Canonical correlations in agricultural research: Method of interpretation used leads to greater reliability of results

Maria Inês Diel[1], Alessandro Dal'Col Lúcio[1], Darlei Michalski Lambrecht[1], Marcos Vinícius Marques Pinheiro[2], Bruno Giacomini Sari[1], Tiago Olivoto[1], Oscar Valeriano Sánchez Valera[3], Patrícia Jesus de Melo[1], Francieli de Lima Tartaglia[1], André Luis Tischler[1], Denise Schmidt[4]

[1] Federal University of Santa Maria, Department of Plant Science, Santa Maria, Rio Grande do Sul, Brazil;

[2]Graduate Program in Agriculture and Environment, State University of Maranhão, São Luís, Maranhão, Brazil;

[3]Zongolica Superior Technological Institute, Zongolica, Veracruz. México;

[4]Department of Agronomic and Environmental Sciences, Federal University of Santa Maria, Frederico Westphalen, Brazil.

Abstract

Canonical correlations analyzes are being used in the agrarian sciences and constitute an important tool in the interpretation of results. This analysis is performed by complicated mathematical equations and it is only possible to use it thanks to the development of computational software, which allows different interpretations of results, and it is up to the researcher to choose according to his knowledge. Canonical correlations can be interpreted using canonical weights, canonical loadings, or canonical cross-loadings. In Brazil, most of the works that use these analyzes interpret the canonical weights. Therefore, this study aims to show, through an analysis of canonical correlations, the best way to interpret the results, so that they are presented in the most reliable way possible. Data from an experiment with two cultivars of biquinho pepper seeded in 5 light spectrums were performed. The variables were root length and volume, plant height, number of leaves, fresh shoot and root mass, shoot dry mass. Two groups of variables were organized, the multicollinearity was determined through the condition number and the inflation factor of the variance. Canonical correlations analysis was carried out, and weights, loadings, and canonical cross-loadings were estimated for the interpretation of the results. After the interpretations, it was defined that the canonical cross-loadings should be preferred for the interpretation of the canonical correlations. Weights or canonical coefficients provide dubious results of relationships between groups of characters and should be avoided.

Keywords: Canonical weights; Canonical loadings; Cross-loadings; Multicollinearity; Methods.

Introduction

Canonical correlation analysis is the most generalized technique of all statistical procedures to determine relationship and interdependence between two or more variables (Basu and Mandal 2010;

Cocozzelli 1990). It is used to determine the relations between groups of variables (Cruz, Regazzi, and Carneiro 2012), that is, the variation in a group of dependent variables analyzed in terms of a group of independent variables (Cocozzelli 1990), and was developed by Hotelling, (1936). This type of analysis allows us to obtain canonical statistical variables, which are linear combinations of two groups of variables (X and Y), and the information contained in the parameters should be focused on the correlation between these new variables (Hair Júnior et al., 2009).

Canonical correlations are determined from the correlation between a linear combination of two groups of variables (Hotelling 1936). Thus, a canonical correlation is one that maximizes the relationship between the linear combinations X and Y with: $X' = [x_1 x_2 ... x_p]$, being the vector of the measures of p characters that constitute the group I and $Y' = [y_1 y_2 ... y_q]$ is the vector of measures of q characters that constitute group II. The first canonical correlation is one that maximizes the relationship between X1 and Y1. The functions X1 and Y1 form the first canonical pair associated with that canonical correlation (Cruz et al. 2012), the highest possible among all these linear combinations. Next, a new pair of linear combinations X2 and Y2. The process continues until a subsequent pair of linear combinations no longer produces a significant correlation (Basu and Mandal 2010), that is until all variation is explained in both sets of variables (Cocozzelli 1990).

This type of analysis is widely used as indirect selection methods for the genetic improvement of plants. However, some precautions should be taken when performing the analyzes, such as the presence of multicollinearity, common in correlations and regression analyses (Blalock 1963; Montgomery and Peck 1982), because when there is the presence of high multicollinearity, errors occur and can produce biased estimates of the effects for the group of variables under study (Montgomery and Peck 1982). To avoid errors, the multicollinearity diagnosis must be performed within each group of variables (Blalock 1963; Cruz et al. 2012).

Many studies have been carried out with canonical correlations analyze, in several areas. Kazmi et al., (2017) calculated the canonical correlation between the groups of metabolic and phenotypic variables for the performance of tomato seeds. Rigão et al., (2009) evaluated through canonical correlation characters of planted tubers, related to those harvested, for the early selection of potato clones.

The interpretation of canonical correlation analysis can be performed in different ways: canonical weights or standardized coefficients, canonical loadings, and canonical cross-loadings (Hair Júnior et al., 1998), and the researcher generally uses the measure that his statistical program used for analysis offers. The canonical coefficients were used for the interpretation by (Alves, Filho, et al. 2017; Alves, Cargnelutti Filho, and Burin 2017; Brum et al. 2011; Rigão et al. 2009). Canonical weights, canonical loadings, and canonical cross-loadings were used for interpretation by (Protasio et al. 2014, 2012).

The interpretation of canonical correlations by canonical weights or standardized coefficients is performed by its magnitude, that is, the greater the magnitude of a certain variable, the greater its contribution, and when they have opposing signs, they exhibit an inverse relationship to each other and will have a direct relationship when the signals have the same signal. Canonical weights are subject to high variability are sensitive to multicollinearity (Hair Júnior et al., 2009).

When interpretation is performed by canonical loadings they measure the simple linear correlation between an original observed variable in the dependent or independent set and the canonical statistical variable of the set; and the higher the coefficient, the more important is the variable for the canonical statistical variable (Hair Júnior et al., 1998). Canonical loadings have advantages over canonical weights, such as lower standard error and greater stability in replicate samples (Meredith 1964). The use of canonical cross-loadings consists of the correlation between the observed dependent variable and the independent canonical statistical variable, and vice versa, providing a direct measure of the relations of dependent-independent variables by eliminating an intermediate step involved in conventional loadings (Hair Júnior et al., 2009).

In Brazil, one of the most used programs for the analysis of data in experimental statistics and quantitative genetics is Genes, this program created at the Federal University of Viçosa presents as an output for the analysis of canonical correlations only the canonical weights (or canonical coefficients) and the canonical loadings (Cruz 2013). The software R, which is one of the most widely used free software in the world for data analysis has implemented some packages that allow the researcher to choose according to their knowledge, what kind of interpretation to do. The package CCA (González et al. 2008) e and the package yacca (Butts 2018) are packages that allow outputs of results with three measures of evaluation of canonical correlations.

Based on the issues discussed above, this study aims to obtain canonical weights, canonical loadings, and canonical cross-loadings, detailing the interpretation, advantages, and disadvantages in each method to answer the best method to interpret the results of studies that use analysis of canonical correlations.

Material and Methods

Local, plant material and culture conditions

For the evaluation, we used data from an experiment conducted at the Federal University of Santa Maria - RS, Frederico Westphalen *campus*. An experiment, with two cultivars of biquinho pepper seeded in 5 light spectra was carried out. For this, seeds of the cultivars of biquinho pepper were used (*Capsicum chinense*) BRS Moema and Airetama Biquinho of brand ISLA®, which were pre-germinated on Germitest® paper in transparent gerbox, kept in a growth chamber at a constant temperature of 25 °C, photoperiod of 12h/12h with an irradiance of 40 µmol m^{-2} s^{-1}, and when 50% of seeds presented protrusion of the root were transplanted in plastic cups of white color (0.3 cm^3) filled with commercial substrate Carolina®.

After emergence and until the beginning of flowering [76 days after emergence (DAE)] the seedlings were kept in a growth room with a constant temperature of 25 ± 2 ºC, photoperiod of 16h/8h, and under different amounts and spectral qualities of light. For this, it was used LED lamps (TEC LED®) white, red, blue, red/blue color (60/40% ratio), and fluorescent lamps (40 W Osram, Brazil). Initially, up to 62 DAE the amount of light was 40 µmol m^{-2} s^{-1}, later this was increased to 80 µmol m^{-2} s^{-1}.

At 76 DAE (beginning of flowering, 50% of flowering plants) the plants were transplanted into black pots (cm^3) filled with soil plus substrate and kept in a greenhouse. Irrigation was performed manually according to the needs of the crop. Plant nutrition was carried out via fertigation.

Experimental design

The experimental design was a randomized block design, in a 2x5 factorial scheme, with two cultivars of biquinho pepper (BRS Moema and Airetama Biquinho) and seedlings from five different light conditions (white LEDs, red LEDs, blue LEDs, red/blue LEDs and fluorescent lamps light), totaling ten treatments with ten replicates per treatment, where each replicate is composed of one plant, totaling 10 plants per treatment.

Variables analyzed

From the emergence to the beginning of flowering (76 DAS) the height of the plant was evaluated, a number of leaves. The number of leaves, height (cm) were evaluated from the emission of the first expanded leaf to the first bifurcation in the main stem of the plant.

At the 76 DAS (before transplant) was evaluated fresh and dry mass (g) of shoot and root, root volume, root length, leaf area.

The fresh mass was calculated shortly after the removal of the substrate from the roots to avoid dehydration of the plant. For the dry mass, the plants were removed from the substrate and then placed in paper bags to be kept in a drying oven. The root volume was calculated after the fresh mass was calculated by displacing the water in a graduated cylinder after submerging the root. The root length was evaluated by measuring the length of the main root (root pivoting).

Statistical analysis

Two groups were defined to verify the relation and interdependence between them. For this, the multicollinearity diagnosis was performed on the linear correlation matrix in each of the groups of variables (X, Y) using the condition number analysis $NC = \frac{\lambda \max}{\lambda \min}$ representing the ratio between the largest and the smallest eigenvalue of the correlation matrix and by the variance inflation factor $VIF = \frac{1}{1 - R_j^2}$, where R_j^2 is the determination coefficient.

With each group of variables presenting low multicollinearity, that is, NC less than 100 and VIF less than 10, we proceeded with canonical correlation analysis. The groups formed were root variables such as root length (CR) and root volume (VR) and a plant growth group: plant height (ALT), leaf number (NF), fresh shoot mass MFPA), fresh root mass (MFR) and dry shoot mass (MSPA).

The number of canonical correlations to be estimated is equal to the number of characters of the smallest group (p or q) and the magnitude decreases in the order in which they are estimated. Thus, a canonical correlation is one that maximizes the relationship between the linear combinations X and Y with: $X' = [x_1 x_2 ... x_p]$, being the vector of the measures of p characters that constitute the group I and $Y' = [y_1 y_2 ... y_q]$ is the vector of measures of q characters that constitute group II (Cruz et al. 2012). Thus, two canonical correlations were estimated.

The estimation of the maximum correlation between linear combinations of characters of groups I and II, where X1 and Y1 are the linear combinations of the characters of groups I and II respectively: $X_1 = a_1x_1 + a_2x_2 + ... + a_px_p$ and $Y_1 = b_1y_1 + b_2y_2 + ... + b_qy_q$ on what: $a' = [a_1a_2...a_p]$ vector 1xp of weights of the group I character and $b' = [b_1b_2...b_q]$ = vector 1xq of II characters weights. In this way, the first canonical correlation will be that which maximizes the relation between X1 and Y1. The functions X1 and Y1 form the first canonical pair associated with that canonical correlation expressed by: $r_1 = \frac{C\hat{o}v(X_1, Y_1)}{\sqrt{\hat{V}(X_1).\hat{V}(Y_1)}}$, being $C\hat{o}v(X_1, Y_1) = a'S_{12}b$, $\hat{V}(X_1) = a'S_{11}a$ and $\hat{V}(Y_1) = b'S_{22}b$ where: S_{11}= matrix pxq of covariance between the characters of group I; S_{22}= matrix qxq of covariance between the characters of group II; S_{12}= matrix pxq of covariance between the characters of group I and II (Cruz et al. 2012). The coefficients or canonical weights are the weights that each variable received for the calculation of the linear combinations, the canonical loadings, which are the correlations between the original variables and their respective canonical statistical variables, and the canonical cross-loadings that represent the correlation between an original variable of a given group and the canonical statistical variable of the other group (Hair Júnior et al., 2009).

The canonical R^2 was calculated by squaring the estimated canonical correlation and represents the correlation between the actual dependent variable and the predicted value, representing the amount of variance in the dependent variable explained by the regression function of the independent variables (Hair Júnior et al., 2009). The significance test was performed through a series of canonical correlations (sequentially) against the null hypothesis that the coefficient tested and all of the following coefficients are zero (Butts 2018). The Wilks' multivariate Lambda significance test (F distribution approximation) was used to evaluate the significance of the canonical roots together.

Statistical analyses were performed at 5% significance. with the help of the yacca (Butts 2018), pracma (Borchers 2018), and faraway packages available in program R.

Results

The multicollinearity diagnosis determined in the correlation matrix in each group of variables resulted in VIF less than 10 and NC less than 100, indicating weak multicollinearity. The results of multicollinearity are important from estimating reliable canonical correlations that reflect the real influence of one group on the other and vice versa. For this study, the results of VIF for the group I was 1.08647 for the two variables, and the NC was 1.78595. For group II, FIV was 3.08 for ALT, 3.52 for NF, 7.40 for MFPA, 1.44 for MFRA, and 10.03 for MSPA. The NC was 54.42.

The first canonical correlation was high and significant and revealed the existence of interdependence among the groups. The canonical R^2 found did not present a very high result, ie, the amount of variance explained (56.25%) between the canonical statistical variables independent and dependent on function 1 was only 56.25% for the groups of characteristics analyzed (Table 1).

Table 01. Canonical correlations, canonical R², and multivariate test of significance.

Function	Canonical correlations	Canonical R²	Total Variance Explained	Degrees of freedom	χ^2	p-valor
1	0.75	0.5625	0.34	10	83.76	9.43E-14
2	0.17	0.0289	0.11	4	3.04	0.54

By interpreting the results by canonical weights (or standardized coefficients) for the first function, we can see that the variables MFRA and MSPA contribute to increasing the variable VR or vice versa, whereas root length (CR) has little or almost none contribution to the independent variables. The second function presented a non-significant canonical correlation and, therefore, does not require interpretation (Table 02).

Table 02. Canonical weights, canonical loadings, and canonical cross-loadings for the two groups of growth variables of biquinho pepper under light spectra.

Variables	**Canonical weights**		**Canonical loadings**		**Canonical cross-loadings**	
	Group I					
	1º	2º	1º	2º	1º	2º
CR*	0.077	-0.432	0.447	-0.895	0.3378	-0.1590
VR	1.608	0.803	0.984	0.176	0.7448	0.0313
	Group II					
ALT	-0.1057	-0.4048	0.0213	-0.1575	0.0161	-0.0280
NF	0.0004	-0.0241	0.1378	0.2346	0.1042	0.0417
MFPA	-0.1528	1.4052	0.0960	0.4651	0.0727	0.0827
MFRA	2.1143	0.1915	0.9610	0.0299	0.7271	0.0053
MSPA	1.2681	-5.2556	0.2716	0.1671	0.2055	0.0297

*CR: root length; VR: root volume; ALT: plant height; NF: leaf number; MFPA: fresh shoot mass; MFRA: fresh root mass; MSPA: dry shoot mass.

The interpretation via canonical loadings must be performed considering the correlation between the variables and their canonical variables. Analyzing the canonical loadings for the canonical variable in group I (dependent) it is observed that the root volume (RV) presented the highest value. For group II (independent), the MFRA variable presented a higher value. The canonical loadings reveal that CR is positively correlated with VR. Besides, the increase in these two variables increases the MFRA, the MSPA, and vice versa. The relationship between ALT, NF, and MFPA was not sufficiently explained by the canonical correlation analysis because the values of the canonical loadings were low.

The results for canonical cross-loadings are similar to canonical loadings. This measure allows for more direct identification of the relationship between the dependent variables - independent. The results show a strong relationship between the MFPA with RV and CR and on the contrary, showing that the increase of VR and CR increases the MFRA and MSPA.

The interpretation of the canonical loadings and the canonical cross-loadings showed the same result, however, the magnitude of the relations decreased for the canonical cross-loadings. Regardless of the interpretation used, it is evidenced that the groups have linear relations and that the characters CR and VR can be used to increase the MFPA, for example. This characteristic is desirable from a photosynthetic point of view because the larger the aerial part in the pepper plants the greater the photosynthetic area in the plant, besides producing more shoots, important for the production of fruits.

Discussion

In the analysis of canonical correlations, the diagnosis of multicollinearity is essential so that the correlation estimates are not biased. The presence of multicollinearity is present in regression and correlation analyzes and this affects the ability to estimate reliable regression or correlation coefficients (Montgomery and Peck, 1982).

The amount of explained variance calculated by the coefficient of canonical correlation squared, called canonical R^2, presented a result of 56.25%, which indicates that although the canonical correlation was significant, it has a little more than 56% explanation of the variance of the dependent variables, and this measure must be taken into account in this type of analysis since the canonical correlation is calculated in such a way as to maximize the relations between them, while R^2 maximizes the explanation or variance between them (Hair Júnior et al., 1998). Akbaş and Takma, (2005) found results that suggest 33.3% of the variance in the Y variables is accounted for by the X variables, and according to the authors, this value indicates that the first correlation has a high significance. Marubayashi Hidalgo et al., (2014), found R^2 of 0.658 and indicated that two sets of variables are highly correlated in all the lines tested revealing an association between the groups of variables.

The interpretation of variables by canonical weights or standardized canonical coefficients allows us to identify the importance of each variable in the calculation of canonical variables, that is, the canonical coefficients give the contribution of each variable in the analysis (Akbaş and Takma, 2005); the canonical weights are similar to the beta coefficients of multiple regression (Protasio et al., 2014).

the canonical weights are similar to the beta coefficients of multiple regression. In this case, the variable CR had low weight computed, this can be attributed to the instability of the canonical coefficients discussed in some studies due to the small sample size or presence of multicollinearity in the data (Akbaş and Takma, 2005; Hair Júnior et al., 2009). In the results of this study, the presence of multicollinearity is not a problem, since the diagnosis of multicollinearity in each group of variables was weak.

When interpreting canonical weights, a measure of shared variance other than square canonical correlations must be used to overcome the uncertainties with this interpretation, and the use of the redundancy index is indicated (Akbaş and Takma, 2005; Hair Júnior et al., 2009). According to Meredith, (1964), if the variables within each set are moderately interrelated, the interpretation of the canonical variables by the weights is practically null, without correcting these correlations.

The interpretation of the canonical correlations using the canonical loadings presented differences of those interpreted by the canonical weights. In this case, the variable CR that had low canonical weight had a high canonical charge. When interpreting the canonical loadings, we can identify the correlation between the variables and their respective canonical variables (Marubayashi Hidalgo et al., 2014), that is, it reflects the shared variance between the observed variable and the linear combination (Hair Júnior et al., 2009).

By the interpretation via canonical cross-loadings the result was similar to the one obtained by the canonical loadings, these indicate the correlation between the variables and the opposite canonical variable (Marubayashi Hidalgo et al., 2014) and can provide a more direct measure of the relationships between dependent and independent variables (Hair Júnior et al., 2009). Similar results between canonical loadings and canonical cross-loadings were found in other works that used these two forms of interpretation (Marubayashi Hidalgo et al., 2014; Protasio et al., 2014) and indicate that the influencing characteristics would be the same using interpretation by canonical loadings or canonical cross-loadings.

In a work carried out in 1975 the problems associated with interpretations of canonical correlations were already discussed, where Lambert and Durand, (1975) reported the dangers when interpreting the analysis of correlations by canonical weights since in showing the contribution of each original variable to the canonical score, they do not show the correlations between the variables.

The weights are used in a way that maximizes the canonical correlations (Meredith, 1964), and if the large correlation is improved, although the correlation with the variables is low, multicollinearity problems may mask the results attributed to canonical weights, and the variables may have little or no correlation between them (Lambert and Durand, 1975). These same authors affirm that the canonical weights are very inadequate and indicate relations between non-existent variables. Most often the interpretation by canonical weights does not provide enough information to fully interpret a canonical effect, because these coefficients still cannot articulate clearly the contributions to the achieved effect (Nimon et al. 2010).

In canonical loadings, the problems related to multicollinearity that affect interpretation by canonical weights are diminished because they show the simple relation with the canonical scores, ignoring the correlation between other variables and the scores, that is, they are not affected by multicollinearity (Lambert and Durand, 1975; Nimon et al., 2010). A limitation of the loadings is that even if they are stable they cannot provide a direct measure of the variation between the variables (Lambert and Durand, 1975).

The canonical charges are preferable to the use of canonical weights, however, the R^2 must be calculated to determine how much of the variance of the observed variable is shared with the canonical scores of this linear combination (Lambert and Durand, 1975). In the case of the present work, for the variable MFRA that presented canonical loadings of 0.9610, we conclude that it presents a high correlation, and raising it to the square, the canonical loadings obtained 92% of the variance in the shared variable. However, the high canonical loadings for the variable CR (0,447) present a low variance of the shared variable with only 19.9%.

Canonical cross-loadings are more reliable than the use of canonical weights or canonical loadings as they provide a more direct measure of the relationships between variables by eliminating an intermediate step involved in canonical loadings (Hair et al., 2009; Lambert and Durand, 1975)

As we have seen here, the interpretations of canonical correlations require great care and attention, since there are many cases in which the results may not be reliable, simply because of the lack of care with the analyzes and their interpretation, besides can be replicating errors in the scientific environment for the lack of knowledge and misuse of this type of analysis. As said by the author's Lambert and Durand, (1975), the aim is not to discourage the use of these analyses, but to make them more accurate and that the researchers use them to make their research more effective, reducing errors of interpretation.

Many works have been published where the interpretation of the analyzes is carried out via canonical weights (coefficients) as in (Alves et al., 2017b, 2016; Bandinelli et al., 2017; Rigão et al., 2009), e even using a standardized structure, they can not fully report the canonical effects (Nimon et al., 2010). Thus, the use of canonical loadings and canonical cross-loadings for the interpretation of canonical correlations is preferable (Takada et al., 1984).

But what method of interpretation should I use? The researcher must use that measure that is available in the computational program for the analysis of data that he dominates, since the analysis of canonical correlation without the aid of software is laborious and can be unsustainable (Nimon et al., 2010), nevertheless one must avoid making interpretations by the canonical weights since they do not present us a measure of relations between the variables (Hair Júnior et al., 2009). The use of canonical cross-loadings is the most indicated, and by the computational advance that facilitates the determination of this measure, it becomes possible to use this measure. When calculating cross-loadings is not possible, the option to be chosen is canonical loadings that are more reliable than canonical weights (Hair et al., 2009; Lambert and Durand, 1975).

Conclusion

Interpretations of canonical correlation analyzes should be made using canonical cross-loadings. Weights or canonical coefficients provide dubious results of relationships between groups of characters and should be avoided.

Acknowledgments

We thank the Coordination for the Improvement of Higher Education Personnel (CAPES) and National Council for Scientific and Technological Development (CNPq) for granting the scholarships to the researchers.

References

Akbaş, Y., and C.¸ Takma. 2005. "Canonical Correlation Analysis for Studying the Relationship between Egg Production Traits and Body Weight, Egg Weight, and Age at Sexual Maturity in Layers." *Czech Journal of Animal Science* 50(4):163–68.

Alves, B. M., A. Cargnelutti Filho, and C. Burin. 2017. "Multicollinearity in Canonical Correlation Analysis in Maize." *Genetics and Molecular Research*: *GMR* 16(1).

Alves, B. M., Alberto Cargnelutti Filho, Cláudia Burin, and Marcos Toebe. 2017. "Linear Associations among Phenological, Morphological, Productive, and Energetic-Nutritional Traits in Corn." *Pesquisa Agropecuaria Brasileira* 52(1):26–35.

Alves, Bruna Mendonça, Alberto Cargnelutti Filho, Marcos Toebe, and Cláudia Burin. 2016. "Linear Relations among Phenological, Morphological, Productive and Protein-Nutritional Traits in Early Maturing and Super-Early Maturing Maize Genotypes." *Journal of Cereal Science* 70:229–39.

Bandinelli, Maurício G., Dilson A. Bisognin, Lindolfo Storck, Francisco S. Gnocato, Paula Kielse, Claudinei Ascoli, Maurício G. Bandinelli, Dilson A. Bisognin, Lindolfo Storck, Francisco S. Gnocato, Paula Kielse, and Claudinei Ascoli. 2017. "Correlação de Caracteres de Planta e Tubérculo Nas Primeiras Gerações de Seleção de Batata." *Horticultura Brasileira* 35(1):82–88.

Basu, A., and A. Mandal. 2010. "Canonical Correlation." p. 52–57 in *International Encyclopedia of Education*.

Blalock, H. M. 1963. "Correlated Independent Variables: The Problem of Multicollinearity." *Social Forces* 42(2):233–37.

Borchers, Hans W. 2018. "Pracma: Practical Numerical Math Functions."

Brandelero, Fernanda Daniela, Betania Brum, Lindolfo Storck, Jessica Cardoso, Talita Slota Kutz, and Thiago de Oliveira Vargas. 2016. "Plant Characters of Broccoli Determinants of Head Production." *Ciência Rural* 46(6):963–69.

Brum, B., S. J. Lopes, L. Storck, A. D. Lucio, P. H. de Oliveira, and M. Milani. 2011. "Canonical Correlation between Variables Seed, Seedling, Plant and Grain Yield in Castor Bean." *Ciencia Rural* 41(3):404–11.

Butts, Carter T. 2018. "Yacca: Yet Another Canonical Correlation Analysis Package." R package version 1.1.1.

Cocozzelli, Carmelo. 1990. "Understanding Canonical Correlation Analysis : Describing Underlying Relations Between Sets of Variables." *Journal of Social Service Research* 13(4):19–42.

Cruz, Cosme Damião. 2013. "GENES: Software Para Análise de Dados Em Estatística Experimental e Em Genética Quantitativa." *Acta Scientiarum. Agronomy* 35(3):271–76.

Cruz, Cosme Damião, Adair José Regazzi, and Pedro Crescencio Souza Carneiro. 2012. *Modelos Biométricos Aplicados Ao Melhoramento Genético*. 4th ed. Viçosa: Editora UFV.

González, Ignacio, Sébastien Déjean, Pascal G. P. Martin, and Alain Baccini. 2008. "CCA: An R Package to Extend Canonical Correlation Analysis." *Journal of Statistical Software* 23(12):1–14.

Hair, J. F., W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham. 2009. *Análise Multivariada de Dados*. 6th ed. Porto Alegre: Bookman.

Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson. 1998. *Multivariate Data Analysis*. v. 1.

Hotelling, Harold. 1936. "Relations Between Two Sets of Variates." *Biometrika* 28(3/4):321.

Kazmi, Rashid H., Leo A. J. Willems, Ronny V. L. Joosen, Noorullah Khan, Wilco Ligterink, and Henk W. M. Hilhorst. 2017. "Metabolomic Analysis of Tomato Seed Germination." *Metabolomics* 13(12):145.

Lambert, Zarrel V., and Richard M. Durand. 1975. "Some Precautions in Using Canonical Analysis." *Journal of Marketing Research* 12(4):468–75.

Marubayashi Hidalgo, André, Luciano Pinheiro da Silva, Rodrigo Reis Mota, and Elias Nunes Martins. 2014. "Canonical-Correlation Analysis Applied to Selection-Index Methodology in Quails." *Livestock Science* 169(C):35–41.

Meredith, William. 1964. "Canonical Correlations with Fallible Data." *Psychometrika* 29(1).

Montgomery, D. C., and E. A. Peck. 1982. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons, Ltd.

Nimon, Kim, Robin K. Henson, and Michael S. Gates. 2010. "Revisiting Interpretation of Canonical Correlation Analysis: A Tutorial and Demonstration of Canonical Commonality Analysis." *Multivariate Behavioral Research* 45(4):702–24.

Protasio, Thiago de Paula, Rosalvo Maciel Guimarães Neto, João de Deus Pereira de Santana, José Benedito Guimarães Júnior, and Paulo Fernando Trugilho. 2014. "Canonical Correlation Analysis of the Characteristics of Charcoal from Qualea Parviflora Mart." *Revista Cerne* 20(1):81–88.

Protásio, Thiago de Paula, Gustavo Henrique Denzin Tonoli, Mário Guimarães, Lina Bufalino, Allan Motta Couto, and Paulo Fernando Trugilho. 2012. "Correlações Canônicas Entre as Características Químicas e Energéticas de Resíduos Lignocelulósicos." *Cerne* 18(3):433–39.

Rigão, Maria Helena, Lindolfo Storck, Dilson Antônio Bisognin, and Sidinei José Lopes. 2009. "Correlação Canônica Entre Caracteres de Tubérculos Para Seleção Precoce de Clones de Batata." *Ciência Rural* 39:2347–53.

Takada, Kenji, Alan A. Lowe, and Vivien K. Freund. 1984. "Canonical Correlations between Masticatory Muscle Orientation and Dentoskeletal Morphology in Children." *Am. J. Orthod.* 86(4):331–41.

Toebe, Marcos, Vanderlei Both, Fabio Rodrigo Thewes, Alberto Cargnelutti Filho, and Auri Brackmann. 2016. "Relações Lineares Entre Caracteres de Frutos de Maçã." *Bragantia* 75(4):396–400.