IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Human Journals
Research Article
August 2021 Vol.:19, Issue:2

Statistical Analysis of Meteorological Elements in The Tapacura-Pernambuco Dam, Brazil

Raimundo Mainar de Medeiros*[1], Manoel Viera de França[1], Wagner Rodolfo de Araújo[2], Fernando Cartaxo Rolim Neto[1], Romildo Morant de Holanda[1], Luciano Marcelo Fallé Saboya[3]

1 Universidade Federal Rural de Pernambuco, Brasil

2 Universidade Estácio de Sá, Brasil

3 Universidade Federal de Campina Grande, Brasil

Submitted: 23 July 2021
Accepted: 31 July 2021
Published: 30 August 2021

HUMAN JOURNALS

www.ijsrm.humanjournals.com

Keywords: Climatology, Linear Model, Density Functions

ABSTRACT

Statistics contribute as a research instrument to the understanding and validation of the climate, in its variables and its applied models, and should be integrated into different fields of academic-scientific knowledge. The objective of this study was to analyze the potential evapotranspiration, real evaporation, cloud cover, mean air temperature, relative air humidity, and total sunshine in annual series (1962-2015) and their correlations of similarities to the Tapacurá dam and its surroundings. The precipitation database used for the analyzes of this work was obtained from the Pernambuco State Agency for Water and Climate, covering the period from 1962 to 2015 for the rainy season, its homogeneity, consistency, and faults fill. The variables worked were the latitude and longitude geographic coordinates in degrees and altitude in meters, besides the data of potential evapotranspiration, real evaporation, cloud cover, average air temperature, relative humidity, and total annual sunshine. The means, standard deviation, coefficient of variance, maximum and absolute values of the study area were calculated. A spreadsheet was used for the graphic elaboration and its analysis. The existence of the temporal variability of the precipitation, evapotranspiration, and annual evaporation for seasonal and for the monthly one in increasing scale, being that its variability is greater during the winter and autumn. We detected fluctuations in the meteorological elements that are due to the meteorological systems acting under the Tapacurá dam and its neighborhood.

INTRODUCTION

Descriptive statistics aims to synthesize series of values of the same nature, thus allowing a global view of the variation of these values, enabling the creation of models, and favoring the integration of knowledge, as a focus in research.

Multiple regression analysis is a statistical technique mainly used for prediction purposes; it consists in determining a mathematical function that seeks to describe the behavior of a certain variable based on more than one character with explanatory power. It helps to understand the behavior of the climate.

The semiarid region has been living in recent years with drought linked to low rainfall in the region, which leads to desertification processes in the area, which has been trying to be remedied using public policies.

Barreto et al. (2015) report that the use of probability functions is directly linked to the nature of the data to which it relates; it is possible to estimate through rainfall data series, calculations of the probability of occurrence of precipitation, an important tool for planning agricultural and human operations.

With the concern to understand the behavior of the climate in its extreme instability, over time and according to spaces, for prognostic purposes or aiming to motivate explanations about its knowledge. The concerns are of significance when mentioning the socioeconomic and environmental effects, in the rural and urban environment, and the population's quality of life.

The climatic principle is formed by a set of highly dynamic elements that interact with the geographic factors of the climate, thus existing a permanent exchange of energy and interdependence according to Jesus (2005).

Melo et al. (2015) studied the spatial variability of monthly rainfall in the State of Paraíba, in the period 1994-2014, obtained by the ACP showed the existence of two rainfall patterns or climatological rainfall stations that explained approximately 84.97% of the total variance of the data. In the mapping of meteorological variables, we used Geostatistics tools to build maps, obtain homogeneous zones, and how to show the behavior of climatology in this period.

Connecting, analyzing, and interpreting content has been disseminated in broad fields of knowledge and, the object of publication in a book, which constituted an approach following Pinto et al. (2008).

According to Moreira et al. (2010) and Pizzato et al. (2012), the use of the average as a dimensioning parameter causes undersizing of irrigation systems, causing damages to the producer.

Martins et al. (2010) observed that the interannual variation in precipitation is less accentuated in the months when α values are higher, and this parameter can be used to determine regular periods of precipitation.

The objective is to analyze the variables potential evapotranspiration, actual evaporation, cloud cover, average air temperature, relative air humidity, and total insolation in annual series (1962-2015) and their correlations of similarities for the Tapacurá dam and its surroundings.

MATERIALS AND METHODS

Multiple linear regression analysis provides the relationship between variables, will be used to adjust linear models for the response (precipitation), or dependent variable, which will be explained by the predictor variables, that is, independent (altitude, longitude, and latitude), the equation of the statistical model by the method of least squares that minimizes the sum of squared errors.

The climate around the Tapecurá dam is of type 'AS' according to the climate classification by Köppen (1928) and following Alvares et al. (2014).

The precipitation database used for the analyzes of this work was obtained from CAPA (Pernambucan Water and Climate Agency) covering the period from 1962 to 2015 for the rainy seasons, it was verified its homogeneity, consistency, and filling gaps. The studied variables were the latitude and longitude of the geographic coordinates in degrees and altitude in meters, in addition to data on potential evapotranspiration, actual evaporation, cloud cover, average air temperature, relative humidity, and total insolation in annual series. The means, standard deviation, coefficient of variance, maximum and minimum absolute values of the study area were calculated. An electronic spreadsheet was used for the graphic elaboration and its analysis.

Table. No. 1: Location of municipalities, geographic coordinates.

Counties	Latitude (°)	Longitude(°)	Altitude(m)
Chã de Alegria	-7,99	-35,21	136
Glória de Goitá	-8,00	-35,31	186
Moreno	-8,11	-35,10	159
Paudalho	-7,53	-35,10	69
S. Lourenço Mata	-7,44	-37,20	616
Vitória S. Antão	-8,83	-35,65	255
Barragem	-7,98	-35,59	236

Source: Medeiros (2021)

Factor analysis

To carry out the Factor Analysis (FA) it was found that the relationships between the variables, using the linear correlation coefficient as a measure of association between each pair of variables. This multivariate technique aimed to reduce the original set of variables, and consequently the dimensionality of the data in a smaller number of linear combinations, which explained the greater variability of the original data by increasing its correlation structure. To this end, the principal component analysis - PCA was used as a method for extracting the factor loadings (MINGOTI, 2005; MANLY, 2008).

PCA consisted of transforming a set of n standardized variables, xi1, xi2, ..., xin into a new set yi1, yi2, ..., yin, in which the yi's are linear functions of the xi's and independent each other. The following properties were verified:

a) If Yij is a major componente

$$Y_{ij} = a_1 x_{i1} + a_2 x_{i2} + \cdots + a_n x_{n1} \quad (1)$$

Onde:

Yij é……..

a1,a2,…,anl are constant

Xi1, Xi2,..., Xnl are variables

b) If Y'ij is another main component, then:

$$Y'_{ij} = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_n x_{n1} \quad (2)$$

$$\sum_j a_j^2 = \sum_j b_j^2 = 1 ; \sum_j a_j b_j = 0 \quad (3)$$

That is, the components are independent.

c) The main components will be obtained by the system solution:

$$\det(R - \lambda_1 I) a = 0 \quad (4)$$

Where:

R = correlation matrix between estimated means,

λ1 = characteristic roots or (eigenvalues) of R,

I = identity matrix of dimension p x p, and

a = characteristic vector (or eigenvector) associated with eigenvalues (CRUZ et al., 1997).

Thus, the R eigenvalues corresponded to the variances of each component and the normalized eigenvectors to the weighting coefficients of the standardized characters. The relative importance of a component was calculated by:

$$\text{Importância de } Y_j = \frac{\lambda_j}{\text{traço}(R)} \quad (5)$$

The choice of the number of factors in the analysis consisted of synthesizing the accumulated variance around 70%. For the interpretation of the factors, in terms of simplification of the data structure, an orthogonal rotation in the varimax type coordinate axes was used. Factor scores were applied using the regression method, as a linear combination of the factors that underwent rotation and standardized measurements according to Manly (1997).

For the validation of the Factor Analysis, the Kaiser-Meyer-Olkin Measure of Adequacy (KMO) test was applied, which served to assess the input value of the variables for the model, and its value allows to provide results ranging from 0.5 to 0.9, providing greater reliability in carrying out this analysis. To find the KMO value, we used the expression:

$$KMO = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j a_{ij}^2} \quad (6)$$

On what:

rij is the level of correlation observed between variables i and j;

aij is the partial correlation coefficient between the same variables, which is simultaneously an estimate of the correlations between the factors. The values of aij approximated to zero resulted from orthogonal factors, as reported by Pereira (2001).

After the detection of the most important components of the variance based on the scores established by the factor analysis, a hierarchical cluster analysis was performed, aiming at the separation or classification of the selected species into a group or into a specific number of mutually exclusive subgroups or clusters (clusters), so that the subgroups formed had characteristics of great internal similarity and great external dissimilarity according to Moori (2002).

To delimit the groups, the Ward method was used, as described by Souza et al. (1997) and as the Euclidean distance dissimilarity measure according to the following expression:

$$d_{ii'} = \left[\sum_{j} \left(f_{ij} - f_{i'j} \right)^2 \right]^{1/2} \quad (7)$$

On what:

dii´= the Euclidean distance between i and i' species; and

fij − fi´j = difference between the scores of the factors of species i and i', for a variable j.

After the cluster analysis for the formation of groups using the Ward method, Fisher's Discriminant Analysis was followed, which consists of discriminating, one by one, the most statistically significant variables to compose the discriminant functions, which are responsible for separating the groups between the municipalities of the Ipojuca river dam. This analysis also allowed us to verify whether the formation of groups generated by the cluster analysis was satisfactory and representative.

The verification of the relevance of the groups formed by the cluster analysis through Fisher's discriminant function was originated according to Hair Jr et al. (2009):

$$Z_{jk} = a + W_1 X_{1k} + W_2 x_{2k} + \cdots + W_n X_{nk} \quad (8)$$

On what:

Zjk = Z score of the discriminant function j for object k;

a = intercept;

Wi = discriminant weight for independent variable i;

Xik = independent variable i for object k;

The method used to identify the most statistically significant variables to compose the discriminant function was stepwise and as a measure of significance for the difference between groups, the Wilks test, which is the standard statistic used to denote the statistical significance of the discriminatory power of the discriminant function. The test value can range from 0.0 to 1.0, considering that the closer to 1.0, the lower the discriminatory power of the function and the closer to 0.0, the greater it will be.

Data analysis was performed with the aid of software, Microsoft Excel 2013 version for Windows 7 and SPSS version 20.

RESULTS AND DISCUSSION

Correlation between meteorological variables

According to the correlation matrix between the meteorological variables per municipality (Table 1), it could be seen that the most correlated variables.

Describing the most significant correlations in Table 2, it is observed that the EVR (Evaporation) parameter has 0.98 in relation to the Rain parameter, indicating that after the occurrence of precipitation the evaporative indices rise. Still, on the EVR, it presents a positive correlation with ETP (evapotranspiration) (0.88), RH (relative air humidity) (0.71), and Cloud (0.63), suggesting that their presence implies the accumulation of precipitable water.

Table. No. 2. Correlation matrix between the meteorological variables of the municipalities belonging to the Ipojuca river dam-PE.

Correlação	Chuva	ETP	EVR	Insolação	Nuvem	Tmédia	UR
Rain	1	0,87	0,98	0,47	0,12	0,83	0,08
ETP		1	0,88	0,38	0,31	0,95	0,13
EVR			1	0,02	0,63	0,10	0,71
Insolation				1	0,07	0,33	0,23
A cloud					1	0,36	-0,33
Average temperature						1	0,16
UR							1

Legend: ETP = Evapotranspiration, EVR = Evaporation, Tmean = Average air temperature, RH = Relative air humidity.

Source: Medeiros (2021)

Another important correlation verified in Table 2 occurs between the factors Tmean and ETP (0.95), indicating that the higher the temperature, the higher the evaporative rates. It also shows that increased rainfall causes greater heat exchange due to temperature variability causing latent heat exchange and sensible heat (0.83). These statistical variabilities occur due to the integrated study of the dam and its surroundings.

The analysis of the frequency of precipitation distribution provides subsidies for water planning, mainly in determining the predominant critical periods of a region or area, whether drought or flood, the knowledge of the behavior of the distribution of precipitation provides us with information aimed at reducing the consequences caused. by the variability of the rainfall regime according to Silva et al. (2013).

For the adequacy of the data to the factor analysis, the KMO test it presented a value of 0.73, which is considered reasonable for the adequacy of the data to this type of analysis. Once the factor analysis was validated, it was possible to observe that the adjusted factor model had 76.3% of the total variability explained by two factors, the first containing 56.6% of the data variance was represented by the variable rain, ETP, EVR, and Tmean, as the second component with 19.7%, containing the variables Cloud and UR. The factor loadings associated with the variables had values above 0.78, that is, each variable is represented by at least 78% within each component, with only the insolation variable being discarded o it has low correlation and variability about the other variables. Noticeably observed in the aforementioned correlation matrix.

Melo et al. (2015) showed that urban, rural and environmental planning, as well as multipurpose water storage, are important. They identified monthly and annual variations and trends in meteorological variables: maximum and minimum air temperatures, precipitation, relative humidity, number of days with rain and total insolation in the municipality of Bom Jesus – PI. The authors used the following theoretical probability distribution functions: Weibull, Lognormal and Logistics to adjust the values of the aforementioned variables. They used the Kolmogorov-Smirnov (KS) test to verify the adjustment of theoretical functions. The results showed an increase in the maximum temperature and a reduction in the minimum temperature, an important condition for the desertification process in the studied area. This condition causes water stress in crops and, as a result, low production, this work corroborates the study in progress.

The works of the authors Gandra et al. (2015) found that for the variable latitude only for July was not significant, for the variables longitude and altitude, the months of February, March, October, November, the winter and autumn periods were not significant. The highest correlation found in Ferreira et al. (2013) was observed during the forecast of rainfall for July, corroborates the study in progress.

Regarding the commonalities, a lower correlation was observed between the cities for the cloud and UR due to the great differences in altitudes, vegetation cover and soil. It is also observed that the greatest variations that occur between the municipalities were ETP, Rain and EVR, which are responsible for representing the greatest correlation between the municipalities of the Tapacurá Dam, being their factors of greatest meteorological performance (Table 3). The concentration and transport of steam caused by altitude and vegetation cover, influencing the water mirror of the entire dam, justify the relationship between the higher values of commonalities and specific variance.

Table. No. 3: Matrix of factor loadings, commonalities, and specific variances according to the meteorological variables obtained in the varimax rotation.

Variáveis	Fator		Comunalidades	Variância específica
	1	2		
Rain	0,94	-	0,90	1,38
ETP	0,95	-	0,91	3,97
EVR	0,93	-	0,89	0,78
Insolation	-	0,78	0,67	0,04
A cloud	0,92	-	0,88	0,67
Average temperature UR	-	0,81	0,69	0,14

Legend: ETP = Evapotranspiration, EVR = Evaporation, Tmean = Average air temperature, RH = Relative air humidity, (-) no information.

Source: Medeiros (2021)

It is observed that the hierarchical grouping related to the cities within the limits of the Ipojuca River dam was subdivided into five groups, demarcated by the Fenon line at 7 Euclidean distance units, illustrated in the dendrogram in (Figure 1).

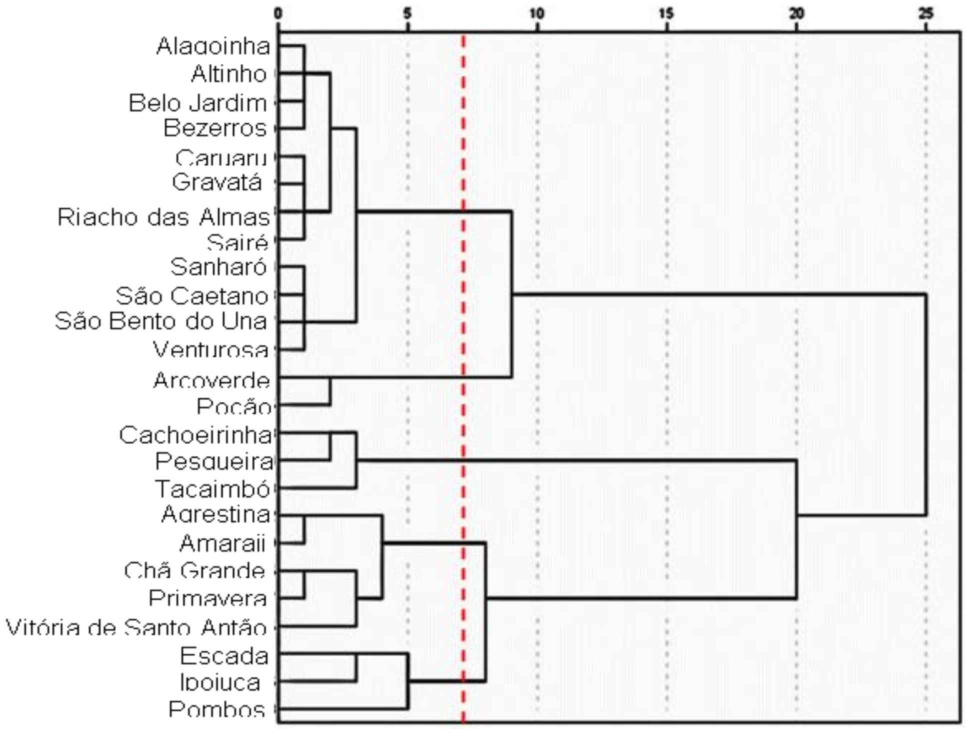

Figure No. 1. Dendrogram containing the hierarchical grouping separating the municipalities into five groups, delimited by the Fenon line cut, dashed in red.

In the first group, which contains the cities of Alagoinha, Altinho, Belo Jardim, Bezerros, Caruaru, Gravatá, Riacho das Almas, Sairé, Sanharó, São Caetano, São Bento do Una, and Venturosa, are grouped because they contain typical characteristics of the Agreste region. In the second group delimited by the municipalities of Arcoverde and Poção there is greater altimetry variability. The third group represents the cities of Cachoeirinha, Pesqueira and Tacaimbó, in the last group the cities of Agrestina, Amaraji, Chã Grande, Primavera, Vitória de Santo Antão, Escada, Ipojuca and Pombos. These grouping orders suggest that group variability is related to the years of rainfall observations for each municipality group.

As a result of the discriminant analysis, it was observed that the initial separation of the groups was confirmed in (Table 4).

Table No. 4: Relevance of the classification of the groups of municipalities formed through discriminant analysis.

Método de Ward	Predição de membros por grupo					Total
	1	2	3	4	5	
1	12 (100)	0	0	0	0	12 (100)
2	0	2 (100)	0	0	0	2 (100)
3	0	0	3 (100)	0	0	3 (100)
4	0	0	0	5 (100)	0	5 (100)
5	0	0	0	0	3 (100)	3 (100)
Total	12	2	3	5	3	25 (100)

Group 1 has a greater predominance of members (12 municipalities), suggesting that there is low rainfall among them, and the other fluctuations are due to the meteorological systems operating under the Tapacurá dam and its vicinity, according to Medeiros (2015).

The descriptive analysis of the data aimed to visually explore the information from the series to detect behaviors and verify the presence of atypical observations, that is, this step aimed to know the analyzed variables. Authors such as Bussab et al. (2002) and Triola (2009) emphasize the need and importance of descriptive analysis of the analyzed variables before any inferential analysis, as facts such as the presence of atypical values, the type of behavior of the analyzed variable and even typing errors in databases, they can distort the results of the inferential analysis, causing incorrect conclusions, a fact that corroborates the article under study.

Through discriminant analysis, according to the groups formed in the cluster analysis, it was observed that the ETP, Cloud, and UR variables were responsible for the separation between the groups, as they were the most statistically representative variables as presented by the method Stepwise test and Wilk's lambda test (Table 5).

Table No. 5: Summary of the Stepwise Method with the canonical discriminant variables selected

Variáveis	Percentual de variância (%)	Wilks' Lambda
ETP	82,9	0,09
Nuvem	15	0,16
UR	2	0,68

Caption: ETP = Evapotranspiration, RH = Relative Air Humidity.

It is observed that the variable related to ETP was the most representative among the three main ones with 82.9%, in second place the Cloud with 15%, and lastly UR with 2%. Because of the results, the existence of specific groups among the municipalities inserted in the Ipojuca River Dam is caused by different rates of ETP, Cloud, and UR.

In many future climate change scenarios, mainly due to increased concentrations of greenhouse gases in the atmosphere, it is assumed that only the mean can change, with the standard deviation remaining unchanged according to Bem-Gai et al. (1998). It was demonstrated by Mearns et al. (1984), Katz (1991), and Katz et al. (1992) that the relative frequency of extreme events depends on changes in the standard deviation and not just the mean. Katz (1991) assumes that a change in a climate variable that has a probability distribution may result in a change in the shape of its distribution.

CONCLUSION

The existence of temporal variability of precipitation, evapotranspiration, and annual evaporation for seasonal and monthly on an increasing scale, and its variability is greater during winter and autumn.

Fluctuations in meteorological elements resulting from meteorological systems acting under the Tapacurá dam and its surroundings were detected.

The smallest behaviours of statistical indicators were for the dry months. It is suggested to apply new models, such as generalized linear models, which relate the random distribution of the dependent variable with the non-random part through a link function.

REFERENCES

1. Alvares, C.A.; Stape, J.L.; Sentles, P.C.; Gonçalves, J.L.M.; Sparovek, G. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift, DOI: http://dx.doi.org/10.1127/0941-2948/2013/0507.2013.
2. Barreto, H.B.F.; Pereira, G.M.; Santos, W.O.; Freire, S G.C.; Maia, P.M.E. Evaluation of probability functions and duration intensity relation – duration intensity – frequency for extreme precipitation in Nepomuceno-MG. Global Science and Technology, v. 8, n. 2, p. 118-127. 2015.
3. Bem-Gai, T.; Bitan, A.; Manes, A.; Alpert, P.; Rubin, S. Spatial and temporal changes in rainfall frequency distribution patterns in Israel. Theor. Appl. Climatol. V.61. p.177-190. 1998.
4. Bussab, W.O.; Morettin, P.O. Basic statistics. 4. ed. São Paulo: Current, 1987. 320 p.
5. Ferreira, FF; Son, C.T.; Oliveira, P.R. Statistical forecast of monthly precipitation in the metropolitan region of Recife. Brazilian Journal of Physical Geography. v.6, n.5, 1517 – 1528. 2013.
6. Gandra, C.F.A.T.; Damé, R.C.F.; Simonete, M.A. Precipitation Prediction from Geographical Coordinates in the State of Rio Grande do Sul. Brazilian Journal of Physical Geography. v.8, p.848-856. 2015.
7. Hair. JR, J.F. et al. Multivariate Data Analysis. Translation Adonai Schlup Sant Anna. 6th edition Bookman, Porto Alegre - RS. p.688, 2009.
8. Manly, BJF Multivariate statistical methods: an introduction. 3. ed. Porto Alegre: Bookman, p.229. 2008.
9. Katz, R.W. Towards a statistical paradigm for climate change. Preprints, 7TH Conference on Applied Climatology, American Meteorological Society, Boston, 1991.
10. Katz, R.W.; Brown, B.G. Extreme Events in a Changing Climate: Variability is more important than averages. Climate Change. v.21, n.3. p.289-302. 1992.
11. Köppen,W.; Geiger,R. “Klimate der Erde. Gotha: Verlag Justus Perthes”. Wall-map 150cmx200cm. 1928.
12. Martins, J.A.; Dallacort, R.; Inoue, M.H.; Santi, A.; Kolling, E.M.; ColettI, A.J. Precipitation probability for the microregion of Tangará da Serra, State of Mato Grosso. Tropical Agricultural Research, v. 40, no. 3, p. 291-296. 2010.
13. Medeiros, R.M. Factors that provoke or inhibit precipitation in the state of Pernambuco. 2016.
14. Moreira, P.S.P.; Dallacort, R.; Magalhães, R.A.; Inoue, M.H.; Stieler, M.C.; Silva, D.J.; Martins, J.A. Distribution and probability of rain occurring in the municipality of Nova Maringá-MT. Journal of Agro-Environmental Sciences, vol. 8, n. 1, p. 9-20. 2010.
15. Melo, V.S.; Medeiros, R.M.; Souza, F.A.S. Analysis of principal components of rainfall in the State of Paraíba for the years 1994 to 2014. Technical Scientific Congress of Engineering and Agronomy. 2015.
16. Melo, V.S.; Medeiros, R.M.; Sousa, F.A.S. Impacts of urban climate in Bom Jesus – PI. Revista Agropecuária Científica no Semiárido, v.12, n.2, p.181-198, 2016.
17. Mingoti, S.A. Data analysis using multivariate statistical methods: an applied approach. Belo Horizonte: UFMG, p.297. 2005.
18. Moori, R.G.; Marcondes, R.C.; Ávila, R.T. Cluster Analysis as a Support Tool for Improving the Quality of Services to Customers. Journal of Contemporary Administration, vol. 6, no. 1, p.23, 2002.
19. Pinto, J.E.S.S.; Aguiar Netto, A.O. Climate, Geography, and Agrometeorology: An Interdisciplinary Approach. 1st ed. UFS. Saint Cristopher. 2008.
20. Pizzato, J.A.; Dallacort, R.; Tieppo, R.C.; Modolo, A.J.; Cremon, C.; MOREIRA, P.S.P. Distribution and probability of precipitation occurrence in Cáceres (MT). Tropical Agricultural Research, v. 42, no. 2, p.137-142. 2012.

21. Silva, I.N.; Oliveira, J.B.; Sources, L.O.; Arraes, J.D.D. Rainfall frequency distribution for the south-central region of Ceará, Brazil. Agricultural Science Magazine. v.44. n.3. p.481-487. 2013.
22. Triola, M.F. Introduction to statistics. Translation: Alfredo Alves de Faria. 7. ed. Rio de Janeiro: LTC, 1999. 410 p.

Image *Author -1*	**Raimundo Mainar de Medeiros** *Federal Rural University of Pernambuco, Brazil*
Image *Author -2*	**Manoel Viera de França** *Federal Rural University of Pernambuco, Brazil*
Image *Author -3*	**Wagner Rodolfo de Araújo** *University Estácio de Sá of Pernambuco, Brazil*
Image *Author -4*	**Fernando Cartaxo Rolim Neto** *Federal Rural University of Pernambuco, Brazil*
Image *Author -5*	**Romildo Morant de Holanda** *Federal Rural University of Pernambuco, Brazil*
Image *Author -6*	**Luciano Marcelo Fallé Saboya** *Federal University of Campina Grande*