Teaching statistics and chemometrics using an open source, free and graphical user interface software.

Roberto Silva de Souza Junior ‡ and Endler Marcel Borges‡*

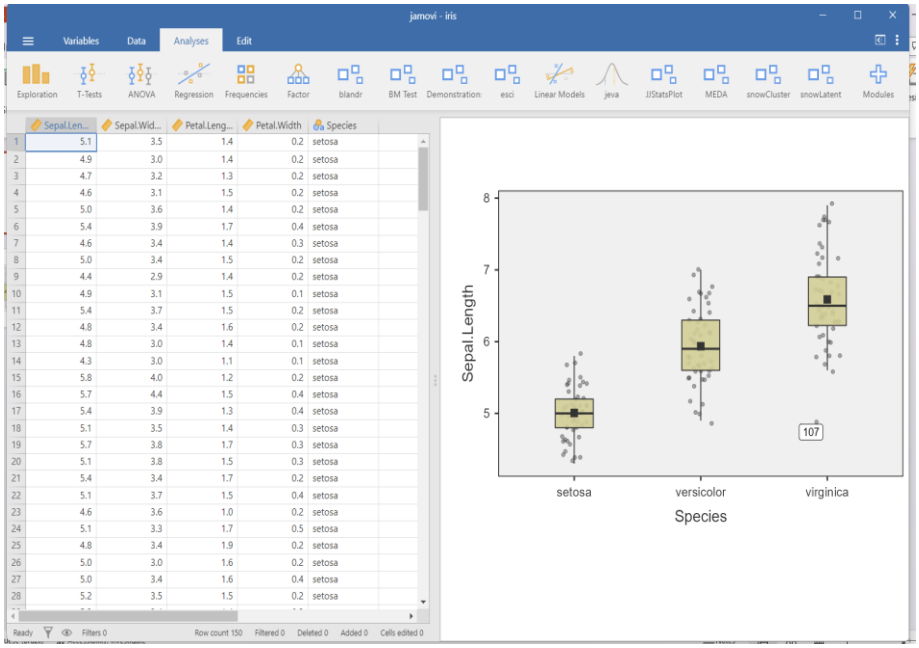
‡*Departamento de Química, Fundação Universidade Regional de Blumenau, FURB, Campus 1, Rua Antônio da Veiga, 140, Victor Konder, 89012-900 Blumenau-SC, Brasil.*

*marcelborgesb@gmail.com embsouza@furb.br

ABSTRACT

This report aims to introduce the fundamental features of the JAMOVI software to academics in the chemistry field for use in undergraduate and graduate-level research. It is freeware with a graphical user interface (GUI) and it is written in the R language. The discussion began on descriptive statistics (mean, median, range, skewness how to check data normality using hypothesis tests (Shapiro-Wilk, Kolmogorov-Smirnov and Anderson-Darling tests). Then, some visual tools for checking data normality were presented (histograms, Q-Q plots, and boxplots). When the data normality was checked, two and more dependent means were compared using parametric tests (t test and ANOVA; Fisher's). When the data was not normally distributed, nonparametric tests were used (Mann-Whitney and Kruskal-Wallis tests). When the data was paired and normally distributed, two and more than two group means were compared using the paired t-test and RMANOVA, respectively. Their nonparametric versions were also used (Wilcoxon and Friedman tests). Means comparisons were also carried out using boxplots and discriminant plots, which provide a visual interpretation beyond the p-values interpretation. In addition, principal component analysis (PCA) was carried out using JAMOVI's plugin MEDA, which builds scores and loading plots. All tests and plots were done easily using JAMOVI's click-and-go interface.

GRAPHICAL ABSTRACT



KEYWORDS
Graduate Education/Research Upper-Division Undergraduate Analytical Chemistry Chemoinformatics Interdisciplinary/Multidisciplinary Chemometrics Computational Chemistry

INTRODUCTION

Statistics are present in all undergraduate chemistry curricula. Normally, it was restricted to basic statistics (mean, median, standard deviation) and some parametric hypothesis tests (t-test for paired and unpaired groups).[1,2]

Even the most used textbooks are restricted to basic statistics and parametric hypothesis tests (t test for paired and unpaired means),[3,4] where nonparametric tests are necessary in many areas.[5]

In the past, statistical calculations were carried out using tedious and time-consuming calculations, where the interpretation of the results was based just on the p-value.

In the 90's, during the beginning of the computer era, spreadsheets were widely used.[6–15] However, spreadsheets were not adequate for statistical calculations and chemometrics and programming software without a graphical user interface (GUI) such as MATLAB,[16] Python,[17–21] R project,[22] had been used.

Since most of the students were not familiar with these programs, freeware GUI software was introduced in chemical education. The R Commander is a GUI that runs into R.[23] It was used to teach statistics and chemometrics.[24,25] There were two freeware GUI software, JASP[26] and JAMOVI, which were written in R language and run outside R environment. The JASP was presented in some recent publications in this Journal. [27–29]

Here, the JAMOVI software but it is not yet widely used. At the submission of this paper, there were few papers published in the American Chemical Society publications.[30–34]

Since students were unfamiliar with programming software, graphical user interface (GUI) software. while software did not have a graphical user interface (GUI).[24] Nowadays, there were several software with GUI, but most of them were paid. In this paper, there are some examples of how to do basic statistics, hypothesis tests, plots, and Principal Component Analysis using JAMOVI.

OVERVIEW

In this article some of the most common examples related to groups comparison (dependent and independent) were presented. These examples were easily resolved using JAMOVI, which is a solution for those that were looking for a free, open source and GUI software.

In the time of artificial intelligence (AI), some key questions were provided to the students (**Box 1**), and they were asked to answer these questions using ChatGPT.[35–41] The responses provided by ChatGPT were shown in the supporting information (ChatGPT.docx). These questions and ChatGPT may be used as guidelines for those who are new to statistics and chemometrics.

There is also a complete guide for JAMOVI in Inglish language.[42] Some questions and its answers were shown in the supporting information (Questions.docx).

Box 1: Question for ChatGPT
1. What was JAMOVI?
2. How can I check for data normality using JAMOVI?
3. Gave me an example of boxplots.
4. How can I make boxpots using JAMOVI?
5. What were the dependent groups?
6. How can I compare dependent groups?
7. What were the independent groups?
8. How can I compare independent groups?
9. What was Principal component analysis (PCA)?
10. Give me examples of Principal component analysis (PCA).
11. In Principal component analysis (PCA), what was the score plot?
12. In Principal component analysis (PCA), what was the loading plot?

MATERIAL AND METHODS

All statistical tests were carried out using JAMOVI (2.3.28). It is freeware written in R language, and it was obtained from its webpage.[43]

The ChatGPT version used was 3.5 (free).[44]

The spreadsheets used in this activity were provided in the supporting information.

Spreadsheet Kale contains the mineral concentration from two Kale genotypes.[45]

Spreadsheet Rice contains the mineral concentrations (As, B, Ba, Ca, Cd, Ce, Co, Cr, Cu, Fe, K, La, Mg, Mn, Mo, P, Pb, Rb, Se and Zn) of organic (n = 18) and ordinary (n = 32) rice samples obtained from the Brazilian retail market.[46]

Spreadsheet FishFeed contains the iron concentration (mg/Kg) in fish feed reference material determined by several laboratories.[47]

Spreadsheet coffee contains concentrations of chemical compounds of green coffee and global score for the coffee brew for three different years.[48]

Spreadsheet Jatropha contains the oil content (%) in Jatropha curcas L. (Euphorbiaceae) grown in 14 different locations.[49]

Spreadsheet Folin-Ciocalteau contains the antioxidant capacity of several beverages determined using the Folin-Ciocalteau test,[50] the test was carried out using a spectrophotometer and digital images obtained using smartphones.[51]

Spreadsheet App, contains the equivalence points determined using a pH meter and smartphone app.[52]

Yadav et al.[53] determined that the fluoride concentrations in real samples using a standard method (smartphone) and a proposed method (UV–Vis Spectroscopy), the data was provided in spreadsheet Fluoride.

Spreadsheet AAI contains the antioxidant activity index (AAI) for several antioxidant compounds determined using different DPPH concentrations.[54] This data was also provided doing the DPPH concentrations was unpaired variables (AAI U).

Ji et al.[55] reports the development of a simple, effective, and high-throughput method combining gas chromatography–tandem mass spectrometry (GC–MS/MS) with either QuEChERS or solid phase extraction (SPE) to determine 147 pesticide residues in traditional Chinese medicines simultaneously. Spreadsheet Recovery contains the recovery of 147 pesticides in three spiking levels.

Crawford & Wang[56] determined acrylamide in olives using commonly used routine methods: liquid chromatography with ultraviolet detection (LC–UV) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) without bromination and LC–UV and gas chromatography–mass

spectrometry (GC–MS) following bromination. The results of Table 4 of the original manuscript were presented in spreadsheet methods.

Spreadsheet Iris contains 150 observations of three iris flower types: Setosa, Versicolor, and Virginica. Four features were measured from each flower in the data: the length and the width of the sepals and petals, in centimeters.[57]

Spreadsheet Smartphone have the phenolic contents in several specific Vietnamese dried tea products and their liquors, it was determined using UV-vis and smartphones.[58]

Spreadsheet Peanut contains the fatty acid profiles of commercially grown Runner-type peanut cultivars (i.e., 10 cultivars, n = 151) collected over two production years (2005 and 2006). Eight major fatty acids were identified in the sample set including palmitic (C16:0), stearic (C18:0), oleic (C18:1, ω9), linoleic (C18:2, ω6), arachidic (C20:0), gondoic (C20:1, ω9), behenic (C22:0), and lignoceric (C24:0) acids. Based on the oleic to linoleic acid (O/L) ratio, these cultivars were denoted as normal, mid-, and high-oleic peanut types.[59]

The spreadsheet Beer has the metal concentration of 35 types of bottled and canned Polish beers.[60]

FIRST STEPS HANDING THE DATASET.

Checking data normality.

The first step analyzing a data set was to check its normality. The data normality may be checked using hypothesis tests and plots such as histograms, boxplots, and Q-Q plots. As an example, the iron concentration in kale samples were used. (spreadsheet Kale.xlsx; supporting information).

Hypothesis test for data normality checking.

Shapiro-Wilk, Kolmogorov-Smirnov and Anderson-Darling tests were hypothesis tests to check the null hypothesis.[61] For the three tests, the null hypothesis is the sample was taken from a population with normal distribution. If the given p-value is less than 0.05, normal distribution can be rejected. The Shapiro-Wilk and Anderson-Darling are more exact than the Kolmogorov-Smirnov test.[62]

Checking if the iron concentration in Kale samples were normally distributed. Shapiro-Wilk and Anderson-Darling tests showed that the data was not normally distributed. The Kolmogorov-Smirnov showed that the data was normally distributed confirming that the Kolmogorov-Smirnov was less sensitive than the other two tests (Table 1).

Table 1: Tests of Normality for the iron concentration in kale cultivars

		statistic	p-value
Fe (mg/100g)	Shapiro-Wilk	0.923	0.022
	Kolmogorov-Smirnov	0.158	0.379
	Anderson-Darling	1.04	0.008

Histograms

Data normality can be checked using visual interpretation of plots. For example, histograms are one of the simplest and most useful ways of visualizing data. It divides up the possible values into bins and then counts the number of observations that fall within each bin.[42] The histogram (Figure 1) clearly shown that the data was not normally distributed.

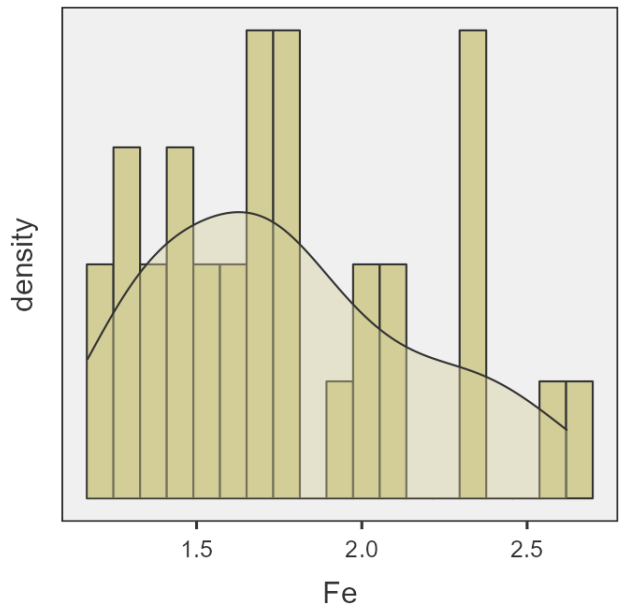


Figure 1: Histogram of iron concentration (mg/g) in Kale genotypes (spreadsheet Kale.xlsx; supporting information)

Quantile-quantile plot

Q–Q plots take the sample data, sort them in ascending order, and then plot them against quantiles calculated from a theoretical distribution. If the data are normally distributed, the points will fall on or close to the 45° reference line.[27] Figure 2 showed that the data was not normally distributed since the points were out of the reference line.

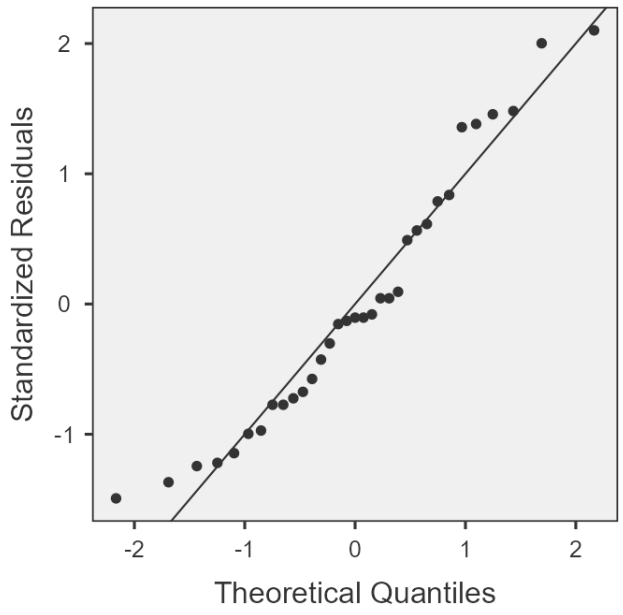


Figure 2: Q–Q plot of iron concentration (mg/100 g) of kale genotypes.

Boxplots

Boxplots are a powerful tool to visualize the data set, it divides the data in quartiles (Q1; first quartile, Q2; second quartile or median, Q3; third quartile). When the median (Q2) moves far from the mean, it was evident that the data was not normally distributed.[27] In this case (Figure 1), the data has right tailed, and the mode moves to the left side. In the boxplot, Q2 moves to the bottom below the mean (Figure 3).

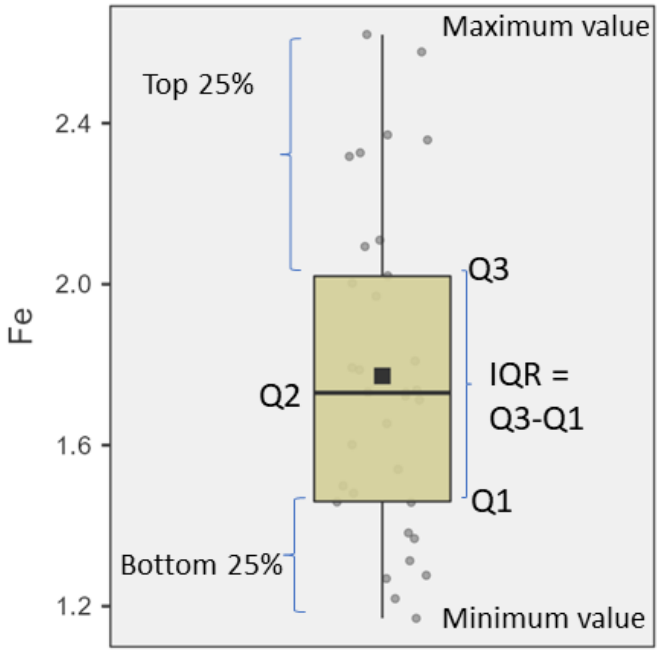


Figure 3: Boxplot of iron concentration in kale genotypes.

COMPARING TWO INDEPENDENT GROUPS

Two independent groups are compared using the t-test for independent samples. Independent groups were groups with no relation, such as comparing the mineral concentrations in kale samples from different genotypes.

The t-test for parametric data

The t-test is parametric test, and it needs normally distributed data, and homogeneity of variances. When the data is normally distributed and the variances are homogeneous, it is known as the Student's test. When the variances are not homogeneous, it is known as the Welch test. When the data is not normally distributed, it is known as the Mann-Whitney test.

When comparing two independent groups, a common mistake was using the Student's test for nonnormally distributed data and nonhomogeneous variances. JAMOVI have the option assumptions check, where data normality is checked using the hypothesis tests and Q-Q plots. Data humanity was also checked hypothesis tests.

As an example of comparing two independent groups, it is presented the comparison of Zn concentrations in organic and ordinary rice samples (Spreadsheet rice.xlsx; supporting information).[46]

The first step was checking data normality. the hypothesis tests showed that the data was normally distributed ((Shapiro-Wilk (p-value 0.293), Kolmogorov-Smirnov (p-value 0.821), and Anderson-Darling (p-value 0.331)). It may also be observed using histograms (Figure 4).

Once that the data was normally distributed, variances were checked, where Levene's test and variance ratio tests showed that the data was normally distributed (p-values 0.675 and 0.540, respectively).

In JAMOVI there are the options group 1 ≠ group 2, group 1 > group 2, and group 1 < group 2. Selecting group 1 ≠ group 2, the hypothesis of equivalence between the two groups were tested.

Once the data was normally distributed and variances were equivalent, the two groups can be compared using the student test, where it shows that the two groups were not equivalent (p-value 0.009). The Welch and the Mann-Whitney tests also showed that the Zn concentrations in the two groups were not equivalent (p-values 0.007 and 0.013, respectively).

By Selecting group 1 > group 2, the hypothesis of group 1 (ordinary rice samples) has larger Zn concentrations than group 2 (organic rice samples) was tested. The Welch and the Mann-Whitney tests provided p-values 0.005, 0.004 and, 0.006, respectively. It showed that ordinary rice samples have larger Zn concentrations than organic rice samples.

Comparisons between the two groups may also be carried out visually using boxplots (Figure 5). It shows that the Zn concentration in ordinary samples were larger than in organic samples, the proximity of the median from the middle of the box showed that the data is normally distributed, the close size of the boxes shows that variances were equivalent, and the position of the boxes shows that the ordinary rice samples have larger Zn concentrations than organic rice samples.

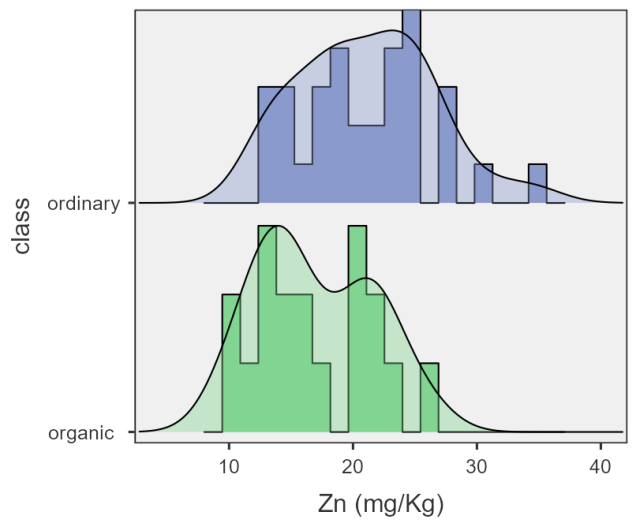


Figure 4: Histograms of Zn concentrations in organic and ordinary rice samples

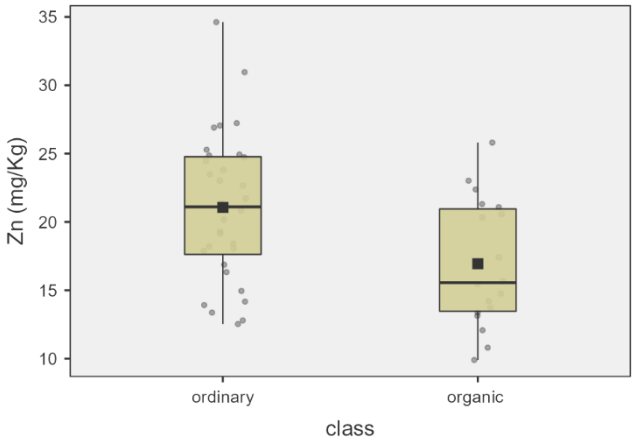


Figure 5: Boxplots of Zn concentrations in ordinary and organic rice samples.

The Mann-Whitney test for nonparametric data

In the case of comparing iron concentration in different kale genotypes, Student, Welch, and Mann-Whitney showed that the iron concentrations in kale genotypes were not equivalent (p-values 0.002, 0.003, and 0.002). Selecting group 1 > group 2, all tests showed that iron concentration in cultivars > germplasm. In this case, the right choice is the nonparametric Mann-Whitney test, but the parametric tests also worked. It can also be observed using boxplots.

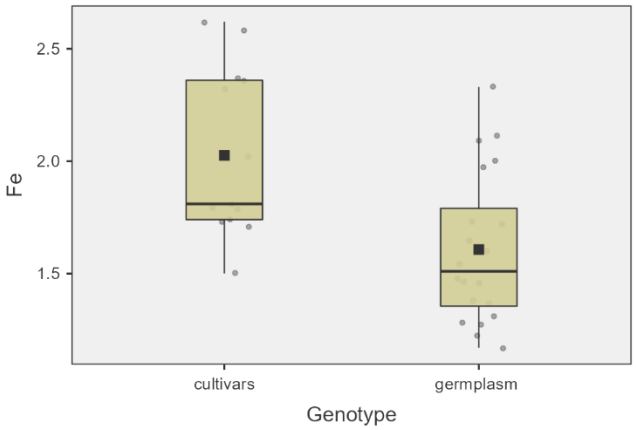


Figure 6: Boxplots of iron concentration in kale samples of different cultivars.

COMPARING MORE THAN TWO INDEPENDENT GROUPS.

More than two independent groups were compared using ANOVA. It is a parametric test that requires normally distributed data and homogeneity of the variances. A common mistake found in literature was considering independent groups as dependent and vice versa. For example, comparing the weight loss of different volunteer groups which did three different diets, comparing the potassium concentration determined by different laboratories in a reference material were examples of multiple independent groups comparison. In those cases, there is one quantitative variable (weight loss and potassium concentration) and one group variable (diets and laboratories), and it is known as one-way ANOVA.

When there are two groups, it is known as two-way ANOVA, for example, comparing the acetyl salicylic acid concentrations determined by five student groups using two methods in aspirin tablets from the same batch. In this case, there are two groups (student groups and methods). ANOVA may have more than two independent groups, but its visualization was difficult, and just one-way and two-way ANOVA were commonly used.

One-way ANOVA

JAMOVI has the option ANOVA, it provides as standard option assumptions check, where data normality is checked using the Shapiro-Wilk test, Q-Q plots, and the variance may be checked using Levene's and Bartlett's test. The ANOVA normally needs distributed data and homogeneity of variances, it is known as Fisher's ANOVA. When homogeneity of variances was not observed, it is used the Welch

correction (Welch's ANOVA). When the data was not parametric, there was also the nonparametric version of ANOVA (Kruskal-Wallis test).

An ANOVA reports whether one or more significant differences among group levels exist, but it does not provide any information about specific group means compared to each other.[63] Post hoc tests are subsequently used to determine where the group differences are.[64]

JAMOVI provides parametric post hoc tests such as Tukey Scheffé, Bonferroni and Holm, where the Tukey correction was the most used. However, all these post-hoc tests were parametric and needed homogeneity of variances. When homogeneity of variances was not observed, there is the Games-Howell post hoc test. JAMOVI also provided a nonparametric post hoc test the Dwass-Steel-Critchlow-Fligner test.

The Kruskal-Wallis test

In this section, it was presented the nonparametric version of ANOVA (the Kruskal-Wallis test). As an example, the Kawamoto et al. data set was analyzed,[47] but this time the, the concentration of magnesium in the reference material which was determined by several laboratories in the reference material was analyzed.

The Shapiro-Wilk (p-value 0.039) and Anderson-Darling (p-value-0.012) tests showed that the data was not normally distributed, while the Kolmogorov-Smirnov test (p-value 0.162) showed that the data was normally distributed. However, the Kolmogorov-Smirnov test is less rigorous than the Shapiro-Wilk and the Anderson-Darling tests.

The Levene's test (p-value 0.002) and the Bartlett's test (p-value < 0.001) showed that variances were not homogeneous. Thus, the best option was the nonparametric test (the Kruskal-Wallis test). However, the parametric tests (Fisher and Welch) and the nonparametric test showed that the magnesium concentration determined by the laboratories were not equivalent (p-value < 0.001).

The differences between laboratories may be observed using the Dwass-Steel-Critchlow-Fligner multiple comparison test. However, it may be difficult to observe differences using a table due to the large number of laboratories which participated the study and differences can be easily observed using discriminant plots (Figure 7). Where it showed that laboratories 4 and 42 presented different results from the others.

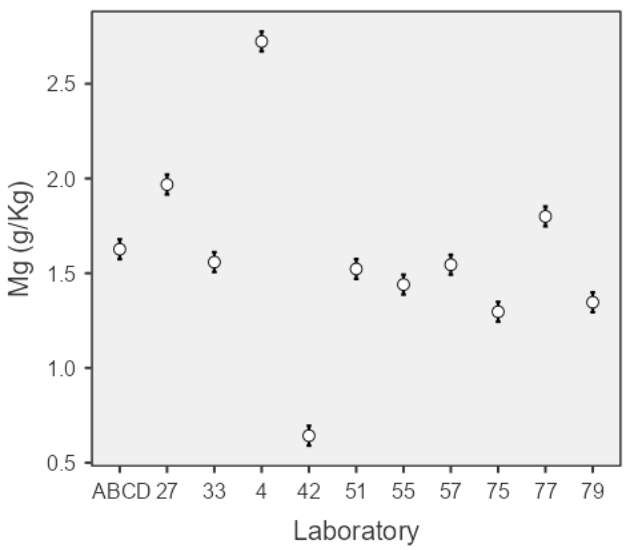


Figure 7: Descriptive plots of magnesium concentrations found in a candidate fish food reference material by different laboratories. Error bards were the confidence intervals at 95% confidence intervals.

Comparing the caffeine content in coffee grown in different seasons.

Comparing the caffeine concentration in coffee samples in three different years (spreadsheet coffee.xlsx; supporting information). The Levene's and Bartlett's (p-values 0.179 and 0.100, respectively) confirmed the homogeneity of variances. The Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling test showed that the data was normally distributed (p-values 0.291, 0.836, and 0.365, respectively).

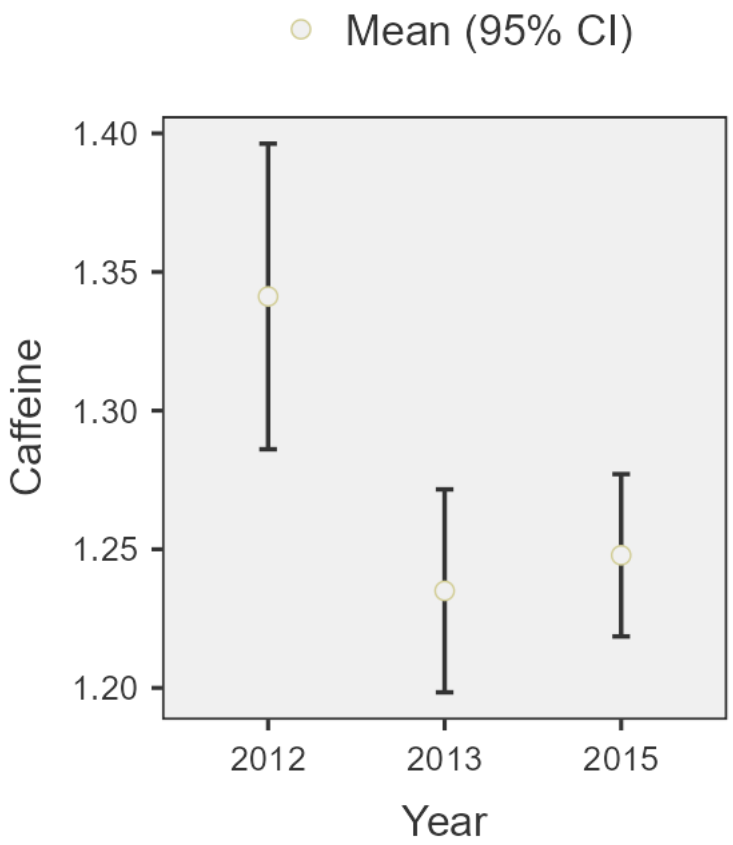
The Fisher ANOVA showed that the means were not equivalent (p-value = 0.001). The Tukey post hoc test showed that the caffeine concentration in 2012 samples were larger than in 2013 and 2015 samples, but coffee samples from 2013 and 2015 had equivalent caffeine concentrations (Table 2).

The observations obtained using Fisher's ANOVA and the Tukey test based on p-values may also be obtained using the discriminant plots (Figure 8) and boxplots (Figure 9).

Table 2: Tukey post hoc test for caffeine concentration in coffee samples from different years.

		2012	2013		2015	
2012	Mean difference	—	0.109	**	0.0965	**
	p-value	—	0.006		0.002	
2013	Mean difference		—		-0.0128	
	p-value		—		0.902	

Note. * p < .05, ** p < .01, *** p < .001



275

Figure 8: Discriminant plots for caffeine concentration in coffee samples from different ears.

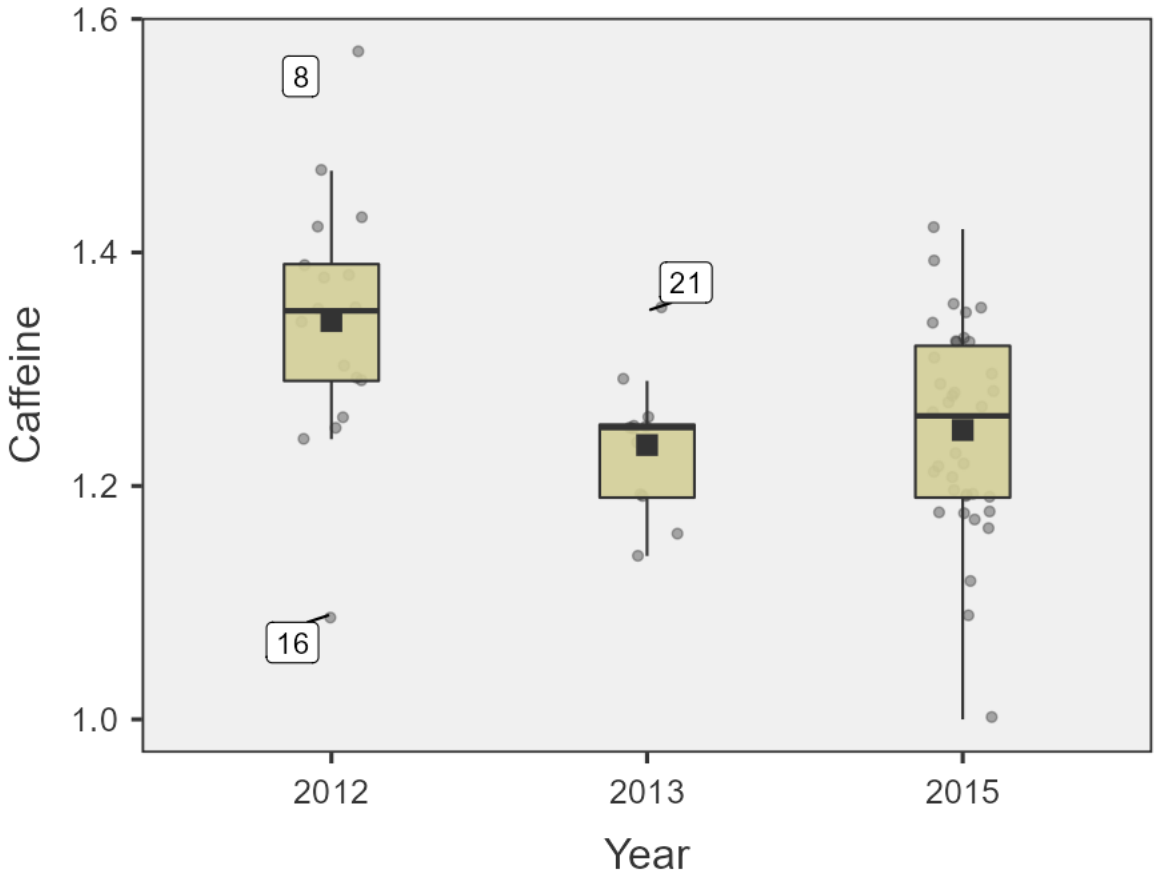


Figure 9: Boxplots for caffeine concentration in coffee samples from different years.

Comparing several independent groups.

In Spreadsheet Jatropha (supporting information) there is the total oil content of 145 Jatropha seeds grown in 14 different locations.[49] The Shapiro-Wilk and Anderson-Darling tests showed that the data was not normally distributed (p-values 0.008 and 0.023). The Levene's and Bartlett's showed that the variances were not homogeneous (p-values < 0.001). Thus, the Kruskal-Wallis test showed that the oil content in plants grown in different regions were not equivalent (p-value < 0.001).The parametric ANOVA (Fisher's and Welch) also provided the same conclusion (p-value < 0.001). The specific difference between locations may be done using the Dwass-Steel-Critchlow-Fligner pairwise comparisons or another post hoc test. However, these tables are huge, and it was hard to observe where the differences were.

Plots such as boxplots and discriminant plots provide a fast comparison between several groups. Looking at the discriminant plot (Figure 10), it was clear that location L1 and L14 provided the maximum and minimum oil content, locations L4, L5, L7, L8, L9 and L10 provide equivalent oil content.

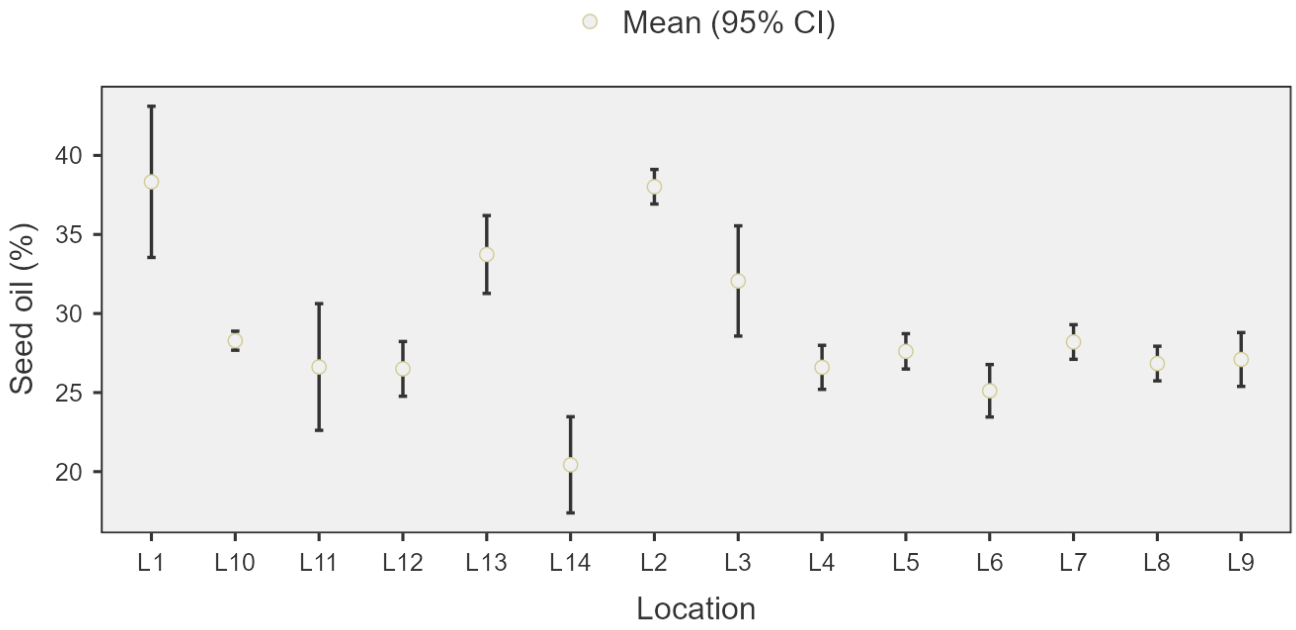


Figure 10: discriminant plot of the oil content in seed grown in different locations.

COMPARING DEPENDENT GROUPS

Paired variables were dependent variables. For example, the antioxidant capacity of different beverages obtained using a proposed and a standard method. Comparing the antioxidant activity index determined for different reducing compounds using different DPPH concentrations. Compare the fluoride concentrations determined in some samples using a standard and a proposed method.

Comparing two dependent groups (t-paired test).

When the data was normally distributed, paired groups were compared using the t test for paired samples. When the data was not normally distributed the Wilcoxon test is recommended.

As an example, the data in spreadsheet Folin-Ciocalteau, which contains the total antioxidant capacity of different beverages, was compared using a paired t test. The Shapiro-Wilk test (p-value = 0.416) and Q-Q plots showed that the data was normally distributed. The t test for paired groups showed that both methods were equivalent (p-value 0.821).

Comparing the results obtained using a pH meter and an app (spreadsheet App; supporting information).[52] The Shapiro-Wilk test showed that the data was normally distributed (p-value = 0.587). The test t for paired samples showed that the methods were not equivalent (p-value = 0.013). The discriminant plot illustrates that the methods were nonequivalent (Figure 11).

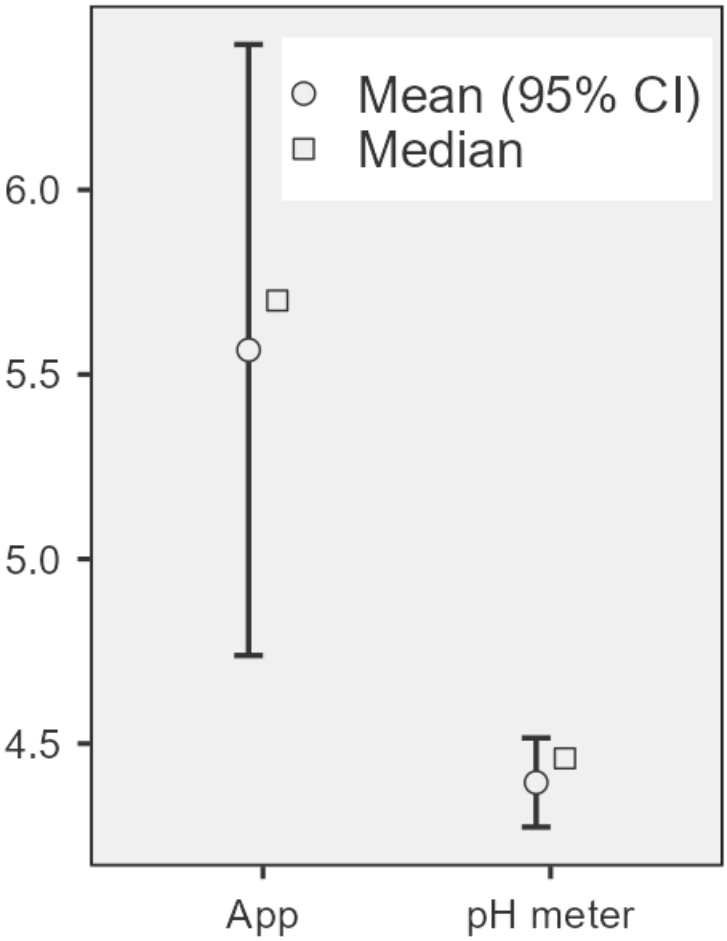


Figure 11; Comparing titration endpoints obtained using an app and a pH meter.

The Wilcoxon test.

As an example, the spreadsheet Fluoride (supporting information) was used. Eight real samples were analyzed using the standard and the prosed method. The Shapiro-Wilk test showed that the data was not normally distributed (p-value = 0.023). The Wilcoxon test showed that both methods were equivalent (p-value = 0.195). The t test for paired samples also showed that both methods were equivalent (p-value = 0.161). JAMOVI also has an add-on, TOSTER,[65,66] that also provides a visual comparison of two dependent groups.

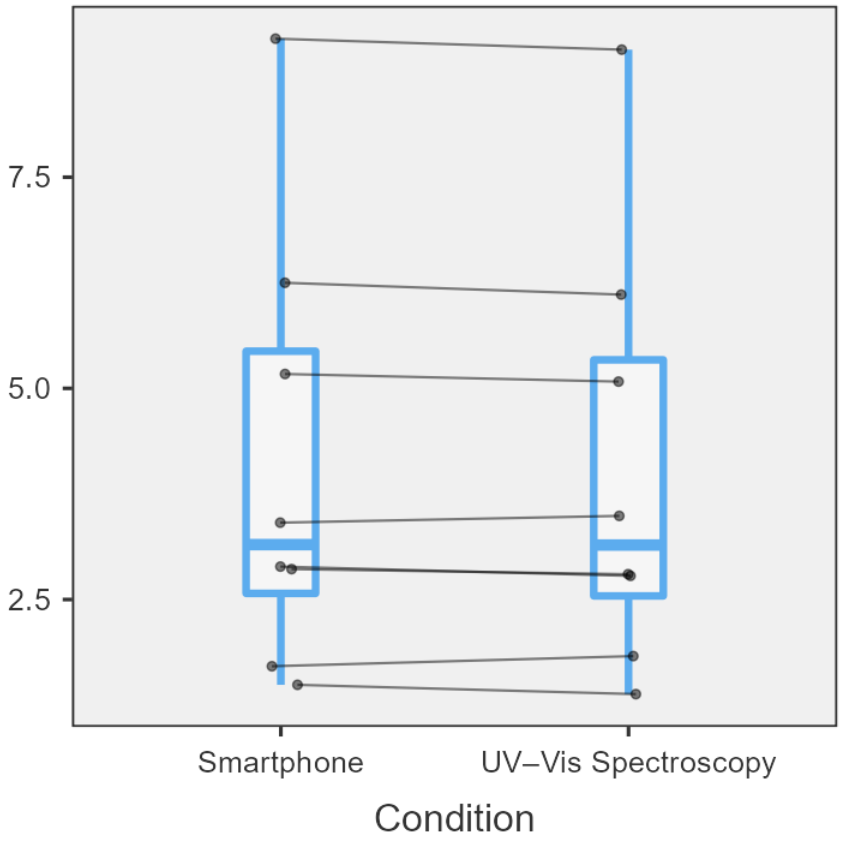


Figure 12: Comparison of two paired groups using TOSTER.

Comparing more than two dependent groups (RMANOVA).

When more than two dependent groups were compared it was used the repeated measure ANOVA (RMANOVA). Godoy & Scherer compared the antioxidant activity index (AAI) for several antioxidant compounds using. The objective was to show that the assay provides equivalent results independent of the amount of DPPH used.[54]

In the study, they considered the AAI measurements of different antioxidant compounds using different DPPH concentrations as independent variables and compared the AAI results using ANOVA.

The ANOVA was carried in spreadsheet AAI U. The three hypothesis tests showed that the data was not normally distributed (p-value < 0.001). Kruskal-Wallis test showed that the AAI values obtained using the three different DPPH concentrations were equivalent (p-value = 0.664).

When the DPPH concentrations were considered as paired variables. Firstly, the variances of paired variables were compared, it was named sphericity. Data sphericity was verified using the Mauchly's test. It confirmed the sphericity of the data (0.962). Then, data normality was checked using Q-Q plots, but it showed that the data was not normally distributed (Figure 13).

The RMANOVA showed that the AAI values were not equivalent (p-value < 0.001). Since the data was not normally distributed, its nonparametric version (the Friedman test) was used.

The Friedman test showed that the AAI values obtained using different DPPH concentration were not equivalent (p-value < 0.001).

The nonparametric post hoc test, Durbin-Conover pairwise comparisons, showed that AAI values obtained using DPPH concentrations of 47.45 and 30.75 µmol/mL were not equivalent (Table 3).

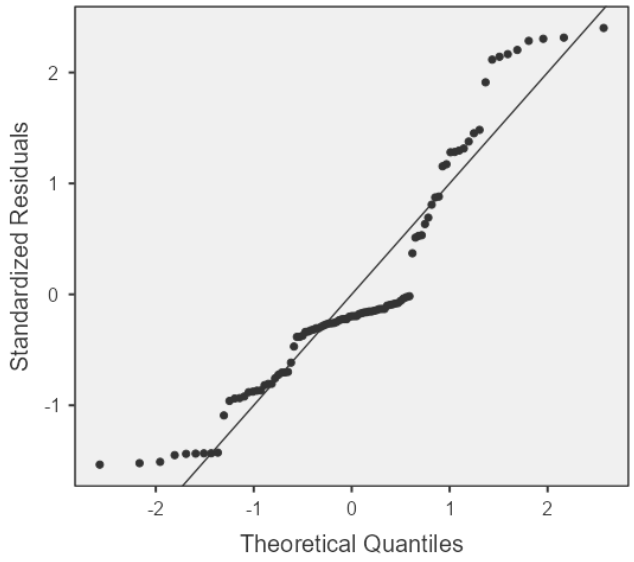


Figure 13: Q-Q plot of AAI values determined using different DPPH concentration.

Table 3: Pairwise Comparisons (Durbin-Conover) of AAI values obtained using different DPPH concentrations

			Statistic	p
AAI 76.89	-	AAI 47.75	1.35	0.182
AAI 76.89	-	AAI 30.75	1.35	0.182
AAI 47.75	-	AAI 30.75	2.70	0.009

The Friedman test.

The Friedman test is a nonparametric version of RMANOVA. It is a valuable tool for method comparison and method validation. For example, Ji et al.[55] evaluated the accuracy and precision of the method, by spiking honeysuckle with a standard pesticide mixture at three different levels (0.1, 0.2, and 1 mg/Kg (n = 6), spreadsheet recovery). The spiking levels effect on recovery was tested using RMANOVA. Mauchly's test confirmed the data sphericity (p-value = 0.127) and the data normality was checked using Q-Q plot (Figure 14). Since the Figure 14 showed that the data deviates from the normality. The effect of spiking level on recovery was analyzed using the Friedman test. The Friedman test showed that there was no effect of spiking on recovery (p-value = 0.173). The parametric RMANOVA also provides the same conclusion that the Friedman test (p-value = 0.263).

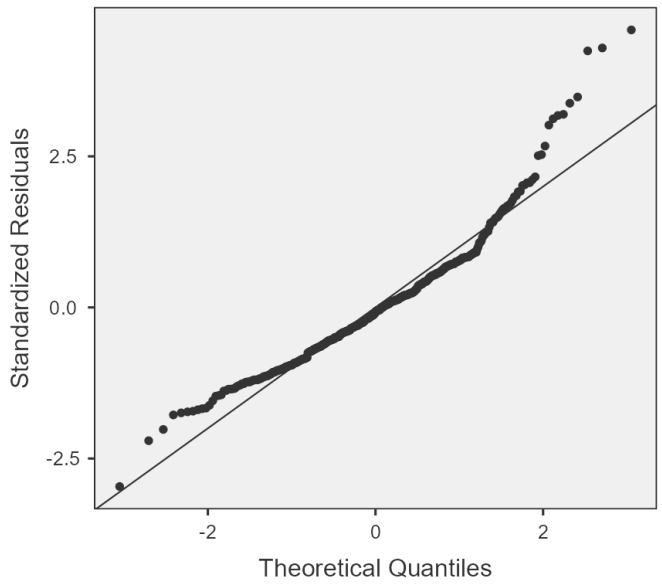


Figure 14: Q-Q plot of the recoveries of 147 pesticides at three spiking levels.

When different samples were analyzed using different methods (n >3), methods are compared using Friedman test, as an example the data provided by Crawford & Wang[56] (spreadsheet methods; supporting information) was analyzed. The Q-Q plots showed that the data deviates from the normality and the Mauchly's test confirmed the data sphericity (p-value = 0.564).

The Friedman test and RMANOVA showed that the methods were not equivalent, p-values 0.019 and 0.017, respectively. Differences between methods were observed using the Durbin-Conover post hoc test (Table 4), it showed that LC-UV method was different from LC-MS/MS and GC-MS with bromination.

Table 4: Pairwise Comparisons (Durbin-Conover)

			Statistic	p
LC-UV	-	LC-MS/MS	3.69	0.001
LC-UV	-	LC–UV with bromination	1.43	0.168
LC-UV	-	GC–MS with bromination	2.50	0.021
LC-MS/MS	-	LC–UV with bromination	2.26	0.035
LC-MS/MS	-	GC–MS with bromination	1.19	0.247
LC–UV with bromination	-	GC–MS with bromination	1.07	0.296

PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) is one of the most important and powerful methods in chemometrics as well as in a wealth of other areas. [24,67–71]

JAMOVI has the plugin MEDA that can generate high-resolution PCA plots, it has the FactoMineR package.[72] The use of this plugin for PCA was described using the spreadsheets Iris and coffee (supporting information).

Comparing Iris flowers using PCA.

The original four variables were combined in two new variables, which were principal components (PC), in the loading plot (Figure 15). Where the first PC1 (Dim 1) holds 72.96 of the variances and PC2 (22.85%). It also showed that sepal length, petal width, and petal length were correlated variables. The iris flowers were projected in this new space (PC1 vs PC2) and it is named the score plot (Figure 16).

In the loading plot (Figure 16) the setosa flowers were well separated from versicolor and virginica flowers, while the last two were not well separated. Looking to the loading plot (Figure 15), it was observed that setosa flowers had larger than versicolor and virginica flowers sepal length, petal width, and petal length.

In this example, there were just four original variables, measured in cm, which were standardized, before the PCA was run. This approach is necessary when variables have different numerical values.

In spreadsheet coffee (supporting information), some variables such as TTA (Total Titratable acidity) were measured in numbers that were much larger than e.g. Kahweol. For example, for TTA, the mean was 303 (mL NaOH 0,1 mol L^{-1} 100 g^{-1} green coffee bean) whereas Kahweol was 0.85 (g 100 g^{-1} green coffee bean). If this difference in scale and possibly offset is not handled, then the PCA model will only focus on variables measured in large numbers. It is desired to model all variables, and there is a preprocessing tool called standardization of variables, which will make each column have the same 'size' so that all variables have an equal opportunity of being modelled.

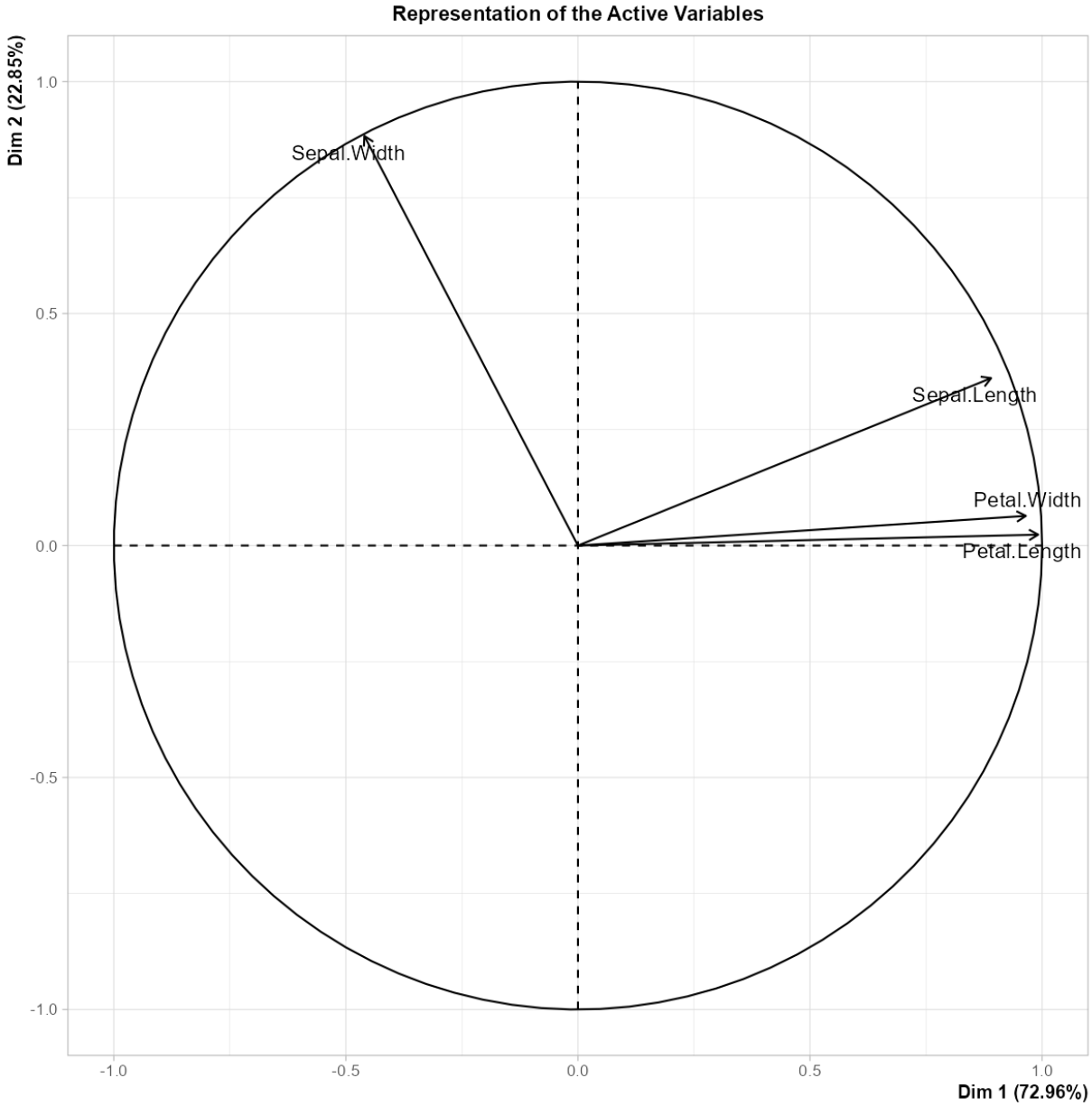


Figure 15: Loading plot of iris flowers.

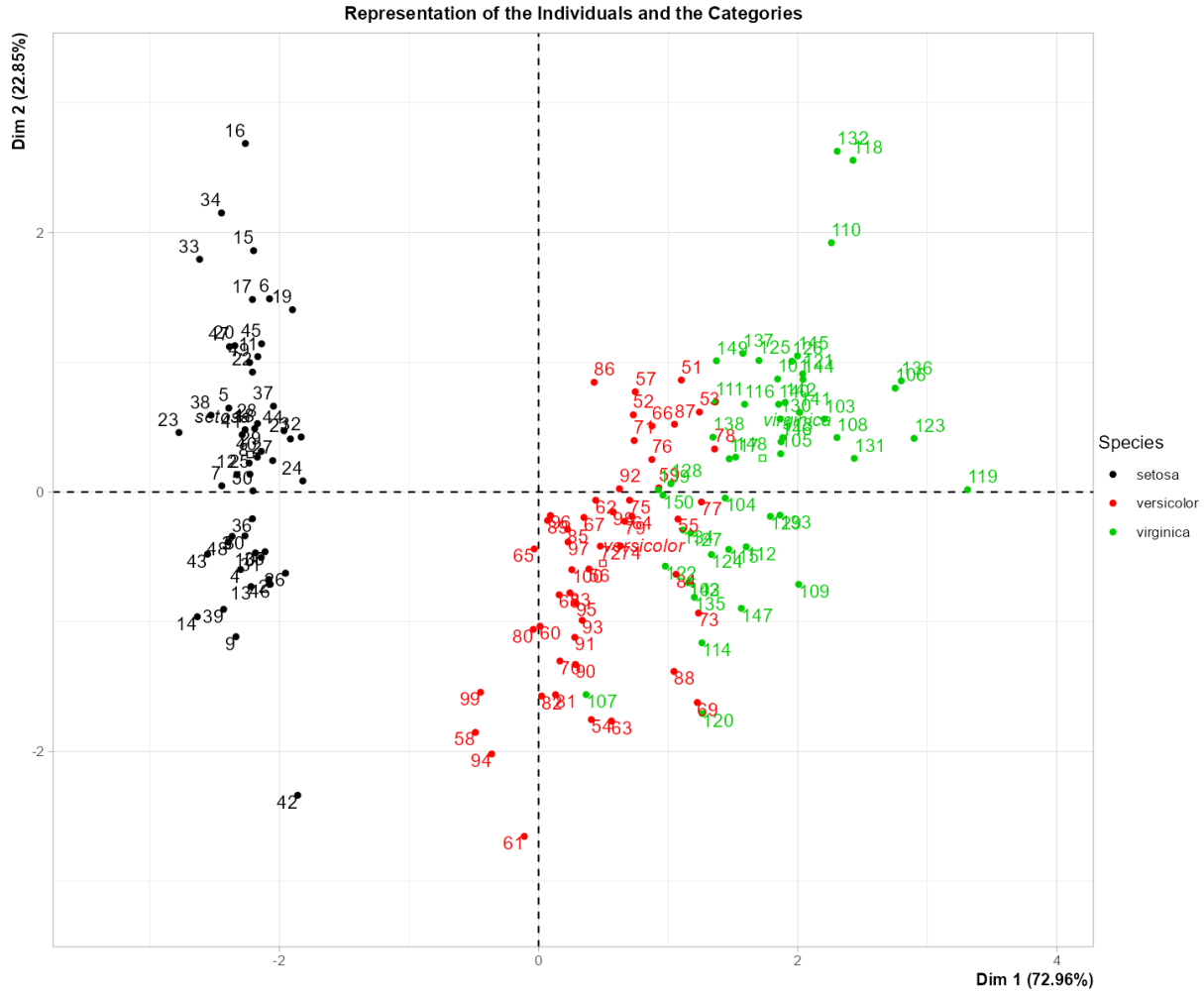


Figure 16: Score plot of iris flowers data set.

Comparing coffee beans from different harvests using PCA.

PCA was run MEDA using spreadsheet coffee (supporting information). The Loading plot (Figure 17) showed that caffeine, protein, and chlorogenic acid were related and inversely related to cafestol and GS (Global Score). The score plot (Figure 18) showed that samples from 2012 were different from those harvest in 2013 and 2015. Samples from 2012 have larger than caffeine, protein, and chlorogenic acid and lower cafestol and GS than samples harvested in 2013 and 2015.

In this case, TTA and GS have numerical values larger than the other variables. If the data was not standardized, the loading plot will account just for these two variables and the position of samples in the score plot dramatically changes (Figure 19). The samples were divided according to TTA and GS values, the PC1 (Dim 1) accounts for 98.52 of the total variances (Figure 19).

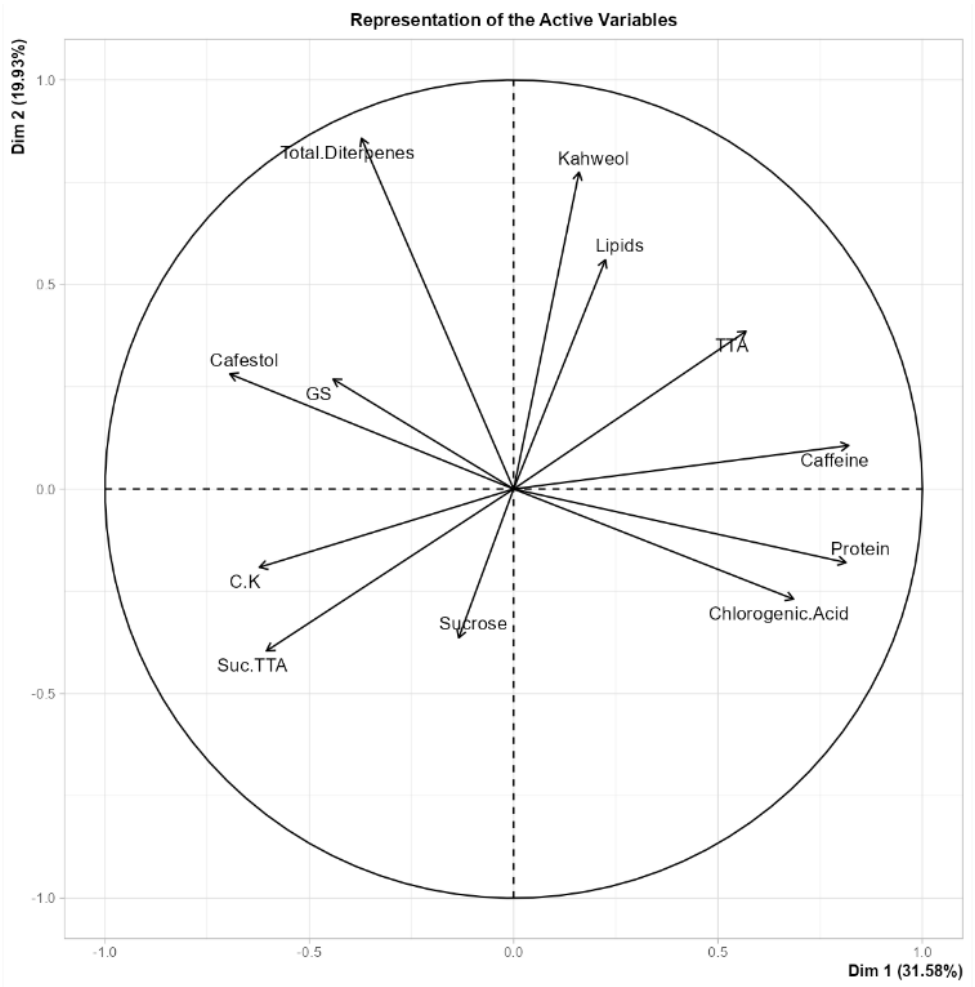


Figure 17: Loading plot of coffee beans from different harvests.

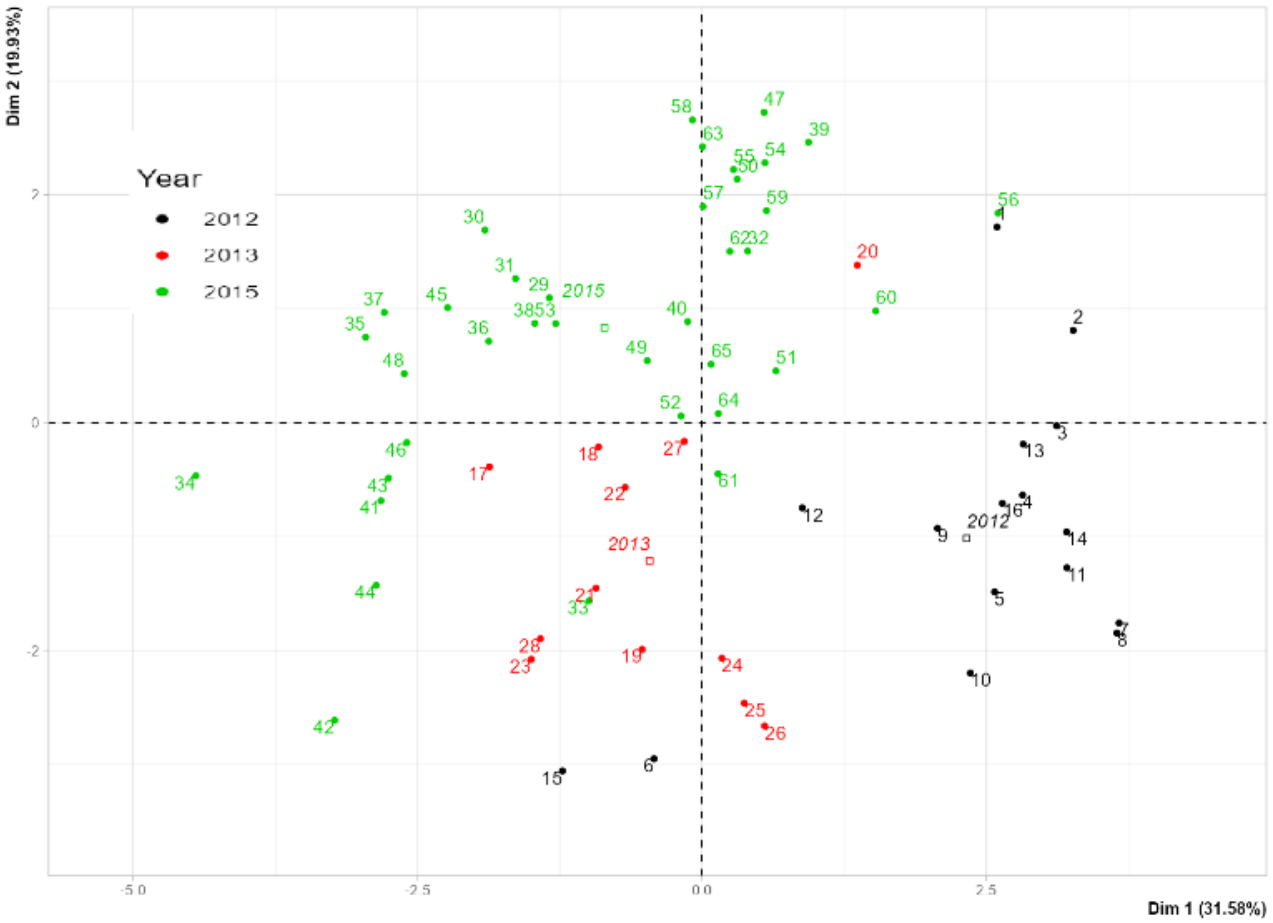
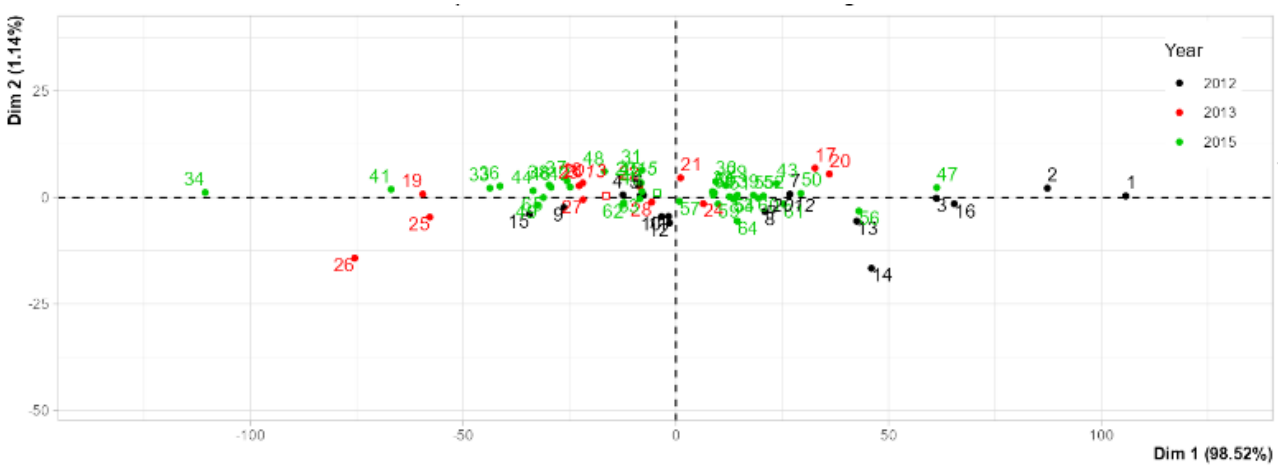


Figure 18: Score plot of coffee beans from different harvests.



410 Figure 19: Scores of coffee beans from different harvests without data standardization

Fatty acid profiles in peanut cultivars

Spreadsheet Peanut has 151 samples and four qualitative variables (cultivars, oil content (O/L), year). PCA can be plotted attributing different colors for each qualitative variable. For example,

different colors were attributed to each cultivar in the score plot (Figure 20), or different colors may be attributed to the oil content (O/L) (Figure 21).

Comparing the score plot (Figure 20) with the loading plot (Figure 22), it was shown that cultivars Tamrun were placed in the left side, according with the loading plot (Figure 22), the cultivars had larger C18:1 and C20:1 concentrations, and smaller C24:0, C22:0, C18:0, C20:0, C18:2, C16:0 than cultivars Georgia which were placed in right side of the score plot.

The score plot (Figure 20) showed that the cultivars C99-R were close to the cultivar Georgia, while the cultivar Flavorunner were close to the cultivars Tamrun. Comparing Figure 20 with Figure 21, it was shown that Tamrun cultivars had larger O/L than the Georgia.

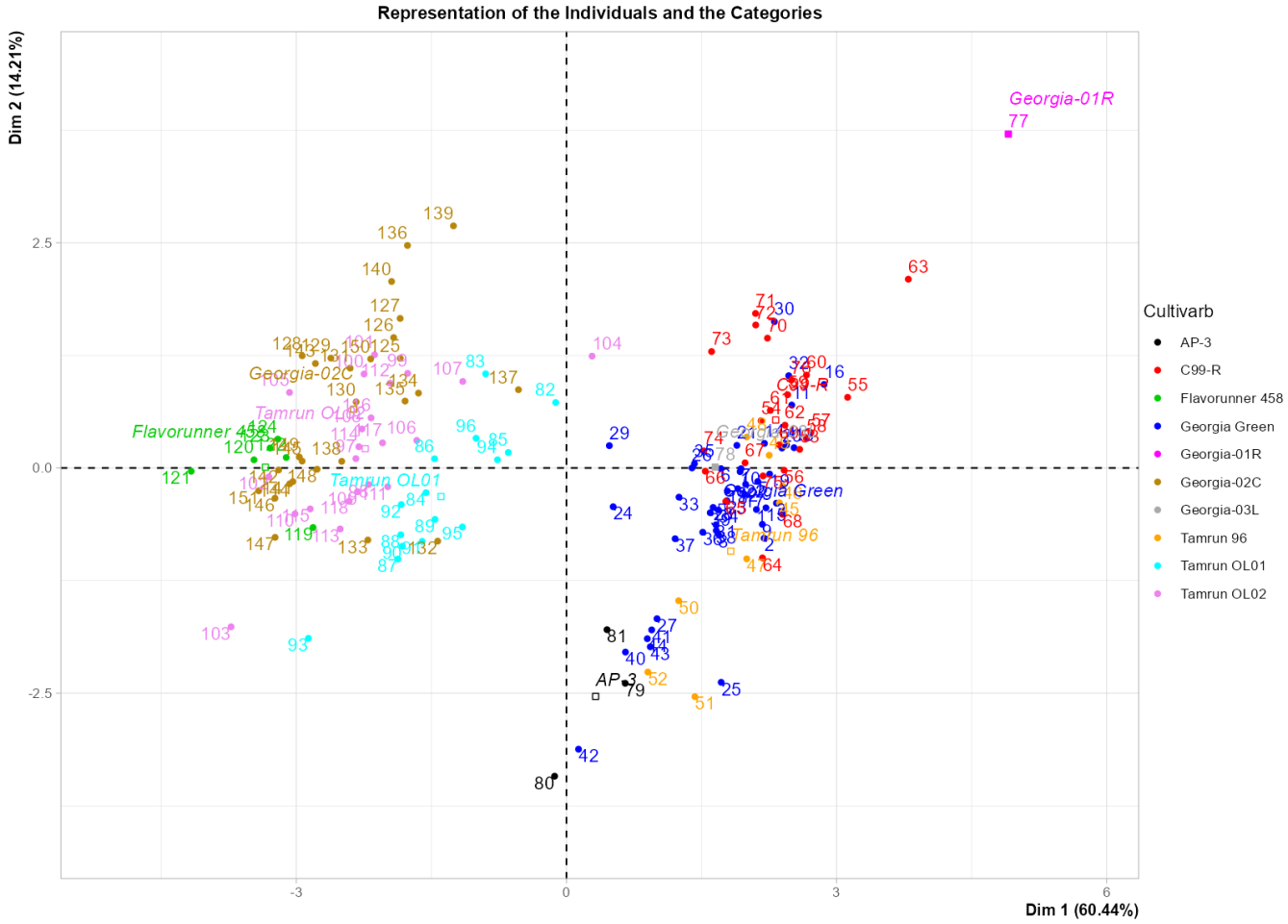


Figure 20: Score plot for peanuts of different cultivars

Figure 21: Score plot for peanuts of different oils content (High, Mid, and Normal)

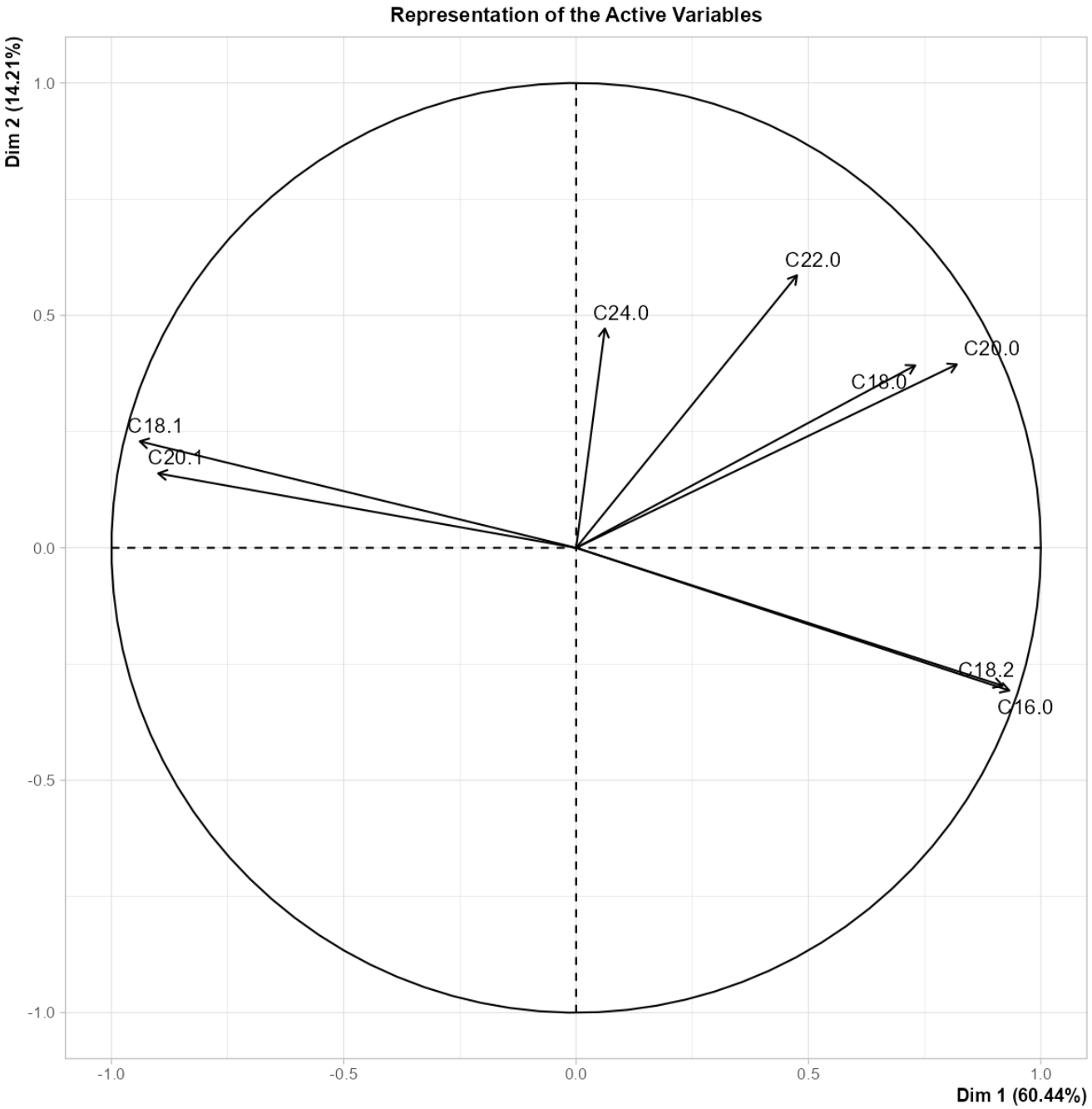


Figure 22: Loading plot for peanuts of different cultivars.

CONCLUSION

Some of the most used statistical tests and chemometric (PCA) tests were described using JAMOVI. In the first place, the data normality was checked using Hypothesis tests (Shapiro-Wilk) and plots (histograms, boxplots and Q-Q plots).

When the data was normally distributed, groups were independent. Group means were compared using t test and ANOVA (Fisher's). When groups were dependent and the data was not normally distributed, it was used for the Mann-Whitney (n = 2) and Kruskal-Wallis test (n > 3).

Once that the data was normally distributed and the groups were dependent, group means were compared using the paired t-test (n = 2) and RMANOVA (n > 2). When the data was not normally distributed and groups were dependent, it was used the Wilcoxon (n = 2) and Friedman tests (n > 2).

The JAMOVI also provided visual interpretation of these tests using boxplots and discriminant plots.

PCA was also easily carried out using JAMOVI's plugin MEDA. Hypothesis tests, plots and PCA can be easily done using the JAMOVI's click and go interface.

ASSOCIATED CONTENT

Supporting Information

The supporting information is provided in the link:

https://drive.google.com/drive/folders/1kyaJr0zkRvc8GTieYnTn33SzlJD4JOM5?usp=drive_link

Spreadsheets (ZIP) used in the examples given in the paper.

ChatGPT (DOCX) contains responses provided by ChatGPT

Questions (DOCX) contains questions related to the manuscript

AUTHOR INFORMATION

Corresponding Author

*E-mail: marcelborgesb@gmail.com; embsouza@furb.br

ACKNOWLEDGMENTS

The authors acknowledge financial support and fellowships from the Brazilian agencies FAPESC (Fundação de Amparo a Pesquisa do Estado de Santa Catarina), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the grant (399402226/2016-0), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

REFERENCES

(1) Schlotter, N. E. A Statistics Curriculum for the Undergraduate Chemistry Major. *J Chem Educ* **2013**, *90* (1), 51–55. https://doi.org/10.1021/ed300334e.

(2) Sanchez, J. M. The Need to Reinforce the Teaching of Basic Descriptive Statistics Required in Reporting Quantitative Laboratory Results: Diagnose of Common Students' Misconceptions. *J Chem Educ* **2023**, *100* (7), 2713–2718. https://doi.org/10.1021/acs.jchemed.3c00394.
(3) Harvey, D. Analytical Chemistry 2.0-an Open-Access Digital Textbook. *Anal Bioanal Chem* **2011**, *399* (1). https://doi.org/10.1007/s00216-010-4316-1.
(4) Harris, D. C.; Lucy, C. A. *Quantitative Chemical Analysis (International Edition)*; 2020.
(5) Granato, D.; de Araújo Calado, V. M.; Jarvis, B. Observations on the Use of Statistical Methods in Food Science and Technology. *Food Research International* **2014**, *55*, 137–149. https://doi.org/10.1016/j.foodres.2013.10.024.
(6) Matsumoto, P. S.; Cao, J. The Development of Computational Thinking in a High School Chemistry Course. *J Chem Educ* **2017**, *94* (9), 1217–1224. https://doi.org/10.1021/acs.jchemed.6b00973.
(7) Prais, M. G. Spreadsheet Exercises for Instrumental Analysis. *J Chem Educ* **1992**, *69* (6), 488. https://doi.org/10.1021/ed069p488.
(8) Edwards, P. A.; McKay, J. B.; Sink, C. W. First-Year Chemistry Laboratory Calculations on a Spreadsheet. *J Chem Educ* **1992**, *69* (8), 648. https://doi.org/10.1021/ed069p648.
(9) Larrabee, C. E.; Sprague, E. D. Spreadsheet Programming in Introductory Physical Chemistry. *J Chem Educ* **1988**, *65* (8), A204. https://doi.org/10.1021/ed065pA204.
(10) Wink, D. J. The Use of Matrix Inversion in Spreadsheet Programs To Obtain Chemical Equations. *J Chem Educ* **1994**, *71* (6), 490. https://doi.org/10.1021/ed071p490.
(11) Mullin, J. L.; Eierman, R. J. Extensive Student Use of a Spreadsheet in Quantitative Analysis. *J Chem Educ* **1990**, *67* (10), 878. https://doi.org/10.1021/ed067p878.
(12) Stone, C. A.; Mumaw, L. D. Practical Experiments in Statistics. *J Chem Educ* **1995**, *72* (6), 518. https://doi.org/10.1021/ed072p518.
(13) Simpson, J. M. Spreadsheet Statistics. *J Chem Educ* **1994**, *71* (4), A88. https://doi.org/10.1021/ed071pA88.
(14) Overway, K. Population versus Sampling Statistics: A Spreadsheet Exercise. *J Chem Educ* **2008**, *85* (5), 749. https://doi.org/10.1021/ed085p749.
(15) Hall, D. R.; D'eon, J. C. How's the Air Out There? Using a National Air Quality Database to Introduce First Year Students to the Fundamentals of Data Analysis. *J Chem Educ* **2023**, *100* (9), 3410–3418. https://doi.org/10.1021/acs.jchemed.3c00333.
(16) Joss, L.; Müller, E. A. Machine Learning for Fluid Property Correlations: Classroom Examples with MATLAB. *J Chem Educ* **2019**, *96* (4), 697–703. https://doi.org/10.1021/acs.jchemed.8b00692.
(17) Lafuente, D.; Cohen, B.; Fiorini, G.; García, A. A.; Bringas, M.; Morzan, E.; Onna, D. A Gentle Introduction to Machine Learning for Chemists: An Undergraduate Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling. *J Chem Educ* **2021**, *98* (9), 2892–2898. https://doi.org/10.1021/acs.jchemed.1c00142.
(18) Stippell, E.; Akimov, A. V.; Prezhdo, O. V. PySyComp: A Symbolic Python Library for the Undergraduate Quantum Chemistry Course. *J Chem Educ* **2023**. https://doi.org/10.1021/acs.jchemed.2c00974.
(19) van Staveren, M. Integrating Python into a Physical Chemistry Lab. *J Chem Educ* **2022**, *99* (7), 2604–2609. https://doi.org/10.1021/acs.jchemed.2c00193.
(20) Menke, E. J. Series of Jupyter Notebooks Using Python for an Analytical Chemistry Course. *J Chem Educ* **2020**, *97* (10), 3899–3903. https://doi.org/10.1021/acs.jchemed.9b01131.

(21) Bravenec, A. D.; Ward, K. D. Interactive Python Notebooks for Physical Chemistry. *J Chem Educ* **2023**, *100* (2), 933–940. https://doi.org/10.1021/acs.jchemed.2c00665.
(22) Dumancas, G. G.; Carreto, N.; Generalao, O.; Ke, G.; Bello, G.; Lubguban, A.; Malaluan, R. Chemometrics for Quantitative Determination of Terpenes Using Attenuated Total Reflectance–Fourier Transform Infrared Spectroscopy: A Pedagogical Laboratory Exercise for Undergraduate Instrumental Analysis Students. *J Chem Educ* **2023**. https://doi.org/10.1021/acs.jchemed.3c00358.
(23) Fox, J. The RCommander: A Basic-Statistics Graphical User Interface to R. *J Stat Softw* **2005**, *14* (9). https://doi.org/10.18637/jss.v014.i09.
(24) Sidou, L. F.; Borges, E. M. Teaching Principal Component Analysis Using a Free and Open Source Software Program and Exercises Applying PCA to Real-World Examples. *J Chem Educ* **2020**, *97* (6), 1666–1676. https://doi.org/10.1021/acs.jchemed.9b00924.
(25) Borges, E. M. Hypothesis Tests and Exploratory Analysis Using R Commander and Factoshiny. *J Chem Educ* **2023**, *100* (1), 267–278. https://doi.org/10.1021/acs.jchemed.2c00155.
(26) Love, J.; Selker, R.; Marsman, M.; Jamil, T.; Dropmann, D.; Verhagen, J.; Ly, A.; Gronau, Q. F.; Smíra, M.; Epskamp, S.; Matzke, D.; Wild, A.; Knight, P.; Rouder, J. N.; Morey, R. D.; Wagenmakers, E.-J. JASP : Graphical Statistical Software for Common Statistical Designs. *J Stat Softw* **2019**, *88* (2). https://doi.org/10.18637/jss.v088.i02.
(27) Marcel Borges, E. Data Visualization Using Boxplots: Comparison of Metalloid, Metal, and Nonmetal Chemical and Physical Properties. *J Chem Educ* **2023**, *100* (7), 2809–2817. https://doi.org/10.1021/acs.jchemed.3c00207.
(28) Filgueiras, M. F.; Borges, E. M. Iron Quantification in Dietary Supplements Using Four Colorimetric Assays. *J Chem Educ* **2022**, *99* (5), 2067–2078. https://doi.org/10.1021/acs.jchemed.1c00869.
(29) Assink Junior, E. J.; de Jesus, P. C.; Borges, E. M. Whey Protein Analysis Using the Lowry Assay and 96-Well-Plate Digital Images Acquired Using Smartphones. *J Chem Educ* **2023**, *100* (6), 2329–2338. https://doi.org/10.1021/acs.jchemed.2c00830.
(30) Bian, Y.; Alem, D.; Beato, F.; Hogenson, T. L.; Yang, X.; Jiang, K.; Cai, J.; Ma, W. W.; Fernandez-Zapico, M.; Tan, A. C.; Lawrence, N. J.; Fleming, J. B.; Yuan, Y.; Xie, H. Development of SOS1 Inhibitor-Based Degraders to Target *KRAS* -Mutant Colorectal Cancer. *J Med Chem* **2022**, *65* (24), 16432–16450. https://doi.org/10.1021/acs.jmedchem.2c01300.
(31) Licato, J. J.; Foster, G. D.; Huff, T. B. Zeolite Composite Materials for the Simultaneous Removal of Pharmaceuticals, Personal Care Products, and Perfluorinated Alkyl Substances in Water Treatment. *ACS ES&T Water* **2022**, *2* (6), 1046–1055. https://doi.org/10.1021/acsestwater.2c00024.
(32) Thagun, C.; Horii, Y.; Mori, M.; Fujita, S.; Ohtani, M.; Tsuchiya, K.; Kodama, Y.; Odahara, M.; Numata, K. Non-Transgenic Gene Modulation *via* Spray Delivery of Nucleic Acid/Peptide Complexes into Plant Nuclei and Chloroplasts. *ACS Nano* **2022**, *16* (3), 3506–3521. https://doi.org/10.1021/acsnano.1c07723.
(33) Anwar, S.; Ali, A.; Ullah, Z.; Binjawhar, D. N.; Sher, H.; Ali, R.; Iqbal, R.; Ali, B.; Ali, I. The Impact of *Trichoderma Harzianum* Together with Copper and Boron on Wheat Yield. *ACS Agricultural Science & Technology* **2023**, *3* (6), 517–527. https://doi.org/10.1021/acsagscitech.3c00076.
(34) Hufziger, K. A.; Farquharson, E. L.; Werner, B. G.; Chen, Q.; Goddard, J. M.; Nugen, S. R. *In Vivo* Capsid Engineering of Bacteriophages for Oriented Surface Conjugation. *ACS Appl Bio Mater* **2022**, *5* (11), 5104–5112. https://doi.org/10.1021/acsabm.2c00428.

(35) Clark, T. M. Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions. *J Chem Educ* **2023**, *100* (5), 1905–1916. https://doi.org/10.1021/acs.jchemed.3c00027.
(36) Tyson, J. Shortcomings of ChatGPT. *J Chem Educ* **2023**, *100* (8), 3098–3101. https://doi.org/10.1021/acs.jchemed.3c00361.
(37) Fergus, S.; Botha, M.; Ostovar, M. Evaluating Academic Answers Generated Using ChatGPT. *J Chem Educ* **2023**, *100* (4), 1672–1675. https://doi.org/10.1021/acs.jchemed.3c00087.
(38) Leon, A. J.; Vidhani, D. ChatGPT Needs a Chemistry Tutor Too. *J Chem Educ* **2023**. https://doi.org/10.1021/acs.jchemed.3c00288.
(39) Humphry, T.; Fuller, A. L. Potential ChatGPT Use in Undergraduate Chemistry Laboratories. *J Chem Educ* **2023**, *100* (4), 1434–1436. https://doi.org/10.1021/acs.jchemed.3c00006.
(40) Exintaris, B.; Karunaratne, N.; Yuriev, E. Metacognition and Critical Thinking: Using ChatGPT-Generated Responses as Prompts for Critique in a Problem-Solving Workshop (SMARTCHEMPer). *J Chem Educ* **2023**, *100* (8), 2972–2980. https://doi.org/10.1021/acs.jchemed.3c00481.
(41) Emenike, M. E.; Emenike, B. U. Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. *J Chem Educ* **2023**, *100* (4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00063.
(42) Navarro, D. J.; Foxcroft, D. R. Learning Statistics with Jamovi. *Learning Statistics with Jamovi: A Tutorial for Psychology Students and Other Beginners* **2022**.
(43) *https://www.jamovi.org/*.
(44) *ChatGPT*.
(45) Reda, T.; Powers, S. E.; Thavarajah, D. Falling into Line: Adaptation of Organically Grown Kale (Brassica Oleracea Var. Acephala) and Kale Relatives to Fall Planting. *Sci Hortic* **2022**, *295*, 110878. https://doi.org/10.1016/j.scienta.2022.110878.
(46) Borges, E. M.; Gelinski, J. M. L. N.; de Oliveira Souza, V. C.; Barbosa Jr., F.; Batista, B. L. Monitoring the Authenticity of Organic Rice via Chemometric Analysis of Elemental Data. *Food Research International* **2015**, *77*, 299–309. https://doi.org/10.1016/j.foodres.2015.06.008.
(47) Kawamoto, M. S.; de Souza, G. B.; de Araujo Nogueira, A. R. Preparation and Evaluation of a New Reference Material for Macro- and Micronutrients in Fish Feed. *Microchemical Journal* **2019**, *149*, 104027. https://doi.org/10.1016/j.microc.2019.104027.
(48) Barbosa, M. de S. G.; Scholz, M. B. dos S.; Kitzberger, C. S. G.; Benassi, M. de T. Correlation between the Composition of Green Arabica Coffee Beans and the Sensory Quality of Coffee Brews. *Food Chem* **2019**, *292*, 275–280. https://doi.org/10.1016/j.foodchem.2019.04.072.
(49) Kaushik, N.; Bhardwaj, D. Screening of Jatropha Curcas Germplasm for Oil Content and Fatty Acid Composition. *Biomass Bioenergy* **2013**, *58*, 210–218. https://doi.org/10.1016/j.biombioe.2013.10.010.
(50) Ledesma, C. M.; Krepsky, L. M.; Borges, E. M. Using a Flatbed Scanner and Automated Digital Image Analysis To Determine the Total Phenolic Content in Beer. *J Chem Educ* **2019**, *96* (10), 2315–2321. https://doi.org/10.1021/acs.jchemed.8b01037.
(51) Abderrahim, M.; M. Arribas, S.; Condezo-Hoyos, L. A Novel High-Throughput Image Based Rapid Folin-Ciocalteau Assay for Assessment of Reducing Capacity in Foods. *Talanta* **2016**, *152*, 82–89. https://doi.org/10.1016/j.talanta.2016.01.051.
(52) Li, J.; O'Neill, M. L.; Pattison, C.; Zhou, J. H.-W.; Ito, J. M.; Wong, C. S. T.; Yu, H.-Z.; Merbouh, N. Mobile App to Quantify PH Strips and Monitor Titrations: Smartphone-Aided

Chemical Education and Classroom Demonstrations. *J Chem Educ* **2023**, *100* (9), 3634–3640. https://doi.org/10.1021/acs.jchemed.3c00227.
(53) Yadav, P.; Laddha, H.; Agarwal, M.; Gupta, R. Fun with Smartphones: Handy Solution for Quantification of Debilitating Fluoride Ions in Drinking Water. *J Chem Educ* **2022**, *99* (7), 2677–2683. https://doi.org/10.1021/acs.jchemed.2c00216.
(54) Scherer, R.; Godoy, H. T. Antioxidant Activity Index (AAI) by the 2,2-Diphenyl-1-Picrylhydrazyl Method. *Food Chem* **2009**, *112* (3), 654–658. https://doi.org/10.1016/j.foodchem.2008.06.026.
(55) Ji, C.; Xiao, L.; Wang, X.; Hua, M. Z.; Wu, Y.; Wang, Y.; Wu, Z.; He, X.; Xu, D.; Zheng, W.; Lu, X. Simultaneous Determination of 147 Pesticide Residues in Traditional Chinese Medicines by GC–MS/MS. *ACS Omega* **2023**, *8* (31), 28663–28673. https://doi.org/10.1021/acsomega.3c03178.
(56) Crawford, L. M.; Wang, S. C. Comparative Study of Four Analytical Methods for the Routine Determination of Acrylamide in Black Ripe Olives. *J Agric Food Chem* **2019**, *67* (46), 12633–12641. https://doi.org/10.1021/acs.jafc.9b00363.
(57) FISHER, R. A. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Ann Eugen* **1936**, *7* (2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.
(58) Minh-Huy, D.; Anh-Dao, L.-T.; Thanh-Nho, N.; Nhon-Duc, L.; Cong-Hau, N. Smartphone-Based Digital Images as a Low-Cost and Simple Colorimetric Approach for the Assessment of Total Phenolic Contents in Several Specific Vietnamese Dried Tea Products and Their Liquors. *Food Chem* **2023**, *401*, 134147. https://doi.org/10.1016/j.foodchem.2022.134147.
(59) Shin, E.-C.; Craft, B. D.; Pegg, R. B.; Phillips, R. D.; Eitenmiller, R. R. Chemometric Approach to Fatty Acid Profiles in Runner-Type Peanut Cultivars by Principal Component Analysis (PCA). *Food Chem* **2010**, *119* (3), 1262–1270. https://doi.org/10.1016/j.foodchem.2009.07.058.
(60) Wyrzykowska, B.; Szymczyk, K.; Ichichashi, H.; Falandysz, J.; Skwarzec, B.; Yamasaki, S. Application of ICP Sector Field MS and Principal Component Analysis for Studying Interdependences among 23 Trace Elements in Polish Beers. *J Agric Food Chem* **2001**, *49* (7), 3425–3431. https://doi.org/10.1021/jf010184g.
(61) Kim, T. K.; Park, J. H. More about the Basic Assumptions of T-Test: Normality and Sample Size. *Korean J Anesthesiol* **2019**, *72* (4), 331–335. https://doi.org/10.4097/kja.d.18.00292.
(62) Lopes, M. D. M.; Castelo Branco, V. T. F.; Soares, J. B. Utilização Dos Testes Estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk Para Verificação Da Normalidade Para Materiais de Pavimentação. *TRANSPORTES* **2013**, *21* (1), 59. https://doi.org/10.4237/transportes.v21i1.566.
(63) Midway, S.; Robertson, M.; Flinn, S.; Kaller, M. Comparing Multiple Comparisons: Practical Guidance for Choosing the Best Multiple Comparisons Test. *PeerJ* **2020**, *8*, e10387. https://doi.org/10.7717/peerj.10387.
(64) Borges, E. M. Hypothesis Tests and Exploratory Analysis Using R Commander and Factoshiny. *J Chem Educ* **2023**, *100* (1), 267–278. https://doi.org/10.1021/acs.jchemed.2c00155.
(65) Macnamara, A.; Schinazi, V. R.; Chen, C.; Coussens, S.; Loetscher, T. The Effect of Age-Related Macular Degeneration on Cognitive Test Performance. *Sci Rep* **2022**, *12* (1), 4033. https://doi.org/10.1038/s41598-022-07924-8.
(66) Chou, R.; Dana, T.; Bougatsos, C.; Grusing, S.; Blazina, I. Screening for Impaired Visual Acuity in Older Adults. *JAMA* **2016**, *315* (9), 915. https://doi.org/10.1001/jama.2016.0783.
(67) Grant St James, A.; Hand, L.; Mills, T.; Song, L.; S. J. Brunt, A.; E. Bergstrom Mann, P.; F. Worrall, A.; I. Stewart, M.; Vallance, C. Exploring Machine Learning in Chemistry through the

Classification of Spectra: An Undergraduate Project. *J Chem Educ* **2023**, *100* (3), 1343–1350. https://doi.org/10.1021/acs.jchemed.2c00682.
(68) Hupp, A. M. Incorporating Chemometric Methods in the Undergraduate Curriculum: A Problem Set Activity for an Upper-Level Analytical Elective Course. *J Chem Educ* **2023**, *100* (3), 1377–1381. https://doi.org/10.1021/acs.jchemed.2c01013.
(69) Bro, R.; Smilde, A. K. Principal Component Analysis. *Anal. Methods* **2014**, *6* (9), 2812–2831. https://doi.org/10.1039/C3AY41907J.
(70) Granato, D.; Santos, J. S.; Escher, G. B.; Ferreira, B. L.; Maggio, R. M. Use of Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) for Multivariate Association between Bioactive Compounds and Functional Properties in Foods: A Critical Perspective. *Trends Food Sci Technol* **2018**, *72*, 83–90. https://doi.org/10.1016/j.tifs.2017.12.006.
(71) Gauthier, J. R.; Burns, D.; Sheng, J.; D'eon, J. C. Exploring the Composition and Authenticity of Honey and Syrup Samples Using Quantitative NMR Spectroscopy and Principal Component Analysis in an Upper-Year Undergraduate Analytical Environmental Course. *J Chem Educ* **2023**, *100* (1). https://doi.org/10.1021/acs.jchemed.2c00501.
(72) Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw* **2008**, *25* (1). https://doi.org/10.18637/jss.v025.i01.