Determination of factors in pepper genotypes for ornamentation Determinação de fatores em genótipos de pimenta para ornamentação Determinación de factores en genotipos de pimiento para ornamentación

Received: 11/17/2020 | Reviewed: 11/19/2020 | Accept: 11/21/2020 | Published: 11/27/2020

# Maria Eduarda da Silva Guimarães ORCID: https://orcid.org/0000-0001-8207-2355 Universidade Federal de Viçosa, Brasil E-mail: maria.eduarda.ufv@gmail.com Ana Carolina Ribeiro de Oliveira ORCID: https://orcid.org/0000-0001-8262-8667 Universidade Federal de Viçosa, Brasil E-mail: kroll olliveira@hotmail.com Ana Izabella Freire ORCID: https://orcid.org/0000-0002-8442-9183 Universidade Federal de Viçosa, Brasil E-mail: anabellafr1987@yahoo.com.br Ariana Mota Pereira ORCID: https://orcid.org/0000-0003-4033-8156 Universidade Federal de Viçosa, Brasil E-mail: ariana.mota@ufv.br **Dreice Nascimento Goncalves** ORCID: https://orcid.org/0000-0003-3896-6890 Universidade Federal de Viçosa, Brasil E-mail: dreicegoncalves@gmail.com Augusto Soares Lins Pantaleão ORCID: https://orcid.org/0000-0001-8351-3183 Universidade Federal de Viçosa, Brasil E-mail: augustopntl@gmail.com João Romero do Amaral Santos de Carvalho Rocha ORCID: https://orcid.org/0000-0002-0976-0917 Universidade Federal de Viçosa, Brasil E-mail: joaoascrocha@gmail.com

#### Mateus de Paula Gomes

ORCID: https://orcid.org/0000-0001-7783-0799 Universidade Federal de Viçosa, Brasil E-mail: mateusgomes2006@yahoo.com.br Abelardo Barreto de Medonça Neto ORCID: https://orcid.org/0000-0003-0190-6392 Universidade Federal de Viçosa, Brasil E-mail: abemendonca@yahoo.com.br **Rusthon Magno Cortez dos Santos** ORCID: https://orcid.org/0000-0002-6967-2871 Universidade Federal de Viçosa, Brasil E-mail: rusthoncortez@hotmail.com Luciana Gomes Soares ORCID: https://orcid.org/0000-0002-5628-6250 Universidade Federal de Viçosa, Brasil E-mail: luci.gomes.soares@gmail.com **Renata Ranielly Pedroza Cruz** ORCID: https://orcid.org/0000-0002-7189-8941 Universidade Federal de Viçosa, Brasil E-mail: renataranielly426@gmail.com Françoise Dalprá Dariva ORCID: https://orcid.org/0000-0002-4259-5212 Universidade Federal de Vicosa, Brasil E-mail: fran\_dariva@hotmail.com **Fernando Luiz Finger** ORCID: https://orcid.org/0000-0002-4046-9634 Universidade Federal de Viçosa, Brasil E-mail: ffinger@ufv.br

## Abstract

The present work had as main objective to use factor analysis to describe the structure of variability of characteristics considered commercially important in ornamental pepper, aiming to summarize the information contained in such variables in a smaller number of latent variables or factors. For that, 12 quantitative traits were evaluated in 29 pepper genotypes

(Capsicum annuum). Of the factors created, 3 presented practical interpretation, grouping a total of 12 variables into factors related to "fruit quality" (8 variables), "plant size" (2 variables) and "plant architecture" (2 variables).

Keywords: Capsicum annuum; Factor analysis; Morpho-agronomic characterization.

#### Resumo

O presente trabalho teve como principal objetivo utilizar análise de fatores para descrever a estrutura de variabilidade de características consideradas comercialmente importantes em pimenta ornamentais, visando resumir a informação contida em tais variáveis em um número menor de variáveis latentes ou fatores. Para isso 12 características quantitativas foram avaliadas em 29 genótipos de pimenta (*Capsicum annuum*). Dos fatores criados, 3 apresentaram interpretação prática, agrupando um total de 12 variáveis em fatores relacionados a "fruit quality" (8 variáveis), "plant size" (2 variáveis) e a "plant architecture" (2 variáveis).

Palavras-chave: Análise de fatores; Capsicum annuum; Caracterização morfo-agronômica.

## Resumen

El presente trabajo tuvo como objetivo principal utilizar el análisis factorial para describir la estructura de variabilidad de características consideradas comercialmente importantes en pimiento ornamental, con el objetivo de resumir la información contenida en dichas variables en un menor número de variables o factores latentes. Para ello, se evaluaron 12 rasgos cuantitativos en 29 genotipos de pimiento (Capsicum annuum). De los factores creados, 3 presentaron interpretación práctica, agrupando un total de 12 variables en factores relacionados con "calidad del fruto" (8 variables), "tamaño de planta" (2 variables) y "arquitectura de planta" (2 variables).

Palabras-clave: Análisis factorial; Capsicum annuum; Caracterización morfoagronómica.

## **1. Introduction**

Pepper is an economically important crop worldwide. In addition to their importance in food, pharmacology, dentistry and medicine, peppers have great potential for ornamentation. Among pepper species, *Capsicum annuum* is the most used in planting with ornamental purposes, due to its small size and the great variability of fruit shapes and colors (Finger et al., 2012). Despite the great variability that exists, few commercial varieties are

used for ornamentation in the country (Vasconcelos et al., 2012). The ornamentals market lacks novelties and new products add competitiveness to the sector and considerably increase the profit margin (Costa et al., 2019). The study of morphological and agronomic characteristics of cultivated plants is important to understand the genetic divergence of the set of germplasm available for use in a breeding program (Elias et al., 2007).

The study of the morphological characters can be carried out individually or simultaneously. Simultaneous analysis allows to conclude for more than one variable and thus better interpret the relationship between them. Multivariate statistics allows the study of complex phenomena, as it performs the treatment of several variables simultaneously (Johnson & Wichern, 2007). Among the multivariate techniques, factor analysis allows to reduce the dimension of the analyzed variables. It makes it possible to group correlated variables into unobservable factors (latent variables), defined through the correlation between variables. After the identification and interpretation of the factors, the latent variables can be predicted and their values used in later analyzes (Silva et al., 2014).

Given the above, this study aimed to verify the empirical structure of morphoagronomic characteristics of *C. annuum* genotypes, so that correlated variables are grouped into a smaller number of latent (interpretable) variables, reducing the dimensionality of the data set.

## 2. Methodology

#### 2.1 Plant materials

The experiment was carried out between the months of November 2017 and March 2018, in a greenhouse, at the Department of Fitotechnics of the Federal University of Viçosa (DFT / UFV). The city of Viçosa, belonging to the state of Minas Gerais, is located at 650 m altitude, latitude 20°45"47" South and longitude 42°49"13" west. According to the Köppen classification, the city's climate is characterized by the cold-dry season between April and August and the hot-rainy season between September and March, with an annual average of 1,341 mm of rainfall and 21,6°C e 14,0°C maximum and minimum temperatures, respectively.

29 genotypes of the species Capsicum annuum were evaluated, selected because they have potential for ornamentation: reduced size; fruits with different shapes; different colors, at different stages of maturation; different positions and quantity of fruits produced; ease of

cultivation; fruit and leaf durability and the ability to grow in containers as a perennial plant (Neitzke et al., 2010). The experiment was installed in a greenhouse in a completely randomized design (DIC), with 29 treatments (genotypes) and five replications, with the experimental unit consisting of one plant per pot.

Sowing was carried out in November due to the requirement of the species for high temperatures. The pepper plant requires high temperatures throughout the cycle, the ideal average monthly temperatures are between 21 ° C and 30 ° C (Rêgo et al., 2011). With the aid of a datalogger, temperature data ranging from 24 °C to 32 °C were collected during the experiment.

The seeds were sown in polystyrene trays, with 200 cells, containing commercial substrate. Two seeds were distributed per cell to ensure that there was no idle cell. The trays were covered with a black cloth to speed up the germination process and kept suspended to favor the natural pruning of the root system by air (Rêgo et al., 2011).

In January, seedlings with four pairs of definitive leaves were transplanted into 800 ml pots and thinning was performed after one week. Manual irrigations were performed three times a day, maintaining the humidity level, being sufficient to start draining at the bottom of the tray (Rêgo et al., 2011). Fertilization and pest and weed control were carried out whenever necessary. The 29 genotypes evaluated are shown in Table 1.

Number of genotypes	Common name			
1	NuMex Big Jim 1			
2	NuMex Big Jim 2			
3	NuMex Conquistador 1			
4	NuMex Conquistador 2			
5	NuMex Espanola Improved 1			
6	NuMex Espanola Improved 2			
7	NuMex Joe E Parker 1			
8	NuMex Joe E Parker 2			
9	NuMex Mirasol 1			
10	NuMex Mirasol 2			
11	NuMex New México 6-4 1			
12	NuMex New México 6-4 2			
13	NuMex Sandia 1			
14	NuMex Sandia 2			
15	NuMex Sweet 1			
16	BGH 1039			
17	BGH 7073			
18	MG 302			
19	Calypso			

**Table 1** – List of the 29 evaluated genotypes of *Capsicum annuum*.

•	and Development, v. 9, n. 11, e58191110348, 2020 5-3409   DOI: http://dx.doi.org/10.33448/rsd-v9i11.10348		
20	Cayenne		
21	Guaraci Cumari do Pará		
22	Jamaica Red		
23	Jamaica Yellow		
24	Peter		
25	Pimenta Doce Italiana		
26	Pimenta Doce Comprida		
27	Pimenta Amarela Comprida		
28	Picante para Vaso		
29	Vulcão		

Source: Authors.

Table 1 identifies the 29 Capsicum annuum genotypes evaluated in the experiment, identified by number and common name. The first 15 genotypes are from New Mexico, the next 3 from the UFV germplasm bank and the remaining 11 are commercial varieties.

## 2.2 Morpho-agronomic characterization

For the morpho-agronomic characterization of the genotypes, the ornamental potential associated with characters of interest for consumption was taken into account, considering that pepper plants can have dual purposes. The descriptors established by the International Plant Genetic Resources Institute for the genus *Capsicum* were taken as a base (IPGRI, 1995).

In March, the fruits were collected and the necessary measurements and evaluations were carried out. For the development of the research, the quantitative method was used, in which the collection of quantitative data was carried out by means of point measurements of quantities using metrology. The collected numbers, with their respective units, generated data sets that were analyzed using mathematical techniques (Pereira et al., 2018).

12 characteristics of plant, flower and fruit were evaluated: FW-fruit weight (grams); FL-fruit length (mm); FD-fruit diameter (mm); TP-thickness of the pericarp (mm); NS/FRnumber of seeds/fruit; FM- fresh matter (grams); DM – Dry matter (grams); CORD-diameter of the corolla (mm); PH-Plant Height (cm); SL-stem length (cm); CAD-canopy diameter (cm); SD-stem diameter (mm). The dimensions measurements were taken using a caliper. The weight data were taken on an analytical balance and the values referring to the quantity, such as number of fruits, were obtained by counting. Fruit measurements were taken in the laboratory.

#### **2.3 Statistical analyses**

The factorial model adopted for an observable  $X_i$  variable, with mean  $\mu_i$  can be represented by the (Johnson & Wichern, 2007):

 $X_i - \mu_i = 1_{i1}F_1 + 1_{i2}F_2 + \ldots + 1_{im}F_m + \epsilon_i$ 

Where: **X**<sub>i</sub> represents the observable variables with mean  $\mu_i$ , i=1,2,...,p e m  $\leq$  p, and m  $\leq$ p, in which p is the number of observable variables; The elements  $\mathbf{1}_{ij}$  refer to the factor loads associated with the i<sup>th</sup> variable **X**<sub>i</sub> and the j<sup>th</sup> common factor;  $F_{j,i}$ , j = 1,2, ... m.  $F_j$  corresponds to common unobservable latent factors;  $\boldsymbol{\epsilon}_i$  are the random errors associated with the i<sup>th</sup> variable **X**<sub>i</sub>.

To measure the adequacy of the analysis, the Kaiser-Meyer-Olkin (KMO) criterion and Bartlett's sphericity test were used (Ferreira, 2011). The number of factors was defined using two criteria. The first criterion was the analysis of the proportion of the total explained variance. An explanation percentage of 70% of the total variability was considered, which according to Ferreira (2011) is sufficient to satisfactorily reduce the data.

The second criterion used was the eigenvalue criterion or Kaiser criterion in which the number of factors will be equal to the number of eigenvalues greater than or equal to 1 (Kaiser, 1958). In addition to the criteria used, the choice of the number of factors (m) took into account the interpretability of the factors and the principle of parsimony (Mingoti, 2005).

The relationship between variables and common factors was made through loadings  $(1_{a})$ , or factorial loads, which represent the correlation between each variable and the respective factors. The loadings values, as well as the simple correlation, vary between -1 and 1 and, the higher the factor load (in module) the greater the correlation between the variable and the respective factor. Making it possible to name the factors based on those variables that are most related to them (Teixeira et al., 2015).

For a better interpretation of the distribution of variables in the respective factors, varimax rotation was used. To evaluate the proportion of each variable explained by the factor to which it belongs and the proportion explained by the random error, the values of commonality were calculated. After identifying and interpreting the factors, the values of the scores for each factor were calculated. Through the scores it is possible to predict the values

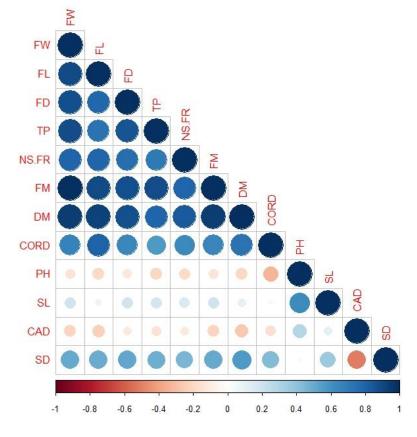
for each sample unit and use these latent variables (factors) in later analyzes (Teixeira et al., 2015).

## 3. Results and Discussion

In order to assess the association structure between the variables, the correlation matrix was calculated. The existence of a correlation between the variables in the data set is of great importance for factor analysis, since this technique aims to identify relationships between variables. This relationship only exists in the presence of a correlation between them (Corrar et al., 2009).

To improve the visualization and interpretation of the correlations between the variables, a correlogram was created represented in Figure 1.

**Figure 1 -** Graph of the sample correlations of the 12 variables referring to the 29 genotypes of *Capsicum annuum* evaluated. FW-fruit weight (grams); FL-fruit length (mm); FD-fruit diameter (mm); TP-thickness of the pericarp (mm); NS/FR-number of seeds/fruit; FM- fresh matter (grams); DM – Dry matter (grams); CORD-diameter of the corolla (mm); PH-Plant Height (cm); SL-stem length (cm); CAD-Canopy diameter (cm); SD-stem diameter (mm).



Source: Authors.

Figure 1 shows the correlogram graph represents the bivariate correlations. It is observed that the positive correlations are blue, with stronger tones for the higher correlations and negative correlations are in pink, with stronger tones for higher correlations. (Silva et al., 2014). It is observed that there are strong and positive correlations between the variables FW, FL, FD, TP, NS/FR, FM, DM e CORD, and between variables PH e SL. In addition to a very strong and negative correlation between the variables CAD e SD. As most of the variables showed high correlations with each other, it can be assumed that the use of factor analysis will be promising.

According to the KMO index (0.74), considered satifastory by criteria of Pallant (2007), which suggests 0.60 as a reasonable value, and with Bartlett's sphericity test, which showed statistical significance ( $\rho < 0.01$ ), it was found that the data are adequate for factor analysis. To determine the number of factors to be used in the factor analysis, it is necessary to calculate the eigenvalues of the correlation matrix, as shown in the Table 2.

Eigenvalue	Eigenvalue (%)	Accumulated variance (%)
7.14	59.48	59.48
1.78	14.85	74.33
1.22	10.13	84.46
0.53	4.42	88.88
0.42	3.47	92.35
0.30	2.51	94.86
0.26	2.20	97.06
0.16	1.31	98.37
0.11	0.96	99.33
0.053	0.44	99.77
0.027	0.229	99.99
0.0002	0.01	100

**Table 2** – Eigenvalue estimates by principal components analysis and the variance proportion explained by them.

Source: Authors.

Table 2 shows the eigenvalues and the accumulated variance of the 12 main components obtained from the genetic correlation matrix, corresponding to the 12 characteristics evaluated. Eigenvalues are numbers that reflect the importance of the factor,

and divided by the sum of all eigenvalues indicate the proportion of the total variability of the data that is explained by the factor (Silva et al., 2014).

Only the first three components are associated with eigenvalues that are greater than one. Thus, according to Kaiser's criteria (Kaiser, 1958), the data can be condensed into three factors. The accumulated variance of the first three components was greater than 80%, indicating that these three factors are sufficient because they represent 84.46% of all variability. Mingoti (2005) suggests that the number of factors to be retained should reflect a value greater than 70% of the original data variability.

The Varimax rotation method was used to give factors greater potential for interpretability, making the factorial solution simpler and more meaningful (Johnson & Wichern, 2007). After varimax rotation, it was observed the formation of 3 factors. The variables were grouped into each factor according to the factor loads, as can be seen in Table 3, where the values of communalities are also presented.

**Table 3** – Factorial loadings after varimax rotation and communalities. FW-fruit weight (grams); FL-fruit length (mm); FD-fruit diameter (mm); TP-thickness of the pericarp (mm); NS/FR-number of seeds/fruit; FM- fresh matter (grams); DM – Dry matter (grams); CORD-diameter of the corolla (mm); PH-Plant Height (cm); SL-stem length (cm); CAD-Canopy diameter (cm); SD-stem diameter (mm).

	Fruit quality	Plant size	Plant architecture	Commonality
FW	0.96	0.0	0.15	0.94
FL	0.92	-0.07	0.16	0.88
FD	0.91	0.08	0.10	0.85
TP	0.88	0.04	0.11	0.79
NS/FR	0.86	0.01	0.07	0.75
FM	0.96	0.06	0.15	0.94
DM	0.95	-0.03	0.22	0.94
CORD	0.76	-0.19	0.15	0.64
PH	-0.19	0.88	-0.16	0.83
SL	0.15	0.91	0.10	0.86
CAD	-0.05	0.21	-0.92	0.89
SD	0.45	0.27	0.74	0.82

Source: Authors.

Table 3 shows the 3 factors obtained by the factor analysis with their respective factor loads for each analyzed variable. The greater the factor load, the greater the correlation of the variable with the factor. Variables with greater loads (greater correlation) within a factor are allocated within it. The values of commonality are also presented for each variable evaluated.

The first factor (F1) was made up by variables related to fruit characteristics (FW, FL, FD, TP, FM, DM, CORD). This result indicates that the variables related to fruits are highly correlated with each other, which makes it possible to denote this factor as "fruit quality". It is observed that all variables have positive correlation values, that is, the higher the value of these variables, the higher the value of the scores of the new variable formed.

The second factor (F2) was composed of variables related to the plant length (PH, SL), denominate "plant size". As in the previous factor for all variables, the loadings were positive, thus, the value of the score for this factor will increase according to the increase in the variables belonging to it. The third factor (F3) can be referred to as "plant architecture" since it grouped two characteristics related to diameter (SD, CAD) that are part of the determination of the architecture and harmony of the plant.

For all variables belonging to the factors that have a practical interpretation, mentioned above, the values of commonality (> 0.60) were acceptable. According to Figueiredo Filho (2010) the values of commonality must be greater than 0.5. The commonality is proportion of variance (or correlation), of each variable explained, by common factors. They can also be interpreted as indexes attributed to the original variables that express, in percentage terms, how much of the variability of each variable is explained by the model adopted. Thus, the higher the commonality values, the better the adjustment of the factorial model (Silva et al., 2014).

The high values of commonality, obtained in this study, demonstrate that the factorial model fits very well to the evaluated data set. The factors obtained represent a large part of the variability of the original data. The reduction in the number of characteristics evaluated in just 3 factors, facilitates the study and understanding of the correlation structure between variables. In addition, reducing the data set facilitates further analysis.

# 4. Conclusion

With the analysis of factors, the 12 characteristics evaluated were reduced to only 3 factors with practical interpretation, with a satisfactory percentage of explained variability.

Subsequently, the factors may have their values predicted for each sample unit, through the scores. These scores can be used for analysis of variance using a smaller set of data. In addition to the use of these colors in obtaining selection indexes free of multicollinearity, since in the factor analysis the highly correlated characteristics are grouped within one factor and among the factors the correlation is low or nonexistent.

## References

Corrar, L. J., Paulo, E. & Dias Filho, J. M. (2009). *Análise multivariada para os cursos de administração, ciências contábeis e economia*. São Paulo: Atlas.

Costa, G., Nascimento, D., Silva, B., Lopes, Â. C. D. A., Carvalho, L. C. B., & Gomes, R. L. F. (2019). Selection of pepper accessions with ornamental potential. *Revista Caatinga*, 32(2), 566-574.

Elias, H. T., Vidigal, M. C. G., Gonela, A. & Vogt, G. A. (2007). Variabilidade genética em germoplasma tradicional de feijão-preto em Santa Catarina. *Pesquisa Agropecuária Brasileira*, 42(10), 1443-1449.

Ferreira, D. F. (2011). Estatística Multivariada. Lavras: Ed. UFLA.

Figueiredo Filho, D. B. & Júnior, J. A. S. (2010). Visão além do alcance: uma introdução à análise fatorial. *Opinião Pública*, 16(1), 160-185.

Finger, F. L., Rêgo, E. R., Segatto, F. B., Nascimento, N. F. F. & Rêgo, M. M. (2012). Produção e potencial de mercado para pimenta ornamental. *Informe Agropecuário*, 33(267), 14-20.

IPGRI. (1995). Descriptores para Capsicum (Capsicum spp). Roma: IPGRI.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200.

Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG.

Neitzke, R. S., Barbieri, R. L., Rodrigues, W. F., Corrêa, I. V. & de Carvalho, F. I. (2010). Dissimilaridade genética entre acessos de pimenta com potencial ornamental. *Horticultura Brasileira*, 28(1), 47-53.

Pallant, J. (2007). SPSS Survival Manual: A Step By Step Guide to Data Analysis Using PSS (Version 15). Maidenhead: Open University Press.

Pereira, A. S., Shitsuka, D. M., Parreira, F. J. & Shitsuka, R. (2018). *Metodologia da pesquisa científica*.[e-book]. Santa Maria. Ed. UAB/NTE/UFSM. Recuperado de https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic\_Computacao\_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.

Rêgo, E. R., Finger, L. F. & Rêgo, M. M. (2011). Produção, genética e melhoramento de pimentas (Capsicum spp.). Recife: Imprima.

Silva, N. C. N., Ferreira, W. L., Cirillo, M. A. & Scallon, J. D. (2014). O uso da análise fatorial na descrição e identificação dos perfis característicos de municípios de Minas Gerais. *Revista Brasileira de Biometria*, 32(2), 201-215.

Teixeira, F. R. F., Nascimento, M., Nascimento, A. C. C., Paixão, D. M., Azevedo, C. F., Silva, F. F., Cruz, C. D., Lopes, P. S. & Guimarães, S. E. F. (2015). Determinação de fatores em características de suínos. *Revista Brasileira de Biometria*, 33(1), 130-138.

Vasconcelos, C. S. D., Barbieri, R. L., Neitzke, R. S., Priori, D., Fischer, S. Z., & Mistura, C.
C. (2012). Determinação da dissimilaridade genética entre acessos de *Capsicum chinense* com base em características de flores. *Revista Ceres*, 59(4), 493-498.

## Percentage of contribution of each author in the manuscript

Maria Eduarda da Silva Guimarães - 15% Ana Carolina Ribeiro de Oliveira - 10% Ana Izabella Freire - 10% Ariana Mota Pereira - 5% Dreice Nascimento Gonçalves - 5% Augusto Soares Lins Pantaleão - 5% João Romero do Amaral Santos de Carvalho Rocha- 5% Mateus de Paula Gomes- 5% Abelardo Barreto de Medonça Neto - 5% Rusthon Magno Cortez dos Santos - 5% Luciana Gomes Soares - 5% Renata Ranielly Pedroza Cruz - 5% Françoise Dalprá Dariva - 5% Fernando Luiz Finger – 15%