ISSN 0103-8478

Classification of the coefficients of variation for sugarcane crops

Classificação dos coeficientes de variação para a cultura da cana de açúcar

Mauricio Farias Couto[I*] Luiz Alexandre Peternelli[I] Márcio Henrique Pereira Barbosa[II]

ABSTRACT

The objective of this paper was to evaluate the range of classification of variation coefficient in experiments with sugarcane. For this study, some of the variables that are mostly used by researchers of the culture were considered. The data was obtained after extensive bibliographical revision from scientific journals, data base from the improvement program of sugarcane from the Federal University of Viçosa, master dissertations, doctorate thesis and technical reports. Classification ranges were proposed according to the Costa method, in which is based the use of a median and pseudo-sigma. The variables tons of stalks per hectare and tons of pol per hectare presented the highest ranges of the classification of the coefficients of variation, while the variable percentage of sucrose presented the lowest range. All of the variables presented specific ranges of classification, showing the need to consider, in the classification of coefficients of variation, the nature of the variable being studied.

Key words: ***Saccharum*** *spp, experimental precision, experimental error.*

RESUMO

O objetivo deste trabalho foi avaliar a faixa de classificação do coeficiente de variação em experimentos com cana de açúcar. Para este estudo, foram consideradas algumas das variáveis mais utilizadas por pesquisadores da cultura. Os dados foram obtidos de extensa revisão bibliográfica em revistas científicas, banco de dados do programa de melhoramento de cana de açúcar da Universidade Federal de Viçosa, dissertações de mestrado, teses de doutorado e relatórios técnicos. Foram propostas faixas de classificação segundo o método de Costa, o qual se baseia no uso da mediana e do pseudo-sigma. As variáveis toneladas de colmos e toneladas de sacarose por hectare apresentaram as maiores faixas de classificação dos coeficientes de variação, enquanto que a variável porcentagem de sacarose apresentou a menor faixa dos coeficientes de classificação. Todas as variáveis apresentaram faixas de classificação específica, evidenciando a necessidade de se considerar, na classificação dos coeficientes de variação, a natureza da variável estudada.

Palavras- chave: ***Saccharum*** *spp, precisão experimental, erro experimental.*

INTRODUCTION

The concern with the quality in the results of any activity is vital in breeding programs. In experiments, the quality of an assay is the factor that indicates reliability of the obtained results. Quality control of the assays is performed at the planning stage, in order to obtain an acceptable level of experimental precision (STORCK & LOPES, 1998).

The precision of an experiment is measured by the magnitudes of the experimental error, defined by STEEL et al., (1996) as the variation due to the effect of non-controlled factors or that occur by chance, at random. Small variations in experimental units, before applying the treatments, cause heterogeneity between parcels, also known as environmental variation or experimental error (RAMALHO et al., 2012).

The existence of a coefficient that estimates experimental precision is crucial, especially in the comparison of scientific papers (SCAPIM et al., 1995). To compare the experimental precision of different experiments it is usually applied the experimental coefficient of variation (CV), in percentages, and obtained by the expression

[I]Departamento de Estatística, Universidade Federal de Viçosa (UFV), 36570-000, Viçosa, MG, Brasil. E-mail: couto_estatistica@hotmail.com.
*Autor para correspondência.
[II]Departamento de Fitotecnia. UFV, Viçosa, MG, Brasil.

Received 05.02.12 Approved 11.21.12 Returned by the author 05.14.13
CR-2012-0096.R1

$$CV = \frac{\sqrt{QME}}{\hat{m}} \times 100$$

In which: QME represent the Mean Square Error and $\hat{m}$ is the estimate of the experimental average. Other measures of the experimental precision are suggested in the literature (CARGNELUTTI FILHO, et al, 2009; 2012), but will not be considered on the present study.

The CV is understood as the estimate of the experimental error in percentage of the estimate of the average. It is one of the most used statistical measures by researchers in the evaluation of experiment precision. However, to know if a particular coefficient of variation is excessively high or low requires experiments with similar data (STEEL et al., 1996).

According to GOMES (2009), in field experiments, if the coefficient of variation is below 10% it is considered low, that is, the experiment has a high precision. From 10 to 20% the CV is considered medium, implying in good precision. From 20 to 30% is considered high, meaning low precision. Finally, if above 30 % it is considered very high, indicating very low precision. The inconvenience of this classification is not taking into consideration the crop studied, the variables in analysis, the heterogeneity of the soil, and the size of the plot, among other factors.

In plant breeding programs, the classification of CV can be useful, for example, to inform the quality of final and intermediary trials of the evaluated crops. In these trials, a set of characters is measured to aid the researcher in the selection and indication of new cultivars (CARVALHO et al., 2003).

The evaluation of CV, as a precision measure in experiments, has been conducted in several types of crops, with some proposing methods to obtain the ranges of classification of CV (ESTEFANEL et al., 1987; GARCIA, 1989; AMARAL et al., 1997, COSTA et al., 2002) and, others defining the ranges of classification (CAMPOS, 1984; SCAPIM et al., 1995; JUDICE et al., 1989; GOMES, 2009).

Specifically for sugarcane, there are not a lot of references on value ranges for CV, despite this crop constitutes in one of the most important options for Brazilian agriculture.

The objective of this paper is to determine the ranges of classification of the coefficient of variation for the variables tons of stalks per hectare, tons of pol per hectare and percentage of sucrose according to the methodology proposed by COSTA et al. (2002), in experiments with sugarcane.

MATERIAL AND METHODS

The data used in this study were obtained by means of a bibliographical revision in scientific journals, in data base of the program of genetic improvement of sugarcane (PMGCA) of the Federal University of Viçosa, and in master dissertations, doctorate thesis and technical reports that contained experiments with sugarcane. The researched journals were: Bragantia (1966-2007), Ciência Rural (1995-2007), Engenharia Agrícola (2004-2007), Pesquisa Agropecuária Brasileira (1999-2007), Scientia Agrícola (1992-2007) and Crop Breeding and Applied Biotechnology (2001-2007). It was found 502 values of CV distributed in three different variables: tons of stalks per hectare (TSH), percentage of sucrose (PSU), and tons of pol per hectare (TPH). TSH is a measure of the total weigh of the stalks in a plot, given in tons; PSU is a measure of the percentage of sucrose of the stalk, obtained from saccharimetric methods at the industry (CONSECANA, 2006); and TPH=(TCH×PSU)/100. Most of the data are from experiments conducted with plant cane and ratoon cane, without distinction for the purpose of the present analysis. Also, it is important to inform that the great majority of the experiments were harvested manually. The rest did not mention about the kind of harvesting.

The values of CV were found directly in the revised articles or later estimated, when possible, through analysis of variance (ANOVA) from the field data. The ANOVA were conducted according to the outline applied in the respective experiments. In these cases the CVs were estimated using the formula: CV=100($\sqrt{QME}/\hat{m}$), in which QME is the mean square of the residue from ANOVA, and $\hat{m}$ is the estimate of the average of the experiment.

For the definition of the ranges of classification of the CVs the method proposed by COSTA et al. (2002) was used. In this method the ranges are based in the use of the median (Md) and the pseudo-sigma (PS) of all the CVs, those in which according to Hoaglin et al. (1983) do not need to have normal distribution. The ranges of classification of the coefficient of variation were defined, as follows:

Low, if CV ≤ (Md-1PS);
Medium, if (Md-1PS)<CV ≤ (Md+1PS);
High, if (Md+1PS)<CV ≤ (Md+2PS);
Very High, if CV>(Md+2PS).

In the proposal of COSTA et al. (2002) the Md is obtained by Md=(Q_1+Q_3)/2, in which Q_1 and Q_3 correspond, respectively, to the first and third quartiles, outlining 25% of each extremity of the distribution; and the PS is obtained by PS=IQR/1.35, in which IQR=Q_3-Q_1 (interquartile range), a robust measurement that indicates how the data are distancing from the median.

According to COSTA et al. (2002) the pseudo-sigma would be the standard deviation that a normal distribution should have in order to produce the same interquartile range of the distribution of the sample data. This interpretation of the pseudo-sigma is justified by the presence of the value 1.35, corresponding the distance between Q_1 and Q_3, which is equivalent to 50% of the data, leaving 25% in each extremity. When IQR is divided by 1.35 the result obtained produces the expected standard deviation if there were a normal standard distribution.

When data have a normal distribution, 68.27% of CVs are included among 1PS+Md; 95.45% of CVs are included among Md 2PS +, and 99, 73% of CVs are included among Md+3PS (SPIGEL, 1993). However the expected frequencies coefficients of variation in the intervals defined by COSTA et al. (2002) matched: 15.86%, 68.27%, 13.59% and 2.28% respectively.

When the data do not have a normal distribution, the use of the pseudo-sigma as a dispersion measurement will be more robust than the classic standard deviation; however if the data have a approximately normal distribution, the pseudo-sigma produces an estimate of variance as good as the one obtained by the classic standard deviation (HOGLIN et al., 1983; BLANXART et al., 1992).

The analyses were conducted using the software R (R Development Core Team, 2009).

RESULTS AND DISCUSSION

In the consulted literature, a greater number of coefficients of variation were found for the variables TSH, TPH and PSU with 205, 131 and 166 values, respectively. The average of the CVs varied between 6.46 and 13.77%. The standard deviations varied between 3.97% and 6.07% (Table 1). Based on maximum and minimum values, it is possible to observe the great magnitude between and within the variables, which indicates influence of different factors in its measurement. This result justifies the need for a specific classification of CV for each variable. The variable that presented the highest variability for the values of the CV was TSH, with standard deviation of 6.07%. The variable with the lowest variability was PSU, presenting a standard deviation of 3.97% (Table 1).

The TPH variable presented the highest limits for the range of classification of the CV when compared to those referring to the other variables (Table 2). MELO et al. (2001) reported that the tons of sugar per hectare (TPH), being the product between TSH and PSU, carries the variation contained in both variables, which would justify the more elevated limits. Taking into consideration the methodology by COSTA et al. (2002) for the variable TPH, experiments with CV that are lower than 7.0% would be classified as of high precision, and experiments with CV between 7.0 and 19.0% would be considered as of good precision. Only those trials with CV scores above 19.0% would be considered as of low precision, and if the CV is very high (greater than 25.0%) they would be considered as of very low precision (Table 2).

The TSH variable presented the second highest range of classification of CV (Table 2). The production of stalks per hectare is a very important variable in the genetic improvement program of sugarcane. It is a variable of complex heritage and that probably presents pleiotropic effects with the other variables. According to FERREIRA et al. (2007) TCH, when estimated indirectly, is a function of the components stalk height (SH), stalk diameter (SD) and stalk number (SN). In this case, a great variation in these yield component variables will directly affect the TSH. It is worth mentioning that the variables SH, SD and SN are subject to measurement errors since they are measured directly in the field, besides suffering variations due to the environment.

Table 1 - Minimum, maximum, average, standard deviation (SD) and total number of observations, related to coefficients of variations obtained in experiments with sugarcane.

Variables[(1)]	Minimum	Maximum	Average	SD	Total of Observations
TSH	4.1	46.67	11.09	6.07	205
TPH	2.89	32.48	13.77	5.84	131
PSU	1.57	23.14	6.46	3.97	166

[(1)]TSH: tons of cane per hectare; TPH: tons of pol per hectare; PSU: Percentage of sucrose.

Table 2 - Limiting values for the classification of the coefficients of variation by variable, according to the method proposed by COSTA et al. (2002), and according to the classification by GOMES (2009). Data from sugarcane experiments.

Classes	Classification of CV(%)			
	Low	Medium	High	Very High
TSH[(1)]	CV < 5,0	5,0 ≤ CV < 15,0	15,0 ≤ CV < 21,0	CV ≥ 21,0
TPH	CV < 7,0	7,0 ≤ CV < 19,0	19,0 ≤ CV < 25,0	CV ≥ 25,0
PSU	CV < 2,0	2,0 ≤ CV < 10,0	10,0 ≤ CV < 13,0	CV ≥ 13,0
Gomes[(2)]	CV < 10	10,0 ≤ CV < 20,0	20,0 ≤ CV < 30	CV ≥ 30,0

[(1)] TSH: tons of stalks per hectare; TPH: tons of pol per hectare; PSU: Percentage of sucrose.
[(2)] Classification according to GOMES (2009).

The classes of trial precision referring to the TSH variable in accordance to the proposed method will be high precision, if the coefficient of variation is inferior to 5.0%. From 5.0 to 15.0% the CV are considered as medium, showing good experimental precision. From 15.0 to 21.0% CVs are considered as high, implying low experimental precision. Above 21.0% the CV scores are considered very high, indicating a very low experimental precision (Table 2).

The PSU variable presented the lowest range of coefficient of variation (Table 2). This fact is justified since, in general, variables measured in laboratories, like the PSU, have lower variations than those determined in the field, subjected to great variability due to environmental factors and difference between evaluators.

Regarding the PSU variable, if the coefficient of variation is inferior to 2.0% it is considered low, that is, the trial has a high precision. From 2.0 to 10.0% the CVs are considered medium, implying in good trial precision. From 10.0 to 13.0% CVs are considered high, indicating low trial precision. Above 13.0% CVs are considered as very high, meaning very low experimental precision (Table 3).

The TSH, TPH and PSU variables presented their own ranges of classification differing from the classification proposed by GOMES (2009). This fact shows the need to take into consideration the specificities of the crop and the analyzed variable obtaining exclusive precision ranges, as it is suggested by CARVALHO et al. (2003), CLEMENTE et al. (2002), COSTA et al. (2002), AMARAL et al. (1997) and SCAPIM et al. (1995).

Another factor to be considered is the comparison of this new classification according to COSTA et al. (2002) with the usual proposed by GOMES (2009). Table 2 demonstrates some of the values considered in this paper as High (15.0%≤CV≤ 21.0%) and Very High (CV≥21.0%) would be considered, respectively as Low (CV<10%) and Medium (10%<CV≤20%) according to the classification by GOMES (2009). This contradicting result has occurred in all of the variables, enlightening the need to consider, in the classification of CV, the nature of the studied variable (Table 2), as suggested by SCAPIM et al. (1995) and AMARAL et al. (1997) and applied in the present study.

Table 3 - Table of value frequency of coefficient of variation belonging to each category (Classes) defined according to the classification proposed by COSTA et al. (2002) (CO), in bold, and according to GOMES (2009) (PG), for three observed variables.

Classes	TSH				TPH				PSU			
	CO	%	PG	%	CO	%	PG	%	CO	%	PG	%
Low	10	4,88	104	50,73	17	12,98	38	29,00	5	3,01	142	85,54
Medium	161	78,54	86	41,95	89	67,94	74	56,49	129	77,71	22	13,25
High	20	9,75	11	5,37	17	12,98	17	12,98	20	12,05	2	1,21
Very High	14	6,83	4	1,95	8	6,11	2	1,53	12	7,23	0	0
Total	205	100	205	100	131	100	131	100	166	100	166	100

[(1)]TSH: tons of stalks per hectare; TPH: tons of pol per hectare; PSU: Percentage of sucrose.

The CV frequencies in the experiments seen in the different classes, according to the classification by COSTA et al. (2002), as well as, the frequencies seen according to the classification by GOMES (2009) are presented on Table 3. It is seen that the herein applied method concentrates more values of CV in the Medium class with 75.50% of the data on average; while in the proposal by GOMES (2009) it classifies most of the values as Low with 56.67% of the data on average.

CONCLUSION

There was a disagreement between the proposed classification and the classification by Pimentel Gomes, showing the need for the analyzed variables to have their own ranges of classification. The magnitude of the coefficients of variation (CV) obtained in the experiments with sugarcane varies according to the nature of the variables. The percentage of sucrose (PSU) presented the lower limits in ranges of CV, while the variable tons of stalks per hectare (TSH) and tons of pol per hectare (TPH) presented the highest limits in ranges of coefficient of variation. The CV upper limits for considering a sugarcane experiment as of good to high precision is 10, 15 and 19% for PSU, TSH and TPH, respectively.

REFERENCES

AMARAL, A.M. do et al. Avaliação do coeficiente de variação como medida de precisão na experimentação com citros. **Pesquisa Agropecuária Brasileira**, Brasília, v.32, n.12, p.1221-1225, dez. 1997. Available from: <http://webnotes.sct.embrapa.br/pab/pab.nsf/FrAnual>. Accessed: Mar. 22, 2012.

CAMPOS, H. de. **Estatística aplicada a experimentação com cana de açúcar**. Piracicaba: FEALQ, 1984. 292p.

CARGNELUTTI FILHO, A. et al. Medidas da precisão experimental em ensaios com genótipos de feijão e de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.44, n.10, p.1225-1231, 2009.

CARGNELUTTI FILHO, A. et al. Medidas de precisão experimental e números de repetições em ensaios de genótipos de arroz irrigado. **Pesquisa Agropecuária Brasileira**, Brasília, v.47, p.336-343, 2012.

CARVALHO, C.G.P. et al. Proposta de classificação dos coeficientes de variação em relação a produtividade e altura da planta de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.38, n.2, p.187-193, 2003. Availale from: <http://webnotes.sct.embrapa.br/pab/pab.nsf/FrAnual>. Accessed: Mar. 22, 2012. doi: 10.1590/S0100-204X2003000200004.

CONSECANA (CONSELHO DOS PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO). **Manual de instruções**. 5.ed. Piracicaba, 2006. 111p. Available from: <http://www.unica.com.br/download.asp?mmdCode=A8D2ABCA-8247-45D1-8720-C14CD485F380>. Accessed: Sept. 28, 2012.

CLEMENTE, A.L.; MUNIZ, J.A. Avaliação dos coeficientes de variação em experimentos com gramínea forrageira. **Ciência e Agrotecnologia**, Lavras, v.26, n.1, p.197-203, 2002. Available from: <http://www.editora.ufla.br/site/_adm/upload/revista/26-1-2002_23.pdf>. Accessed: Mar. 22, 2012.

COSTA, N.H. de A.D. et al. Novo método de classificação de coeficiente de variação para a cultura do arroz de terras altas. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.3, p.243-249, 2002. Available from: <http://webnotes.sct.embrapa.br/pab/pab.nsf/FrAnual>. Accessed: Mar. 22, 2012. doi: 10.1590/S0100-204X2002000300003.

ESTEFANEL, V. et al. Avaliação do coeficiente de variação de experimentos com algumas culturas agrícolas. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 2.; REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 32., 1987, Londrina, PR. **Anais...** Londrina: Universidade Estadual de Londrina, 1987. p.115-131.

GARCIA, C.H. **Tabelas para classificação do coeficiente de variação**. Piracicaba: Ipef, 1989. 12p. (Circular técnica, 171).

GOMES, F.P. **Curso de estatística experimental**. 15.ed. Piracicaba: Esalq, 2009. 477p.

HOAGLIN, D.C. et al. **Understanding and exploratory data analysis**. New York: J Wiley, 1983. 447p.

JUDICE, M.G. et al. Avaliação do coeficiente de variação na experimentação com suínos. **Ciência e Agrotecnologia**, v.23, n.1. p.170-173, 1999. Disponível em: <http://www.editora.ufla.br/site/_adm/upload/revista/23-1-1999_22.pdf>. Acesso em: Mar. 22, 2012.

R DEVELOPMENT CORE TEAM. **R**: a language and enviroviment for statistical computing. Austria: R-Fundation for Statistical Computing, 2009. Available from: <http://www.R-project.org>. Accessed: Sept. 2009.

RAMALHO, M.A.P. et al. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2012. p.305.

RESENDE, M.D.V.; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**. v.37, n.3, p.182-194, 2007. Available from: <http://www.revistas.ufg.br/index.php/pat/article/view/1867/1773>. Accessed: Mar. 22, 2012.

SCAPIM, C.A. et al. Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.5, p.683-686, 1995. Available from: <http://webnotes.sct.embrapa.br/pab/pab.nsf/FrAnual>. Accessed: Mar. 22, 2012.

SPIEGEL, M.R. **Estatística**. 3. ed. São Paulo: Markon Books, 1993. 643p.

STEEL, R.G.D. et al. **Principles and procedures of statistics**: a biometrical approach. 3.ed. New York: McGraw-Hill Book, 1996. 672p.

STORCK, L.; LOPES.S.J. **Experimentação II**. 2.ed. Santa Maria: editora ufsm, 1998. 217p.