

Sample Size to Estimate the Mean and Median of Traits in Canola

Alberto Cargnelutti Filho¹, Cleiton Antonio Wartha², Jéssica Andiana Kleinpaul³,
Ismael Mario Marcio Neu³ & Daniela Lixinski Silveira³

¹ Department of Crop Science, Federal University of Santa Maria, Santa Maria, Brazil

² Student of Agronomy, Federal University of Santa Maria, Santa Maria, Brazil

³ Postgraduate Program in Agronomy, Federal University of Santa Maria, Santa Maria, Brazil

Correspondence: Alberto Cargnelutti Filho, Department of Crop Science, Federal University of Santa Maria, Avenida Roraima, nº 1000, Bairro Camobi, CEP: 97105-900, Santa Maria, RS, Brazil. Tel: 55-55-3220-8899. E-mail: alberto.cargnelutti.filho@gmail.com

Received: July 4, 2018

Accepted: August 24, 2018

Online Published: October 15, 2018

doi:10.5539/jas.v10n11p123

URL: <https://doi.org/10.5539/jas.v10n11p123>

Abstract

The aim of this study was to determine the sample size (*i.e.*, number of plants) required to estimate the mean and median of canola (*Brassica napus* L.) traits of the Hyola 61, Hyola 76, and Hyola 433 hybrids with precision levels. At 124 days after sowing, 225 plants of each hybrid were randomly collected. In each plant, morphological (plant height) and productive traits (number of siliques, fresh matter of siliques, fresh matter of aerial part without siliques, fresh matter of aerial part, dry matter of siliques, dry matter of aerial part without siliques, and dry matter of aerial part) were measured. For each trait, measures of central tendency, variability, skewness, and kurtosis were calculated. Sample size was determined by resampling with replacement of 10,000 resamples. The sample size required for the estimation of measures of central tendency (mean and median) varies between traits and hybrids. Productive traits required larger sample sizes in relation to the morphological traits. Larger sample sizes are required for the hybrids Hyola 433, Hyola 61, and Hyola 76, in this sequence. In order to estimate the mean of canola traits of the Hyola 61, Hyola 76 e Hyola 433 hybrids with the amplitude of the confidence interval of 95% equal to 30% of the estimated mean, 208 plants are required. Whereas 661 plants are necessary to estimate the median with the same precision.

Keywords: *Brassica napus* L., sample design, experimental precision, reference sample

1. Introduction

Canola (*Brassica napus* L.) belongs to the Brassicaceae family (Tomm, Wiethölter, Dalmago, & Santos, 2009). Canola seeds are utilized as edible oil for human consumption, livestock feed and biofuel feedstock, being the second most prominent oil seed crop in the world (Li et al., 2015). In Brazil, canola is used for grain production during winter cool-season in crop rotation systems. Canola can act controlling root diseases in annual crops by mechanisms of biofumigation and absence of host (Angus et al., 2015). Other synergistic effects on total system yield can be observed, such as weed growth suppression and flexibility in chemical weed control options. Therefore, besides grain yield and oil content, sample dimensioning of dry matter traits is relevant to assess its cover crop potential.

In Agricultural Sciences research, the number of treatments, number of replicates, plot size and shape, and sample size are important factors to be considered in the planning of field crop experiments. The inadequate sizing of these factors can cause problems in data statistical analysis (Storck, Garcia, Lopes, & Estefanel, 2016), such as high coefficient of variation, and consequently low experimental precision and reliability of the research results. Cargnelutti Filho et al. (2015a) investigated the optimal plot size and the number of replicates for combinations of number of treatments and levels of experimental precision to evaluate the fresh matter of canola (*Brassica napus* L.). The authors provided substantial information for experimental designs with canola.

Given that the size (area) of the experiment (number of treatments × number of replicates × plot size) is defined for a particular crop; the questioning regarding the number of plants (sample size) that should be evaluated in each plot of each treatment commonly arises. Reduced plant number may not be representative, whereas excessive plant number may be unnecessary and impractical. Thus, sample sizing enough plants for a certain

precision is essential in the estimation of statistical measures such as measures of central tendency (mean and median).

In symmetric data distributions of a given trait, the mean is an appropriate measure of central tendency. However, in asymmetric distributions (positive or negative), the median separates the ordered set of data into two parts (50% below and 50% above the median) and is a measure of position indicated to represent the central tendency of the dataset. The mean has been extensively used in Agricultural Sciences research but the median would be more adequate in situations of asymmetric data distribution.

Sample sizing studies (determination of the number of plants) can be performed based on data collected in plants of a uniformity trial (blank experiment). Considering that in the experimental area of a uniformity trial with a given crop, all procedures (soil preparation, fertilization, sowing, pest and disease control and evaluations) are carried out uniformly and disregarding the genetic variability between the plants, the remaining variability represents the random (environmental) variation occurring between the plants in the experimental area.

Sampling in a uniformity trial should contemplate the maximum variation that may exist between plants in order to adequately represent the plant population and to enable the study of sample size. Evidently, extrapolations of the sample size for other experimental areas are questionable due to the existence of environmental variations. However, the results of these studies can be used as a reference for the planning of other researches and assist the researchers in defining the number of sampled plants.

Resampling with reposition is an appropriate procedure for sample sizing and exhibits independence of the data probability distribution (Ferreira, 2009). This procedure was used to determine the sample size for the mean estimation in agricultural crops, such as millet (Kleinpaul et al., 2017) and rye (Bandeira et al., 2018a) and for the estimation of mean and median of white lupine (Burin, Cargnelutti Filho, Toebe, Alves, & Fick, 2014) and flax (Cargnelutti Filho et al., 2018). Meantime, a methodology based on Student's t distribution (Bussab & Morettin, 2017) has been used to determine the sample size for the mean estimation of forage turnip (Cargnelutti Filho et al., 2014), black oat (Cargnelutti Filho et al., 2015b), and rye (Bandeira et al., 2018b). These researches have demonstrated sample size variability between traits, cultivars, sowing dates, evaluation times, agricultural years, and between measures of central tendency (median and mean). These studies revealed promising aspects of the correct sample sizing.

Other researches involving aspects related to sample sizes have been performed with canola to forecast Sclerotinia stem rot (Turkington, Morrall, & Baker, 1988); determinate seed chlorophyll content (Daun & Symons, 2000); semiquantitative detection of genetically modified seeds (Emslie, Whaites, Griffiths, & Murby, 2007); and for evaluating the temporal-spatial distribution of the cabbage aphid on canola (Nematollahi, Fathipour, Talebi, Karimzadeh, & Zalucki, 2014).

We assumed that the number of plants for the estimation of mean and median traits differs between canola hybrids. Thus, the aim of this study was to determine the sample size (*i.e.*, number of plants) required to estimate the mean and median of canola (*Brassica napus* L.) traits of the Hyola 61, Hyola 76, and Hyola 433 hybrids with precision levels.

2. Material and Methods

Three uniformity trials (experiment without treatment, which the crop and all procedures performed during the experiment are homogeneous in the experimental area) were carried out with canola (*Brassica napus* L.) in experimental area of 45 m × 60 m (2,700 m²) in southern Brazil, located at 29°42'S, 53°49'W, with 95 m altitude. According to Köppen climate classification, the climate of the region is Cfa, subtropical humid, with warm summers and without dry season defined (Heldwein, Buriol, & Streck, 2009). The type of soil of the local trial is classified as sandy loam typic Paleudalf (Santos et al., 2013).

Each canola hybrid Hyola 61, Hyola 76, and Hyola 433 was evaluated in a uniformity trial of 15 m × 60 m (900 m²). In the three uniformity trials, sowing was performed in rows spaced by 0.50 m between rows on June 13, 2013. Plant emergence was observed on June 20, 2013. A basic fertilization of 8 kg ha⁻¹ N, 80 kg ha⁻¹ of P₂O₅, and 80 kg ha⁻¹ of K₂O (only NPK in proportion 02-20-20) was carried out on June 13, 2013. The number of plants was counted and the density obtained was respectively of 320,267, 351,067, and 410,800 plants ha⁻¹ for Hyola 61, Hyola 76, and Hyola 433 hybrids (Cargnelutti Filho et al., 2015a).

In each uniformity trial, an area of 15 m × 15 m (225 m²) was demarcated. Then, 225 plants were randomly selected in each uniformity trial on October 15, 2013 (124 days after sowing). These plants were distanced in 1 meter in a matrix of 15 rows and 15 columns. The plants were in the grain maturation stage at this time. The plants were cut at the soil surface. Thereupon, the plant height (PH) of each plant was measured in cm. In

addition, the number of siliques per plant (NS) were counted. In each plant, the siliques were removed. The fresh matter of siliques (FMS), in g plant^{-1} , the fresh matter of aerial part without siliques (FMWS), in g plant^{-1} , and the fresh matter of aerial part (FM = FMS + FMWS), in g plant^{-1} , were obtained by weighing. After drying in an oven, the weight of dry matter of siliques (DMS), in g plant^{-1} , dry matter of aerial part without siliques (DMWS), in g plant^{-1} , and dry matter of aerial part (DM = DMS + DMWS), in g plant^{-1} were obtained. In this study, the PH was considered a morphological trait. Furthermore, NS, FMS, FMWS, FM, DMS, DMWS, and DM were considered productive traits.

For the eight measured traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM), the calculated statistics were: minimum, percentiles 1, 2.5, and 25, median (percentile 50), percentiles 75, 97.5, and 99, maximum, range, mean, variance, standard deviation, standard error, coefficient of variation, coefficient of skewness, coefficient of kurtosis, and p-value of normality *Kolmogorov-Smirnov* test. For each trait, the means of the hybrids (Hyola 61 versus Hyola 76, Hyola 61 versus Hyola 433, Hyola 76 versus Hyola 433) were compared by *Student's t*-test (two-tailed) for independent samples at 5% significance. Subsequently, 1,999 sample sizes were planned for each trait. The initial sample size was two plants and the other ones were obtained with the addition of one plant up to 2,000 plants. Therefore, the following sample sizes were planned for the simulations: 2, 3, 4, ... 2,000 plants for each trait.

Confidence intervals for the mean and the median can be constructed by resampling, regardless of the probability distribution of the sample data. Thus, for each sample size planned in each trait, there were 10,000 resampling with replacement. For each resample, mean and median were estimated. Thereby, for each sample size of each trait, 10,000 estimates of the mean and 10,000 estimates of the median were obtained (Ferreira, 2009) and percentiles 2.5 and 97.5 were calculated. Then, the amplitude of the confidence interval of 95% for the mean and the median for each sample size of each trait was calculated through the difference between the percentile 97.5 and percentile 2.5. Hereafter, the sample size (*i.e.*, number of plants) was determined to estimate the mean and median of each trait with precision levels. The initial size (*i.e.*, two plants) was used for beginning this determination. The sample size was considered as the number of plants from which the amplitude of the confidence interval of 95% was less than or equal to 15% (greater precision), 16%, 17%, ... 30% (minor precision) respectively of the estimated mean and median. These experimental precisions were considered adequate to estimate the mean and the median.

The mean and percentiles 2.5 and 97.5 of the 10,000 means of dry matter of siliques (DMS) and the 10,000 medians of the fresh matter of aerial part (FM) of each sample size were plotted on graphs, for hybrid Hyola 433. These two traits were plotted as a function of the greater sample size needed. In the graphical representation of the mean and the limits of the confidence interval of 95%, an interval of forty plants was used for better visual representation. We started with forty plants because smaller sizes than this one are not of interest since they have large confidence intervals (low precision). The statistical analysis was performed using Microsoft Office Excel® and R software (R development Core Team, 2018).

3. Results and Discussion

The mean of the PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM traits was greater in the Hyola 76 hybrid, intermediate in Hyola 433, and smaller in Hyola 61, which demonstrates superior agronomic performance of Hyola 76 hybrid (Table 1). In the assessment of six canola hybrids (Hyola 433, Hyola 50, Hyola 61, Hyola 76, Hyola 571 CL, and Hyola 575 CL), Rigon et al. (2017) concluded that Hyola 76 hybrid presented higher number of days of flowering, grain yield and oil, evidencing the superior performance of this hybrid, as verified in our study.

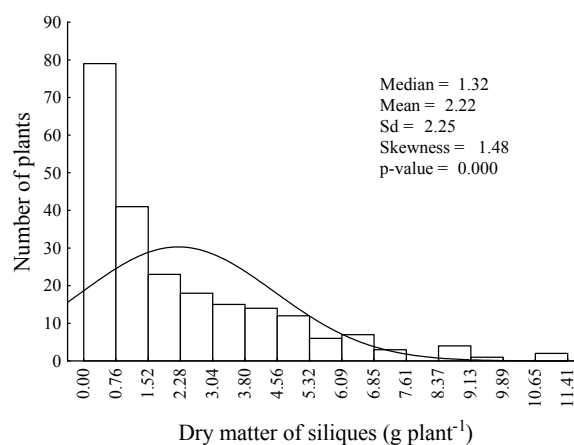
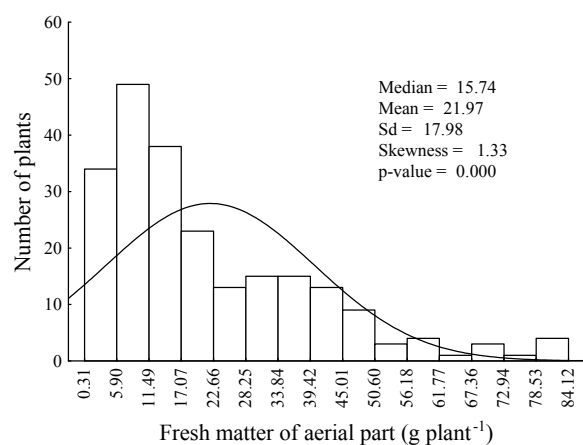
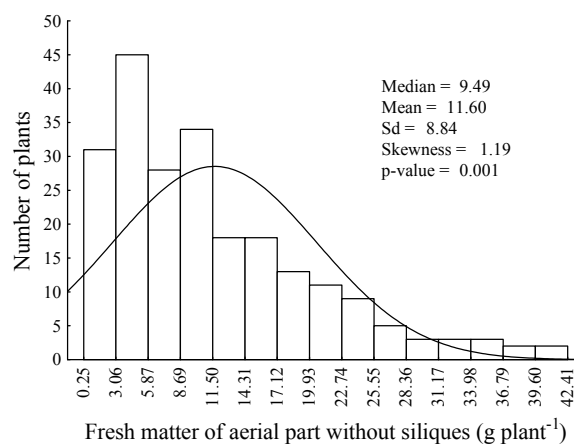
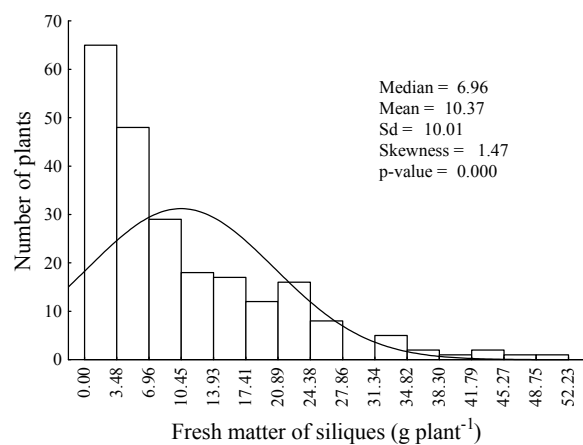
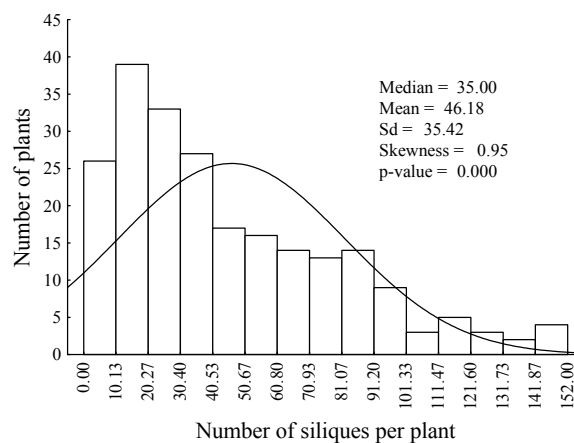
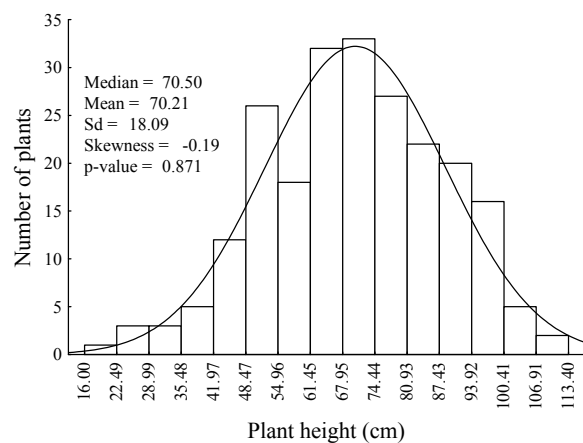
Table 1. Minimum, percentiles 1, 2.5, 25, median (50 percentile), 75, 97.5, and 99, maximum, range, mean, variance, standard deviation, standard error, coefficients of variation (CV), skewness, kurtosis, and p-value of the *Kolmogorov-Smirnov* normality test of traits⁽¹⁾ measured in 225 plants of canola (*Brassica napus* L.) of each hybrid (Hyola 61, Hyola 76, and Hyola 433)

Statistics	PH	NS	FMS	FMWS	FM	DMS	DMWS	DM
	--- cm ---		----- g plant^{-1} -----					
<i>Hyola 61</i>								
Minimum	16.00	0.00	0.00	0.25	0.31	0.00	0.08	0.10
Percentile 1	27.00	0.00	0.00	1.27	1.55	0.00	0.26	0.32
Percentile 2.5	34.06	1.00	0.05	1.49	2.61	0.02	0.39	0.53
Percentile 25	57.20	19.00	2.89	4.85	9.02	0.54	1.07	1.74

Median	70.50	35.00	6.96	9.49	15.74	1.32	2.07	3.24
Percentile 75	84.00	69.00	15.52	16.14	31.71	3.47	3.35	7.05
Percentile 97.5	101.40	133.20	36.32	34.17	70.89	8.51	6.66	15.16
Percentile 99	104.52	149.04	43.30	38.87	82.42	9.54	7.91	17.81
Maximum	113.40	152.00	52.23	42.41	84.12	11.41	8.59	19.14
Range	97.40	152.00	52.23	42.16	83.81	11.41	8.51	19.05
Mean	70.21 c	46.18 c	10.37 c	11.60 c	21.97 c	2.22 c	2.44 c	4.66 c
Variance	327.31	1254.50	100.25	78.18	323.17	5.08	3.07	14.84
Standard deviation	18.09	35.42	10.01	8.84	17.98	2.25	1.75	3.85
Standard error	1.21	2.36	0.67	0.59	1.20	0.15	0.12	0.26
CV (%)	25.77	76.69	96.58	76.23	81.84	101.58	71.82	82.74
Skewness ⁽²⁾	-0.19ns	0.95*	1.47*	1.19*	1.33*	1.48*	1.08*	1.35*
Kurtosis ⁽³⁾	-0.27ns	0.31ns	2.30*	1.08*	1.58*	2.16*	0.94*	1.72*
p-value	0.871	0.000	0.000	0.001	0.000	0.000	0.018	0.000
<hr/>								
<i>Hyola 76</i>								
Minimum	34.00	0.00	0.00	0.49	0.49	0.00	0.08	0.08
Percentile 1	59.08	3.00	0.44	2.56	3.11	0.09	0.60	0.70
Percentile 2.5	65.90	6.00	0.81	3.94	6.77	0.14	0.91	1.32
Percentile 25	110.20	54.00	8.50	17.24	25.65	1.54	3.74	5.60
Median	122.00	97.00	16.94	26.94	44.72	3.58	5.92	9.77
Percentile 75	131.00	151.00	29.66	40.58	69.33	7.79	9.19	16.77
Percentile 97.5	143.50	276.60	53.73	65.55	115.35	15.33	16.08	29.83
Percentile 99	149.26	324.36	64.69	73.82	138.61	17.24	17.96	34.26
Maximum	154.50	338.00	73.65	82.77	150.51	18.91	19.41	38.32
Range	120.50	338.00	73.65	82.28	150.01	18.91	19.34	38.24
Mean	118.48 a	106.90 a	20.69 a	29.83 a	50.52 a	5.04 a	6.76 a	11.80 a
Variance	365.86	4,972.09	243.43	277.06	980.68	18.41	16.10	64.48
Standard deviation	19.13	70.51	15.60	16.65	31.32	4.29	4.01	8.03
Standard error	1.28	4.70	1.04	1.11	2.09	0.29	0.27	0.54
CV (%)	16.14	65.96	75.42	55.80	61.99	85.16	59.37	68.08
Skewness ⁽²⁾	-1.36*	0.86*	0.92*	0.62*	0.75*	1.05*	0.76*	0.89*
Kurtosis ⁽³⁾	2.60*	0.61ns	0.38ns	-0.02ns	0.09ns	0.47ns	0.21ns	0.25ns
p-value	0.008	0.024	0.011	0.144	0.008	0.001	0.060	0.009
<hr/>								
<i>Hyola 433</i>								
Minimum	22.60	0.00	0.00	1.06	1.06	0.00	0.21	0.21
Percentile 1	39.20	0.00	0.00	1.24	1.36	0.00	0.27	0.37
Percentile 2.5	46.80	1.60	0.11	1.37	2.22	0.04	0.39	0.47
Percentile 25	74.60	25.00	3.60	5.32	9.50	1.03	1.37	2.62
Median	89.20	52.00	9.36	10.46	20.36	2.76	2.72	5.43
Percentile 75	103.30	103.00	19.40	23.03	43.81	5.77	5.95	11.78
Percentile 97.5	133.12	272.40	60.80	65.99	128.34	17.49	16.37	31.82
Percentile 99	143.31	312.88	78.64	78.29	151.18	23.64	18.64	41.33
Maximum	150.20	407.00	87.86	82.48	170.27	27.13	20.30	45.98
Range	127.60	407.00	87.86	81.41	169.21	27.13	20.09	45.77
Mean	89.20 b	76.73 b	15.04 b	16.93 b	31.98 b	4.38 a	4.21 b	8.59 b
Variance	459.81	5,428.44	275.20	278.37	1,071.97	23.82	16.56	75.61
Standard deviation	21.44	73.68	16.59	16.68	32.74	4.88	4.07	8.70
Standard error	1.43	4.91	1.11	1.11	2.18	0.33	0.27	0.58
CV (%)	24.04	96.02	110.29	98.52	102.39	111.43	96.60	101.21
Skewness ⁽²⁾	0.01ns	1.66*	2.01*	1.78*	1.86*	2.12*	1.73*	1.87*
Kurtosis ⁽³⁾	0.39ns	2.92*	4.67*	3.29*	3.75*	5.35*	3.04*	3.82*
p-value	0.677	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note. ⁽¹⁾ PH: plant height; NS: number of siliques; FMS: Fresh matter of siliques; FMWS: fresh matter of aerial part without siliques; FM: fresh matter of aerial part; DMS: dry matter of siliques; DMWS: dry matter of aerial part without siliques; and DM: dry matter of aerial part. For each trait, means of the Hyola 61, Hyola 76, and Hyola 433 hybrids not followed by the same letter in the column differ by *Student's* t-test (two-tailed) for independent samples at 5% significance, with 448 degrees of freedom. ⁽²⁾ * Asymmetry differs from zero by *Student's* t-test at 5% probability level. ^{ns} Non-significant. ⁽³⁾ * Kurtosis differs from zero by *Student's* t-test at 5% probability level. ^{ns} Non-significant.

Based on the percentiles, proximity of the mean and median estimates, asymmetry and kurtosis coefficients, p-value of the *Kolmogorov-Smirnov* test and frequency histograms for the three hybrids (Hyola 61, Hyola 76, and Hyola 433), we observed that the morphological trait (plant height), except for the plant height of the Hyola 76 hybrid, exhibited better adherence to the normal distribution in relation to the seven productive traits (NS, FMS, FMWS, FM, DMS, DMWS, and DM) (Table 1; Figures 1, 2 and 3).



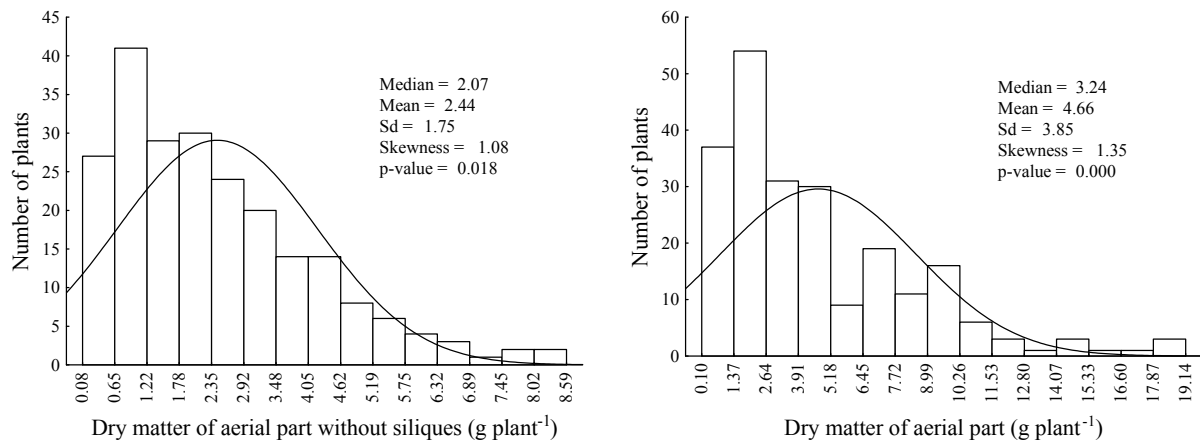
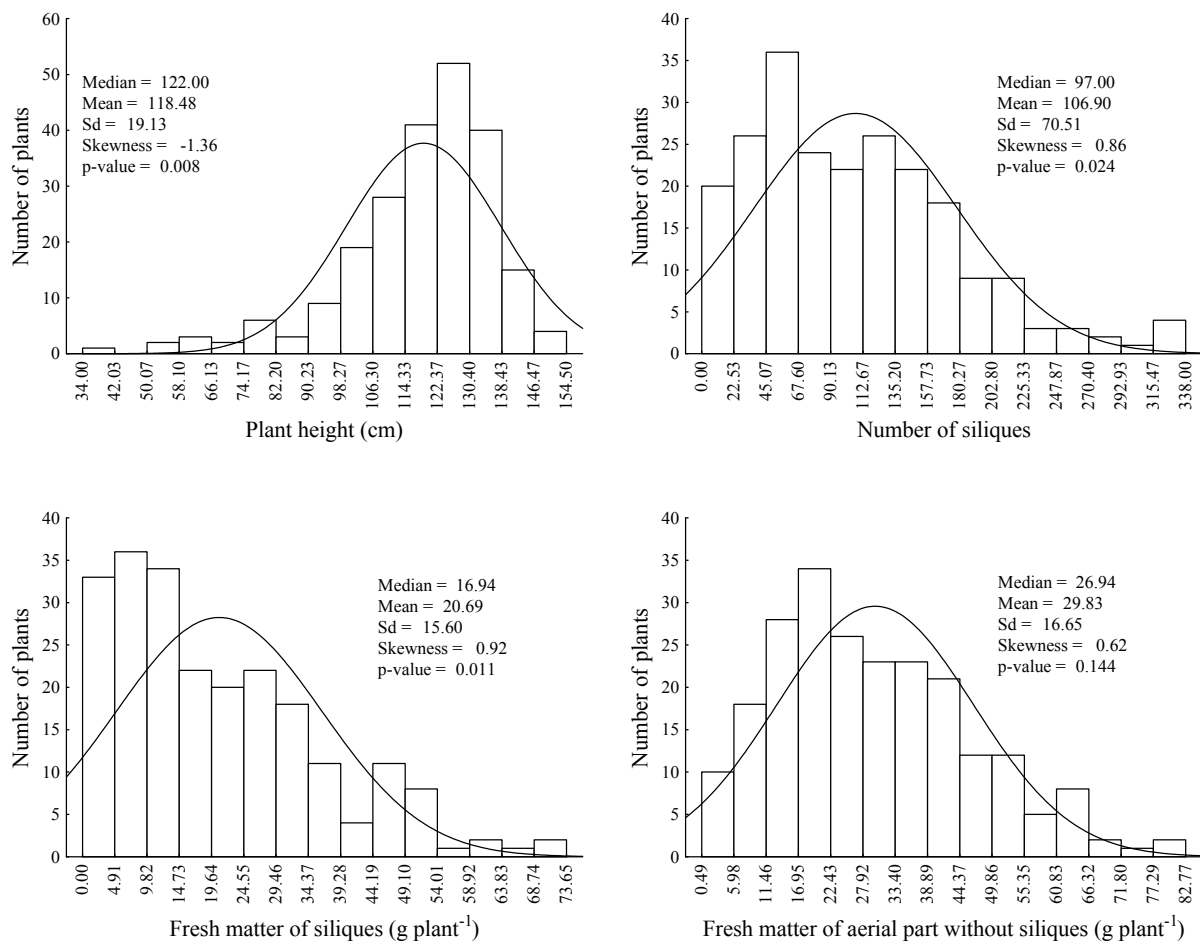


Figure 1. Frequency histograms of traits measured in 225 plants of canola (*Brassica napus* L.) hybrid Hyola 61. In the histograms, the line represents the normal distribution curve. Median (percentile 50), mean, Standard error (Sd), skewness, and p-value of the *Kolmogorov-Smirnov* normality test of the traits



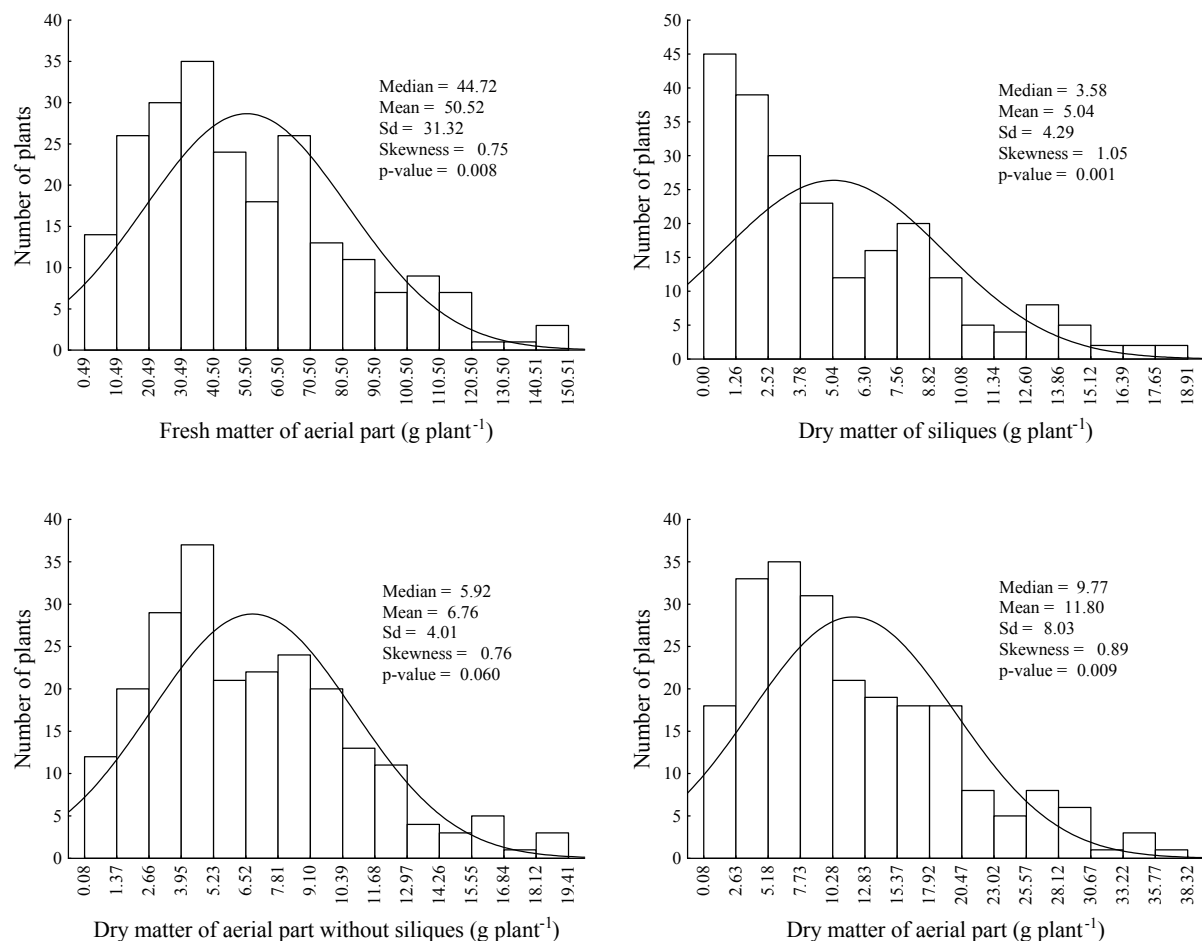
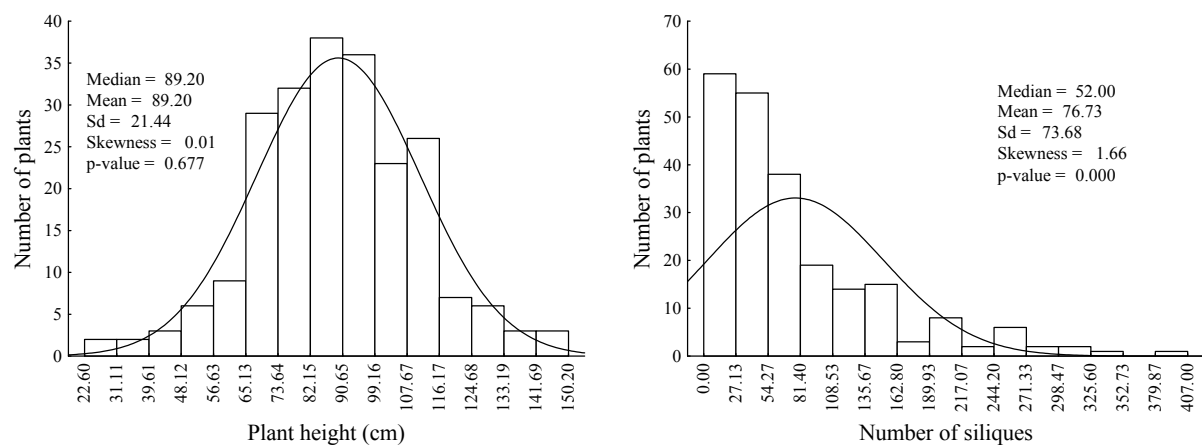


Figure 2. Frequency histograms of traits measured in 225 plants of canola (*Brassica napus* L.) hybrid Hyola 76. In the histograms, the line represents the normal distribution curve. Median (percentile 50), mean, Standard error (Sd), skewness, and p-value of the *Kolmogorov-Smirnov* normality test of the traits



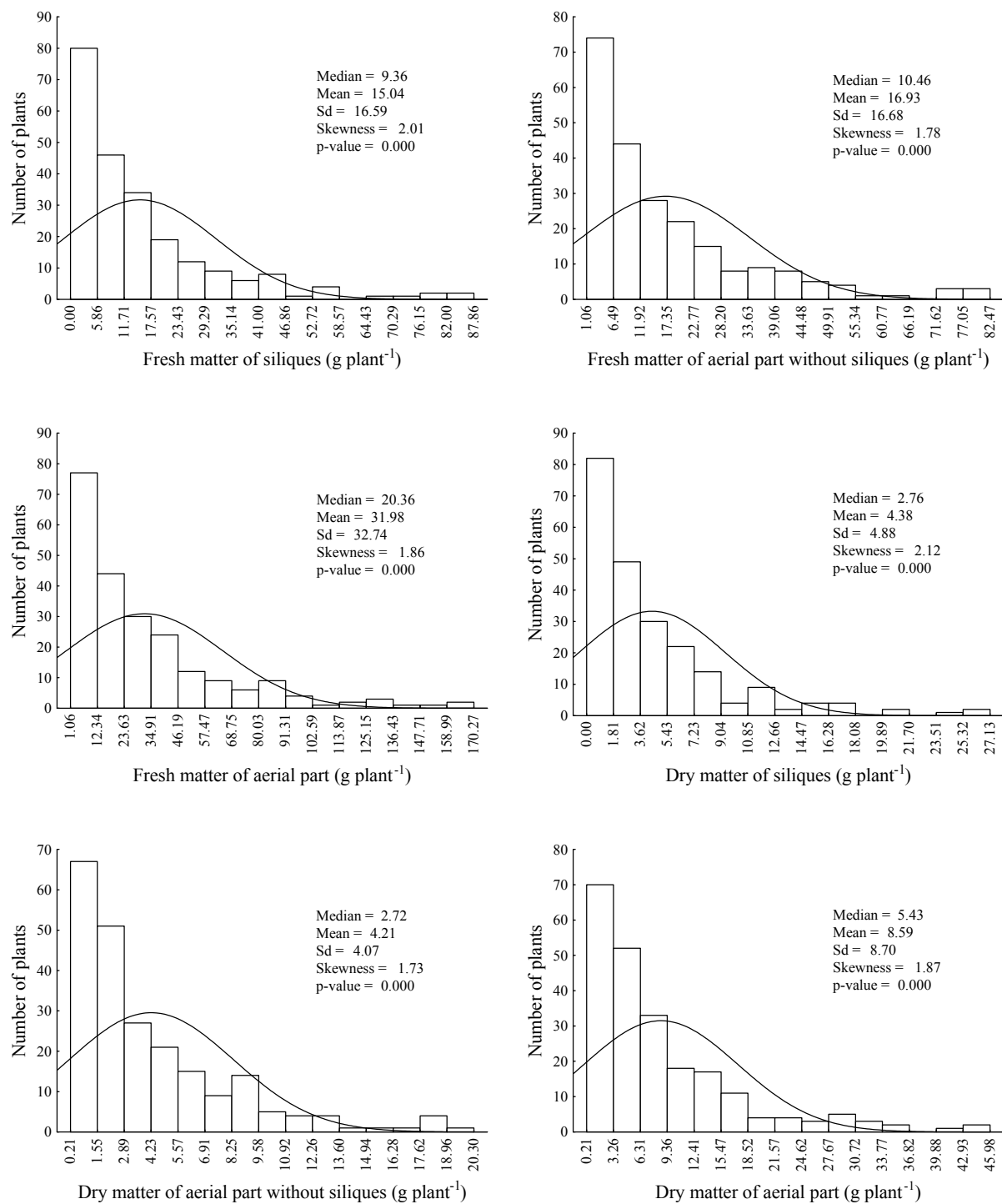


Figure 3. Frequency histograms of traits measured in 225 plants of canola (*Brassica napus* L.) hybrid Hyola 433. In the histograms, the line represents the normal distribution curve. Median (percentile 50), mean, Standard error (Sd), skewness, and p-value of the *Kolmogorov-Smirnov* normality test of the traits

The high p-values of the *Kolmogorov-Smirnov* test for plant height of hybrids Hyola 61 (p-value = 0.871) and Hyola 433 (p-value = 0.677) and the asymmetry and kurtosis coefficients not different from zero respectively demonstrate symmetric and mesokurtic distribution. Consequently, the mean is suitable as a measure of central tendency of these data. On the other hand, for plant height of Hyola 76 hybrid, the low p-value of the *Kolmogorov-Smirnov* test (p-value = 0.008), the asymmetry coefficient lower than zero (negative asymmetry)

and the kurtosis coefficient greater than zero (leptokurtic distribution) indicates that the median is more adequate than the mean as a measure of position. Therefore, determining the sample size for the mean and median estimation is relevant for this context.

The seven productive traits of the three hybrids (Hyola 61, Hyola 76, and Hyola 433) presented mean greater than median, asymmetry greater than zero and low p-value of the *Kolmogorov-Smirnov* test ($p\text{-value} \leq 0.144$) (Table 1 and Figures 1, 2 and 3). This reveals that these traits do not fit the normal distribution and consequently the median is more reliable than the mean as a measure of position. The positive asymmetry of productive traits may be associated with possible right-positioned unilateral outliers. These outliers contribute to the mean being greater than the median. The possible outliers were maintained in the study in order to contemplate the data variability, conferring credibility of the sample size study. Removal of outliers could underestimate the sample size.

Based on the amplitude, variance, standard deviation, standard error, and coefficient of variation, greater variability of the productive traits (NS, FMS, FMWS, FM, DMS, DMWS, and DM) in relation to the morphological trait (PH) was found (Table 1). This suggests that a larger sample size is required for productive than for morphological traits.

The coefficients of variation of the productive traits of Hyola 61 ($71.82\% \leq CV \leq 101.58\%$, mean = 83.92%), Hyola 76 ($55.80\% \leq CV \leq 85.16\%$, mean = 67.40%), and Hyola 433 ($96.02\% \leq CV \leq 111.43\%$, mean = 102.35%) were greater in relation to the morphological trait plant height (25.77%, 16.14%, and 24.04%, respectively for Hyola 61, Hyola 76, and Hyola 433 hybrids) (Table 1). Therefore, based on these measures of variability for the same precision level, larger sample size is expected to measure the productive traits in relation to the morphological one. In addition, differentiated sample size is expected between hybrids, *i.e.*, larger sample size for Hyola 433, intermediate for Hyola 61, and smaller for Hyola 76. Larger coefficients of variation for productive traits in relation to morphological traits were also found in white lupine (Burin et al., 2014), forage turnip (Cargnelutti Filho et al., 2014), black oat (Cargnelutti Filho et al., 2015b), and flax (Cargnelutti Filho et al., 2018).

The sample size (number of plants) for mean estimation of these eight traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM) of canola hybrids with amplitude of confidence interval of 95% equals to 15% of the mean estimate (higher precision) ranged from 46 to 685 for Hyola 61, 18 to 480 for Hyola 76, and 40 to 815 for Hyola 433, respectively for the traits plant height and dry matter of siliques (Table 2). At the other extreme, the sample size (number of plants) for mean estimation of these eight traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM) of canola hybrids with amplitude of confidence interval of 95% equals to 30% of the mean estimate (lower precision) ranged from 12 to 175 for Hyola 61, 5 to 125 for Hyola 76, and 10 to 208 for Hyola 433, respectively for the traits plant height and dry matter of siliques. Sample size variability between traits for mean estimation was also verified in agricultural crops, such as white lupine (Burin et al., 2014), forage turnip (Cargnelutti Filho et al., 2014), black oat (Cargnelutti Filho et al., 2015b), millet (Kleinpaul et al., 2017), rye (Bandeira et al., 2018a, Bandeira et al., 2018b), and flax (Cargnelutti Filho et al., 2018). Among the traits for the three hybrids, larger sample sizes are required for the productive traits (NS, FMS, FMWS, FM, DMS, DMWS, and DM) in relation to the morphological one (PH), being explained by greater CV values by the productive traits in relation to the morphological one. Larger sample sizes of the productive traits in relation to the morphological traits were also obtained in white lupine (Burin et al., 2014), forage turnip (Cargnelutti Filho et al., 2014), black oat (Cargnelutti Filho et al., 2015b), and flax (Cargnelutti Filho et al., 2018).

Table 2. Sample size (*i.e.*, number of plants) to estimate the mean of canola (*Brassica napus* L.) traits⁽¹⁾ of each hybrid (Hyola 61, Hyola 76, and Hyola 433) for the amplitudes of the confidence interval of 95% equal to 15%; 16%; 17%; ...; 30% of the estimated mean

Amplitude	PH	NS	FMS	FMWS	FM	DMS	DMWS	DM
	--- cm ---	----- g plant ⁻¹ -----						
<i>Hyola61</i>								
15%	46	394	624	391	452	685	340	454
16%	40	341	548	346	390	602	303	408
17%	36	298	478	298	345	530	270	355
18%	32	271	440	274	313	474	240	317
19%	28	246	389	244	273	430	214	285
20%	26	221	347	222	254	389	192	257

21%	23	201	323	203	229	351	179	235
22%	22	182	292	184	212	320	164	217
23%	20	165	263	167	192	296	148	194
24%	18	156	241	157	175	271	137	180
25%	17	142	223	140	162	253	125	163
26%	15	134	208	130	147	235	116	155
27%	14	123	193	120	140	211	108	139
28%	14	115	179	113	131	196	101	134
29%	13	106	169	103	122	187	94	125
30%	12	101	160	100	112	175	86	114
<i>Hyola76</i>								
15%	18	292	376	211	257	480	230	311
16%	16	255	336	188	226	431	210	272
17%	14	225	295	165	202	373	186	244
18%	13	200	267	142	177	326	163	213
19%	12	186	240	132	163	296	147	194
20%	11	165	215	116	144	274	134	174
21%	9	152	197	107	132	251	121	158
22%	9	136	174	98	123	226	111	148
23%	8	126	165	87	111	206	103	134
24%	7	117	152	82	102	190	93	123
25%	7	108	136	75	96	174	85	114
26%	6	98	128	70	87	159	79	101
27%	6	90	118	65	81	156	75	97
28%	6	84	108	61	74	141	68	92
29%	5	80	104	57	71	129	64	83
30%	5	74	95	54	66	125	62	77
<i>Hyola433</i>								
15%	40	611	805	647	697	815	628	684
16%	34	536	694	570	618	739	533	592
17%	31	477	633	500	549	633	486	535
18%	28	423	562	446	479	575	438	471
19%	25	380	500	399	438	517	387	423
20%	23	349	458	357	388	469	345	375
21%	21	312	409	335	352	413	319	349
22%	19	288	379	304	320	381	290	318
23%	17	265	345	277	300	353	267	291
24%	15	245	315	254	281	324	241	272
25%	14	223	296	234	245	297	229	243
26%	14	203	271	223	236	275	206	231
27%	13	193	255	199	216	254	192	210
28%	12	179	227	187	204	241	174	193
29%	11	168	216	175	189	228	170	183
30%	10	152	205	156	178	208	160	172

Note. ⁽¹⁾ PH: plant height; NS: number of siliques; FMS: Fresh matter of siliques; FMWS: fresh matter of aerial part without siliques; FM: fresh matter of aerial part; DMS: dry matter of siliques; DMWS: dry matter of aerial part without siliques; and DM: dry matter of aerial part.

For the same precision level, sample sizes for mean estimation of these eight traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM) were decreasing in the following order of hybrids: Hyola 433, Hyola 61, and Hyola 76 (Table 2). These results are explained by the lower CV values of Hyola 76, intermediate of Hyola 61, and greater of Hyola 433. Therefore, in experiments with the three canola hybrids, Hyola 76 would be preferred because it allows experiments with a greater experimental precision (lower variation) when compared to the Hyola 433 and Hyola 61 hybrids. In addition, the means of Hyola 76 traits exhibited superior performance to Hyola 433 and

Hyola 61 hybrids (Table 1). Sample size variability among cultivars was also observed in rye (Bandeira et al., 2018a, Bandeira et al., 2018b).

The sample size (number of plants) for median estimation of these eight traits (PH, NS, FMS, FMWS, FM, DMMS, DMWS, and DM) of canola hybrids with amplitude of confidence interval of 95% equals to 30% of the median estimate (lower precision in this study) ranged from 18 to 642 for Hyola 61 and from 5 to 421 for Hyola 76, respectively for plant height and dry matter of siliques (Table 3). For Hyola 433, sample size oscillated between 14 for plant height (PH) and 661 for fresh matter of aerial part (FM). Sample size variability between traits for median estimation was also observed in agricultural crops, such as white lupine (Burin et al., 2014) and flax (Cargnelutti Filho et al., 2018). Among the traits for the three hybrids, larger sample sizes were found for productive traits (NS, FMS, FMWS, FM, DMS, DMWS, and DM) in relation to the morphological one (PH), as verified in white lupine (Burin et al., 2014) and flax (Cargnelutti Filho et al., 2018).

Table 3. Sample size (*i.e.*, number of plants) to estimate the median of canola (*Brassica napus* L.) traits⁽¹⁾ of each hybrid (Hyola 61, Hyola 76, and Hyola 433) for the amplitudes of the confidence interval of 95% equal to 15%; 16%; 17%; ...; 30% of the estimated median

Amplitude	PH	NS	FMS	FMWS	FM	DMS	DMWS	DM
	--- cm ---		----- g plant ⁻¹ -----					
<i>Hyola61</i>								
15%	74	524	1,083	633	537	-(²)	627	1,559
16%	63	498	1,083	622	536	-	539	1,559
17%	54	498	1,083	622	519	-	498	1,545
18%	45	419	1,050	622	519	-	429	1,545
19%	43	419	1,050	529	519	1,522	388	1,545
20%	40	402	857	502	519	1,139	301	1,466
21%	36	402	857	413	506	1,101	301	1,466
22%	31	402	819	413	506	1,101	286	1,466
23%	27	309	810	362	506	1,090	286	1,122
24%	25	309	810	290	506	835	265	1,068
25%	23	309	643	259	488	835	248	863
26%	22	280	643	240	421	814	219	843
27%	21	280	643	222	406	804	216	802
28%	20	280	630	195	361	804	203	667
29%	20	261	630	184	361	657	203	637
30%	18	256	630	176	342	642	190	535
<i>Hyola76</i>								
15%	16	1,038	1,076	500	808	1,135	1,038	1,107
16%	14	622	833	482	808	1,078	839	1,107
17%	14	513	649	367	649	851	642	1,107
18%	12	490	639	338	649	809	614	1,068
19%	10	441	626	301	610	801	502	1,068
20%	10	441	624	278	610	675	369	1,068
21%	9	424	537	231	508	659	369	853
22%	8	414	517	201	435	659	352	669
23%	8	367	504	201	398	626	334	533
24%	6	367	504	189	311	626	306	508
25%	6	352	429	188	254	618	284	425
26%	6	342	408	170	205	529	248	412
27%	6	301	371	157	193	514	214	408
28%	6	297	354	152	173	502	189	352
29%	5	294	313	140	170	421	189	308
30%	5	290	301	123	159	421	189	263
<i>Hyola433</i>								
15%	57	-	1,545	-	1,569	-	-	-
16%	52	-	1,489	1,571	1,492	1,595	1,567	1,599
17%	46	-	1,489	1,539	1,149	1,514	1,557	1,528

18%	42	-	1,077	1,539	1,128	1,111	1,557	1,452
19%	38	-	831	1,468	1,110	845	1,522	1,107
20%	34	-	831	1,463	1,064	835	1,494	1,107
21%	31	-	831	1,093	873	824	1,466	1,050
22%	28	1,548	802	1,086	858	782	863	1,050
23%	26	1,472	802	821	858	535	802	1,050
24%	24	813	802	792	823	517	672	845
25%	22	623	620	792	823	517	653	671
26%	20	623	535	622	776	502	636	614
27%	18	515	523	523	661	492	533	513
28%	16	492	518	514	661	398	525	510
29%	16	365	496	500	661	364	504	504
30%	14	365	439	445	661	342	428	437

Note. ⁽¹⁾ PH: plant height; NS: number of siliques; FMS: Fresh matter of siliques; FMWS: fresh matter of aerial part without siliques; FM: fresh matter of aerial part; DMS: dry matter of siliques; DMWS: dry matter of aerial part without siliques; and DM: dry matter of aerial part. ⁽²⁾ The sample size was not calculated (> 2,000 plants).

For the same precision, the sample sizes for median estimation of these eight traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM) were decreasing in the following order of hybrids: Hyola 433, Hyola 61, and Hyola 76 (Table 3). These results reinforce the previous assertion that in experiments with the three canola hybrids, the Hyola 76 would be preferred, since the hybrid allows experiments with greater experimental precision (lower variation) when compared to Hyola 433 and Hyola 61 hybrids.

With the exception of plant height (PH) of the Hyola 76 hybrid, with similar sample sizes for the mean and median estimation, the remaining cases (3 hybrids \times 8 traits) evidenced that larger sample size is required for median estimation in relation to the mean estimation (Tables 2 and 3). In white lupine (Burin et al., 2014) and in flax (Cargnelutti Filho et al., 2018), larger sample sizes were also found for median estimation in relation to the mean.

Considering the sample size variability between traits (PH, NS, FMS, FMWS, FM, DMS, DMWS, and DM), hybrids (Hyola 61, Hyola 76, and Hyola 433) and measures of central tendency (mean and median), choosing the desired precision and assuming the largest sample size to estimate measures of central tendency of interest (medium or medium) is recommended in order to contemplate all situations.

For example, if the researcher decides to estimate the mean with precision level of 30%, 208 plants would be sufficient for these eight traits and three hybrids (Table 2, Figure 4A). Thus, when planning a field experiment in a completely randomized experimental design, 208 plants per treatment should be evaluated to estimate the mean of each treatment with 30% of precision. If the experiment is planned with five replicates per treatment, 42 plants per replicate ($208/5 = 41.6 \cong 42$) would be sampled, *i.e.*, 42 plants per plot (210 plants per treatment). In addition, if 10 treatments would be evaluated in the experiment, 2,100 plants would be sampled (210 per treatment). For the same precision (30% of the mean estimate), if the researcher designs the experiment only with the Hyola 76 hybrid (better performing hybrid with more precision), 125 plants per treatment ($125/5 = 25$ plants per plot) would be enough for these eight traits (Table 2). In the experimental area, the plants would be randomly taken in each plot of each treatment.

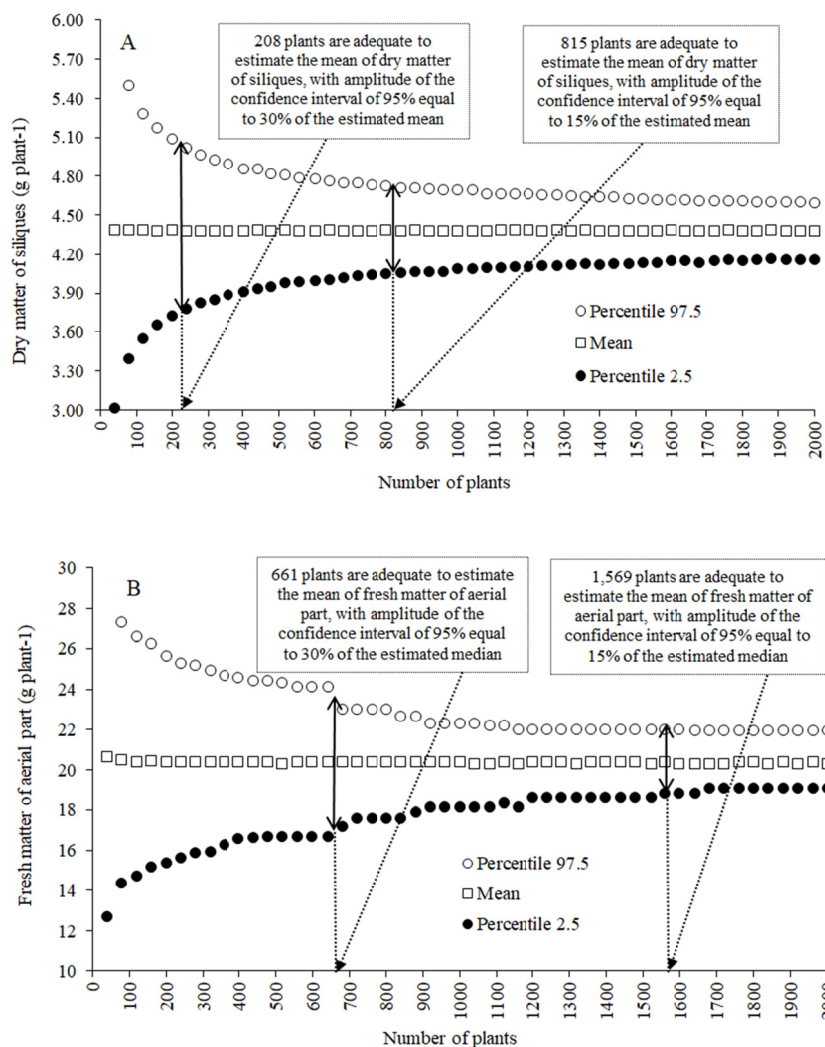


Figure 4. Percentile 2.5, mean, and percentile 97.5 of the 10,000 means of the dry matter of siliques (A) and 10,000 medians of the fresh matter of aerial part (B), obtained by resampling (based on 225 plants of canola (*Brassica napus* L.), hybrid Hyola 433) in each sample size (*i.e.*, number of plants)

In another scenario, if the researcher decides to estimate the median with precision level of 30%, 661 plants would be sufficient for these eight traits and three hybrids (Table 3, Figure 4B). Therefore, when planning a field experiment in a completely randomized experimental design, 661 plants per treatment should be evaluated to estimate the median of each treatment with 30% of precision. If the experiment is planned with five replicates per treatment, 132 plants per replicate ($661/5 = 132.2 \approx 132$) would be sampled, *i.e.*, 132 plants per plot (660 plants per treatment). In addition, if 10 treatments would be evaluated in the experiment, 6,600 plants would be sampled (660 per treatment). For the same precision (30% of the mean estimate), if the researcher designs the experiment only with the Hyola 76 hybrid (better performing hybrid with more precision), 421 plants per treatment ($421/5 = 84.2 \approx 84$ plants per plot) would be enough for these eight traits (Table 3).

In practice, the results of this study allow the researcher to choose the appropriate sample size to estimate the measures of central tendency (mean or median) of these traits and hybrids for a given precision (Tables 2 and 3). Therefore, the sample size information provided in this study along with the results published by Cargnelutti Filho et al. (2015a) on the optimum plot size and the number of replicates for combinations of treatments and experimental precision levels serve as a reference for experimental designs with canola (*Brassica napus* L.).

4. Conclusions

The sample size (number of plants) required for the estimation of measures of central tendency (mean and median) varies among canola traits and hybrids.

Larger sample sizes are required for the productive traits in relation to the morphological trait.

Larger sample sizes are required for the hybrids in this order Hyola 433, Hyola 61, and Hyola 76.

For mean estimation of canola traits of the Hyola 61, Hyola 76, and Hyola 433 hybrids with the amplitude of the confidence interval of 95% equal to 30% of the estimated mean, 208 plants are required.

For median estimation of canola traits of the Hyola 61, Hyola 76, and Hyola 433 hybrids with the amplitude of the confidence interval of 95% equal to 30% of the estimated median, 661 plants are required.

Acknowledgements

We thank the National Council for Scientific and Technological Development (CNPq-Processes 401045/2016-1 and 304652/2017-2) and the Coordination for the Improvement of Higher Education Personnel (Capes) for granting scholarships. Embrapa researcher, Gilberto Omar Tomm, for the donation of canola (*Brassica napus* L.) seeds. Students who assisted in data collection.

References

- Angus, J. F., Kirkegaard, J. A., Hunt, J. R., Ryan, M. H., Ohlander, L., & Peoples, M. B. (2015). Break crops and rotations for wheat. *Crop and Pasture Science*, 66(6), 523-552. <https://doi.org/10.1071/CP14252>
- Bandeira, C. T., Cargnelutti Filho, A., Carini, F., Schabarum, D. E., Kleinpaul, J. A., & Pezzini, R. V. (2018a). Sample sufficiency for estimation of the mean of rye traits at flowering stage. *Journal of Agricultural Science*, 10(3), 178-186. <https://doi.org/10.5539/jas.v10n3p178>
- Bandeira, C. T., Cargnelutti Filho, A., Follmann, D. N., Bem, C. M., Wartha, C. A., & Thomasi, R. M. (2018b). Sample size to estimate the mean of morphological traits of rye cultivars in sowing dates and evaluation times. *Semina: Ciências Agrárias*, 39(2), 521-532. <https://doi.org/10.5433/1679-0359.2018v39n2p521>
- Burin, C., Cargnelutti Filho, A., Toebe, M., Alves, B. M., & Fick, A. L. (2014). Sample size for estimating the average and median character of white lupine (*Lupinus albus* L.). *Comunicata Scientiae*, 5(2), 205-212.
- Bussab, W. O., & Morettin, P. A. (2017). *Estatística Básica* (p. 576). São Paulo: Saraiva.
- Cargnelutti Filho, A., Alves, B. M., Burin, C., Kleinpaul, J. A., Silveira, D. L., & Simões, F. M. (2015a). Plot size and number of repetitions in canola. *Bragantia*, 74(2), 176-183. <https://doi.org/10.1590/1678-4499.0420>
- Cargnelutti Filho, A., Alves, B. M., Santos, G. O., Wartha, C. A., Kleinpaul, J. A., & Silveira, D. L. (2018). Sample size to estimate the mean and median of traits in flax. *Revista Brasileira de Ciências Agrárias*, 13(1), e5492. <https://doi.org/10.5039/agraria.v13i1a5505>
- Cargnelutti Filho, A., Facco, G., Lúcio, A. D., Toebe, M., Burin, C., Fick, A. L., & Neu, I. M. M. (2014). Sample size to estimate the average of morphological and productive traits of turnip. *Ciência Rural*, 44(2), 223-227. <https://doi.org/10.1590/S0103-84782014000200005>
- Cargnelutti Filho, A., Toebe, M., Alves, B. M., Burin, C., Santos, G. O., Facco, G., & Neu, I. M. M. (2015b). Sample size to evaluate morphological and productive characters in black oat in evaluation times. *Ciência Rural*, 45(1), 9-13. <https://doi.org/10.1590/0103-8478cr20140504>
- Daun, J. K., & Symons, S. (2000). How green is green? Sampling and perception in assessing green seeds and chlorophyll in canola. *Journal of the American Oil Chemists' Society*, 77(11), 1209-1213. <https://doi.org/10.1007/s11746-000-0188-0>
- Emslie, K. R., Whaites, L., Griffiths, K. R., & Murby, E. J. (2007). Sampling plan and test protocol for the semiquantitative detection of genetically modified canola (*Brassica napus*) seed in bulk canola seed. *Journal of Agricultural and Food Chemistry*, 55(11), 4414-4421.
- Ferreira, D. F. (2009). *Estatística básica* (p. 664). Lavras: UFLA.
- Heldwein, A. B., Buriol, G. A., & Streck, N. A. (2009). O clima de Santa Maria. *Ciência e Ambiente*, 38(1), 43-58.
- Kleinpaul, J. A., Cargnelutti Filho, A., Alves, B. M., Burin, C., Neu, I. M. M., Silveira, D. L., & Simões, F. M. (2017). Sample size to estimate the mean of millet traits in evaluation times. *Revista Brasileira de Milho e Sorgo*, 16(2), 51-262.
- Li, L., Long, Y., Zhang, L., Dalton-Morgan, J., Batley, J., Yu, L., ... Li, M. (2015). Genome wide analysis of flowering time trait in multiple environments via high-throughput genotyping technique in *Brassica napus* L. *Plos One*, 10(3), p.e0119425. <https://doi.org/10.1371/journal.pone.0119425>

- Nematollahi, M. R., Fathipour, Y., Talebi, A. A., Karimzadeh, J., & Zalucki, M. P. (2014). Sampling procedure and temporal-spatial distribution of the cabbage aphid, *Brevicoryne brassicae* L. (Hemiptera: Aphididae), on Canola. *Journal of Agricultural Science and Technology*, 16(6), 1241-1252.
- R development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org>
- Rigon, C. A. G., Goergen, A. B., Bordin, R., Pilla, R. B., Zanatta, T. P., Silva, V. R., & Tomm, G. O. (2017). Agronomic characteristics, yield of oil and protein of rapeseed in different sowing dates. *Revista Brasileira de Tecnologia Agropecuária*, 1(2), 124-132.
- Santos, H. G., Jacomine, P. K. T., Anjos, L. H. C., Oliveira, V. A., Oliveira, J. B., Coelho, M. R., ... Cunha, T. J. F. (2013). *Sistema brasileiro de classificação de solos* (p. 353). Brasília: Embrapa.
- Storck, L., Garcia, D. C., Lopes, S. J., & Estefanel, V. (2016). *Experimentação vegetal* (p. 198). Santa Maria: UFSM.
- Tomm, G. O., Wiethölter, S., Dalmago, G. A., & Santos, H. P. (2009). *Tecnologia para produção de canola no Rio Grande do Sul* (p. 41). Passo Fundo: Embrapa Trigo.
- Turkington, T. K., Morrall, R. A. A., & Baker, R. J. (1988). Sample sizes in relation to forecasting sclerotinia stem rot of canola. *Canadian Journal of Plant Pathology*, 10(2), 159-165. <https://doi.org/10.1080/07060668809501748>

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/>).