REVISTA BRASILEIRA DE FRUTICULTURA
Published by the Brazilian Society of Fruit Crops
ISSN 0100-2945

Harvest And Postharvest

Original Article - Edited by: Alessandro Dal'col Lúcio

Sample size for the physical and physico-chemical characteristics of the cashew

Adroaldo Guimarães Rossetti[1]*, Carlos Farley Herbster Moura[1], Ebenézer de Oliveira Silva[1], Francisco das Chagas Vidal Neto[1], Laiza Brito Ribeiro[2], Ana Carolina Pinto de Almeida[2]

[1]Embrapa Agroindústria Tropical, Fortaleza, Ceará, Brazil.
[2]Federal University of Ceara, Brazil.
*Corresponding author. E-mail addresses: adroaldo.rossetti@embrapa.br.

Abstract – The aim of this study was to estimate sample size for the physical and physico-chemical characteristics of the peduncles and nuts of the cashew (*Anacardium occidentale* L.) as an aid in research using classical analytical methods and near-infrared (NIR) spectroscopy. Sample sizes were estimated by means of uniform stratified sampling, comprising six strata (S1, S2, ..., S6) corresponding to the clones A+C 276/1, BRS 226, CCP 76, HI58-92-2, PRO 553/2 and PRO 555/2. Determining the size of the sample for each stratum depends on the variance of the characteristic under evaluation and the accepted margin of error *B* of the estimates, or on the desired precision in the results. The greater the variance of the stratum, the greater the size of the sample, the smaller the acceptable margin of error of the estimates, and the greater the desired precision of the results. For an error *B* = 0.2 g, the sample size for nuts of the A+C276/1 clone (stratum S1), with variance s^2 = 5.1568, was n_1 = 30 nuts. For firmness, with *B* = 0.50 N, the sample size for the PRO 555/2 clone (S6) was n_6 = 22 peduncles. For vitamin C, which presented the highest variances, the sample size of the BRS 226 clone (S2), with *B* = 10 mg100^{-1}g, was n_2 = 38 peduncles. The variables pH, soluble solids, titratable acidity and soluble sugars presented the smallest variances, giving the smallest sample sizes for each of the clones.

Keywords: sample sizing, uniform stratified random sampling, sampling accuracy, near-infrared spectroscopy.

Revista Brasileira de Fruticultura, v.44, n.6, e-614. DOI: https://dx.doi.org/10.1590/0100-29452022614
Received 08 Jun, 2022 • Accepted 01 Nov, 2022 • Published Nov/Dec, 2022

All the contents of this journal, except where otherwise noted, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0).

Tamanho de amostras de características físicas e físico-químicas de caju

Resumo - O objetivo deste trabalho foi estimar tamanhos de amostras de características físicas e físico-químicas de pedúnculos e castanhas-de-caju (*Anacardium occidentale* L.) para apoio às pesquisas, por métodos analíticos clássicos e por espectroscopia no infravermelho próximo (NIR). Os tamanhos de amostras foram estimados por amostragem estratificada uniforme, constituída de seis estratos (S1, S2, ..., S6) correspondentes aos clones: A+C 276/1, BRS 226, CCP 76, HI58-92-2, PRO 553/2 e PRO 555/2. O dimensionamento da amostra de cada estrato depende da variância da característica avaliada e da margem de erro *B* admitida nas estimativas ou na precisão desejada nos resultados. O tamanho da amostra será maior quanto maior a variância do estrato, menor a margem de erro admitida nas estimativas e maior a precisão desejada nos resultados. Para um erro *B* = 0,2 g, o tamanho da amostra de castanhas do clone A+C276/1 (estrato S1), com variância s^2 = 5,1568, foi n_1 = 30 castanhas. Para firmeza, com *B* = 0,50 N, o tamanho da amostra do clone PRO 555/2 (S6) foi n_6 = 22 pedúnculos. Para vitamina C, que apresentou as maiores variâncias, o tamanho da amostra do clone BRS 226 (S2), com *B* = 10 mg100^{-1}g, foi n_2 = 38 pedúnculos. As variáveis pH, sólidos solúveis, acidez titulável e açúcares solúveis apresentaram as menores variâncias, possibilitando os menores tamanhos de amostras para todos os clones.

Termos de indexação: dimensionamento de amostra, amostragem aleatória estratificada uniforme, precisão de amostragem, espectroscopia no infravermelho próximo.

Introduction

In addition to the size, precocity and production of cashew trees, the characteristics of the nut and the peduncle are also considered fundamental as parameters that define desirable genotypes. Quality characteristics, such as the weight and firmness of the nut, pH, soluble solids (SS °Brix), titratable acidity (TA%), the soluble solids/titratable acidity ratio (SS/TA), vitamin C (mg100^{-1}g), and soluble sugars (%) from the stalk, among others, are critical factors in evaluating the quality of the cashew (nut and peduncle).

As the cashew tree produces large amounts of cashews during each crop, evaluating these characteristics in each of the cashews from all the plants in the experiment, and even from each plant in one experimental unit (plot), would not be practical. Even if it were feasible, it would involve a long and costly period of collection, weighing and laboratory evaluation, all of which are nearly always subject to error.

Several studies that involve evaluating these characteristics in the nut and peduncle of the cashew for the most varied of purposes, have used samples of various sizes without explaining how the respective values were arrived at, or the level of precision of the results. Others mention, inaccurately or vaguely, or don’t mention at all, the sample size used, which can weaken any conclusions reached.

In view of this, it is essential that appropriate sampling techniques and adequate sample sizes be used as reliable estimators of these characteristics, which, according to Rossetti and Andrade (2002), contribute to the precision of estimates of this type of parameter. However, a sample study is always subject to some degree of uncertainty since only part of the population is assessed. This uncertainty can be reduced by having more sample units (or larger samples) and using better measuring instruments, which usually takes both time and money.

This requires that sample sizes be estimated and the desired level of precision in the results be defined to indicate the sampling accuracy or the acceptable size of any error in the sampling process, and the probability of this error occurring in the sampling plan or in the accuracy of the sampling process (Cochran, 1977). Unfortunately, many studies found in the literature that involve sampling do not take such important principles into account.

As in other crops, near-infrared (NIR) spectroscopy is already used for the non-destructive characterisation of the cashew peduncle to evaluate certain characteristics of interest to agroindustry and to fresh consumption, and to develop multivariate calibration models. Again, the size of the samples is often variable, imprecise, vague and/or not even mentioned.

Lopes et al. (2012) used 15 peduncles, divided into three replications of five peduncles each, in a physical and physico-chemical characterisation employing classical analytical methods. Souza (2012) used cashew peduncles "from a plantation located in the city of Alto Paraíso, Rondônia" to study the physico-chemical characteristics of cashew pulp, both in natura and industrialised, but did not mention the number of peduncles that were used.

Souza et al. (2019) used three replications of fifteen cashews each to investigate the effects of the individual and joint application of gibberellic acid (GA3, 180 mg L^{-1}) and aminoethoxyvinylglycine (AVG, 180 mg L^{-1}) on pre-harvest shedding in the cashew, and its effects on the physical and physico-chemical quality of the peduncle and nut. Almeida et al. (2018) evaluated the influence of the region, year of production and genotype on the physical characteristics of peduncles from three dwarf cashew clones. From the total of 30 plants per clone, "around 30 ripe cashews were harvested" and evaluated in the laboratory.

Evangelista (2017) evaluated 25 dwarf cashew clones for post-harvest quality in the peduncle to indicate clones to produce cashews intended for fresh consumption and/or processing based on the qualitative attributes of the peduncle. "Between 15 and 40 cashews from each clone" were used and the following characteristics evaluated: weight (total, peduncle and nut), size (apical and basal diameters and length), firmness, pH, titratable acidity (TA%), soluble solids (SS), SS/TA ratio, soluble sugars (%), total extractable polyphenols and vitamin C.

Silva (2019) analysed the post-harvest quality of peduncles from 25 dwarf cashew clones, with a view to fresh consumption and/or processing. "Approximately 30 fruits were harvested per clone" and, among other variables, their firmness (N), pH, titratable acidity (TA%), soluble solids (SS °Brix) and SS/TA ratio were evaluated.

Alves Filho et al. (2019) used 21 cashew plants chosen from the set of accessions of the Cashew Germplasm Bank (BAG-Caju) to evaluate the physico-chemical and nutritional characteristics of cashew peduncles, with the aim of developing an advanced tool for rapid phenotyping based on near infrared spectroscopy (NIR). With the same goal, Silva et al. (2021) used four replications of samples of unspecified size to "develop a method of monitoring the vitamin C content of acerola, based on the near infrared (NIR)".

Ribeiro et al. (2016) collected 129 samples of unspecified size of the peduncles from eight dwarf cashew clones to investigate the potential of using the NIR spectrometer as a useful, robust, low-cost and non-destructive alternative for predicting quality parameters in the cashew. Samamad et al. (2018) used 23 samples of peduncles from 17 dwarf cashew genotypes to estimate, using the near-infrared (NIR) spectrometer, quality characteristics of the cashew, such as firmness, pH, soluble sugars (%), soluble solids (SS), titratable acidity (TA), SS/TA ratio (flavour) and vitamin C (VC), without giving the size of the samples.

Galdino et al. (2019) analysed the peduncles of 25 dwarf cashew clones to infer the physical and physico-chemical characteristics and functional compounds, and suggest the best clones for fresh consumption and/or processing based on a sample of 12 cashew plants harvested from plots of four plants. Neves et al. (2020) collected pseudo-fruit from a cashew plant and analysed the physico-chemical characteristics of the cashew, such as pH, acidity, ethanol concentration, SS and sensory qualities for the production of fermented cashew using the pulp of the fresh pseudo-fruit. They do not state the number of peduncles used.

Freitas et al. (2020) evaluated the physico-chemical and sensory qualities of the peduncles of five cashew clones to aid in the selection of new materials suitable for commercialisation as table fruit or for processing. They evaluated titratable acidity, pH, soluble solids, SS/acidity ratio, ascorbic acid and total polyphenols in peduncles harvested from three lots, without stating how many peduncles were used. Magwaza et al. (2012) made an excellent review of the progress of non-destructive evaluations of citrus fruits using NIR spectroscopy, but again did not mention sample size or the degree of sampling accuracy.

Silveira et al. (2018) developed protocols to evaluate the physical and physico-chemical characteristics of the bioactive compounds and antioxidant activity in the cashew peduncle. For the physical evaluations, they recommend that "the analyses should be determined in samples comprising at least 21 fresh peduncles per plant, which should be evaluated individually". "To carry out the physico-chemical analysis, each sample should comprise at least 21 peduncles, divided into a minimum of seven replications, each of three peduncles". They do not give the sampling methodology used to estimate the sample size, nor the level of sampling precision or the margin of error allowed in the results, which may weaken application of their protocols in other studies.

Although the use of sampling in research with perennial crops is common, few studies have explained the sampling methodology used to estimate sample size and the level of sampling precision, and/or the allowed/acceptable margin of error in the sampling process. Using these methodological principles, Rossetti et al. (2014) estimated sample size to assess the average weight of cashew nuts. Similarly, Rossetti et al. (2019, 2020, 2021) used this methodology to estimate the size of cashew nut samples based on the weight of the nut and the type of clone.

The aim of this study was to estimate samples size in the peduncle and nut of the cashew (*Anacardium occidentale* L.) from dwarf cashew clones in evaluating physical and physico-chemical characteristics using classical analytical methods and near-infrared (NIR) spectroscopy.

Material and Methods

Six dwarf cashew clones (A+C276/1, BRS 226, CCP 76, HI58-92-2, PRO 553/2 and PRO 555/2) were used in the research, selected from among those showing the best performance. The experiment was conducted in the Experimental Area of Embrapa Agroindústria Tropical in Pacajus, Ceará, (4°11'26.62'' S; 38°29'50.78'' W; altitude 60 m), and included 25 genotypes, three blocks (replications) and four plants per plot. During the 2018, 2019 and 2020 seasons, 45 cashews were collected from each clone in each crop; these were taken from three plants selected from each plot in each of the blocks, all at the same stage of maturation (ready for harvest).

In the post-harvest laboratory, the following characteristics were evaluated in the peduncle using an ultra-compact near-infrared spectrometer (MicroNIR) connected to a laptop: weight of the nut (g), firmness (N), potential hydrogen (pH), soluble solid content (SS °Brix), titratable acidity (TA%), soluble solids/titratable acidity ratio (SS/TA), vitamin C ($mg100g^{-1}$) and soluble sugars (%). At the same time, the weight of the nuts (g) was determined individually using a semi-analytical balance (Mark, Model 3100, with a maximum capacity of 3,100 g and a precision of 0.01 g).

The firmness of the peduncle (N) was evaluated using a Bishop model FT 011 manual penetrometer equipped with an 8 mm diameter flat tip, by means of two readings taken on opposite sides of the basal area. The pH, soluble solids (SS °Brix) and titratable acidity (TA%) were determined as recommended by AOAC (2005). Vitamin C was obtained by direct titration with Tillman's solution (2,6 dichloro-phenol-indophenol – DFI 0.02%), and expressed in mg100g^{-1}. The soluble sugar content (%) was obtained using the anthrone method, as per Yemm and Willis (1954).

The research followed the uniform stratified random sampling methodology, which consists in subdividing the population into homogeneous subgroups (strata), in such a way that there is homogeneity within the strata and heterogeneity between them (Ryan, 2013; Scheaffer et al., 2011; Cochran, 1977). The sampling plan consisted of six strata (S1, S2, ..., S6), with each stratum represented by one clone: A+C 276/1, BRS 226, CCP 76, HI58-92-2, PRO 553/2 and PRO 555/2. The constituted strata were as postulated by Scheaffer et al. (2011) and Pfeffermann and Rao (2009), i.e. large strata, each the same size.

After the laboratory evaluation, the hypothesis of normality for each variable/characteristic in each stratum was proved using the Shapiro and Wilk (1965) method; the statistical analysis was carried out using the Statistical Analysis System software (SAS, 2009). At a probability level of $\alpha = 0.05$, the quantile of the standard normal is approximately 2.0. As such, the variance from the mean in the precision of the sampling plan is associated with the maximum margin of error *B* allowed in the estimates or with the desired levels of precision in the results, i.e. $\sigma^2 = B^2/4$. Under these conditions, and considering that the costs per observation were the same for each stratum, the allocation or sample size n_i, of the i-th stratum (Neyman allocation), as per Scheaffer et al. (2011) and Pfeffermann and Rao (2009), is obtained by:

$$n_i = n\left(\frac{N_i \sigma_i}{\sum_{k=1}^{L} N_k \sigma_k}\right)$$

Equation 1.

where:
n_i: is the sample size for the i-th stratum: (i = 1, 2, ..., 6);
n: is the total sample size;
N_i: is the size of the i-th stratum: (i = 1, 2, ..., 6);
σ_i: is the standard deviation of the i-th stratum (i = 1, 2, ..., 6);
L: is the number of strata, in this case 6 (six clones);
N_k: is the sample size for the k-th stratum (k = 1, 2, ..., 6);
σ_k: is the standard deviation of the sample of the k-th stratum (k = 1, 2, ..., 6).

Under these conditions, i = k, so:

$$n_i = n\left(\frac{\sigma_i}{\sum_{i=1}^{L} \sigma_i}\right)$$

Equation 2.

where, according to Scheaffer et al. (2011), the total sample size *n* is obtained by:

$$n = \frac{\left(\sum_{k=1}^{L} N_k \sigma_k\right)^2}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2}.$$

Equation 3.

As i = k:

$$n \approx \frac{\left(\sum_{i=1}^{L} N_i \sigma_i\right)^2}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2}.$$

Equation 4.

where:
N: is the number of sampling units in the population: $N = N_1 + N_2 + \ldots + N_L$;
σ_i^2: is the variance of the i-th stratum (i = 1, 2, ..., 6);
D: is the estimator of the fixed variance from the mean in the precision of the sampling plan, associated with the maximum margin of error allowed in the estimates: $D = B^2/4$.

Is should be noted that the sample size n_i in each stratum (clone) and the total size of sample *n* presuppose a knowledge of the position and dispersion values of the variables, the object of the study; these were estimated, and the results shown in Table 1.

Table 1. Observed values (minimum, maximum, mean) and measurements of variability of the physical and physico-chemical variables of the six dwarf cashew clones under study.

Clone (stratum)	Variable	Minimum	Maximum	Mean ($\bar{x}$)	Variance (s^2)	S. deviation (s)	CV (%)[1]
A+C276/1 (S1)	Firmness (N)	3.61	16.34	9.56	13.8857	3.7264	38.99
	Nut mass (g)	4.88	19.35	10.28	5.1568	2.2709	22.09
	pH	3.97	4.99	4.39	0.0714	0.2672	6.08
	SS (°Brix)	9.90	14.15	12.38	1.0803	1.0394	8.39
	TA (%)	0.15	0.82	0.40	0.0326	0.1804	45.34
	SS/TA	13.59	81.87	38.38	331.7207	18.2132	47.45
	Vitamin C ($mg100^{-1}g$)	142.57	953.74	285.63	28673.5938	169.3328	59.28
	Soluble sugars (%)	5.18	9.31	7.26	1.0525	1.0259	14.14
BRS 226 (S2)	Firmness (N)	4.45	14.46	9.06	10.5093	3.2418	35.79
	Nut mass (g)	6.91	13.26	9.93	1.4612	1.2088	12.17
	pH	3.80	4.47	4.10	0.0186	0.1364	3.33
	SS (°Brix)	7.85	14.35	12.50	0.7796	0.8829	7.06
	TA (%)	0.17	0.55	0.33	0.0126	0.1122	34.29
	SS/TA	21.47	71.88	43.11	241.7178	15.5473	36.06
	Vitamin C ($mg100^{-1}g$)	241.63	793.30	430.66	31571.2669	177.6831	41.26
	Soluble sugars (%)	4.86	13.88	8.61	5.1179	2.2623	26.28
CCP 76 (S3)	Firmness (N)	3.06	13.77	8.85	12.9241	3.5950	40.61
	Nut mass (g)	4.91	9.68	8.24	1.1092	1.0532	12.79
	pH	4.05	4.65	4.33	0.0213	0.1460	3.37
	SS (°Brix)	11.10	14.90	12.56	0.7894	0.8885	7.07
	TA (%)	0.16	0.59	0.23	0.0103	0.1016	43.45
	SS/TA	20.19	80.76	59.78	232.5568	15.2498	25.51
	Vitamin C ($mg100^{-1}g$)	210.97	781.64	379.77	32011.5497	178.9177	47.11
	Soluble sugars (%)	5.44	12.89	8.97	2.7017	1.6437	18.33
HI58-92-2 (S4)	Firmness (N)	4.45	14.74	9.21	12.3061	3.5080	38.10
	Nut mass (g)	8.76	14.28	12.11	2.0969	1.4481	11.96
	pH	3.39	4.66	4.34	0.0450	0.2122	4.89
	SS (°Brix)	12.10	14.95	13.19	0.4483	0.6696	5.08
	TA (%)	0.14	0.56	0.22	0.0080	0.0894	39.91
	SS/TA	22.12	91.16	64.87	299.6144	17.3094	26.68
	Vitamin C ($mg100^{-1}g$)	218.71	636.29	327.15	14287.8761	119.5319	36.54
	Soluble sugars (%)	5.95	17.63	9.23	3.8698	1.9672	21.32
PRO 553/2 (S5)	Firmness (N)	4.45	13.63	8.74	8.5886	2.9306	33.52
	Nut mass (g)	5.71	12.16	10.31	2.0513	1.4322	13.89
	pH	3.89	5.01	4.48	0.0777	0.2787	6.23
	SS (°Brix)	9.40	13.80	11.68	0.9570	0.9783	8.37
	TA (%)	0.11	0.64	0.29	0.0160	0.1265	43.00
	SS/TA	14.84	111.54	47.47	428.5640	20.7018	43.61
	Vitamin C ($mg100^{-1}g$)	84.26	652.93	290.04	36292.2192	190.5052	65.68
	Soluble sugars (%)	3.16	9.39	6.45	1.8978	1.3776	21.37
PRO 555/2 (S6)	Firmness (N)	4.31	17.80	9.96	16.2357	4.0294	40.44
	Nut mass (g)	9.21	16.47	12.48	2.2061	1.4853	11.90
	pH	4.20	5.24	4.56	0.0543	0.2331	5.11
	SS (°Brix)	9.40	14.00	12.05	0.6424	0.8015	6.65
	TA (%)	0.12	0.95	0.33	0.0425	0.2062	63.26
	SS/TA	13.05	103.20	49.04	560.0752	23.6659	48.25
	Vitamin C ($mg100^{-1}g$)	151.93	663.00	299.87	24529.4182	156.6187	52.23
	Soluble sugars (%)	3.68	9.81	7.31	2.3420	1.5304	20.94

[1]CV (%): Coefficient of variation expressed as a percentage.

Results and Discussion

Once the variances (s^2) and the respective standard deviations (s) of the variables of each stratum (Table 1) had been estimated and substituted in equations 2 and 4, the sample sizes n_i (i = 1, 2, ..., 6) and the total sample size n of each stratum were estimated for the maximum margin of error *B* allowed in the estimates or the desired precision of the results.

It can be seen (Table 1) that cashew clones are variable in firmness and other variables, both within each clone (stratum) and between clones, in line with EMBRAPA AGROINDÚSTRIA TROPICAL (2020). However, in this case, the Shapiro and Wilk (1965) test showed that each variable had a normally distribution.

As the values set for *B* should be related to the unit of measurement of the phenomenon under study, care must be taken when they are chosen. In this respect, values of *B* were simulated for each variable under evaluation. For firmness, for example, a value for *B* from 0.25 (N) to 2.00 (N) was allowed; for the weight of the nut, as the values are very small, a value from 0.1 g to 1.0 g was accepted, and so on (Tables 2 to 4).

Sample sizes for cashew nuts

Assuming *B* = 0.2 g, for example, as the maximum margin of error accepted in the estimates or the desired level of precision of the results, the sample size for nuts of clone A+C276/1 (stratum S1), which had the greatest variance s^2 = 5.1568 (Table 1), was n_1 = 30 nuts (Table 2). For the same margin of error, the sample size of clone CCP 76 (stratum S3), which had the smallest variance s^2 = 1.1092 (Table 1), was n_3 = 14 nuts. Under the same conditions, for clone PRO 555/2 (stratum S6), of intermediate variance s^2 = 2.2061, the sample size was n_6 = 20 nuts.

In this case, based on stratified sampling methodology, which states that the total sample size (n) for the independent random variables, as in the present case, is the sum of the sample sizes, the total sample size would be n = 64 nuts. Consequently, considering the samples from each of the clones (strata), the total sample size for all the clones would be n = 118 nuts (Table 2).

Note that, for the same margin of error allowed in the estimates or desired precision in the results, the sample size varies as a function of the variance of the weight of the nut from the clone (stratum); that is, smaller sample sizes are obtained from strata with smaller variances. The sample size of cashew nuts for the strata under study varied as a function of the variance of the weight of the nut from each stratum and the margin of error allowed or tolerated in the estimates or in the desired level of precision in the results (Table 2). Therefore, the smaller the margin of error (*B*) allowed in the estimates or the greater the level of precision desired in the

Table 2. Sample sizes of total cashew nuts (n) per clone/stratum (n_1, n_2, ..., n_6), for nut weight (g), as a function of the maximum margin of error *B* (g) allowed in the estimates or in the precision of the results.

Error B (g)	n	Clone/Strata (S_i: i=1, 2, ..., 6)					
		A+C276/1S1 (n_1)	BRS 226 S2 (n_2)	CCP 76 S3 (n_3)	HI58-92-2 S4 (n_4)	PRO553/2 S5 (n_5)	PRO555/2 S6 (n_6)
0.1	197	50	27	23	32	32	33
0.2	118	30	16	14	19	19	20
0.3	71	18	10	8	12	11	12
0.4	45	12	6	5	7	7	8
0.5	31	8	4	4	5	5	5
0.6	22	6	3	2	4	3	4
0.7	17	4	2	2	3	3	3
0.8	13	3	2	2	2	2	2
0.9	10	2	1	1	2	2	2
1.0	8	2	1	1	1	1	2

results, the larger the sample size regardless of the clone (stratum), as such agreeing with the literature.

It should be noted that with a margin of error $B \geq 0.9$ g, even with variances from 1.1092 to 2.2061, for strata S2 to S5 (Table 1) the sample size was only one nut (Table 2), which, from a practical point of view, is neither representative nor to be recommended. For example, in this case, the margin of error $B \geq 0.9$ g is equivalent to more than 10% of the average weight of the nuts (8.24 g) from the S3 stratum (Table 1). Therefore, margins of error with high proportions relative to the mean value of the variable under evaluation are to be avoided.

Sample sizes for the physico-chemical characteristics of cashew peduncles

If when evaluating the firmness of the peduncle the maximum margin of error adopted is B = 0.75 N, for example (Table 3), the sample size n_5 = 9 peduncles is obtained with the PRO 553/2 clone (stratum S5), which showed the smallest variance s^2 = 8.5886 (Table 1). For the same margin of error, the largest sample sizes, n_6 = 13 and n_1 = 12 peduncles, were obtained with the PRO 555/2 (stratum S6) and A+C276/1 (stratum S1) clones, whose variances were, respectively, s^2 = 16.2357 and s^2 = 13.8857. Under similar conditions, given the similarity of the variances s^2 = 10.5093 (stratum S2), s^2 = 12.3061 (stratum S4) and s^2 = 12.9241 (stratum S3) (Table 1), the sample sizes were similar (n_2 = 10 and n_3 = n_4 = 11 peduncles) (Table 3).

Vitamin C was the characteristic that showed the greatest variation among the clones (strata). The PRO553/2 (S5) clone ranged from 84.26 mg100^{-1}g to 652.93 mg100^{-1}g with a mean of 290.04 mg100^{-1}g and CV = 65.68%, while A+C276/1 (S1) ranged from 142.57 mg100^{-1}g to 953.74 mg100^{-1}g with a mean of 285.63 mg100^{-1}g and CV = 59.28%. The clones with the lowest variation indices were HI58-92-2 (S4) and BRS 226 (S2), whose variation was, respectively, 218.71 to 636.29 and 241.63 to 793.30 (mg100^{-1}g), with mean values of 327.15 and 430.66 (mg100^{-1}g) and CVs of 36.54%, and 41.26% (Table 1); even so, these values were still quite high.

In view of this, the variable, vitamin C, registered the highest variances for the clones under study, and as such, also the largest sample sizes. Setting B = 10 mg100^{-1}g as the smallest margin of error, for the PRO553/2 clone (stratum S5), whose variance was s^2 = 36292.2192 (Table 1), the sample size was n_5 = 41 peduncles (Table 4). For the CCP 76 clone (stratum S3), where the observed variance was s^2 = 32011.5497, the sample size was n_3 = 39 peduncles with the same margin of error. In this case, the total sample size would be n = 80 peduncles. Clearly, with greater margins of error B, the samples are smaller, as can be seen in Table 4.

Table 3. Sample sizes of total cashew peduncles (n) per clone/stratum (n_1, n_2, ..., n_6), for firmness (N), as a function of the maximum margin of error B (N) allowed in the estimates and/or in the precision of the results.

Error B (N)	n	Clone/Strata (S_i: i=1, 2, ..., 6)					
		A+C276/1 S1 (n_1)	BRS 226 S2 (n_2)	CCP 76 S3 (n_3)	HI58-92-2 S4 (n_4)	PRO553/2 S5 (n_5)	PRO555/2 S6 (n_6)
0.25	200	36	31	34	33	28	38
0.50	114	20	18	19	19	16	22
0.75	66	12	10	11	11	9	13
1.00	42	7	7	7	7	6	8
1.25	29	5	5	5	5	4	5
1.50	21	4	3	4	3	3	4
1.75	16	3	2	3	3	2	3
2.00	12	2	2	2	2	2	2

Table 4. Sample sizes of total cashew peduncles (n), per clone/stratum (n_1, n_2, ..., n_6), for vitamin C ($mg100^{-1}g$), as a function of the maximum margin of error *B* ($mg100^{-1}g$) allowed in the estimates and/or in the precision of the results.

Error B ($mg100^{-1}g$)	n	Clone/Strata (S_i: i=1, 2, ..., 6)					
		A+C276/1 S1 (n_1)	BRS 226 S2 (n_2)	CCP 76 S3 (n_3)	HI58-92-2 S4 (n_4)	PRO553/2 S5 (n_5)	PRO555/2 S6 (n_6)
10	214	36	38	39	26	41	34
20	135	23	24	25	16	26	21
30	84	15	15	15	10	16	13
40	55	9	10	10	7	10	9
50	38	6	7	7	5	7	6
60	27	5	5	5	3	5	4

The variables pH, soluble solids (SS), titratable acidity (TA) and soluble sugars (%) showed the greatest uniformity, and as a result, the smallest variances for each of the clones under study (Table 1). Therefore, even with the smallest margins of error *B*, it is also possible to obtain smaller sample sizes for each of the clones (strata). For pH, for example, where the variances ranged from 0.0186 for BRS 226 (S2) to 0.0777 for PRO 553/2 (S5) (Table 1), a sample of n = 17 peduncles can be obtained, with a margin of error *B* = 0.1.

For titratable acidity (TA), with variances from 0.0080 for the HI58-92-2 (S4) clone to 0.0425 for PRO 555/2 (S6), the same margin of error *B* would give a sample of size n = 7 peduncles. For soluble solids (SS °Brix), where the variances ranged from 0.4483 for HI58-92-2 (S4) to 1.0803 for A+C276/1 (S1), samples are obtained of size n = 42 and n = 12 peduncles, with a margin of error *B* = 0.25 and *B* = 0.50, respectively. In the case of soluble sugars, where the variances varied from 1.0525 for the A+C276/1 clone (S1) to 5.1179 for the BRS 226 clone (S2), with a margin of error *B* = 1.0%, a sample of size of n = 11 peduncles would be sufficient.

The results obtained in this study agree with those found by Rossetti et al. (2020, 2021) sizing cashew nut samples using this methodology. When comparing sampling methods used in similar situations to those of the present research, Sabino and Villaça (1999) and Rossetti (2001) among others, concluded that the use of this technique afforded better results than those found using other sampling methods. Compared with the examples presented by Ryan (2013) and Scheaffer et al. (2011) regarding a reduction in sampling error and the precision of the estimates of sample sizes with this methodology, they are consistent with the literature.

Furthermore, they agree with Thompson (2012), Pfeffermann and Rao (2009) and Ryan (2013), among others, that three factors influence sample size: (a) confidence level (the greater the confidence level, the larger the sample size); (b) maximum error allowed in the estimates (the smaller the acceptable error, the larger the sample size); and (c) the variability of the phenomenon being investigated (the greater the variability, the larger the sample size).

Based on the variable, vitamin C, which had the highest variance, it can be inferred that for a study involving each of these characteristics together, allowing a margin of error of *B* = $30mg100^{-1}g$ for vitamin C, *B* = 0.75 N for firmness, and *B* = 0.3 g for the weight of the nut, a sample size of 15 cashews (nut and peduncle) each for of the BRS 226, CCP 76 and PRO553/2 clones, and of 18 and 12 each for A+C276/1 and HI58-92-2, respectively, would be reasonable (Tables 2 to 4).

In the same way, with greater precision desired in the results, i.e. with smaller margins of error, e.g. *B* = $10mg100^{-1}g$ for vitamin C, *B* = 0.25 N for firmness, and *B* = 0.1 g for the weight of the nut, for example, a sample size of 38 cashews (nut and peduncle) for the BRS

226 clone, 39 for CCP 76, 41 for PRO553/2, 34 for PRO 555/2, 50 for A+C276/1 and 32 for HI58-92-2, respectively, would be recommended (Tables 2 to 4). In either case, the total sample size (n) cannot be determined by the sum of the sample sizes as they are not independent random variables.

The results shown in Tables 2 to 4, and the remarks concerning the variables pH, soluble solids (SS), titratable acidity (TA) and soluble sugars (%), can serve as a basis for choosing the sample size to be used in research that requires sampling of the physical and physico-chemical characteristics of cashew based on the margin of error that the researcher accepts as reasonable or desires for the precision of the results of the research he/she undertakes, using both classical analytical methods and near infrared (NIR) spectroscopy. The information in this study can only be used for other genotypes if the variances of the variables under evaluation are identical to those obtained here, observing the error margins accepted/allowed by the researcher.

Procedures for cashew collection to determine sample sizes for the physical and physico-chemical characteristics of the cashew

For the purpose of standardising the sample size in the laboratory (selection of cashews that were possibly damaged during transport or otherwise), it is prudent to collect around 20% more cashews than the required sample size.

Ripe cashews should be harvested once they are ready for harvest (same stage as maturation).

Cashews should be harvested in proportion to the required sample size, from all the plants selected in the plots of each of the blocks from all the genotypes and years/crops that make up the research population.

Conclusions

Cashew sampling, for the purpose of evaluating the physical and physico-chemical characteristics of cashew populations should be based on genotype.

The sample size, whether for studies using analytical methods or by NIR spectroscopy, should be estimated based on the variable/characteristic of interest.

The sample size, for the same accepted margin of error, varied according to genotype and the characteristic under evaluation.

The sample size should be chosen based on the accepted margin of error and the desired degree of precision in the result.

The characteristics with the lowest variance for each of the clones were pH, soluble solids, titratable acidity and soluble sugars; the greatest was vitamin C.

Acknowledgment

The authors wish to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for granting scientific initiation scholarships to Laiza Brito Ribeiro and Ana Carolina Pinto de Almeida, both co-authors of this study.

References

ALMEIDA, M.L.B; MOURA, C.F.H.; INNECCO, R.; SILVEIRA, M.R.S. da. Características físicas de pedúnculos de clones de cajueiro-anão (*Anacardium occidentale* L.) produzidos em função da variação ambiental e temporal. **Revista Colombiana de Ciencias Hortícolas**, Tunja-Boyacá, v.12, n.1, p.41-9, 2018.

ALVES FILHO, E.; SILVA, L.M.; LIMA, Y.; RIBEIRO, P.; SILVA, E.; ZOCOLO, G.; CANUTO, K.; MORAIS, S.; CASTRO, A.C.; BRITO, E. Metabolomic variability of different genotypes of cashew by LC-Ms and correlation with Near-Infrared Spectroscopy as a tool for fast phenotyping. **Metabolites**, Nevada, v.9, n.121, p.1-19, 2019.

AOAC - Association of Official Analytical Chemists. **Official methods of analysis of AOAC International**. 21.ed. Rockville, 2019. Disponível em: https://www.aoac.org/official-methods-of-analysis-21st-edition-2019/. Acesso em: 22 dez. 2021.

EMBRAPA AGROINDÚSTRIA TROPICAL. **Clones de cajueiro**. Fortaleza: Embrapa Agroindústria Tropical, 2020. 1 Folder. Disponível em: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1125932. Acesso em: 22 dez. 2021.

COCHRAN, W.G. **Sampling techniques**. 3.ed. New York: John Wiley & Sons, 1977. 428 p.

EVANGELISTA, J. da S.B. **Seleção de clones de Cajueiro-Anão por meio da qualidade de pedúnculos**. 2017. 76 f. Dissertação (Mestrado Agronomia) - Universidade Federal do Ceará, Fortaleza, 2017. Disponível em: http://www.repositorio.ufc.br/handle/riufc/30579. Acesso em: 22 dez. 2021.

FREITAS, A.S. de; DANTAS, B. dos S.; ARAÚJO, Í.M. da S.; GARRUTI, D. dos S. Suitability of peduncles of new cashew tree clones for comercial purposes. **Revista Ciência Agronômica**, Fortaleza, v.51, n.3, p.1-8, 2020.

GALDINO, A.G. da S.; RIBEIRO, L.B.; MOURA, C.F.H.; AMORIM, F.F.V.R. de; INNECCO, R. Physical analysis of peduncles of dwarf cashew clones for consumption, processed or natural. **Revista Colombiana de Ciencias Hortícolas**, Tunja-Boyacá, v.13, n.2, p.200-7, 2019.

LOPES, M.M. de A.; MIRANDA, M.R.A. de; MOURA, C.F.H.; ENÉAS FILHO, J. Bioactive compounds and total antioxidant capacity of cashew apples (*Anacardium occidentale* L.) during the ripening of early dwarf cashew clones. **Ciência Agrotécnica**, Lavras, v.36, n.3, p.325-32, 2012.

MAGWAZA, L.S.; OPARA, U.L.; NIEUWOUDT, H.; CRONJE, P.J.R.; SAEYS, W.; NICOLAÏ, B. NIR spectroscopy applications for internal and external quality analysis of citrus fruit - a review. **Food and Bioprocess Technology**, New York, v.5, n.2, p.525-44, 2012.

NEVES, W.B.P.; HOLANDA JÚNIOR, L.M. de; MELO, R.P.F.; OLIVEIRA, M.N. de M.; CARMO, S.K.S. Produção e caracterização físico-química do fermentado de caju.**Brazilian Journal of Development**, Curitiba, v.6, n.9, p.68106-20, 2020.

PFEFFERMANN, D.; RAO, C.R. (ed.). **Sample surveys**: design, methods and applications. North Holland: Elsevier, 2009. 722 p. (Handbook of Statistics, 29 A).

RIBEIRO, L.P.D.; SILVA, A.P.M. da; LIMA, A.A.de; SILVA, E. de O.; RINNAN, A.; PASQUINI, C. Non-destructive determination of quality traits of cashew apples (*Anacardium occidentale*, L.) using a portable near infrared spectrophotometer. **Journal of Near Infrared Spectroscopy**, Susssex, v.24, n.1, p.77-82, 2016.

ROSSETTI, A.G. Precisão experimental e tamanho da área de experimentos de campo com fruteiras e outras plantas perenes arbóreas em função da unidade experimental e do número de repetições. **Revista Brasileira de Fruticultura**, Jaboticabal, v.23, n.3, p.704-8, 2001.

ROSSETTI, A.G.; ANDRADE, D.F. A amostragem na avaliação de experimentos de campo com frutíferas perenes arbóreas. *I*n: REUNIÃO REGIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA, 34., 2002, Fortaleza. **Programa e Resumos** [...]. Fortaleza: Universidade Federal do Ceará - Departamento de Estatística e Matemática Aplicada, 2002. v.1. p.3.

ROSSETTI, A.G.; VIDAL NETO, F.das C.; BARROS, L.de M. **Amostragem e tamanho de amostra de castanha-de-caju como suporte às pesquisas com o cajueiro**. Fortaleza: Embrapa Agroindústria Tropical, 2021. 31 p.(Boletim de Pesquisa e Desenvolvimento, 223). Disponível em: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1135759. Acesso em: 21 dez. 2021.

ROSSETTI, A.G.; VIDAL NETO, F. das C.; BARROS, L. de M. Sampling of cashew nuts from cashew tree clones. **Revista Brasileira de Fruticultura**, Jaboticabal, v.42, n.1, p.1-9, 2020.

ROSSETTI, A.G.; VIDAL NETO, F. das C.; BARROS, L. de M. Sampling in cashew nuts as an aid to research in the genetic improvement of the cashew tree. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v.54, p.e00057, 2019.

ROSSETTI, A.G.; VIDAL NETO, F. das C.; MOISÉS, S.M.M.; COUTINHO, V.T.R. Tamanho de amostra para estimar o peso médio de castanha-de-caju. *In*: CONGRESSO BRASILEIRO DE FRUTICULTURA, 23., 2014, Cuiabá. **Resumos** [...]. Campinas: Sociedade Brasileira de Fruticultura, 2014. v.1, p.1-4. 1 CD-ROM.

RYAN, T.P. **Sample size determination and power**. New Jersey: Wiley, 2013. 404 p.

SABINO, C.M.; VILLAÇA, R. Estudo comparativo de métodos de amostragem de comunidades de Costão. **Revista Brasileira de Biologia**, São Carlos, v.59, n.3, p.407-419, 1999.

SAMAMAD, N.T.I.; RIBEIRO, L.P.D.; LOPES, M.M. de A.; PUSCHMANN, R.; SILVA, E. de O. Near infrared spectroscopy, a suitable tool for fast phenotyping - The case of cashew genetic improvement. **Scientia Horticulturae**, New York, v.238 p.363-8, 2018.

SAS INSTITUTE. **User's guide**: SAS/STAT. Version 9.2. Cary, 2009.

SCHEAFFER, R.L.; MENDENHALL III, W.; OTT, R.L.; GEROW, K.G. **Elementary survey sampling**. 7.ed. Boston: Cengage Learning, 2011. 436 p.

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, Cambridge, v.52, n.3/4, p.591-611, 1965.

SILVA, F.M.R. **Qualidade pós-colheita do pedúnculo de clones de Cajueiro-Anão cultivados em Pacajus-CE**. 2019. 72 f. Dissertação (Mestrado Tecnologia de Alimentos) - Instituto Federal do Ceará, Campus Limoeiro do Norte, Limoeiro do Norte, 2019. Disponível em: https://ifce.edu.br/limoeirodonorte/arquivos_pgta/dissertacoes/silva_f_m_r_2019_pgta_ifce.pdf. Acesso em: 21 dez. 2021.

SILVA, L.M.H. da; RIBEIRO, L.P.D.; COSTA, B.C.; SILVA, E.O.; MIRANDA, M.R.A. de. Monitoring vitamin C extraction using multivariate calibration models by NIR. **Revista Ciência Agronômica**, Fortaleza, v.52, n.1, p.e20196949, 2021.

SILVEIRA, M.R.S. da; OSTER, A.H.; MOURA, C.F.H.; SILVA, E. de O.; SILVA, L.M.A.E; SOUSA, A.E.D. de. **Protocolos para avaliação das características físicas e físico-químicas, dos compostos bioativos e atividade antioxidante do pedúnculo do caju**. Fortaleza: Embrapa Agroindústria Tropical, 2018.43 p. (Documentos, 182). Disponível em: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/175343/1/DOC18004.pdf. Acesso em: 22 dez. 2021.

SOUZA, K.O; SILVEIRA, A.G.; LOPES, M.M.A.; MOURA, C.F.H.; SILVA, E.O.; AYALA-ZAVALA, J.F.; SOARES, L.S.P.; MIRANDA, M.R.A. AVG and GA_3 prevent preharvest fruit drop and enhance postharvest quality of 'BRS 189' cashew. **Scientia Horticulturae**, New York, v.257, p.2-8. 2019.

SOUZA, L.G. de. **Análise físico-química da polpa do caju (*Anacardium occidentale* L.) nas formas in natura e industrializada do município de Alto Paraíso/RO**. 2012. 36 f. Monografia (Licenciatura em Química) - Faculdade de Educação e Meio Ambiente, Campus Porto Velho, Fundação Universidade Federal de Rondônia, Ariquemes, 2012.

THOMPSON, S.K. **Sampling**. 3.ed. New Jersey: John Wiley & Sons, 2012. 472 p.

YEMM, E.W.; WILLIS, A.J. The estimation of carbohydrates in plant extracts by anthrone. **Biochemical Journal**, London, v.57, n.3, p.508-14, 1954.