P Varela, T Næs



Suppl. Fig. S4a. Results for strawberry cultivar paired comparisons of Festival (FE) with Yvahé (YV; left column), with Yurí (YR; center column), and with Guenoa (GU; right column) in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel paired difference scores (white dot) and TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

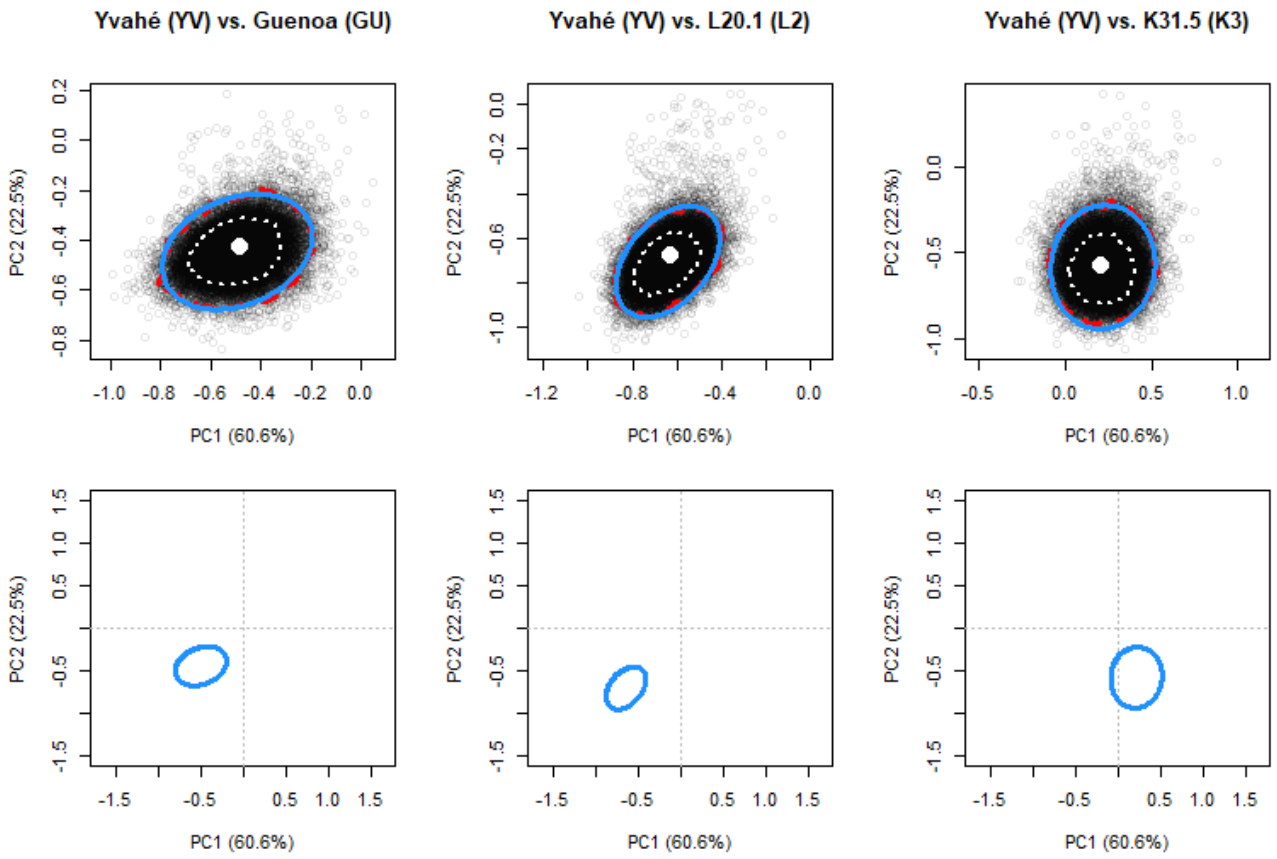
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 4) JC Castura, P Varela, T Næs



Suppl. Fig. S4b. Results for strawberry cultivar paired comparisons of Festival (FE) with L20.1 (L2; left column) and with K31.5 (K3; center column), and of Yvahé (YV) vs Yurí (YR; right column) in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel paired difference scores (white dot) and TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

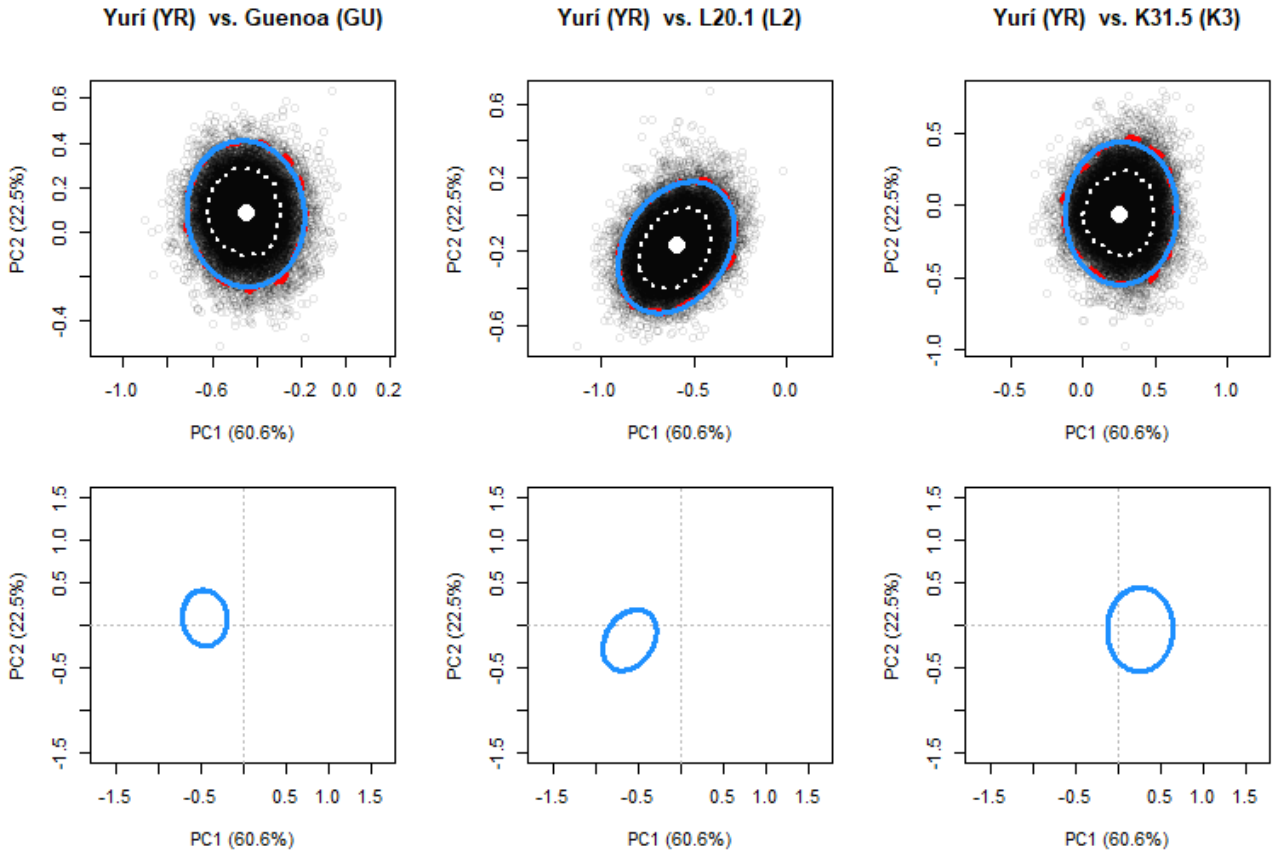
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 4) JC Castura, P Varela, T Næs



Suppl. Fig. S4c. Results for strawberry cultivar paired comparisons of Yvahé (YV) with Guenoa (GU; left column), with L20.1 (L2; center column), and with K31.5 (K3; right column) in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel paired difference scores (white dot) and TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

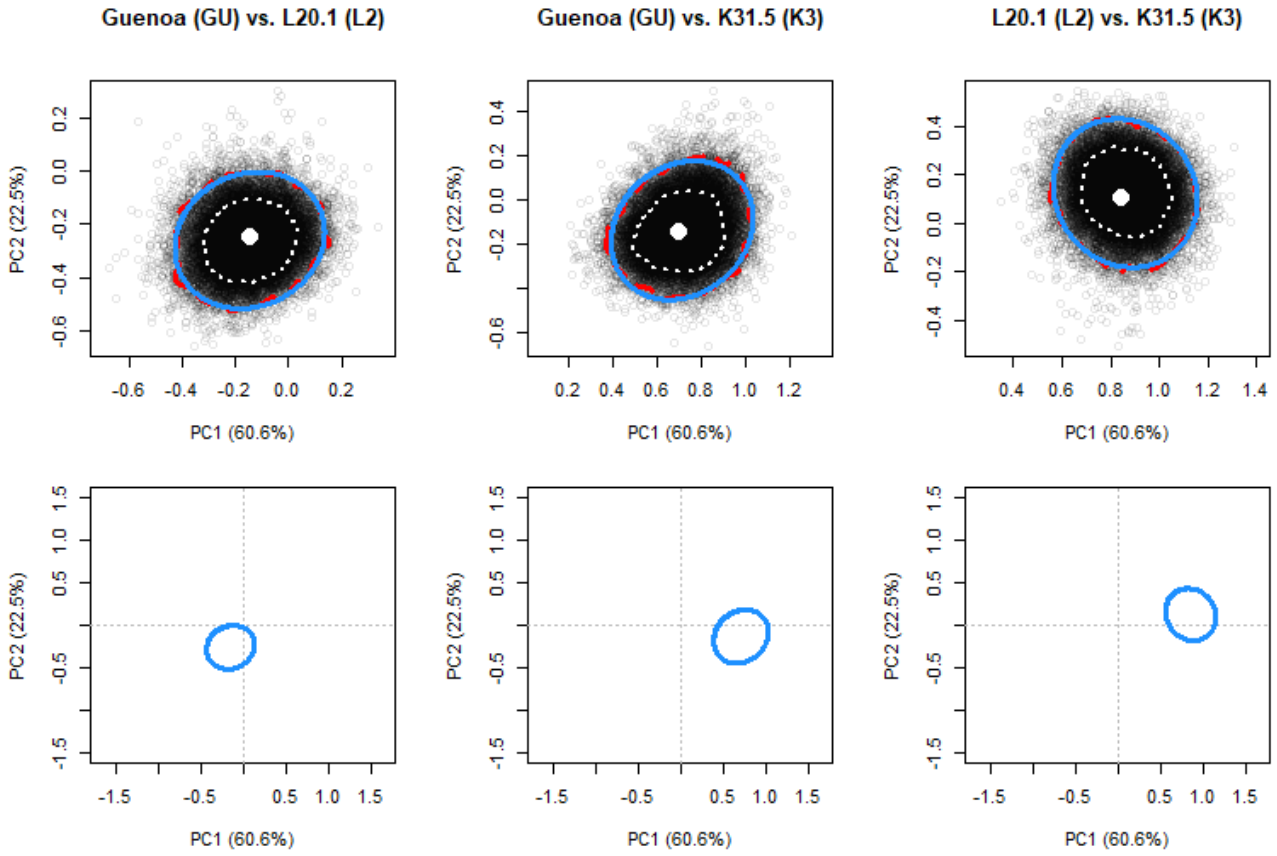
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 4) JC Castura, P Varela, T Næs



Suppl. Fig. S4d. Results for strawberry cultivar paired comparisons of Yurí (YR) with Guenoa (GU; left column), with L20.1 (L2; center column), and with K31.5 (K3; right column) in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel paired difference scores (white dot) and TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843

Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis
(eComponent; Suppl. Fig. 4) JC Castura, P Varela, T Næs



Suppl. Fig. S4e. Results for strawberry cultivar paired comparisons of Guenoa (GU) with L20.1; left column) and with K31.5 (K3; center column), and of L20.1 (L2) vs K31.5 (K3; right column) in the PC1 vs PC2 plane. Each plot in the top row zooms in on the real-panel paired difference scores (white dot) and TTB-derived scores (grey points) which are overlaid by the 95% confidence ellipse (blue line), 95% density contour (red line), and the 68% density contour (white dotted line). The 95% density contour is often obscured by the 95% confidence ellipse. Plots in the bottom row zoom out to show the origin and the position of the 95% confidence ellipse in the plane.

© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
See the published manuscript in *Food Quality and Preference* for all quotes and citations: https://doi.org/10.1016/j.foodqual.2023.104843