Food Research International 75 (2015) 270–280

Contents lists available at ScienceDirect

Food Research International

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

The use of statistical software in food science and technology: Advantages, limitations and misuses

Cleiton Antônio Nunes [a], Verônica Ortiz Alvarenga [b], Anderson de Souza Sant'Ana [b], Jânio Sousa Santos [c], Daniel Granato [c,*]

[a] Department of Food Science, Federal University of Lavras, CP 3037, 37200-000 Lavras, Minas Gerais, Brazil
[b] Department of Food Science, University of Campinas, Campinas, São Paulo, Brazil, Rua Monteiro Lobato, 80, 13083-862 Campinas, São Paulo, Brazil
[c] Department of Food Engineering, State University of Ponta Grossa, Av. Carlos Cavalcanti, 4748, 84030-900 Ponta Grossa, Brazil

ARTICLE INFO

Article history:
Received 23 February 2015
Received in revised form 7 June 2015
Accepted 8 June 2015
Available online 11 June 2015

Keywords:
Applied statistics
ANOVA
Chemometrics
Software

ABSTRACT

Strict requirements of scientific journals allied to the need to prove the experimental data are (in)significant from the statistical standpoint have led to a steep increase in the use and development of statistical software. In this aspect, it is observed that the increasing number of software tools and packages and their wide usage has created a generation of 'click and go' users, who are eagerly destined to obtain the *p*-values and multivariate graphs (projection of samples and variables on the factor plane), but have no idea on how the statistical parameters are calculated and the theoretical and practical reasons he/she performed such tests. However, in this paper, some published examples are listed and discussed in detail to provide a holistic insight (positive points and limitations) about the uses and misuses of some statistical methods using different available statistical software. Additionally, a brief description of several commercial and free statistical software is made highlighting their advantages and limitations.

1. Introduction

It is noteworthy that the use of mathematical and statistical methods, including chemometrics and many other statistical methods/algorithms, in food science and technology has increased steeply in the last 20 years (Fig. 1), and this trend is clearly followed by the development of different algorithms and computational software that facilitates the widespread use of those methods. This trend can be attributable to the low cost of computers and the increasing capacity of processing techniques to analyze complex and high volumes of experimental results. In addition, there is a concern by software developers in providing computational packages with user-friendly interfaces. In this sense, mathematicians and statisticians have optimized and developed new methods to solve problems in different areas, such as medicine, chemistry, agronomy, biology, food science and technology, among others. These methods or models have been implemented in computational packages (software) and scientists have been benefited from their use as the manual calculation is time-consuming and usually imprecise (for large datasets). As well outlined by Granato and Calado (2014), it is still common to find researchers using inefficient statistical methods to analyze experimental results, especially experiments that do not follow an experimental design (when the design would allow much better understanding of the generated data). In addition, as well noted by Passari, Soares, and Bruns (2011), many researchers still have difficulties in understanding and interpreting crucial statistical concepts and results from computational software. This is an aspect of concern as contemporary issues demand the use of multidisciplinary approaches, which are usually based on the use of computational software and on the understanding of statistical and modeling concepts.

Computational tools available can be used not only to run statistical analysis such as univariate and bivariate tests as well as multivariate calibration and development of complex models, but also to run simulations of different scenarios considering a set of inputs or simply making predictions for specific data sets or conditions. Conducting a quick search in the most reputable scientific databases (Pubmed, ScienceDirect, Scopus), it is possible to observe that statistical methods have gained a huge space in different areas (Fig. 1B), and there is a great interest in using these approaches in many scientific fields (i.e., microbiology, nutrition, optimization of products and processes, food chemistry, food technologies). In addition, several modeling tools for predictions of microbial behavior, stability of chemical compounds during food processing and for simulations are available and become well known and increasingly applied in the field (Corradini and Peleg,

* Corresponding author. Tel.: +55 42 3220 3268.
E-mail address: granatod@gmail.com (D. Granato).

http://dx.doi.org/10.1016/j.foodres.2015.06.011
0963-9969/

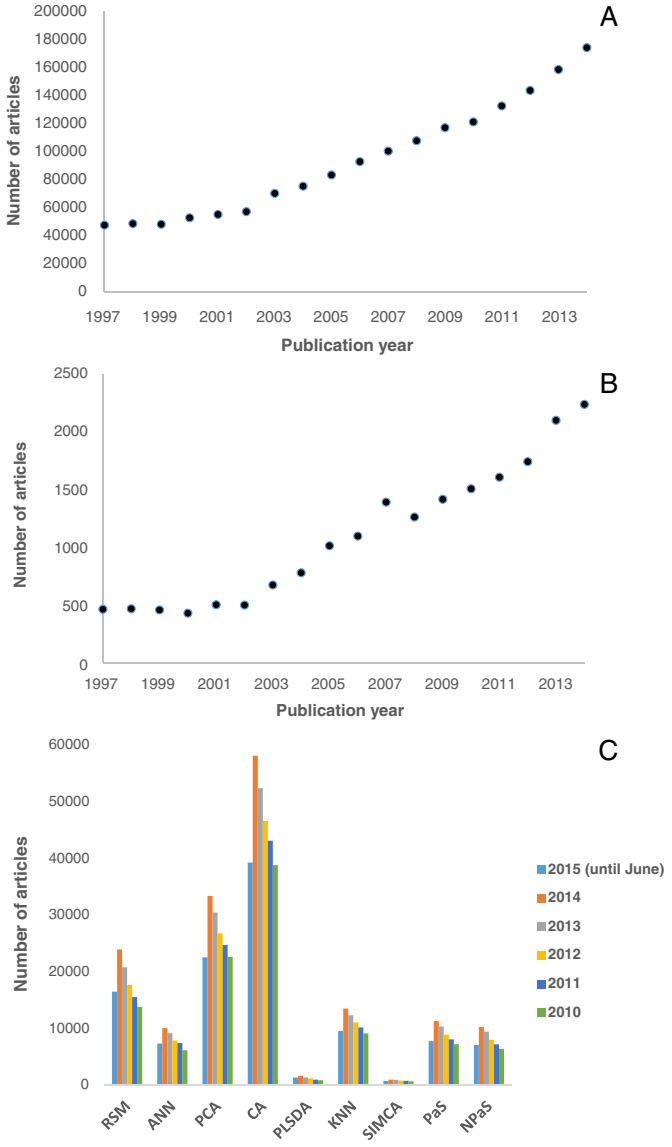

Fig. 1. Number of published articles in ScienceDirect from 1997 to 2014 using the term 'statistical analysis' (A), 'chemometrics' (B), and some other statistical techniques used in data treatment (from 2010 to June 2015) (C). Note: RSM = response surface methodology; ANN = artificial neural networks; PCA = principal component analysis; CA = cluster analysis; PLSDA = partial least-squares discriminant analysis; KNN = k-nearest neighbors; SIMCA = soft independent modeling of class analogy; PaS = parametric statistics; NPaS = non-parametric statistics.

2006, van Boekel, 2008, Tenenhaus-Aziza and Ellouze, 2015). These facts highlight the relevance of statistical and computational tools in the advancements and understanding of important aspects dealing with quality and safety problems of foods.

Despite recent increased awareness of the benefits of using these tools, it is important to mention that they have not been widely applied by all scientists due to many technical reasons. Among them, we could highlight the lack of explanation on how these techniques can be used and how they should be used with different types of experimental data. Another debatable issue is that most statistical books are aimed to mathematicians and chemists, which means that the technical language used by authors is somewhat difficult to understand and a deep knowledge on basic statistics and mathematical modeling are required to use these books as references. Notwithstanding, it is very important to stand out the importance of using accurate statistical techniques of data from food science, including sensory evaluation, development of food products/processes and food microbiology, as well as in quality control in food industries. In this sense, the concept of statistics applied in food research is widespread, but concern exists on how the statistical analyses have been carried out, especially when computational software is used, and how researchers have interpreted and reported the results. Based on this concern, the objective of this paper is to give emphasis on the use and misuses of several commercial and free statistical software for conduction of statistical analysis. Nonetheless, in this article, we will focus on the aspects related to advantages, limitations and applications of some statistical software used in food science and technology.

2. Analysis of the literature: does the use of software represent a good strategy for the food scientist?

There is no right answer to this general question as the use of software may represent an excellent and required strategy in some cases, such as multivariate calibration using partial least-square regression, but may be unnecessary to calculate the means for three measurements, for example. Therefore, answering this question mainly depends on the type of study and characteristics of the generated data (*i.e.*, quantitative, qualitative).

2.1. Statistical software: concepts, advantages and limitations

Many statistical software are available for the data analysis. Although there are a large variety of free software (freely downloaded and used in their fully-functional mode), such as OpenStat, SOFA, EpiInfo, ViSta, and PSPP, many statistical software need to be purchased in order to be used, and licenses, overall, need to be upgraded from time to time, which is somehow a limiting factor for young scientists and students. Some of these commercial software are Statistica, Stata, Unscrambler, Minitab, SAS, Pirouette, Design-Expert, Matlab, MathCad, Statistical Package for the Social Sciences = SPSS, Origin, Microsoft Excel, among others.

Because of costs involved in acquiring those commercial statistical packages, currently there is an increasing and encouraging demand for the use of free software. Such programs include public domain and may be used for statistical and/or mathematical analysis. Some of these packages are: *R* and [PSPP], composed of many researchers from all over the world, IDAMS (UNESCO), EpiInfo (Centers for Disease Control and Prevention, USA), BROffice (Brazil), Chemoface and SensoMaker (UFLA, Brazil), and Action (Statcamp, Brazil), among others. The advantages of using these programs in food research are clear: graphs with a high definition can be generated in seconds, simulations of results can also be performed using a small amount of time, software represents an auxiliary educational means of teaching theoretical requirements to students with varying knowledge of mathematics and statistics, and complex algorithms and equations are also made in seconds.

All these free and commercial mathematical and statistical software are able to perform a large amount of calculations as well as algorithms to solve different types of problems in food research, such as descriptive and inferential statistics, design of experiments, and multivariate statistical techniques. Additionally, they are widely used to build databases, fix spreadsheets, design and analyze experiments, and optimize products and processes (Teófilo & Ferreira, 2006). Among the main functions and use of statistical programs, the most common methods are: cross-confirmation of data in different databases, recoding of variables for analysis, descriptive statistics including frequency calculation, calculation of position measurements (average, median, quartiles, mode), measures of dispersion (range, variance, standard error, standard deviation, coefficient of variation), analysis of data normality and homoscedasticity, parametric statistics, non-parametric statistics, and multivariate statistics.

Although many of these methods are easily performed by statistical packages due to their friendly interface, before using them, it is a

reasonable idea to know how the methods in the different areas of statistics (inferential, descriptive, probabilistic) actually work, bearing in mind their advantages and limitations, ensuring the right choice. Then, this specific software can be employed to the best advantage to choose the most suitable test and always taking into account all the necessary assumptions are met, so that the right conclusions can be stated. It is also important to stress that the scientist needs to know how these packages make the calculations.

In this sense, no statistical/mathematical software has the ability to perform all the methods and tests needed by a scientist. Thus, reading the tutorial guide of each software is highly recommended. In general, usually two or more software are required to analyze different experimental data. Then, limitations in using these software also need to be stated: lack of interest of the scientist in understanding how the statistics are calculated, many packages do not work if there are missing data, by using multiple programs may lead to conflicting outputs, and some of these packages, such as *R* and EpiInfo, require intermediate knowledge of programming, which limits their use.

Table 1 briefly describes the most commonly used statistical packages in food science, technology and engineering, with special attention to accessibility, limitations and overall positive characteristics and features.

As the use of statistical and mathematical methods is usually divided into univariate (graph analysis, descriptive statistics), bivariate (correlation, linear regression analysis) and multivariate methods (exploratory, class-modeling, and classification methods), we listed some characteristics of such methods in the following sections.

2.2. Univariate and bivariate methods

In food science and technology, statistics may be used for different purposes: design of experiments, modeling of response variables using response surface methodology, and recently, the application of multivariate statistical methods has been more widely spread (Alezandro, Granato, Lajolo, & Genovese, 2011; Besten et al., 2013). In this sense, as mentioned by Granato, Calado, and Jarvis (2014), undoubtedly, descriptive analysis (basic statistics: means, median, correlation, linear regression, standard deviation, among others) followed by inferential statistics (*i.e.*, analysis of variances and multiple comparison of means) are the most frequently used methods.

With respect to this issue, most papers do not demonstrate in the 'Material and Methods' section if the statistical assumptions prior to the application of inferential methods were evaluated, that is, *normality* and *homoscedasticity*. Testing whether a set of data (usually $n \geq 4$ observations) follows a normal distribution is of major relevance for statistical procedures, namely parametric tests. Normality can be assessed visually, through histograms and plots (such as boxplot, Q–Q and P–P plots) and through normality tests such as Kolmogorov–Smirnov (K–S) test, Anderson–Darling test and Shapiro–Wilk test, among others (Ghasemi & Zahediasl, 2012; Schoder, Himmelmann, & Wilhelm, 2006). Further, homoscedasticity is an important feature for parametric tests when conducting inferential tests (i.e. multiple comparison of means). This is relevant because homoscedasticity of data means that different groups present the similar standard deviations. This is important because otherwise the chances of getting false positives can be higher than the α-value established at the beginning of the experiment ($P < 0.05$ or 5%) (McDonald, 2014). There is no consensus with regard on how to deal with the lack of homoscedasticity (*i.e.*, heteroscedasticity); however, one possible way to reduce the deviation is the use of transformations such as $\log_{10}$, square root, natural log, among others (McDonald, 2014). These approaches have been widely used by food scientists to stabilize variance in data obtained in their experiments (Alber & Schaffner, 1992; Sant'Ana, Franco, & Schaffner, 2012; Schaffner, 1998). Before these aspects are considered, for example, when correlation analysis needs to be carried out, normality of data should be assessed. In this case the question is: how should correlation coefficient be calculated: by using the mean values or all replicates? Other requirements such as number of experimental points should also be considered. The same question applies when linear regression analysis is performed, especially when a calibration curve is plotted. Souza and Junqueira (2005), Granato, Calado, and Jarvis (2014) and Cozzolino (2014) made useful comments on the application of these methods in food research and listed the prerequisites needed to be checked prior to the analysis. Answering the above question is simple: using all replicates to generate regression models or calculate the *r*-value and its significance (*p*-value) makes a more realistic correlation coefficient. Imagine an analyst assessing $n = 5$ apple juices, in triplicate, for antioxidant activity and quantifying the content of total (+)-catechin. The correlation analysis using the mean values for both responses ($n = 5$ each), rendered $r = 0.89$ and $p = 0.02$. However, if the triplicate values for both responses ($n = 15$ each) were used, $r = 0.49$ and $p = 0.08$ were obtained. From this example, it is obvious that the correlation coefficient is dependent on the number of observations. In some fields, such as food microbiology, when dealing with microbial behavior under stress conditions (interface between growth and no-growth), in which variability is high; the number of replicates and data points must be carefully appraised. Therefore, it is advisable to always mention how the *r*-values and *p*-values were calculated.

Many authors do not pay attention to homoscedasticity of data when samples/treatments need to be compared using inferential tests, that is, if the variances among treatments are equivalent from the statistical standpoint. As well outlined by our previous work (Granato, Calado, & Jarvis, 2014), various formal tests can be used to check for homogeneity of variances, such as Hartley, Cochran, Bartlett, Levene, and Brown-Forsythe. Overall, the same test must be applied for all data sets in order to be coherent and consistent, once these tests use different ways to calculate the statistics. For example, consider the following fact: a scientist needs to compare the ascorbic acid content in three different fruit juices (A, B, and C), in which the three independent replicate results (mg/250 mL) were: A (10.55; 10.56; 10.52), B (10.02; 10.10; 10.03), and C (11.25; 11.09; 12.20), respectively. If one applies the Levene test (using Statistica v.7 software), the *F*-value = 12.26 and *p*-value = 0.0076 are obtained, whereas if the Brown-Forsythe test is carried out, *F*-value = 1.41 and *p*-value = 0.3142 are obtained. For the first test, data would be considered heteroscedastic and for the latter test data would be regarded as homoscedastic. Consistency in the use of the same test throughout the study is required.

The development and reengineering of products and optimization of processes in food research is also another important application of statistics. There are two main options to conduct these studies: using a design of an experiment coupled to response surface methodology (Bas & Boyaci, 2007; Bassani, Nunes, & Granato, 2014; Domínguez-Perles, Teixeira, Rosa, & Barros, 2014; Farris & Piergiovanni, 2009; Granato, Castro, Ellendersen, & Masson, 2010; Granato, Grevink, Zielinski, Nunes, & van Ruth, 2014) or by using the 'one variable at a time' approach, that is, the application of random levels of selected factors (*i.e.* ingredients or process parameters) (Bassett et al., 2014; Boobier, Baker, & Davis, 2006; Granato, Castro, Piekarski, Benincá, & Masson, 2011; Haj-Isa & Carvalho, 2011). However, when the researcher uses the 'one variable at a time' approach to develop products and processes, the main effects of factors and their interactions cannot be calculated and the relationship between the response and the factors cannot be estimated. Another disadvantage of this approach is the high number of experiments required to conduct the research, leading to a more time-consuming methodology and investment in reagents/chemicals. More importantly, when this approach is used, the solution obtained usually does not represent the 'optimal' conditions to obtain the desired conditions. In order to avoid these technical drawbacks, Bezerra, Santelli, Oliveira, Villar, and Escaleira (2008) and Granato and Calado (2014) explain in detail the types of experimental designs and their practical applications together with the steps on how to analyze data. It is important to stress that in some areas of food science, such

Table 1
Brief description of some statistical packages used in food science and technology research.

Software	Website	Accessibility	Advantages and limitations
Microsoft Excel	http://products.office.com/en-us/excel	Commercial, no internet access	- *Advantages*: - Easy to conduct basic statistics (descriptive and inferential) - Possible to analyze data using the add-in 'Data Analysis' - Correlation and linear regression analysis are included - F-test for two variances - Limitations: - It does not include normality and homoscedasticity tests - It does not include multivariate statistical methods - Graphical representations are not of high-quality - Correlation analysis cannot be evaluated by methods other than Pearson's method
Statistica	www.statsoft.com	Commercial, no internet access	- Advantages: - Descriptive and inferential analysis can be easily carried out - Multivariate exploratory and classification techniques, including principal component analysis, clustering techniques, and data mining techniques (including neural networks) are included - Multiple and linear regressions can be performed with detection of outliers and residual analysis - Possibility to perform non-parametric analysis - Quality control charts can be made - Software has a user's guide manual explaining the rationale of each test - Graphical representations are of high quality - Limitations: - Very expensive to acquire add-ins for all statistical analysis; - User must know *a priori* the principle of the tests - For multivariate analysis, the software only performs auto-scaling as data pre-processing - Up to version 9, Welch test (ANOVA procedure for non-homoscedastic data) cannot be applied
Pirouette	www.infometrix.com	Commercial, no internet access	- Advantages: - Graphical outputs are excellent - Multivariate exploratory, class-modeling and classification techniques can be performed - Models can be validated using leave-one-out cross-validation - Various pre-processing methods can be applied and analysis are made simultaneously - Limitations: - It is not easy to perform external validation: the user must separate data randomly using other packages (Excel is an option) to perform the external-validation - Software dedicated to multivariate statistical techniques
Matlab	www.mathworks.com/products/matlab/	Commercial, no internet access	- Advantages: - A large number of tests (uni- and multivariate) can be implemented - Users can implement algorithms in order to improve the analyses - The software has a very elaborated and useful help - The help presents illustrative examples - Features can be improved by additional toolboxes - Users can contribute with free code/toolboxes *via* File Exchange (http://www.mathworks.com/matlabcentral/fileexchange/) - High quality graphics can be obtained - Limitations: - It requires intermediate knowledge of programming, limiting the number of users
Action	www.portalaction.com.br	Free, downloadable	- Advantages: - Both descriptive and inferential (both parametric and non-parametric) tests can be used - Homoscedasticity and normality tests are included - Software has a user's guide manual explaining the rationale of each test and how to perform the analysis - Quality control charts and statistics regarding metrology can be made - Design of experiments (response surface methodology) is included - Limitations: - Only multivariate analysis of variances (MANOVA) and cluster analysis can be performed as multivariate statistics - It requires the use of Microsoft Excel to be run.
R-Project	www.r-project.org/	Free, downloadable	- Advantages: - It incorporates all of the standard statistical tests, models, and analyses - It is able to perform both univariate or multivariate tests - Graphical capabilities surpasses most other statistical packages - It is open source and has been reviewed by many statisticians and computational scientists - Features can be improved by additional packages - Limitations: - It requires intermediate knowledge of programming, limiting the number of users
Chemoface	www.ufla.br/chemoface	Free, downloadable	- Advantages: - It has an user-friendly interface - Various pre-processing methods can be easily applied - It is easy to perform cross-validation and external validation - The is a tool for data plot which is useful mainly to view continuous data, such as spectra and chromatograms - There are tools to detect outliers in multivariate calibration models - The software has a user guide with examples - Limitations:

(continued on next page)

Table 1 (*continued*)

Software	Website	Accessibility	Advantages and limitations
Sensomaker	www.ufla.br/sensomaker	Free, downloadable	- Software dedicated to multivariate statistical techniques - Graphic editing is limited - Advantages: - It has an user-friendly interface - There are interfaces to collect data from sensory tests - Both univariate or multivariate analysis can be performed - Parallel Factor Analysis (PARAFAC) can be applied to build preference maps - The software has a user guide with examples - Limitations: - Software dedicated to sensory analysis - Graphic editing is limited

as food and process development and optimization, not using a design of experiments represents an old-fashioned and not exact approach to perform a scientific study.

DOE has also been applied to study the effects of some independent variables (factors) on selected responses (screening methods). The use of screening experimental designs, such as the 2^2 and 2^3 types, is useful because through this approach, factors that have less significant effects may be unconsidered in a future study (Silva, Sant'Ana, & Massaguer, 2010). For example, if the aim of a study was intended to verify the effects of time and temperature of extraction of bioactive compounds from a food matrix and the data analysis showed that the temperature of extraction had a non-significant effect ($p = 0.445$) on the content of bioactive compounds. Thus, in a future work, the researcher may use other factors (*i.e.*, particle size, presence of agitation, pH of extraction) to optimize the system and obtain extracts with a high content of bioactive components.

In general, when DOE is used in combination with response surface methodology to analyze the experimental data, multiple regression equations relating to the independent variables with the response (y) are obtained and one can use this mathematical equation to predict the quantitative value of the response, inside the range of the tested values, for values (of the independent factors) not tested in the experiment. For industrial purposes, this methodology can be of great interest as the data can be used to have a quantitative idea about the impact of changing the factors on the response (Badoei-Dalfard & Karami, 2013; Joshi, Yadav, & Desai, 2008; Pedro, Granato, & Rosso, in press).

Another important observation about the design of experiments must be made: regardless of the design and/or mathematical/statistical method used to analyze data, the statistical software and its version must always be stated in the report. This is required once some software calculate the parameters different from the other packages. For example, if the homogeneity of variances is assessed by the Levene test using Action (Estatcamp, Brazil) package, the output will be different from the result obtained by Statistica software (Statsoft, USA). This is because Action calculates the statistics using the median (and not the mean value) while Statistica uses the means. In this specific case, in Statistica, the Brown-Forsythe test uses the median to assess homoscedasticity. Once there are numerous statistical packages, both from free and commercial versions, the researcher needs to know the theoretical principles involved in the calculation of each parameter prior to the use of the computational software to analyze data.

To be consistent, a consideration regarding the number of replicates should be made: usually authors state the 'analyses were performed in triplicate', but sometimes, this information is not totally clear: independent samples means the process of making a beer, for example, must be performed three times to obtain three products. If the scientist produces only one beer and the analysis of ethanol is performed three times, this situation is not an example of true/genuine replicates. The ideal situation is that the beer is produced three times and each beer must be analyzed in triplicate for a certain attribute (*i.e.*, ethanol content), in which the mean value of the attribute for each beer must be averaged (Fig. 2). The concept of 'triplicates' is simple: using two alcohol contents in beer (*i.e.*, 4.80 and 4.90), it is easy to calculate the mean value ($\bar{x} = (4.80 + 4.90)/2 = 4.85$). The reason to perform the analysis in triplicate is based on the fact that measuring three times the same property is an acceptable standard concerning the precision and the work to be performed. According to Passari et al. (2011), the mean value of triplicate measurements is the best estimate of the analyte in the sample, and the standard deviation is the best estimate of the experimental error. This can be also explained by the confidence interval of the mean value shown in Eq. (1):

$$\mu = \bar{x} \pm t_{n-1} \frac{s}{\sqrt{n}} \quad (1)$$

whereby $\bar{x}$ is the mean value (with n measurements), s is the standard deviation from n measurements, and t represents the critical value derived from the Student-t distribution with $n - 1$ degree of freedom. Using this equation and the Student-t distribution, it is possible to infer that the larger the number of measurements (replicates), the higher the precision. For example, at a 95% confidence level, the t-values (based on $n - 1$ degree of freedom) are: 12.71, 4.30, 3.18, and 2.78 when n goes from 2 to 5. Therefore, when the analyst performs the experiment in triplicate rather than in duplicate, the precision is highly improved.

Regarding analysis of variances (ANOVA), there are several types of approaches that can be used: one-way, two-way, three-way, main effects, repeated measures, multivariate, and factorial ANOVA. However, few papers actually state which ANOVA was used in the experiments and this information is required so other scientists can use the exact data treatment listed in earlier works (Polanco-Lugo, Dávila-Ortiz, Betancur-Ancona, & Chel-Guerrero, 2014). Aiming at comparing the means of the three juices (A, B, and C), if one-way ANOVA was applied after Levene test, which would be wrong, an F-value = 13.77 and p-value = 0.0057 would be attained, but if the Welch-ANOVA was used, which is the advisable test to be performed for heteroscedastic data, an F-value = 132.96 and p-value = 0.0009 would be obtained. Obviously, differences were obtained by using both ANOVA procedures, and one may say that the type of ANOVA employed is irrelevant. However, the right analysis of variances should always be chosen to express the

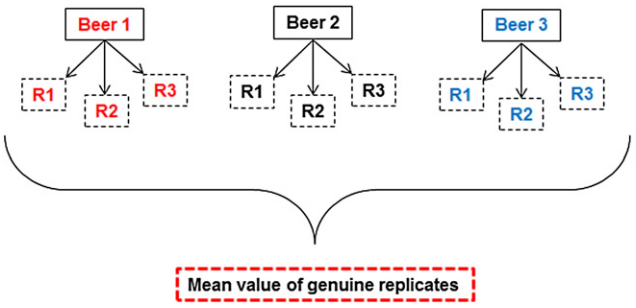

Fig. 2. Concept of genuine replicates (R): $R_{1 \to 3}$ should be summed and one mean value is obtained for one beer sample. Then, the mean value of beer 1, beer 2, and beer 3 should be averaged to obtain the mean value of the analyte.

statistical significance and this is dependent on the data normality and also on the homoscedasticity.

When the ANOVA highlights a probability value (*p*-value) below the stipulated α-value, usually 0.05 or 0.01, a *post-hoc* test is usually applied to check for differences among means, and Tukey's test is shown to be the most preferred option in food research (Galvão, Narain, & Nigam, 2014; Pasqualin Cavalheiro et al., 2014). In this sense, adopting $\alpha = 0.05$ is less restrictive than $\alpha = 0.01$ or $\alpha = 0.001$. As in food research the inferential tests are usually applied to check for differences among treatments (food samples, methods, processes, among others), an $\alpha = 0.05$ is preferred (Andrade et al., 2014; Bortolotto, Bueno, Braga, Barbosa, & Sanzovo, 2014; Ferrari, Clerici, & Chang, 2014). In pharmaceutical and medical sciences, stricter α-values (0.01 or 0.001) are generally used (Wanigasinghe, Arambepola, Ranganathan, & Muhandiram, 2014). In practice, the use of Tukey test seems to be based on 'tradition' rather than taking on its effectiveness in discriminating differences among treatments/samples. Granato, Calado, and Jarvis (2014) stated that Tukey's test may be a good option depending on the data characteristics but overall its use is not recommended as it is not robust, that is, often fails in detecting difference among means ($n \geq 3$). Thus, Duncan's or Fisher's least significance difference tests are recommended in all science fields (Mezquita, Barragán-Huerta, Ramírez, & Hinojosa, 2014). Consider that a scientist measured the total copper chelating power of some food extracts and obtained the following data (expressed as %): Extract 1: 48.23; 47.56; 48.68; Extract 2: 46.58; 46.89; 47.06; and Extract 3: 52.36; 53.74; 52.99. The Levene test (using Statistica v.7 software) was used to assess the equality of variances and the *p*-value was 0.4484. Once the hypothesis of equality of variances was accepted, one-way ANOVA was used to check for differences between extracts and a *p*-value < 0.0001 was obtained. Then, the efficiency of different multiple tests to compare the means was assessed and it was possible to observe that by using Fisher LSD and Duncan tests, the means of all treatments seemed to be different ($p < 0.05$), while by using Scheffé, Tukey, and Bonferroni tests, Extract 1 (means = 48.16%) and *2* (means = 46.84%) were statistically similar but different from Extract 3 (means = 53.03%). By analyzing these results, we emphasize the importance of understanding how these tests work prior to deciding which one will be used to analyze experimental data. Regardless of this, as mentioned earlier, Duncan and Tukey tests should be used consistently.

2.3. Misuses on chemometrics

Chemometric techniques are increasingly being applied not only in food science and technology but also in related fields such as experimental nutrition (Alezandro, Granato, & Genovese, 2013; Corrêa et al., 2014; Domingo, Tirelli, Nunes, Guerreiro, & Pinto, 2014; Nunes, 2014; Zielinski, Alberti et al., 2014), and nowadays ready-to-use toolboxes are available to develop models using various techniques, including pattern recognition, classification and multivariate calibration. Chemometrics enable a multivariate analysis of complex data and the extraction of relevant information is clearly observed. However, getting significant results requires not only meaningful data but also rational analysis and understanding of the purpose of the analysis. Many sophisticated chemometric techniques are applied by uninitiated scientists who do not have a fundamental knowledge of the capabilities and the limits of the chosen method (Kjeldahl & Bro, 2010; Pretsch & Wilkins, 2006). Next, considerations about common misuses on data preprocessing, pattern recognition, classification and regression models are carried out.

2.3.1. Data preprocessing

An important step for multivariate data analysis is the data preprocessing. It is used in order to reduce or remove random or systematic sources of variation in the data set. There are a lot of preprocessing techniques, but mean-centering and autoscaling are the most widely employed methods. However, the misuse of these techniques can result in outputs with different possible ways of interpretations. In this sense, another observation should be made regarding the use of chemometric tools: sequential publications from some authors have evidenced the lack of study in the pre-treatment (pre-processing) of data, that is, only one method is usually used and results are solely discussed using that specific approach. Understanding the importance of testing different preprocessing methods, it is advisable to test at least three different approaches (*i.e.*, mean-centering, autoscaling, smoothing, among others), compare results and choose the option that fits best the objectives of the work. Following this, some authors have compared the results and choosing the most suitable approach for different types of analytes and food matrices (Capuano et al., 2014; Özdestan et al., 2013; Tres, Heenan, & van Ruth, 2014).

Mean-centering is accomplished by subtracting the mean of a variable column from every element in the column. After this, the variable has a mean of zero, the data are shifted by mean and the center of the data becomes the new origin; consequently the information about the origin is lost, but the distances between the data points remain unchanged (Varmuza & Filzmoser, 2009). Mean-centering is useful for data including offsets since the purpose of centering is to remove this feature. It can be useful if different variables have different means, but usually do not have a major influence on pattern recognition and classification results (Brereton, 2009). For a dataset containing variables measured in different units of different magnitude, the different ranges manifest themselves in the modeling of the data, where the variables with little variation will not be modeled to a significant degree. Mean-centering does not remove these scale differences, but if the mean-centered variables are scaled to have the same standard deviation, *i.e.*, autoscaling the data, these differences disappear.

Autoscaling consists of dividing each mean-centering element in a column by the standard deviation of this variable. Autoscaling shifts the centroid of the data points to the origin and changes the scaling of the axes. Consequently the relative distances between the data points are changed and causes a blow-up (highlight) of variables with small values (Varmuza & Filzmoser, 2009). If a dataset is autoscaled, all variables have the same variance; every variable has the same opportunity

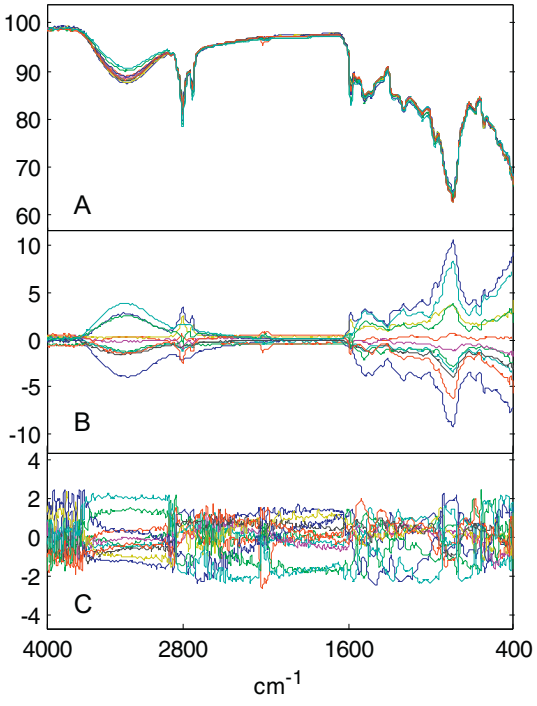

Fig. 3. Mid-infrared spectra using no preprocessing (A), mean-centering (B) and autoscaling (C).

of entering the model. On the other hand, if continuous data (such as spectra and chromatograms) are autoscaled, every variable, *i.e.,* real signal or noise from baseline, has the same standard deviation. This blow-up of the noise can impair dramatically the quality of the model. These effects can be visualized in Fig. 3, which spectral data were mean-centered and autoscaled. The blow-up effect of variables with intensities close to the baseline (2800–1600 cm^{-1}) can be observed, while in the mean-centering the baseline is not highlighted and its magnitude is preserved

2.3.2. Principal component analysis

One of the most used chemometric methods in food research is the Principal Component Analysis (PCA). The PCA scores plot is commonly used in order to grouping samples based on their similarities or dissimilarities usually using a 2D or 3D projection of the samples. In addition, an interpretation about how the variables influence this pattern is achieved through a loadings plot. However, some practices can contribute to misinterpreted results. Firstly, axes scales in scores plot must be comparable to avoid misinterpretations, *i.e.*, to say that two samples are similar based on their position, the axes on the scores plot must have comparable (equivalent) scales. Likewise, stretching the axes must be avoided (Geladi, Manley, & Lestander, 2003; Kjeldahl & Bro, 2010). Indeed most software does not produce plots with comparable scales because the plotting mostly aims filling the plot area. It is also important to note the amount of variance described by each principal component (PC). In the scores plot of the Fig. 4A, sample C is far from B along PC2, *i.e.*, sample C seems to be so different from B like E or F. However PC2 retains a very low amount of variance; so if this is considered, the difference between C and B is much less significant that the difference between C and E or F. This can be verified when PC2 and PC1 are in equivalent scales (Fig. 4B). Another point is that when interpreting the loadings plot, which represents the contribution (load) of a variable to discriminate the samples along the PC axes. A variable with large loading value in a PC contributes expressively to this PC; but if the amount of variance for this PC is low, the variable may not be truly important (Beebe, Pell, & Seasholtz, 1998). In the loadings plot presented in Fig. 4C, variable 5 seems to greatly influence PC2 because of its large loading value. However, in practice, this PC represents a very low amount of variance in relation to PC1.

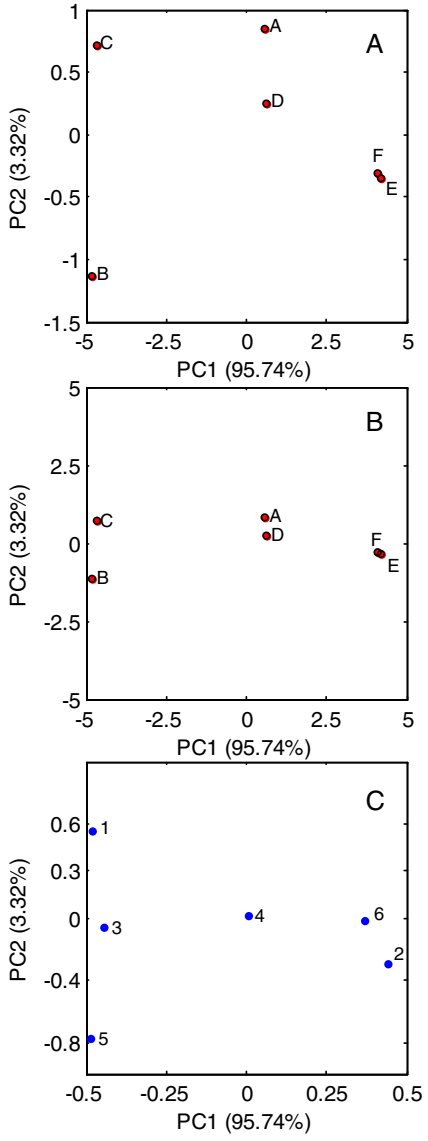

Fig. 4. PCA scores filled plot (A), scores with equivalent scales (B) and loadings plot (C).

2.3.3. Discriminant analysis

Classification methods, such as Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA) have been widely used in food research in order to obtain classification models (Granato, Oliveira, Caruso, Nagato, & Alaburda, 2014; Macatelli et al., 2009). In LDA or PLS-DA a function is used to model descriptors against categorical dependent variables. Singularity problems arise for high-dimensional data if the variables are highly correlating or if less samples than variables are available. In this case, the information contained in the descriptors can be summarized by latent variables that allow for dimension reduction, which can be achieved by PLS-DA (Varmuza & Filzmoser, 2009). These methods use a special *y* variable with a binary "dummy" system and the predicted *y* is not a quantitative value like PLS regression models; it is a number which indicates a class. Therefore, Root Mean Square Errors (RMSE) are not valid as a performance parameter for these classification models. RMSE measures the error in terms of deviations from a reference value, and larger deviations contribute more to the RMSE. So, this approach does not consider the classification rules and the actual class borders (Kjeldahl & Bro, 2010). If a class with *y*-value initially defined as 1 is predicted by the model as class 2, this will represent a numerical error of 1. Likewise, if this class 1 is predicted as class 3, the numerical error is 2. If this approach is considered, the prediction error from class 1 to class 3 is more significant than class 2. Indeed this is not true and irrelevant because what's important is that the model fails to predict the membership of class 1. The performance of the model can be better measured by the percentage of samples correctly classified (%CC) (Kjeldahl & Bro, 2010). As an example, Table 2 presents two cases with the same %SS, but if the RMSE (calculated by the difference between numerical values for actual and predicted classes) approach is considered, case B would be better that case A, while in fact they have equal performance.

Another critical point to be considered in classification models is the total number of samples. If the number of samples in the dataset is small, the percentage of samples correctly classified can be very

Table 2
Two cases showing the effect of RMSE approach to evaluate the performance of classification models with equal %CC.

Case A		Case B	
Actual class	Predicted class	Actual class	Predicted class
1	1	1	1
1	3	1	2
2	2	2	2
2	1	2	1
3	1	3	2
3	3	3	3
%CC	50	%CC	50
RMSE	0.8	RMSE	0.5

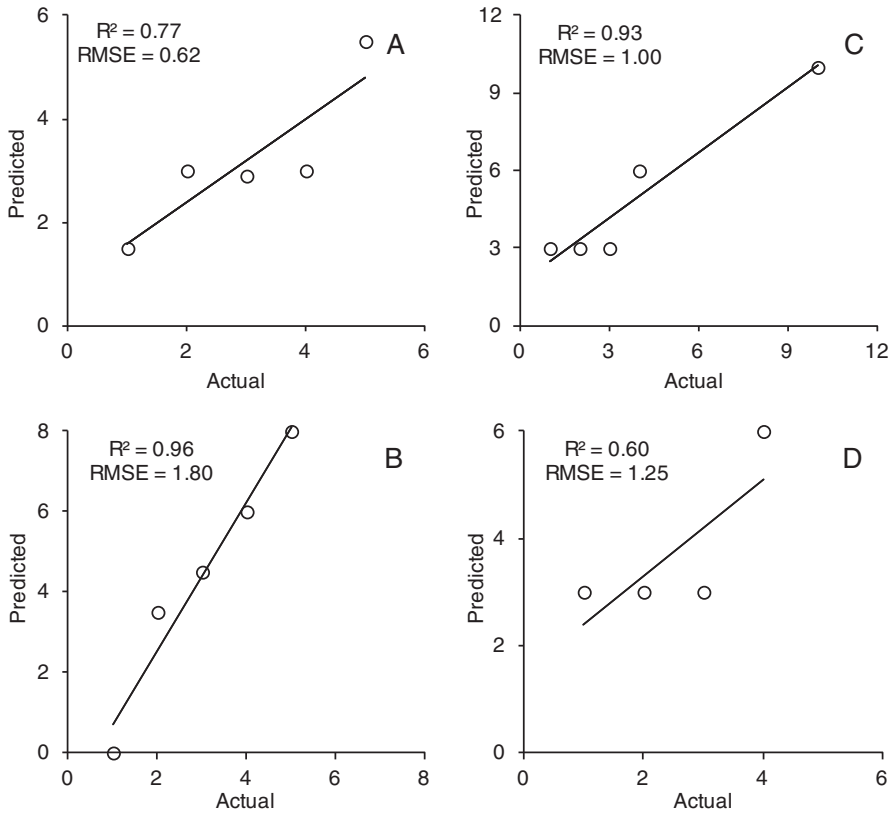

Fig. 5. Correlation between actual and measured values with low R^2 and low RMSE (A), high R^2 and high RMSE (B), high R^2 with (C) and without (D) extreme value far from most data.

sensitive to random results. Brereton (2009) reported an example that illustrates this problem: if a coin is tossed 10 times, and it obtained 8 heads and 2 tails, does this evidence that the coin is biased? Analogously if a predictive model is tested on 10 samples and 8 are correctly classified, does this evidence that this model really has a good performance? So the percentage of samples correctly classified must be critically used to measure the model performance on small datasets. If the number of samples is sufficiently large, the central limit theorem can be applied.

2.3.4. Coefficient of determination (R^2)

Squared correlation coefficient (or coefficient of determination — R^2) is a performance parameter commonly used to check the concordance between actual and predicted values in multivariate calibration models. In general it is uncritically assumed that high R^2 implies in good concordance between actual and predicted values. As an example, Fig. 5 shows that a high R^2 does not necessarily indicate high quality of the data. The correlation between actual and measured values in Fig. 5A presents a low R^2 in comparison with the data in Fig. 5B. However it is verified that the prediction error (RMSE) is a lot higher for case B, even presenting a higher R^2. Another critical point is the distribution of the intervals among the tested values. It is common to verify datasets with extreme values far from most values, which can induce an "untrue" high R^2. Fig. 5C shows an actual-predicted correlation with an extreme value presenting an elevated R^2, but if this extreme value is removed from the dataset, the R^2 assumes a very low value, in addition to a higher RMSE. Sometimes R^2 is close to 1 merely because the number of observations (samples) is low in comparison to the number of model parameters; therefore adjusted R^2, which "penalizes" R^2 for the low number of degree of freedom, is also used. In the example, adjusted R^2 values also present similar behavior, which was 0.69, 0.94, 0.90, and 0.40 for the cases A, B, C, and D of the Fig. 5 respectively. So, it is prudent to check RMSE values when evaluating R^2. Graphs correlating actual and predicted values are also useful to avoid misinterpretations of R^2 and RMSE. Some useful references report the important details on how how to evaluate the statistical quality of proposed models based on R^2 values (Badertscher & Pretsch, 2006; Hibbert, 2005).

2.3.5. Clustering methods

Clustering techniques are employed in a high frequency in quality control programs, provenancing of various specialty foods, to detect adulteration and therefore to attest authenticity and also to analyze the physicochemical, sensory, chemical, nutritional value, and rheological properties of various foods and beverages (Dahimi et al., 2014; Souza et al., 2011; Torres, Garbelotti, & Neto, 2006; Yudthavorasit, Wongravee, & Leepipatpiboon, 2014; Zielinski, Haminiuk, Alberti et al., 2014). In summary, cluster analysis is an exploratory multivariate technique used to explore the data structure and overall characteristics when little (or even none) information about group structure is available (Ares, 2014). The most used type of CA is the hierarchical approach (HCA). HCA is based on the determination of a distance between objects (degree of similarity/dissimilarity) and the application of an agglomerative (amalgamation) method to establish clusters of *n*-objects. Objects

Table 3
Data to illustrate the use of different similarity indices and amalgamation rules in hierarchical cluster analysis.

Product	Feature 1	Feature 2	Feature 3	Feature 4
A	20	4.56	100.36	55.56
B	10	2.22	90.25	54.25
C	5	2.00	77.36	20.13
D	7	0.89	55.47	43.25
E	7.5	0.69	23.35	40.12
F	9	1.66	71.23	29.98
G	15.6	3.34	80.73	17.56

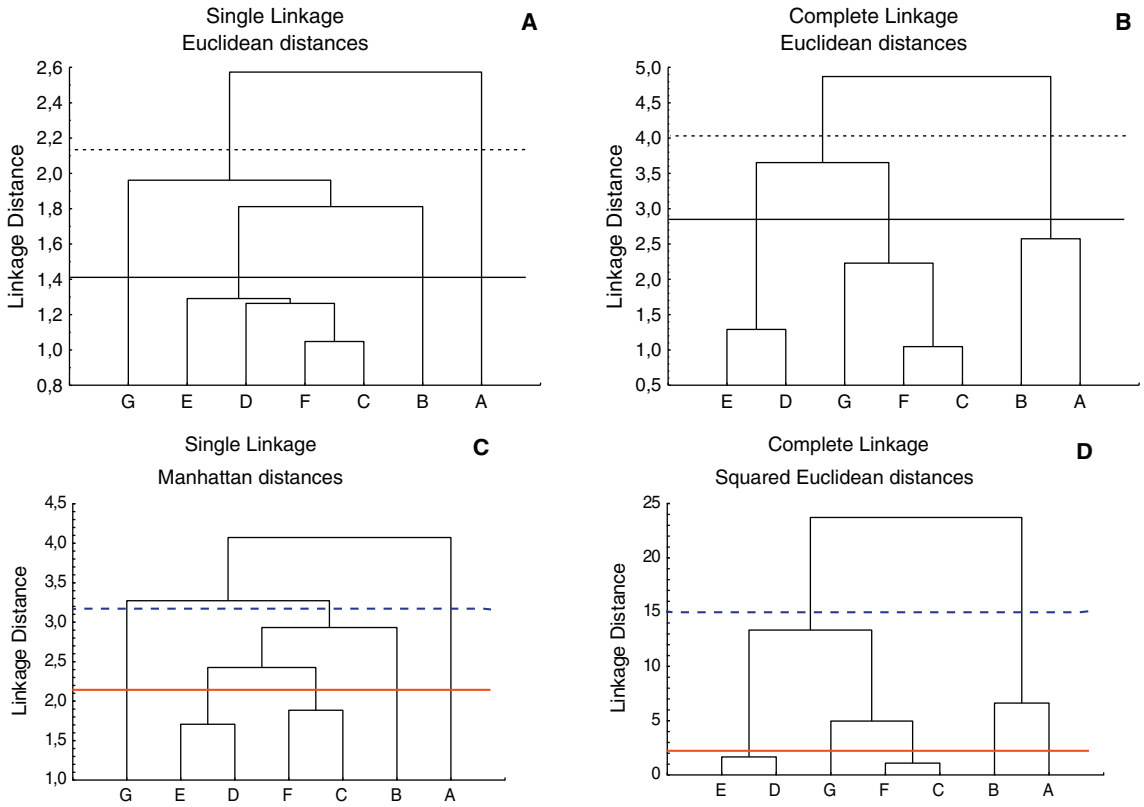

Fig. 6. Dendrograms obtained by hierarchical cluster analysis by means of different agglomerative methods and similarity distances to autoscaled data in Table 3: A) single linkage and Euclidean distance; B) complete linkage and Euclidean distance; C) single linkage and Manhattan distance; D) complete linkage and squared Euclidean distance. Blue and red lines in each graph represent suggestions on how the interpret the software outcomes.

in the same group present similar features among themselves, while a distinct difference is observed as compared to the objects pertaining to other clusters. Once there are many ways to calculate the degree of similarity/dissimilarity between objects, such as Euclidean, squared Euclidean, and Manhattan distances, it is advisable to test which one suits better the purpose of the work. Likewise, one should test which amalgamation method is better to discriminate samples, and there are many ways to do that, but simple linkage, complete linkage, average linkage, and Ward's method are most widely employed. Ares (2014) and Zielinski, Haminiuk, Nunes et al. (2014) explained in details the usability and application of these methods.

To illustrate the use and application of these methods, consider the data presented in Table 3. In this example, data were autoscaled and different methods were applied to obtain dendrograms for samples (Fig. 6). As expected, different results are obtained when different amalgamation rules and similarity distances are used. In this figure it is possible to note that depending on the selected method to agglomerate samples into groups, distinct outputs can be obtained. From the statistical standpoint, all projections are meaningful; however, the choice for any of these projections should be based on the judgment and expertise of the analyst. Thus, it is appropriate to stress that the final solution of HCA is not unique and depends not only on the methods employed to generate the dendrograms but also on the distance selected by the analyst (see differences in results if lines are in blue or in red are chosen). From a practical point of view, tree-clustering, Ward's method and Euclidean distances are usually used to form homogenous groups of objects, that is, groups with similar number of objects (Ginon, Ares, Issanchou, Laboissière, & Deliza, 2014; Rocha, Deliza, Corrêa, Carmo, & Abboud, 2013; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015).

More details on the application of other multivariate statistical techniques in food science and technology are comprehensively described elsewhere (Zielinski, Haminiuk, Nunes, et al., 2014) and practical examples on the use of multivariate statistical methods in many fields of food science, such as food technology, microbiology, sensory studies, and clinical protocols (Alezandro et al., 2011; Peltier, Visalli, & Schlich, 2015; Solieri, Bianchi, Mottolese, Lemmetti, & Giudici, 2014; Yang & Rose, 2014).

3. Perspectives and final comments

In this article, we highlighted that some positive characteristics and limitations of some commonly used free and commercial statistical packages. Some examples on how mathematical and statistical methods should be actually performed were also provided and described. In this sense, authors encourage the food research scientist to learn about different mathematical and statistical methods prior to choosing the most suitable software and statistical/mathematical approach to analyze experimental data. In this aspect, we strongly discourage those that use computational software to 'click and go' and do not know what was obtained or how to interpret holistically the outputs. In order to avoid this conduct, a series of published guidelines on statistical methods in food science and technology was listed and a critical evaluation on some of the most frequently used methods was made.

Acknowledgements

D. Granato thanks CAPES/PNPD for a Post-Doctoral fellowship. A.S. Sant'Ana shows gratitude to Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de

Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for the financial support of projects undertaken at the Laboratory of Quantitative Food Microbiology, University of Campinas.

References

Alber, S.A., & Schaffner, D.W. (1992). Evaluation of data transformations used with the square root and schoolfield models for predicting bacterial growth rate. *Applied and Environmental Microbiology*, *58*(10), 3337–3342.

Alezandro, M.R., Granato, D., & Genovese, M.I. (2013). Jaboticaba (*Myrciaria jaboticaba* (Vell.) Berg), a Brazilian grape-like fruit, improves plasma lipid profile in streptozotocin-mediated oxidative stress in diabetic rats. *Food Research International*, *54*, 650–659.

Alezandro, M.R., Granato, D., Lajolo, F.M., & Genovese, M.I. (2011). Nutrition aspects of second generation soy foods. *Journal of Agricultural and Food Chemistry*, *59*, 5490–5497.

Andrade, M.C.N., Jesus, J.P.F., Vieira, F.R., Viana, S.R.F., Spoto, M.H.F., & Minhoni, M.T.A. (2014). Effect of gamma irradiation on the nutritional quality of *Agaricus bisporus* strains cultivated in different composts. *Anais da Academia Brasileira de Ciências*, *86*(2), 897–906.

Ares, G. (2014). Cluster analysis: application in food science and technology. (Org.) In D. Granato, & G. Ares (Eds.), (1st ed.). *Mathematical and Statistical Approaches in Food Science and Technology*, *1*. (pp. 103–119). . Oxford: Wiley Blackwell.

Badertscher, M., & Pretsch, E. (2006). Bad results from good data. *Trends in Analytical Chemistry*, *25*, 1131–1138.

Badoei-Dalfard, A., & Karami, Z. (2013). Screening and isolation of an organic solvent tolerant-protease from *Bacillus* sp. JER02: Activity optimization by response surface methodology. *Journal of Molecular Catalysis B: Enzymatic*, *89*, 15–23.

Bas, D., & Boyaci, I.H. (2007). Modeling and optimization I: usability of response surface methodology. *Journal of Food Engineering*, *78*, 836–845.

Bassani, B.C., Nunes, D.S., & Granato, D. (2014). Optimization of phenolics and flavonoids extraction conditions and antioxidant activity of roasted yerba-mate leaves (*Ilex paraguariensis* A. St.-Hil., Aquifoliaceae) using response surface methodology. *Annals of the Brazilian Academy of Sciences*, *86*, 923–933.

Bassett, M.N., Pérez-Palacios, T., Cipriano, I., Cardoso, P., Ferreira, I.M.P.L.O., Samman, N., et al. (2014). Development of bread with NaCl reduction and calcium fortification: study of its quality characteristics. *Journal of Food Quality*, *37*, 107–116.

Beebe, K.R., Pell, R.J., & Seasholtz, M.B. (1998). *Chemometrics: a practical guide*. New York: John Wiley & Sons (348 pp.).

Besten, M.A., Nunes, D.S., Wisniewski, A., Jr., Sens, S.L., Granato, D., Simionatto, E.L., et al. (2013). Chemical composition of volatiles from male and female specimens of *Baccharis trimera* collected in two distant regions of southern Brazil: a comparative study using chemometrics. *Quimica Nova*, *36*, 1096–1100.

Bezerra, M.A., Santelli, R.E., Oliveira, E.P., Villar, L.S., & Escaleira, L.A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, *76*, 965–977.

Boobier, W.J., Baker, J.S., & Davis, B. (2006). Development of a healthy biscuit: an alternative approach to biscuit manufacture. *Nutrition Journal*, *5*, 1–7.

Bortolotto, O.C., Bueno, A.F., Braga, K., Barbosa, G.C., & Sanzovo, A. (2014). Biological characteristics of *Heliothis virescens* fed with Bt-soybean MON 87701 × MON 89788 and its conventional isoline. *Anais da Academia Brasileira de Ciências*, *86*(2), 973–980.

Brereton, R. (2009). *Chemometrics for pattern recognition*. Chichester: John Wiley & Sons (522 pp.).

Capuano, E., van der Veer, G., Boerrigter-Eenling, R., Elgersma, A., Rademaker, J., Sterian, A., et al. (2014). Verification of fresh grass feeding, pasture grazing and organic farming by cows farm milk fatty acid profile. *Food Chemistry*, *164*, 234–241.

Corradini, M. G., & Peleg, M. (2006). On modeling and simulating transitions between microbial growth and inactivation or vice versa. *International Journal of Food Microbiology*, *108*(1), 22–35.

Corrêa, S.C., Pinheiro, A.C.M., Siqueira, H.E., Carvalho, E.M., Nunes, C.A., & Vilas Boas, E.V.B. (2014). Prediction of the sensory acceptance of fruits by physical and physical–chemical parameters using multivariate models. *LWT — Food Science and Technology*, *59*(2), 666–672. http://dx.doi.org/10.1016/j.lwt.2014.07.042.

Cozzolino, D. (2014). The use of correlation, association and regression to analyse processes and products. (Org.) In D. Granato, & G. Ares (Eds.), (1st ed.). *Mathematical and Statistical Approaches in Food Science and Technology*, *1*. (pp. 19–30). . Oxford: Wiley Blackwell.

Dahimi, O., Rahim, A.A., Abdulkarim, S.M., Hassan, M.S., Hashari, S.B.T.Z., Mashitoh, A.S., et al. (2014). Multivariate statistical analysis treatment of DSC thermal properties for animal fat adulteration. *Food Chemistry*, *158*, 132–138.

Domingo, E., Tirelli, A.A., Nunes, C.A., Guerreiro, M.C., & Pinto, S.M. (2014). Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: a review. *Food Research International*, *60*, 131–139.

Domínguez-Perles, R., Teixeira, A.I., Rosa, E., & Barros, A.I. (2014). Assessment of (poly)-phenols in grape (*Vitis vinifera* L.) stems by using food/pharma industry compatible solvents and Response Surface Methodology. *Food Chemistry*, *164*, 339–346.

Farris, F., & Piergiovanni, L. (2009). Optimization of manufacture of almond paste cookies using response surface methodology. *Journal of Food Process Engineering*, *32*, 64–87.

Ferrari, M.C., Clerici, M.T.P.S., & Chang, Y.K. (2014). A comparative study among methods used for wheat flour analysis and for measurements of gluten properties using the Wheat Gluten Quality Analyser (WGQA). *Food Science and Technology (Campinas)*, *34*(2), 235–242.

Galvão, M.S., Narain, N., & Nigam, N. (2014). Influence of different cultivars on oil quality and chemical characteristics of avocado fruit. *Food Science and Technology (Campinas)*, *34*(3), 539–546.

Geladi, P., Manley, M., & Lestander, T. (2003). Scatter plotting in multivariate data analysis. *Journal of Chemometrics*, *17*, 503–511.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489.

Ginon, E., Ares, G., Issanchou, S., Laboissière, L.H.E.S., & Deliza, R. (2014). Identifying motives underlying wine purchase decisions: results from an exploratory free listing task with Burgundy wine consumers. *Food Research International*, *62*, 860–867.

Granato, D., & Calado, V.M.A. (2014). The use and importance of design of experiments (DOE) in process modelling in food science and technology. (Org.) In D. Granato, & G. Ares (Eds.), (1st ed.). *Mathematical and Statistical Approaches in Food Science and Technology*, *1*, . Oxford: Wiley Blackwell (536 pp.).

Granato, D., Calado, V.M.A., & Jarvis, B. (2014). Observations on the use of statistical methods in food science and technology. *Food Research International*, *55*, 137–149.

Granato, D., Castro, I.A., Ellendersen, L.S.N., & Masson, M.L. (2010). Physical stability assessment and sensory optimization of a dairy-free emulsion using response surface methodology. *Journal of Food Science*, *73*, 149–155.

Granato, D., Castro, I.A., Piekarski, F.V.B.W., Benincá, C., & Masson, M.L. (2011). Influence of passion fruit juice on colour stability and sensory acceptability of non-sugar yacon-based pastes. *Brazilian Archives of Biology and Technology*, *54*, 149–159.

Granato, D., Grevink, R., Zielinski, A.A.F., Nunes, D.S., & van Ruth, S.M. (2014). Analytical strategy coupled with response surface methodology to maximize the extraction of antioxidants from green, yellow, and red teas (*Camellia sinensis* var *sinensis*). *Journal of Agricultural and Food Chemistry*, *62*(42), 10283–10296.

Granato, D., Oliveira, C.C., Caruso, M.S.F., Nagato, L.A.F., & Alaburda, J. (2014). Feasibility of different chemometric techniques to differentiate commercial Brazilian sugarcane spirits based on chemical markers. *Food Research International*, *60*, 212–217.

Haj-Isa, N.M.A., & Carvalho, E.S. (2011). Development of biscuits enriched with merluza. *Ciência e Tecnologia de Alimentos*, *31*, 313–318.

Hibbert, D.B. (2005). Further comments on the (mis-)use of r for testing the linearity of calibration functions. *Accreditation and Quality Assurance*, *10*, 300–301.

Joshi, S., Yadav, S., & Desai, A.J. (2008). Application of response-surface methodology to evaluate the optimum medium components for the enhanced production of lichenysin by *Bacillus licheniformis* R^2. *Biochemical Engineering Journal*, *41*(2), 122–127.

Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics*, *24*, 558–564.

Macatelli, M., Akkermans, W., Koot, A., Buchgraber, M., Paterson, A., & van Ruth, S. (2009). Verification of the geographical origin of European butters using PTR-MS. *Journal of Food Composition and Analysis*, *22*, 169–175.

McDonald, J.H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, Maryland: Sparky House Publishing.

Mezquita, P.C., Barragán-Huerta, B.E., Ramírez, J.P., & Hinojosa, C.O. (2014). Stability of astaxanthin in yogurt used to simulate apricot color, under refrigeration. *Food Science and Technology (Campinas)*, *34*(3), 559–565.

Nunes, C.A. (2014). Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Research International*, *60*, 255–261.

Özdestan, Ö., van Ruth, S.M., Alewijn, M., Koot, A., Romano, A., Cappellin, L., et al. (2013). Differentiation of specialty coffees by proton transfer reaction-mass spectrometry. *Food Research International*, *53*, 433–439.

Pasqualin Cavalheiro, C., Lüdtke, F.L., Santi Stefanello, F., Kubota, E.H., Terra, N.N., & Fries, L.L.M. (2014). Replacement of mechanically deboned chicken meat with its protein hydrolysate in mortadella-type sausages. *Food Science and Technology (Campinas)*, *34*(3), 478–484.

Passari, L.M.Z.G., Soares, P.K., & Bruns, R.E. (2011). Estatística aplicada à química: dez dúvidas comuns. *Quimica Nova*, *34*, 888–892.

Pedro, A.C., Granato, D., & Rosso, N.D. (2016). Extraction of anthocyanins and polyphenols from black rice (*Oryza sativa* L.) by modeling and assessing their reversibility and stability. *Food Chemistry*. http://dx.doi.org/10.1016/j.foodchem.2015.02.045 (in press).

Peltier, C., Visalli, M., & Schlich, P. (2015). Comparison of canonical variate analysis and principal component analysis on 422 descriptive sensory studies. *Food Quality and Preference*, *40*, 326–333.

Polanco-Lugo, E., Dávila-Ortiz, G., Betancur-Ancona, D.A., & Chel-Guerrero, L.A. (2014). Effects of sequential enzymatic hydrolysis on structural, bioactive and functional properties of *Phaseolus lunatus* protein isolate. *Food Science and Technology (Campinas)*, *34*(3), 441–448.

Pretsch, E., & Wilkins, C.L. (2006). Use and abuse of chemometrics. *Trends in Analytical Chemistry*, *25* (Editorial).

Rocha, M.C., Deliza, R., Corrêa, F.M., Carmo, M.G.F., & Abboud, A.C.S. (2013). A study to guide breeding of new cultivars of organic cherry tomato following a consumer-driven approach. *Food Research International*, *51*, 265–273.

Ropodi, A.I., Pavlidis, D.E., Mohareb, F., Panagou, E.Z., & Nychas, G.J.E. (2015). Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International*, *67*, 12–18.

Sant'Ana, A.S., Franco, B.D.G.M., & Schaffner, D.W. (2012). Modeling the growth rate and lag time of different strains of *Salmonella enterica* And *Listeria monocytogenes* in ready-to-eat lettuce. *Food Microbiology*, *30*(1), 267–273.

Schaffner, D.W. (1998). Predictive food microbiology Gedanken experiment: why do microbial growth data require a transformation? *Food Microbiology*, *15*(2), 185–189.

Schoder, V., Himmelmann, A., & Wilhelm, K.P. (2006). Preliminary testing for normality: some statistical aspects of a common concept. *Clinical and Experimental Dermatology*, *31*(6), 757–761.

Silva, A.R., Sant'Ana, A.S., & Massaguer, P.R. (2010). Modelling the lag time and growth rate of Aspergillus section Nigri IOC 4573 in mango nectar as a function of temperature and Ph. *Journal of Applied Microbiology*, *109*, 1105–1116.

Solieri, L., Bianchi, A., Mottolese, G., Lemmetti, F., & Giudici, P. (2014). Tailoring the probiotic potential of non-starter *Lactobacillus* strains from ripened Parmigiano Reggiano cheese by *in vitro* screening and principal component analysis. *Food Microbiology*, *38*, 240–249.

Souza, S.S., Cruz, A.G., Walter, E.H.M., Faria, J.A.F., Celeghini, R.M.S., Ferreira, M.M.C., et al. (2011). Monitoring the authenticity of Brazilian UHT milk: a chemometric approach. *Food Chemistry*, *124*, 692–695.

Souza, S.V.C., & Junqueira, R.G. (2005). A procedure to assess linearity by ordinary least squares method. *Analytica Chimica Acta*, *552*, 25–35.

Tenenhaus-Aziza, F., & Ellouze, M. (2015). Software for predictive microbiology and risk assessment: A description and comparison of tools presented at the ICPMF8 Software Fair. *Food Microbiology*, *45*, 290–299.

Teófilo, R. F., & Ferreira, M. M. C. (2006). Chemometrics II: spreadsheets for experimental design calculations, a tutorial. *Química Nova*, *29*(2), 338–350.

Torres, E.A.F.S., Garbelotti, M.L., & Neto, J.M.M. (2006). The application of hierarchical clusters analysis to the study of the composition of foods. *Food Chemistry*, *99*, 622–629.

Tres, A., Heenan, S.P., & van Ruth, S. (2014). Authentication of dried distilled grain with solubles (DDGS) by fatty acid and volatile profiling. *LWT — Food Science and Technology*, *59*, 215–221.

Van, Boekel, & Martinus, A. J. S. (2008). Kinetic Modeling of Food Quality: A Critical Review. *Comprehensive Reviews in Food Science and Food Safety*, *144*, 158.

Varmuza, K., & Filzmoser, P. (2009). *Introduction to multivariate statistical analysis in chemometrics*. Boca Raton: CRC Press (336 pp.).

Wanigasinghe, J., Arambepola, C., Ranganathan, S.S., & Muhandiram, E.C. (2014). The efficacy of moderate-to-high dose oral prednisolone versus low-to-moderate dose intramuscular corticotropin for improvement of hypsarrhythmia in West syndrome: a randomized, single-blind, parallel clinical trial. *Pediatric Neurology*, *51*(1), 24–30.

Yang, J., & Rose, D.J. (2014). Long-term dietary pattern of fecal donor correlates with butyrate production and markers of protein fermentation during *in vitro* fecal fermentation. *Nutrition Research*, *34*(9), 749–759.

Yudthavorasit, S., Wongravee, K., & Leepipatpiboon, N. (2014). Characteristic fingerprint based on gingerol derivative analysis for discrimination of ginger (*Zingiber officinale*) according to geographical origin using HPLC-DAD combined with chemometrics. *Food Chemistry*, 101–111.

Zielinski, A.A., Alberti, A., Braga, C.M., Silva, K.M., Canteri, M.H.G., Mafra, L.I., et al. (2014). Effect of mash maceration and ripening stage of apples on phenolic compounds and antioxidant power of cloudy juices: a study using chemometrics. *LWT — Food Science and Technology*, *57*, 223–229.

Zielinski, A.A.F., Haminiuk, C.W.I., Alberti, A., Nogueira, A., Demiate, I.M., & Granato, D. (2014). A comparative study of the phenolic compounds and the *in vitro* antioxidant activity of different Brazilian teas using multivariate statistical techniques. *Food Research International*, *60*, 246–254.

Zielinski, A.A.F., Haminiuk, C.W.I., Nunes, C.A., Schnitzler, E., van Ruth, S.M., & Granato, D. (2014). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline. *Comprehensive Reviews in Food Science and Food Safety*, *13*, 300–316.