

Exploratory Analysis of Biometric Data Concerning Characteristics of Urucum (*Bixa orellana* L.) in the Northeast of Brazil

Janilson Pinheiro de Assis¹, Roberto Pequeno de Sousa¹, Paulo César Ferreira Linhares¹,
Eudes de Almeida Cardoso¹, José Aluisio de Araújo Paula², Maria Francisca Soares Pereira³,
Lunara de Sousa Alves⁴, Francisco Laíres Cavalcante⁵ & Ariana da Mota Oliveira⁵

¹ Jitirana Research Group, Department of Agronomic and Forestry Sciences, Federal Rural Semi-Arid University, Mossoró, RN, Brazil

² Brazilian Service to Support at Micro and Small Enterprises of Rio Grande do Norte (SEBRAE-RN), Mossoró, RN, Brazil

³ Green Group of Agroecology and Bees, PB, Brazil

⁴ Federal University of Paraíba, Campus Areia, Areia, PB, Brazil

⁵ Federal University of Campina Grande, Campus Pombal, PB, Brazil

Correspondence: Janilson Pinheiro de Assis, Jitirana Research Group, Department of Agronomic and Forestry Sciences, Federal Rural Semi-Arid University, Mossoró, RN, 59625-900, Brazil. E-mail: janilson@ufersa.edu.br

Received: January 10, 2019

Accepted: February 19, 2019

Online Published: April 15, 2019

doi:10.5539/jas.v11n5p225

URL: <https://doi.org/10.5539/jas.v11n5p225>

Abstract

Urucum is a plant adapted to the soil and climate conditions of the semi-arid region. This study evaluates the biometry of urucum seeds. Twenty seeds of annatto were collected in an area of native vegetation with presence of the species, located in the Mossoró Mountains in the municipality of Mossoró, State of Rio Grande do Norte, Northeast Brazil in July 2017 and taken to the plant breeding laboratory, where the following characteristics were evaluated: (a) morphological characterization of the seed being determined the length and width in millimeters, of 200 seeds well developed, with the aid of a pachymeter with precision of 0.1mm and (b) weight of the seed expressed in grams. Descriptive and graphical analyzes were carried out using the statistical software R. The length and width showed a small range of variation, resulting in excellent and reasonable values of coefficients of variation, respectively. We found a regular degree of symmetry and a mesokurtic distribution for length and width of seeds. There was no significant linear correlation between length and width. The features of urucum seeds did not fit to the normal distribution of probability.

Keywords: medical species, descriptive statistics, inference

1. Introduction

The urucum or urucu (*Bixa orellana* L.), a shrub species native from tropical America, has gained importance in the world agricultural market every year and is cultivated in the tropics all over the world (Mercadante & Pfander, 1998). The fruits are ovoid capsules containing 30 to 40 seeds (Lorenzi, 1998). The urucum produces particular atoxic coloring substances with the indication as a hypolipidemic agent, which makes it one of the few pigments allowed by the World Health Organization (Silva et al., 2006).

The species has multiple uses, such as ornamental and medicinal plant and soil restorer. Urucum is a pioneer species native to the Amazon Rainforest. Its seeds are used to produce condiments and tinctures widely used in industry. The fast growth of urucum allows the cultivation together with other species in degraded areas destined to reforestation (Lorenzi, 1998).

Morphological analyzes of fruits and seeds help to understand the germination process and the characterization of vigor and viability (Mathuem & Lopes, 2007). Biometric analyzes comprise an essential tool to detect genetic variability within and between populations and to define the relationships between the variability and environmental factors, thus contributing to genetic improvement programs (Gusmão et al., 2006).

Biometrics of fruits and seed are central in studies of species, providing data for differentiation of species within the same genus. For example, Carpanezzi and Marques (1981) showed that the seeds of *Hymenaea courbaril* L.

weight almost twice the seeds of *Heritiera parvifolia*. Also, the seeds characteristics influence its patterns of dispersion and establishment of seedlings (Fenner, 1993), which are used to differentiate pioneer and non-pioneer species in tropical forests (C. C. Baskin & J. M. Baskin, 1998). In most shrubs and trees, there is an inverse relation between seed size and the number of seeds per fruit (Carvalho et al., 1998).

Urucum stands out for its therapeutic potential and as food. However, there is no record of its exploration in the region of Mossoró, RN, as well as, studies of urucum biometry, showing the need for studies concerning this species.

The objective of this study was to evaluate the use of statistical techniques of exploratory data analysis, the biometric variables related to the seeds of annatto, as support for studies comparing their characteristics when submitted to different environments and aiming at genetic studies and plant breeding.

2. Materials and Methods

The study was carried out in Mossoró, RN, geographic coordinates: 5°11' S and 37°20' W at 18 m of altitude, with an annual mean temperature of 27.5 °C and relative humidity of 68.9% (Carmo Filho et al., 1991). According to Köppen's classification, the climate in Mossoró is BSw^h, hot and dry. Twenty seeds of annatto were collected in an area of native vegetation with presence of the species, located in the Mossoró Mountains in the municipality of Mossoró, State of Rio Grande do Norte, Northeast Brazil in July 2017 and taken to the plant breeding laboratory, where the following characteristics were evaluated: (a) morphological characterization of the seed being determined the length and width in millimeters, of 200 seeds well developed, with the aid of a pachymeter with precision of 0.1 mm and (b) weight of the seed expressed in grams. The descriptive analysis and graphs were done with the software R version 3.1.1 (2018).

3. Results and Discussion

To determine and compare quantitative aspects of the distributions of values of the variables seeds length and width we based the analysis on specialized biometric literature (Ferreira, 2005; Spiegel & Stephens, 2009; Oliveira et al., 2009; Bussab & Morettin, 2010; Zar, 2010; Claudio & Stein, 2011; Cecon et al., 2012). Thus, we adopted the exploratory data analysis using frequency distributions, box plots, as well as the statistical estimators of the variables under study, which are the main descriptive and inferential statistical measures, such as arithmetic mean, median, total range, variance, standard deviation, standard error of mean, coefficient of variation, asymmetry coefficient, kurtosis coefficient, quartiles and interquartile deviation. We used the Pearson correlation coefficient and statistical inference such as the hypothesis tests, T-test or Z-test, at a significance level of 5% probability, based on Student's t-distribution and Normal distribution, respectively, for the construction of confidence intervals with 95% and 99% probability (Tables 1 to 5 and Figures 1 to 9).

Table 1. Frequency distribution of seed length (mm) of Urucum (*Bixa orellana* L.)

Classes (Length in mm)	f_i	X_i	$f\%$
[2.99-3.25)	1	3.12	0.50
[3.25-3.51)	3	3.38	1.49
[3.51-3.77)	9	3.64	4.46
[3.77-4.03)	20	3.90	9.90
[4.03-4.28)	40	4.15	19.80
[4.28-4.54)	64	4.41	31.68
[4.54-4.80)	49	4.67	24.26
[4.80-5.06)	16	4.93	7.92
Total	200	-	100

Table 2. Frequency distribution of seed width (mm) of Urucum (*Bixa orellana* L.)

Classes (Width in mm)	f_i	X_i	$f\%$
[1.73-2.00)	7	1.87	3.47
[2.00-2.27)	25	2.14	12.38
[2.27-2.55)	26	2.41	12.87
[2.55-2.82)	26	2.68	12.87
[2.82-3.09)	34	2.96	16.83
[3.09-3.36)	31	3.23	15.35
[3.36-3.64)	21	3.50	10.40
[3.64-3.91)	24	3.77	11.88
[3.91-4.18)	8	4.04	3.96
Total	200	-	100

Table 3. Frequency distribution for milligram weight of urucum (*Bixa orellana* L.) seeds

Classes (Weight in milligrams)	f_i	X_i	$f\%$
[5.40-9.94)	5	7.67	2.48
[9.94-14.47)	16	12.20	7.92
[14.47-19.01)	17	16.74	8.42
[19.01-23.54)	48	21.27	23.76
[23.54-28.08)	49	25.81	24.26
[28.08-32.61)	46	30.35	22.77
[32.61-37.15)	15	34.88	7.43
[37.15-41.68)	3	39.42	1.49
[41.68-46.22)	3	43.95	1.49
Total	200	-	100

Table 4. Exploratory data analysis (descriptive statistics) and inductive analysis of the variables: length in millimeters, width in millimeters and weight in milligrams of 200 urucum (*Bixa orellana* L.) seeds

Sample statistic or estimator	Length (mm)	Width (mm)	Weight (mg)
Sample size (number of seeds)	200	200	200
Minimum value	2.99	1.73	5.40
Maximum value	5.05	4.17	46.20
Total Amplitude	2.06	2.44	40.80
Arithmetic Mean	4.36	2.96	24.61
Median or Second Quartile	4.40	2.98	25.40
Variance	0.12	0.33	50.42
First quartile	4.15	2.51	20.25
Third quartile	4.60	3.37	29.15
Standard deviation	0.35	0.58	7.10
Mean standard error	0.03	0.04	0.50
Coefficient of variation (%)	8.07	19.43	28.85
Asymmetry Coefficient	-0.73	0.07	-0.14
Curtose Coefficient	0.84	-0.91	0.23
Interquartile Range (IIQ)	0.45	0.86	8.90
Z Test for the mean at 0.1% probability	145.33***	74.00***	49.22***
Confidence Interval for the average 95% probability	4.30 to 4.42	2.88 to 3.04	23.63 to 25.59
Theoretical Normal Distribution or Gaussian	P value = 0.004	P value = 0.10	P value = 0.64
Probabilities, D'Agostino-Pearson's Test	No Normal Distribution Adjustment	Normal Distribution Adjustment	Normal Distribution Adjustment

Table 5. Results of the descriptive and inductive statistical analysis of Pearson's simple linear correlation coefficient (r), between the (x , y) length and width pairs in millimeters (mm) and the weight in milligrams of 200 sample units for urucum (*Bixa orellana* L.) seeds, Mossoró, RN, Brazil, 2018

Variables studied	Length	Width	Weight
Length	1.00	-	-
Width	0.14	1.00	-
Weight	-0.01	-0.06	1.00
Confidence Intervals [I.C.] for Pearson's simple linear correlation coefficient (r) at 95% and 99% confidence probability			
	Length \times Width	Length \times Weight	Width \times Weight
	IC _(95%) : 0.01 to 0.28	IC _(95%) : -0.15 to 0.12	IC _(95%) : -0.14 to 0.13
	Length \times Width	Length \times Weight	Width \times Weight
	IC _(99%) : -0.06 to 0.30	IC _(99%) : -0.29 to 0.07	IC _(99%) : -0.19 to 0.18
P-value of the Student's t-parametric test for the simple linear correlation coefficient of Pearson r	0.09	0.11	0.95

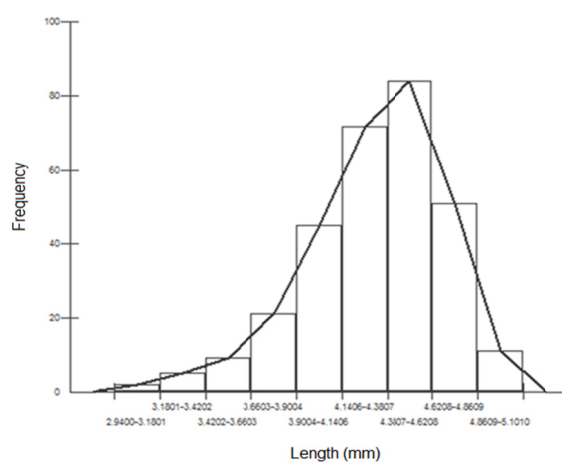


Figure 1. Histogram and polygon of frequencies representative of the length distribution (mm) of urucum seeds (*Bixa orellana* L.)

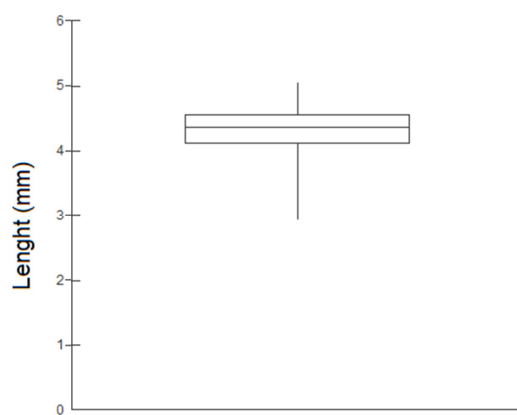


Figure 2. Boxplot using median and quartiles of length of seeds (mm) of Urucum (*Bixa orellana* L.)

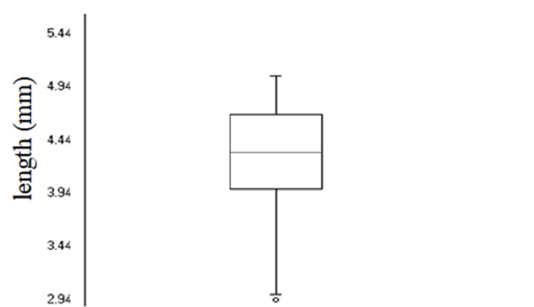


Figure 3. Box plot using the mean and standard deviation of seeds length (mm) of urucum (*Bixa orellana* L.)

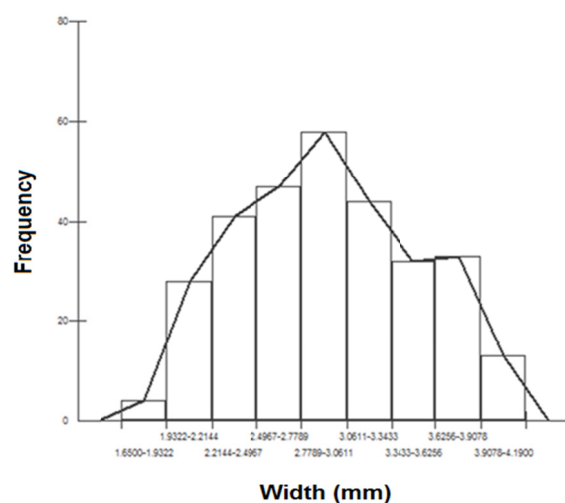


Figure 4. Histogram and polygon of frequencies of the width (mm) distribution of urucum seeds (*Bixa orellana* L.)

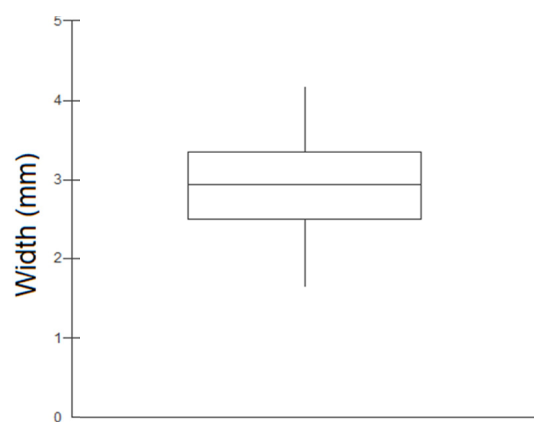


Figure 5. Box plot using median and quartiles of the seeds width (mm) of urucum (*Bixa orellana* L.)

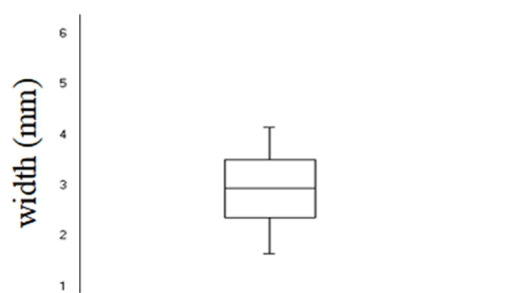


Figure 6. Box plot using the mean and standard deviation of seed width (mm) of Urucum (*Bixa orellana* L.)

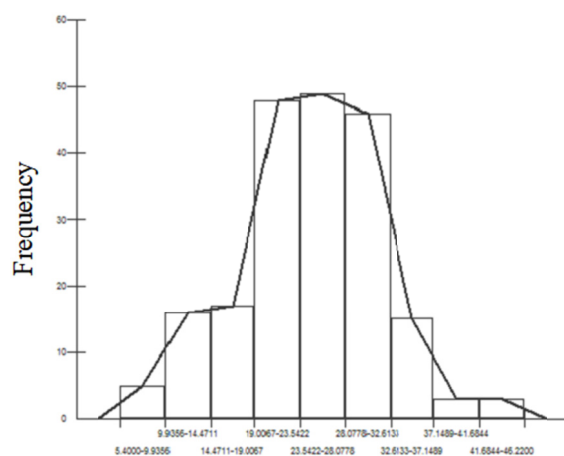


Figure 7. Histogram of frequencies and polygon of frequencies representative of the weight distribution in milligrams of urucum seeds (*Bixa orellana* L.)

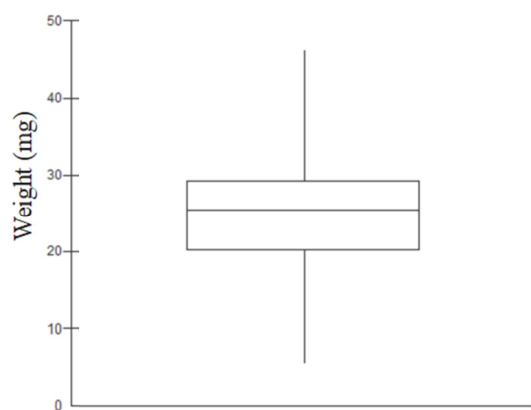


Figure 8. Box plot with median and quartiles referring to the weight in milligrams of urucum (*Bixa orellana* L.) seeds

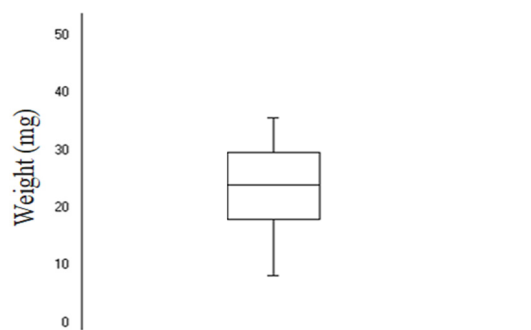


Figure 9. Box plot with mean and standard deviation referring to the weight in milligrams of urucum (*Bixa orellana* L.) seeds

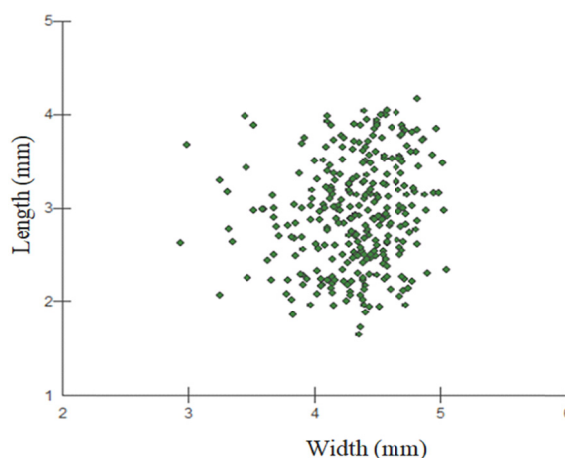


Figure 10. Pearson linear correlation dispersion diagram for length and width (mm) of 200 urucum seeds (*Bixa orellana* L.)

The length and width of seeds had a small amplitude of variation and a greater amplitude for the weight in milligrams (Table 4), as well as an optimum, good and regular value respectively for the coefficients of variation, which shows a high, medium and regular degree, respectively, of homogeneity of these characteristics evaluated, the length being much less dispersed than the width and the weight of the seeds. There is not a great similarity in the location of the length and the width of the seeds, since the means 4.36 and 2.96 mm for the average length and the average width, respectively, are not very close. It is worth noting that the calculation of the average does not have to make real sense. However, it can be stated that the seeds have a center of gravity in the distribution of these variables, which shows that the length of the seeds is close to 4.36 millimeters and the width around 2.96 millimeters. The relative variation between the variables measured by the coefficients of variation, in this case, was the most adequate, since they are independent of the magnitude of the variable, as well as of its unit of measure, being, therefore, the most correct procedure. The coefficients of variation (8.07%, 19.43% e 28.85% for length, width and the weight, respectively), thus showing a lower relative variability for seed length when compared to the width and weight, thus presenting a significant difference magnitude. Therefore, these coefficients were the most adequate to measure and compare the relative variability between analyzed characteristics, since they are independent of the magnitude of the variables and their units (Ferreira, 2005; Oliveira et al., 2009; Zar, 2010; Fonseca & Martins, 2012).

Other results obtained, as can be seen in Table 4, were that the mean and median were very close to the length, width and weight, which indicated a reasonable degree of symmetry for these characteristics, which were respectively -0.73, 0.07 and -0.14. On the other hand, the coefficients of kurtosis that were 0.84, -0.91 and 0.23, respectively, showed that these characteristics presented mesocuric distribution, and the characterization of the degree of asymmetry and kurtosis of a distribution according to Spiegel and Stephens (2009), Zar (2010), Andrade and Ogliari (2010), Casella and Berger (2010), and Fonseca and Martins (2012), can not be done by looking only at measures of position or central tendency such as the average and the median, but also through the asymmetry and kurtosis coefficients, as well as through frequency histograms and polygons (Figures 1 and 4).

Regarding quantiles, the first quartile shows that 25% of the lowest values for the length, width and weight of the seeds reach a maximum of 4.15 mm, 2.51 mm and 20.25 mg, respectively. On the other hand, showed that the 25% of seed lengths, widths and weights are represented by at least 4.60 mm, 3.37 mm and 29.15 mg, respectively (Figueiredo et al., 2007). In this case the interquartile range obtained, which serves to verify the dispersion of the data in relation to the median and thus to identify the presence of outlier's data, was 0.45 mm for the length of 0.86 mm for the width and 8.90 milligrams for seed weight (Table 4).

The degree of association or simple linear correlation, whose values range from -1 to +1, including zero, and its magnitude is dimensionless, between length and seed width of urucum was 0.14 (Table 4), showing a non-significant positive relation, as evidenced by the result of Student's t-test (Ferreira, 2005; Figueiredo et al., 2007; Oliveira et al., 2009; Spiegel & Stephens, 2009; Bussab & Morettin, 2010; Zar, 2010; Claudio & Stein, 2011; Cecon et al., 2012; Costa, 2012).

From a graphic, visual or geometric point of view, the dispersion diagram (Figure 7), showed high dispersion which confirms the lack of correlation between length and width found in the low value of correlation coefficient (0.14). The length and width variables of urucum seed did not fit to the theoretical or special density distribution of normal or bivariate Gaussian probability so that all the conclusions obtained through statistical inference are not guaranteed or assured by this assumption (Table 4).

Reporting the quantiles comprises an excellent way to illustrate the dispersion of a distribution. Researchers are more familiar with the type of quantile called percentile because of its use in standardized tests. When a test score is reported within the 90th percentile, 90% of the scores are smaller than it, and 10% are higher. Contrary to variance and standard deviation, quantile values do not depend on arithmetic mean or median values. When distributions are asymmetric or have outliers, which are extreme values that are not characteristic of the distribution quantile box plots can portray the distribution of data more accurately than mean and standard deviation.

We obtained high values at the Z test, both for seed length and seed width, concluding that the mean values of these characteristics were highly significant (Table 4).

According to the results shown in Table 5 by means of descriptive and inferential statistical analyzes, it was verified that there is no simple linear correlation of Pearson r statistically significant as a function of the values obtained for or are low myths, between variables length in millimeters with width also in millimeters, length in millimeters with weight in grams and width in millimeters and weight in grams. In addition, inferential statistical analyzes by constructing confidence intervals with 95% and 99% probability, showed that in repeated sampling there is a high reliability that these results occur or are confirmed in at least ninety nine times the on the other hand also in function of the application of the Student's t-parametric test to or, it was verified that this produced levels described or p-value for said test of results with quite high probability values not allowing the researchers reject the hypothesis of nullity that the true population coefficient of simple linear correlation of Pearson ρ between these evaluated variables is null, that is, the hypothesis of nullity $H_0: \rho = 0$ is rejected. It is worth noting that the null hypothesis used in this work will always be equal to zero in order to guarantee the symmetry of the sample distribution of the estimate or r so that it can be modeled through the curve of a theoretical distribution of Student's t-probability, since if one assumes a hypothesis that this theoretical population coefficient is different from zero, that is, $H_0: \rho \neq 0$ would have to be applied a Fisher zeta transform to be able to guarantee such symmetry of the sample distribution of the simple linear correlation coefficient of Pearson r , and thus allow the construction of the confidence intervals and the application of the parametric tests of Student's hypothesis t (Fonseca & Martins, 2012). These results reinforce the need to conduct other works repeated in time and space to verify convergent or divergent results guiding researchers in the evaluation of genetics and plant breeding as well as yield performance of this plant species to make it viable as a crop commercial.

Boxplots, using the median and the quartiles are a widely used chart in the biological and medical sciences, showing the median, first and third quartiles. It also shows the lowest and highest scores through the lower and upper boundaries of vertical straight lines, which originate from the first and third quartiles, respectively. According to the results obtained in Figure 2, using the median and quartiles, a strong concentration of the seed length data in millimeters around the central value in the case the median was observed, which evidenced a high homogeneity of these observations, which makes its analysis and current and subsequent interpretation, including the making of statistical inferences such as the construction of confidence intervals and the application of hypothesis test tests, as well as the adjustment of regression models for estimation and forecasting purposes.

Using the mean and standard deviation, the box plot graph, similar to the previous graph, shows the mean and standard deviation in the Box. It also shows the lowest and highest scores through the upper and lower boundary

of vertical straight lines where the presence of outliers can also be verified (Figure 3). In the case of seed length, only an atypical value occurred, showing a high similarity among the grouped observations, except for the presence of this unusual observation. Also according to the results obtained in Figure 3, most seed length data clumped around the central value, in the case the mean, which suggests a high homogeneity of these observations.

According to the results obtained in Figure 5, a strong concentration of Urucum seed width data in millimeters around the central (median) value was observed, showing a high homogeneity (Figure 5). Using the mean and standard deviation (Figure 6), the presence of atypical data was not observed, showing a significant similarity in the pooled observations and a strong concentration of the seed width data around the central (mean) value, which also evidenced the high homogeneity.

Regarding seed weight using the median and quartiles (Figure 8) and mean and standard deviation (Figure 9), a large concentration of data was found around the central value, which evidenced a strong homogeneity of these observations. We also did not verify the presence of atypical data, showing a relevant similarity between the pooled observations.

In general, the results of the descriptive measures of location, variability, asymmetry and kurtosis can serve as a basis for future studies of descriptive analysis and statistical inference, for the comparison of different environments, genetic improvement studies, grouping of experiments in joint analysis, in stability analysis of cultivars, as well as in the construction of so-called components of variance. (Ferreira, 2005; Figueiredo et al., 2007; Oliveira et al., 2009; Spiegel & Stephens, 2009; Bussab & Morettin, 2010; Casella & Berger, 2010; Zar, 2010; Claudio & Stein, 2011; Cecon et al., 2012; Costa, 2012).

Standard deviation and variance are special cases of what statisticians and physicists call the central momentum (CM). Central moment comprises the mean deviation of all observations in a data set from the mean of the observations, raised to the power r . The first central moment ($r = 1$) measures the sum of the differences between each observation minus the sample mean (arithmetic), which is always equal to zero. The second central moment ($r = 2$) is the variance. The third central moment ($r = 3$), divided by the standard deviation to the cube (s^3), is the asymmetry. The asymmetry describes how the sample differs from the form of a symmetric distribution.

A normal distribution has an asymmetry coefficient equal to zero. A distribution in which the value of the asymmetry coefficient is greater than zero has asymmetry to the right, that is, there is a long tail of larger observations at the right of the mean. However, if the asymmetry coefficient is less than zero it has asymmetry on the left, there is a long tail of smaller observations to the left of the mean. The kurtosis or flattening has its basis in the fourth central moment ($r = 4$), measuring the extent or peak at which the probability density is distributed in the tails versus the center of the distribution.

The distribution is classified as heavy tail or light tail when compared to a standard normal distribution (mesokurtic). Aggregate or platykurtic (flattened) distributions have a kurtosis coefficient less than zero, compared to the normal distribution, meaning more mass of probability in the center of the distribution and less probability at the tails. In contrast, leptokurtic (tapered) distributions have a kurtosis coefficient greater than zero. Leptokurtic distributions have less mass of probability in the center and tails of relatively heavy probabilities (Gotelli & Ellison, 2011).

According to Gotelli and Ellison (2011), the law of large numbers proves that for an infinitely large number of observations, the formula $(\sum_{i=1}^n Y_i)/n$ is an approximation of the population mean μ , where, $Y_n = [Y_i]$ is the sample of size n of a random variable Y with expected value $E(Y)$. Similarly, the variance of y_n is σ^2/n . Since the standard deviation is just the square root of the variance, y_n is given by $\sqrt{\sigma^2/n} = \sigma/\sqrt{n}$, what is the same as the standard error of the mean. Therefore, we have an estimate of the standard deviation of the population mean, which is the standard error.

If the conclusions based on a single sample are representative of the entire population, it is recommended using the standard error of the mean. However, if the samples limit the conclusions, it is better to use the sample standard deviation.

Large observational surveys covering large spatial scales with a substantial number of samples are likely representative of the population of interest as a whole, and the standard error of the mean should be used. Small, controlled experiments with few replicates are likely based on a single cluster, and possibly unrepresentative of individuals, as a consequence, the standard deviation should be used to characterize or measure the degree of absolute dispersion of the samples.

4. Conclusions

The length and width of seeds had a small amplitude of variation and a greater amplitude for the weight in milligrams, as well as an optimum, good and regular value respectively for the coefficients of variation, which shows a high, medium and regular degree, respectively, of homogeneity of these characteristics evaluated, the length being much less dispersed than the width and the weight of the seeds.

There was a regular degree of symmetry and a mesocuric distribution for the length, width and weight of the seeds.

The characteristics of seed length, width and weight of the seeds did not present significant correlation and presented mean values with highly significant differences.

It was verified that only the data regarding the width and the weight of the annatto seeds evaluated in this work were adjusted to the normal distribution of probabilities.

The results of the descriptive measures of location, variability, asymmetry, and kurtosis can aid in future studies of descriptive analysis and statistical inference, for the comparison of different environments, studies of plant genetic improvement, and support criteria used for grouping of experiments in analysis, study stability of cultivars, in multivariate analysis, as well as support the construction of so-called components of variance.

References

- Andrade D. F., & Ogliari, P. J. (2010). *Statistics for the agricultural and biological sciences with notions of experimentation* (3rd ed., p. 470). Florianópolis: Editora da UFSC.
- Baskin C. C., & Baskin, J. M. (1998). *Seeds: Ecology, biogeography and evolution of dormancy and Germination*. London: Academic Press. <https://doi.org/10.1016/B978-012080260-9/50007-5>
- Bussab, W. O., & Morettin, P. A. (2010). *Basic statistics* (6th ed., p. 540). São Paulo: Saraiva.
- Casella, G., & Berger R. L. (2010). *Statistical inference* (p. 612). São Paulo: Editora Cengage Learning.
- Carmo Filho, F., & Oliveira, O. F. (1995). *Mossoró: A municipality of the northeastern semi-arid region, climatic characterization and floristic aspect* (p. 62). Mossoró: ESAM.
- Carpanezzi, A. A., & Marques, L. C. T. (1981). *Germination of jutaí-açu (Hymenaea courbaril L.) and jutaí-mirim (H. parvifolia Huber) seeds scarified with commercial sulfuric acid* (p. 15, Circular Técnica 19). EMBRAPA-CPATU, Belém.
- Carvalho, J. E. U., Nascimento, W. M. O., & Muller, C. H. (1998). Physical characteristics and seed germination of fruit species native to the Amazon. *Research Bulletin*, 203.
- Cecon, P. R., Silva, A. R. S., Nascimento, M., & Ferreira, A. (2012). *Statistical methods* (p. 229). Viçosa: Editora UFV.
- Claudio, L. C., & Stein, C. E. (2011). *Descriptive statistics and probability theory* (2nd ed., p. 213). Blumenau: Edifurb.
- Costa, G. G. O. (2012). *Course of inferential statistics and probabilities: theory and practice* (p. 370). São Paulo: Editora Atlas.
- Fenner, M. (1993). *Seed ecology* (p. 151). London: Chapman & Hall.
- Ferreira, D. F. (2005). *Basic statistics* (p. 625). Lavras: UFLA.
- Figueiredo, F., Figueiredo, A., Ramos, A., & Teles, P. (2007). *Descriptive statistics and probabilities - solved and proposed problems with R. applications Escolar Editora* (p. 420). Lisboa: Portugal.
- Fonseca, J. S., & Martins, G. A. (2012). *Statistical course* (6th ed., p. 320). Reimp. São Paulo: Atlas.
- Gotelli, N. J., & Ellison, A. M. (2010). *Principles of statistics in ecology* (p. 532). São Paulo: Artmed Editora.
- Gusmão, E., Vieira, F. A., & Júnior, E. M. F. (2006). Fruit biometry and murici endocarps (*Byrsonima verbascifolia* Rich, Ex. A. Juss). *Revista Cerne*, 12(1), 84-91.
- Lorenzi, H. (1998). Brazilian trees: Manual of identification and cultivation of native tree plants of Brazil. *New Odessa: Plantarum*, 2(1), 352.
- Mathus, M. T., & Lopes, J. C. (2007). Morphology of fruits and seedlings and seed germination of *Erythrina variegata* L. *Revista Brasileira de Sementes*, 29(3), 8.

- Mercadante, A. Z., & Pfander, H. (1998). Carotenoids from annatto: A review. *Recent Research Developments in Agriculture and Food Chemistry*, 2(1), 79-91.
- Oliveira, M. S. (2009). *Introduction to statistics* (p. 329). Lavras, MG: Publisher of the Federal University of Lavras (UFLA).
- Rao, K. V. K., Schwartz, S. A., Nair, H. K., Aalinkeel, R., Mahajan, S., Chawda, R., & Nair, M. P. N. (2004). Plant derived products as a source of cellular growth inhibitory phytochemicals on PC-3M, DU-145 and LNCaP prostate cancer cell lines. *Current Science*, 87(1), 1585-1588.
- R Version 3.1.1. (2014). Viena, Áustria: Foundation for Statistical Computing.
- Silva, J. H. V., Silva, E. L., Filho, J. J., Ribeiro, M. L. G., & Costa, F. G. P. (2006). Residue of annatto seed (*Bixa orellana* L.) as dye of yolk, skin, beak and ovary of laying hens evaluated by two analytical methods. *Ciência e Agrotecnologia*, 30(5), 988-994. <https://doi.org/10.1590/S1413-70542006000500024>
- Spiegel, M. R., & Stephens, L. J. (2009). *Statistics* (4th ed., p. 597). Porto Alegre: Bookman.
- Zar, J. H. (2010). *Biostatistical analysis* (5th ed., p. 944). New York: Prentice Hall.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).