# Insertion of Distributed Photovoltaic Generation in Brazil: A Correlation Analysis between Socioeconomic and Geographic Aspects

**Marcelle Feitoza Bassi Costa\*, João Alberto Neves dos Santos**

Departament of Civil Engineering, Fluminense Federal University, Rio de Janeiro, Brazil. *Email: mbassi@id.uff.br

**ABSTRACT**

Brazil has relevant regional differences in social, economic, and geographical terms, and the technological development of each region is affected by the dynamics of this inequality. The objective of this paper is to measure the degree of relationship between socioeconomic and geographic variables, with the level of insertion of residential photovoltaic (PV) generation technology in each federative unit in Brazil. The variables selected for this study were: (1) Human development index; (2) residential expenditure on electricity; (3) number of PV companies; (4) PAYBACK; and (5) solar irradiation index, based on the Brazilian Solar Energy Atlas developed by the National Institute for Space Research (INPE) in 2006. Five hypotheses were established. The methodology used to test these hypotheses was through the statistical analysis of Pearson's correlation coefficient. Based on the obtained results, it can be concluded that the solar irradiation index is not a significant variable in the insertion process of residential PV distributed generation around the country; and where the residential expenditure on electricity is higher, are also the places that had the smallest installed PV systems. This may help the perception of the importance of each of the variables to increase the use of PV energy in Brazil.

**Keywords:** Correlation, Social Economics, Photovoltaic Distributed Generation
**JEL Classifications:** C1, A130

## 1. INTRODUCTION

A distributed generation represents a new configuration for the electricity industry. In Brazil, this model was introduced in 2012 and regulated by ANEEL (National Electric Energy Agency) from Normative Resolution No. 482/2012, which established the general conditions for access of distributed micro and mini generation to distribution systems (EPE, 2014). In 2015, the regulation was improved to make the connection process faster and thus expand access to distributed generation to a more significant number of consumer units (MME, EPE, 2018).

Since then, it has been a growing insertion of these systems in the Brazilian electricity matrix, especially the generation from photovoltaic (PV) solar energy, which has been the fastest-growing source in recent years.

In addition to regulatory frameworks, Brazil has a series of government programs and policies to support the development of the PV industry, ranging from fostering the national industry - through the Progressive Nationalization Plan (PNP) - to tax incentives such as the agreement. CONFAZ 101/97 and 16/2015 and Law No. 13,169/2015 (MDIC, SDCI, DECOI, 2018).

In the private sector, associations like ABINEE (Brazilian Association of Electrical and Electronic Industry) and ABSOLAR (Brazilian Association of PV Solar Energy) promote the PV sector by bringing together companies from various segments of the

production chain, besides to submitting proposals of improving public policies for the sector (MDIC, SDCI, DECOI, 2018).

All these measures contributed to the advancements in residential PV micro and mini-generation in the country, which since 2012 has been showing significant growth (Figure 1).

Thus, this article aims to verify the measure of the relationship between socioeconomic and geographical variables, with the level of insertion of residential PV distributed generation (PV) technology surveyed in each Federative Unit (FU) in Brazil. For this, the following variables were selected: (1) The human development index (HDI) of each FU; (2) monthly residential expenditure on electricity; (3) the number of PV companies operating in the region; (4) PAYBACK expected by the PV system investment, and (5) the level of solar radiation in each region. For each selected variable a hypothesis was also established, and these hypotheses were tested through the statistical analysis of Pearson's correlation coefficient ($R_{XY}$). The hypotheses established were as follows:

Hypothesis 1: The higher the HDI in a given FU, the higher the number of PV systems installed;

Hypothesis 2: The higher the expenditure on electricity in a given FU, the higher the number of PV systems installed;

Hypothesis 3: The higher the number of solar PV companies in a given FU, the higher the number of PV systems installed;

Hypothesis 4: The lower the PAYBACK of the PV system investment in a given FU, the higher the number of PV systems installed; and

Hypothesis 5: The higher the solar irradiation rate in a given FU, the higher the number of PV systems installed.

The accomplishment of this research is justified by four fundamental questions: (i) It examines a subject not yet adequately identified in the literature, in researches that aim to explore a retrospective analysis of the growth of the distributed generation of PV in the country, mainly considering the specific aspects of each one region; (ii) discusses a theme that, due to the significant cost values involved in projects that use PV distributed generation, implies an identification of the degree of relationship between socioeconomic and geographic variables, in order to avoid investments that are doomed to failure; (iii) provides increased knowledge about a newly introduced technology in the country, requiring studies on its feasibility, opportunities and challenges. Besides, increased data availability enables a more

robust statistical analysis to be performed; and (iv) it allows the development of exploratory data analysis and the application of statistical methods to analyze the dynamics of the insertion of this technology over the last years, due to the existence of more than 40 thousand distributed generation PV connections in the country.

The contribution of this study is to fill a gap that exists today in Brazil on the issue of the use of PV energy, besides fostering an interdisciplinary perspective, whose focus lies on the discussion of a practical problem that is limited to the intersection between the areas of Engineering, Architecture, Urbanism, Economics and Administration. Thus, the results of this research contribute to give knowledge about an implicit subject to the construction projects, because the real estate projects are susceptible to several risk factors that make the realization of investments even more difficult, having a close relationship with Economics studies. Civil Engineering, Architecture and Administration activities. According to Kowaltowski and Neves (2016. p. 58), "these areas impact and are impacted by other fields of knowledge of engineering, technology, urbanism and the applied social sciences in general."

In addition, the skills required for training in economics, engineering and administration, aggregate interdisciplinary knowledge (anthropological, social, economic, environmental, artistic, constructive, technological, structural, graphic, among others), whose skills, academics and future professionals they need to dominate with a generalist profile, to be effective in their areas of activity and to enter the labor market Batistello et al. (2019). Thus, the approach used in this paper seeks to exploit this gap identified in previous academic research, as well as promoting the basis for future research on this topic.

## 2. THEORETICAL REFERENCE

### 2.1. PV Distributed Generation Technology Insertion in Brazil

As a country with a continental dimension, Brazil presents relevant regional differences in economic, social and geographical terms, and some of these differences impact the insertion process of these new technologies in each region.

When we look at the growth curve of the number of residential PV plants (UFV) installed by region (Figure 2), we notice significant differences in the angles of the curves. While the Southeast and the South show strong growth, the North, Northeast and Midwest regions show a moderate growth curve.

Only in 2014, the first residential PV systems installations appear in the Northern region, and only in 2017 do all states in the region include PV installations. Acre, Amapá, and Roraima were the last states to have PV systems installed.

However, regardless of the region, the level of insertion of residential PV systems is still deficient across the country. According to the study by Lange (2012), the national potential of this sector is 32820 MW. Being 2215 MW in the North region; 8840 MW in the Northeast region; 2075 MW in the Midwest region; 14055 MW in the Southeast region; and 5005 MW in

**Figure 1:** The number of residential UFV installations (2012-2018)
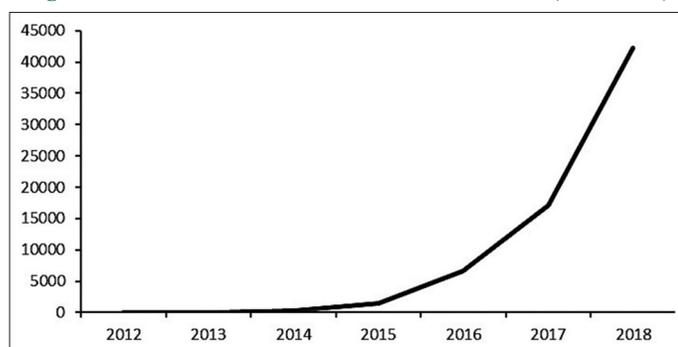


Source: BIG/ANEEL (2019). Elaborated by the authors

the South region. That is, by the end of 2018, the number of residential PV systems represented only 0.63% of the national potential (Table 1).

In 2018, Minas Gerais was the state with the highest number of residential PV installations, followed by São Paulo and the Rio Grande do Sul. In Figure 3, it can be seen that the five states that presented the most significant number of PV installations belong to the Southern and Southeast, being notorious the concentration of residential PV systems in these two states. The map below (Figure 4) clearly illustrates this scenario.

## 2.2. Pearson Correlation Coefficient

Correlation analysis is used to determine the degree of relationship between two different variables. According to Mukaka (2012),

**Table 1: Residential UFV insertion level by region (2018)**

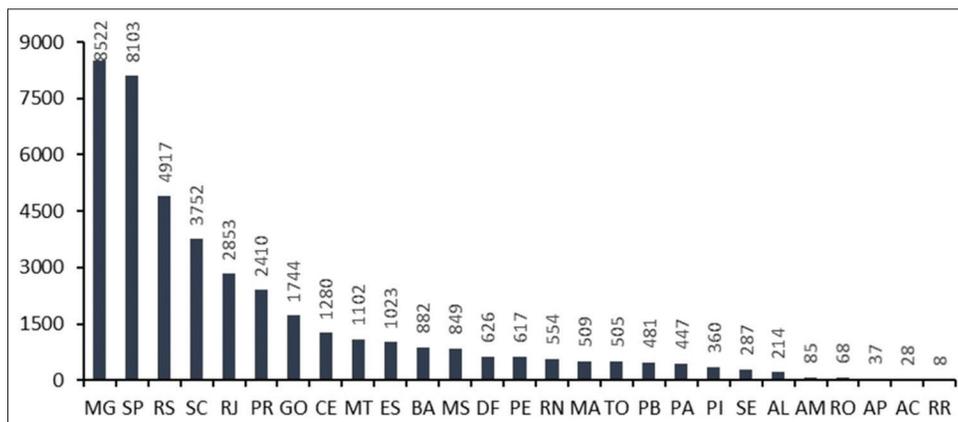| Region | Potential (MW) | Installed power 2018 (MW) | Insertion level (%) |
|--------|---------------|---------------------------|---------------------|
| North | 2215 | 7.7 | 0.35 |
| Northeast | 8840 | 32.2 | 0.36 |
| Midwest | 2705 | 22.5 | 0.83 |
| Southeast | 14055 | 90.6 | 0.64 |
| South | 5005 | 52.5 | 1.05 |
| Total | 32820 | 205.5 | 0.63 |

Source: Technical note DEA 19/14, EPE (2014) - Based on Lange (2012)

**Figure 2:** Number of residential UFV installations by region (2012-2018)
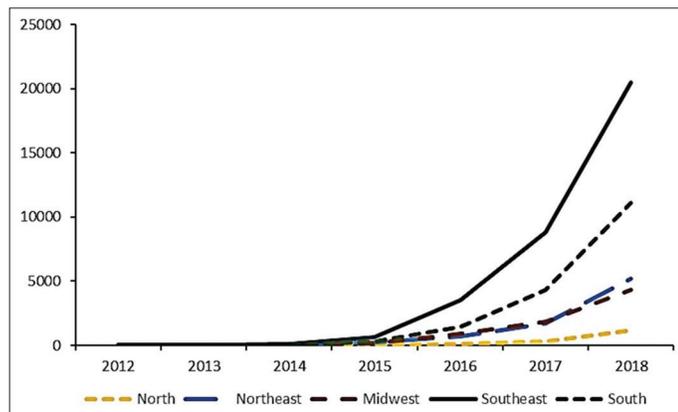


Source: BIG/ANEEL (2019). Elaborated by the authors

being a simple method to be calculated and interpreted, it is, therefore, one of the most used statistical methods in scientific research (Taylor, 1990).

To understand the relationship between two variables $X$ and $Y$, one must first understand the concept of covariance. According to Guimarães (2017), covariance is a statistic that indicates "how two variables vary together" (Guimarães, 2017. p. 3), and what is the linear relationship between them. The covariance formula is given by:
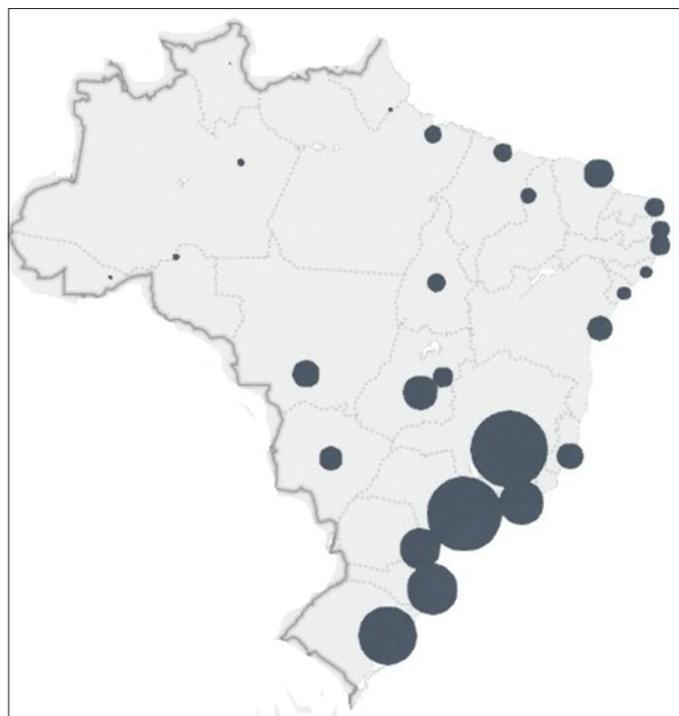
$$C(X,Y) = \frac{\sum_i^n (X_i - \bar{X}) \times (Y_i - \bar{Y})}{n} \quad (1)$$

However, although covariance is an appropriate statistic for measuring the linear relationship between two variables, it is not adequate to measure the degree of relationship between variables, since it is influenced by the units of measurement of each variable (which can be metered, kilometers, kilograms, centimeters). Thus, to avoid the influence of the order of each variable, the covariance is divided by the standard deviation of $X$ and $Y$, giving rise to the Pearson correlation coefficient (Guimarães, 2017).

$$R_{xy} = \frac{\sum_i^n (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \times \sqrt{\sum_i^n (Y_i - \bar{Y})^2}} \quad (2)$$

In statistical terms, Pearson's correlation coefficient indicates the direction and intensity of the linear relationship between two quantitative variables. Its value is dimensionless and ranges from −1 to +1. If the coefficient is a positive number, the variables are directly related (i.e., if the value of one increase, the value of the other variable also tends to increase). On the other hand, if the coefficient is a negative number, the variables are inversely related (that is, if one value increases, the other tends to decrease). When the coefficient is 0, it means that there is no linear correlation between the variables. When the coefficient is −1 or +1, it means that there is a perfectly linear correlation (Mukaka, 2012). However, there is no single classification on intermediate values. In this work, we chose to use the classification used by Dancey and Reidy (2006), according to the Table 2 below.

**Figure 3:** The number of residential UFV installations per UF (2018)



Source: BIG/ANEEL (2019). Elaborated by the authors

**Figure 4:** The number of residential UFV installations per state (2018)



Source: BIG/ANEEL (2019). Elaborated by the authors

**Table 2: Pearson's correlation coefficient classification**

| Correlation coefficient | Correlation |
|---|---|
| $R_{XY}=1$ | Perfectly positive |
| $0.7 \leq R_{XY} < 1$ | Strong positive |
| $0.5 \leq R_{XY} < 0.8$ | Moderate positive |
| $0.1 \leq R_{XY} < 0.4$ | Weak positive |
| $0 \leq R_{XY} < 0.1$ | Weak positive |
| $0$ | Null |
| $-0.1 \leq R_{XY} < 0$ | Very weak negative |
| $-0.4 \leq R_{XY} < -0.1$ | Weak negative |
| $-0.7 \leq R_{XY} < -0.4$ | Moderate negative |
| $-1 \leq R_{XY} < -0.7$ | Strong negative |
| $R_{XY}=-1$ | Perfectly negative |

Source: Dancey and Reidy (2006)

In graphical terms, the best way to illustrate the relationship pattern between the two variables is through scatter plots. Geometrically, the scatter diagram presents the points formed by the values of the two variables under analysis, one represented by the *X*-axis and the other by the *Y*-axis (Moore, 2003). According to Guimarães (2017), the dispersion diagram is the best method to analyze the occurrence of trends among the data, the grouping between the variables and the existence of the outliers.

Figure 5 presents some scatter diagrams and their respective correlation classifications.

Pearson's correlation coefficient does not differentiate independent and dependent variables, so "the correlation value between *X* and *Y* is the same between *Y* and *X*" (Filho and da Silva Júnior, 2009. p. 121). Moreover, it is essential to highlight that the correlation coefficient does not indicate causality. Thus, it cannot be said that one variable varies as a function of the other.

# 3. METHODOLOGY

Given the characteristics presented and the proposed objective, the present work is characterized as ex-post-facto research, exploratory, and with a quantitative approach. Ex-post facto because it is a study done after the fact, so that the researcher could not interfere in the identified variables (Almeida, 2011); exploratory because its primary purpose is to develop or clarify concepts on a subject little explored (Gil, 2008); and quantitative approach by using the statistical tool to measure the relationships between variables, thus presenting a quantifiable result (Fonseca, 2002).

## 3.1. Research Phases

The study contemplated in this article adopts a methodology based on exploratory research, where the chosen strategies follow the quantitative method approach, which implies a data collection and analysis procedure, combined with quantitative techniques in the same research design, a task that requires researcher is familiar with working with quantitative analysis of numerical data in a single study (Creswell and Plano, 2011; Creswell, 2012; 2009). To meet the proposed objectives, the method was divided into several stages, and the development of this research guided by a process based on the following methodological procedures listed in Figure 6.

The statistical tool used in this work was Pearson's correlation coefficient ($R_{XY}$) because through this basic technique of descriptive and inferential statistics it is possible to identify the degree of relationship between two variables, and thus determine which variables are potentially crucial to understanding the phenomenon studied (Rodrigues, 2012).

The analysis performed considered only residential-class PV distributed generation systems. This choice was made because this class represents almost 75% of all distributed generation present in the country in 2018. Also, according to Garcez (2017), non-residential systems are highly heterogeneous and end up distorting the results found.

The following will be presented in detail on how the data collection and treatment used in this work were performed.

## 3.2. Data Collect

The data necessary for the development of this research were collected at ANEEL (2018), EPE (2018), IBGE (2010), Greener (2019) and INPE (2017), and structured in cross-sectional format.

ANEEL provides the number of residential PV installations in its public database. From there it is possible to obtain information about the consumer units with distributed generation by selecting the source of generation, the location (region, states or municipalities), the type of generation, the consumption class, the voltage group, and the period connection. In this paper, we select the following information: generation source was solar radiation; the location was by state; the type of generation was by UFV (PV Plants); the consumption class was the residential class; the voltage group was B1; the connection period was between 2012 and 2018.

The variable selected to characterize the social context was obtained from the website of the Brazilian Institute of Geography and Statistics (IBGE). We chose to use the HDI because this is an index capable of portraying the quality of life of a region through multidimensional aspects, including the population's level of income, health, and education.

As for geographical aspects, what stands out most concerning the potential for solar generation is the weather condition of each region (Miranda, 2013). Thus, the variable used to represent this dimension was the average annual value of the total daily solar radiation available in the Brazilian Solar Energy Atlas developed by INPE in 2017.

To represent the economic context, we chose to use three variables, which are: the monthly expenditure on electricity; the number of PV companies operating in the region; and the PAYBACK expected by the PV system investment.

Data regarding the number of companies and PAYBACK were obtained directly from the Strategic Study of the Distributed Generation PV Market, prepared by Greener (Research and Consulting Company specializing in PV solar energy).

**Figure 5:** Correlation classification through the scatter plot



Source: Rodrigues (2012)

The monthly residential expenditure on electricity per state - as it is not available - was calculated from the average residential consumption between 2013 and 2017, and the value of the last tariff available by ANEEL (March, 2019). Taxes were not considered in the calculation since PIS (Social Integration Program), and COFINS (Contribution to Social Security Financing) vary monthly and therefore change the final price charged for energy. Thus, the calculation of monthly expenditure on electricity per FU is given by:

$$A_m = \frac{\left(\sum_{2013}^{2017} A_m\right)}{5} \tag{3}$$

$$T = TUSD + TE \tag{4}$$

$$MEE = C_m \times T \tag{5}$$

Where:
$A_m$ = Average power consumption (kWh)
$T$ = Electricity tariff (R$/kWh)
$TUSD$ = Distribution system use tariff
$TE$ = Energy tariff
$MEE$ = Monthly expenditure on electricity (R$/month)

The following Table 3 summarizes all the variables used in this work.

### 3.3. Data Processing and Analysis

Because it is a parametric test, the data used to calculate the Pearson correlation coefficient should be normally distributed. According to Filho and da Silva Júnior (2009), this assumption is especially important in small samples ($n < 40$). As this study is stratified by FUs, our sample size is 27 data; it is considered a small sample.

The Shapiro-Wilk test was performed using the SPSS software to verify the normality condition, where the significance level (P-value) that measures the degree of agreement between the data and the null hypothesis ($H_0$) is provided. In this case, we consider $H_0$ to be a normal distribution. If $P \leq 0.05$, it rejects $H_0$, considering that it is not a normal distribution. If $P > 0.05$, accepts $H_0$, considering that the distribution is normal.

From this test, it was observed that the variables amount of PV installations (P = 0.000), solar irradiation index (P = 0.002), number of companies (P = 0.000), and monthly expenditure with electrical energy (P = 0.025) did not present normal distribution. Thus, to fulfill the requirements of Pearson's coefficient analysis, logarithmic transformations (Ln) were employed for these non-normalized
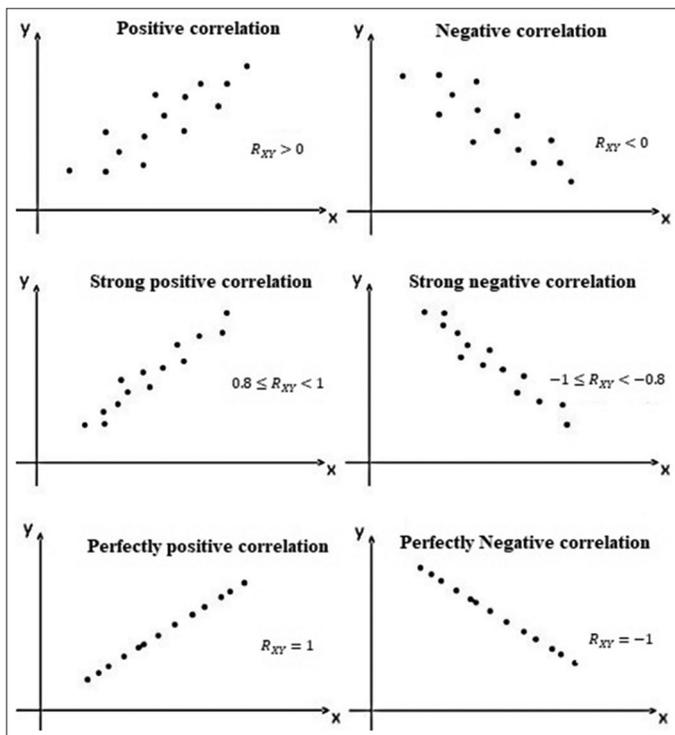
**Figure 6:** Research methodology
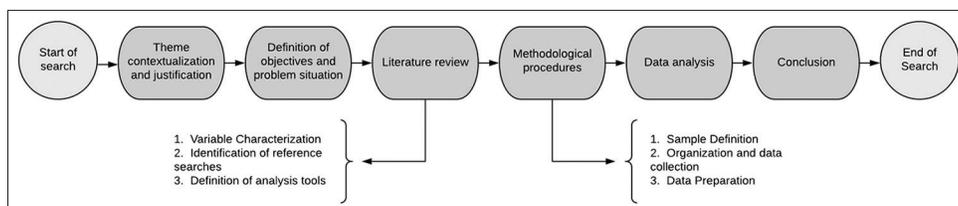
**Table 3: Description of variables**

| Variable | Description | Data source |
|---|---|---|
| Dependent variable * | | |
| Num_PV_sys | Number of micro and mini residential PV systems installed | ANEEL, 2018 |
| Independent variables | | |
| HDI | Municipal human development index | IBGE, 2010 |
| Expenditure | Monthly residential expenditure on electricity (R$/month) | - |
| Solar_Index | Global average solar radiation (kWh/m$^2$) | ATLAS Solar, 2018 |
| Num_PV_comp | Number of PV companies operating in the State (units) | Greener, 2019 |
| PAYBACK | Expected PAYBACK time with PV energy investment (years) | Greener, 2019 |

*Correlation analysis does not differentiate dependent variables from independent ones, as, according to Filho and da Silva Júnior (2009), "the correlation value between $X$ and $Y$ is the same between $Y$ and $X$." However, in this study, we chose to keep the nomenclature of the variables for convenience but respecting the non-causality effect. PAYBACK: The PAYBACK considered in the Greener study is simple PAYBACK

variables. According to Pino (2014), this type of transformation results in better approximations for the normal distribution.

The Pearson correlation matrix was built using the SPSS software to identify the association between the variables. From this coefficient, it was possible to conclude the meaning and intensity of the relationship between variables $X$ and $Y$.

In addition to the coefficient matrix, correlation diagrams were also drawn to verify the trend of the data. If the data approaches a straight line, it can be said that there is a linear relationship. If the slope of the line is increasing, the correlation is positive. If the slope is decreasing, the correlation is negative. The intensity of the correlation will be evaluated according to the proximity of the points with the line (Correa, 2003).

# 4. RESULTS

The analysis of the correlation coefficients indicated that the number of PV installations is positively related to the HDI and the number of PV companies operating in the region (Figures 7 and 9), while negatively related to the expenditure monthly and the PAYBACK (Figures 8 -10). The solar index (Figure 11) presented a feeble deficient degree of correlation (−0.026) and also a non-significant statistical result (P > 0.05) (Table 4).

## 4.1. Results Related to Hypothesis H$_1$
Regarding Hypothesis 1, the research results indicate that the higher the HDI in a given FU, the greater the number of PV systems installed, since the calculated Pearson Coefficient reached a moderate value of 0.485, is significant, because calculated P = 0.005, lower than the limit value of 0.05. This shows that human development in the various federation units can leverage the search for more PV facilities.

## 4.2. Results Related to Hypothesis H$_2$
Regarding Hypothesis 2, the research results presented a negative Pearson Coefficient (−0.441) statistically significant (P = 0.011). It shows that as lower as expenditure is in a given FU, greater the number of PV systems installed.

This result indicates a process of socioeconomic unequal insertion, where places that would benefit most from PV generation, are the places with the least installed systems.

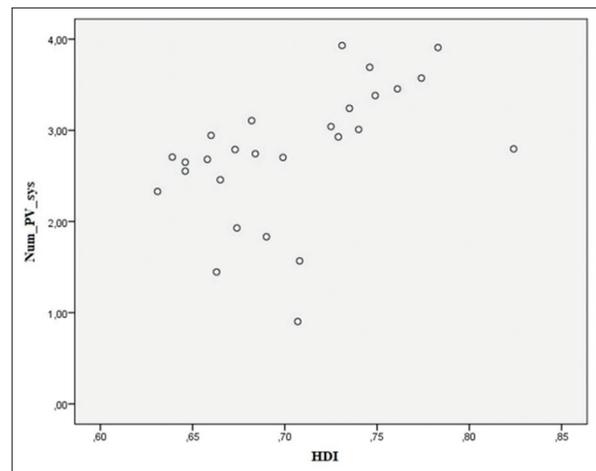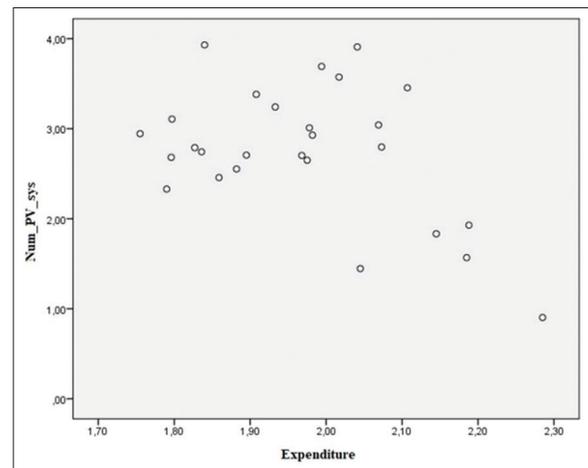**Figure 7:** HDI × number of photovoltaic systems installed



**Figure 8:** Expenditure × Number of PV systems installed



## 4.3. Results Related to Hypothesis H$_3$
As for hypothesis 3, the results indicate that the greater the number of PV companies operating in a given state, the greater is the number of installed systems. In this case, the Pearson coefficient reached an extremely high value of 0.953 and statistically significant (P = 0.000). This result confirms the relevance of the productive sector performance in the development of the PV distributed generation market in the country.

**Figure 9:** N° of PV companies × number of PV systems installed
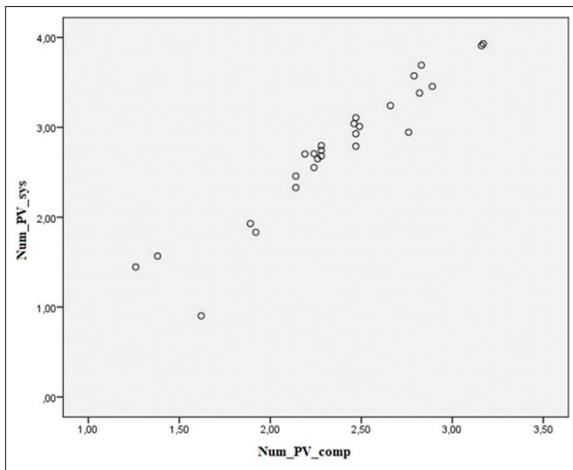


**Figure 10:** PAYBACK × number of PV systems installed
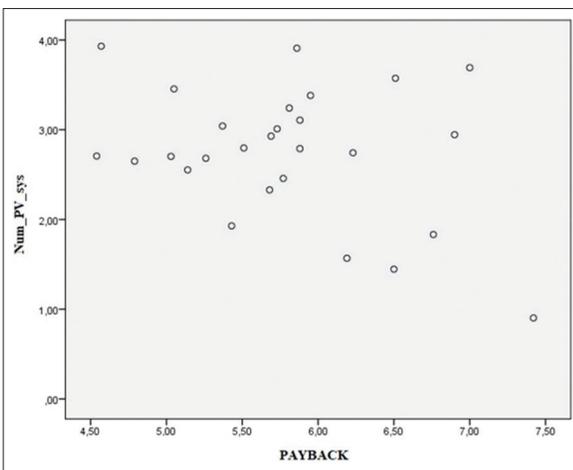


**Figure 11:** Solar index × number of PV systems installed
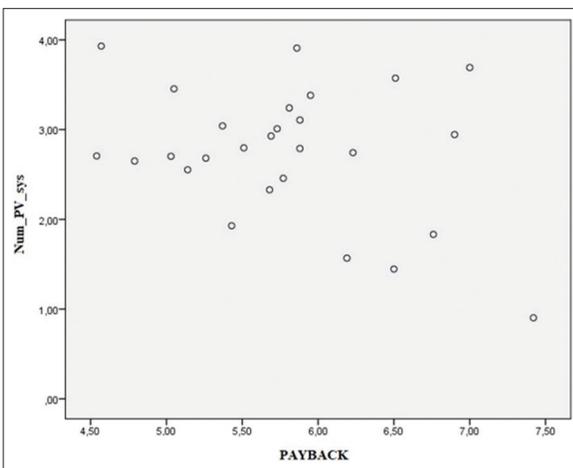


### 4.4. Results Related to Hypothesis $H_4$

Regarding hypothesis 4, the results indicate that the shorter the PAYBACK time for the investment in PV systems (PAYBACK), the greater the number of installed systems. Pearson's coefficient found presented a value of −0.324 and a P = 0.05. This result indicates a weak relationship between the variables and a significance value within the established limit.

### 4.5. Results Related to Hypothesis $H_5$

Regarding Hypothesis 5, the research results show that there is no correlation between the solar irradiation index of a given FU and the number of PV systems installed since the calculated Pearson Coefficient was extremely low at −0.026 and not significant because the calculated P = 0.448, greater than the limit value of 0.05. This shows that the number of PV installations is not related to the solar irradiation index.

## 5. ANALYSIS AND FINAL CONSIDERATIONS

From the analysis performed above, we found that sites with higher HDI and a more significant number of PV companies also presented larger amounts of installed systems, thus confirming hypotheses 1 and 3. These results confirm the direct relationship between the level of socioeconomic development and the performance of the productive sector of each state with the insertion level of distributed generation PV.

Locations with the lowest PAYBACK time for PV systems investment (PAYBACK) had the most significant number of installed systems, confirming the financial sensitivity of the PV solar energy business in the country. The degree of relationship, although low ($R = 0.324$), confirms hypothesis 4 assumed in this paper.

Regarding the solar irradiation index, we found that, besides presenting a feeble degree of correlation with the number of PV installations, its result was not statistically significant ($P > 0.05$), which leads us to refute hypothesis 5, and, consider that the success of distributed PV generation is not related to the availability of natural resources.

As for the monthly expenditure on electricity, we found that the higher this value, the lower the number of PV installations, thus refuting hypothesis 2. This result is counterintuitive and goes against common sense because the expectation is that the higher the spent-on electricity, the more likely the consumer is to install a PV system. However, caution must be exercised when interpreting this result. First, it must be emphasized that correlation analysis does not indicate causality. So, while we may conclude that the amount of PV installations decreases as electricity expenditure increases, we cannot say that high electricity expenditure causes a decrease in the number of PV installations.

In 2007, the research by Aguiar et al. (2007) stated that the tariff and the level of electricity consumption depended on the household income pattern and the poorer a region, the lower these values would be. However, according to the EPE Statistical Yearbook of Electric Energy (2018), between 2013 and 2017 the North region had the highest residential consumption of electricity, consuming on average 214 kWh, while the national average was 173 kWh in the same period (Table 5).

Besides, according to a study conducted by Abrace (Brazilian Association of Large Industrial Energy Consumers and Free Consumers), between 2014 and 2017 the North region presented

**Table 4: Pearson correlation coefficient matrix**

| Variable | Num_PV_sys | HDI | Expenditure | Solar_Index | Num_PV_comp | PAYBACK |
|---|---|---|---|---|---|---|
| Num_PV_sys | 1.000 | | | | | |
| HDI | 0.485 | 1.000 | | | | |
| Expenditure | −0.441 | 0.397 | 1.000 | | | |
| Solar_Index | −0.026 | −0.435 | −0.569 | 1.000 | | |
| Num_PV_comp | 0.953 | 0.465 | −0.409 | −0.003 | 1.000 | |
| PAYBACK | −0.324 | 0.126 | 0.272 | −0.324 | −0.231 | 1.000 |
| Sig. * (unilateral) | | | | | | |
|   HDI | 0.005 | | | | | |
|   Expenditure | 0.011 | 0.020 | | | | |
|   Solar_Index | 0.448 | 0.012 | 0.001 | | | |
|   Num_PV_comp | 0.000 | 0.007 | 0.017 | 0.495 | | |
|   PAYBACK | 0.050 | 0.266 | 0.085 | 0.050 | 0.123 | |

*Considers the correlation between being statistically significant if P<0.05. That is, the probability that the correlation coefficient happened by chance must be <0.05

**Table 5: Average residential consumption by region and FU (kWh/month)**

| | 2013 | 2014 | 2015 | 2016 | 2017 | Average (2013-2017) |
|---|---|---|---|---|---|---|
| **North** | **201.2** | **212.9** | **220.1** | **220.4** | **212.6** | **213.1** |
| Rondônia | 224.7 | 232.2 | 231.7 | 218.5 | 222.9 | 226.0 |
| Acre | 171.5 | 176.9 | 185.8 | 184.9 | 186.1 | 181.0 |
| Amazonas | 211.6 | 228.4 | 237.3 | 217.6 | 197.9 | 218.6 |
| Roraima | 275.9 | 302.4 | 319.5 | 315.0 | 305.9 | 303.7 |
| Pará | 126.3 | 141.7 | 143.9 | 148.9 | 142.6 | 140.7 |
| Amapá | 258.0 | 263.6 | 271.9 | 292.0 | 266.8 | 270.5 |
| Tocantins | 140.6 | 145.2 | 150.7 | 165.7 | 165.9 | 153.6 |
| **Northeast** | **118.7** | **122.0** | **122.4** | **122.1** | **120.8** | **121.2** |
| Maranhão | 112.8 | 118.8 | 120.8 | 123.7 | 121.9 | 119.6 |
| Piauí | 114.4 | 117.1 | 127.5 | 125.8 | 125.6 | 122.1 |
| Ceará | 124.2 | 129.4 | 125.6 | 130.8 | 126.1 | 127.2 |
| Rio Grande do Norte | 139.8 | 143.5 | 142.9 | 143.6 | 142.5 | 142.5 |
| Paraíba | 112.0 | 115.5 | 114.8 | 115.2 | 115.1 | 114.5 |
| Pernambuco | 129.9 | 131.2 | 130.0 | 127.4 | 125.2 | 128.8 |
| Alagoas | 113.4 | 116.9 | 114.9 | 105.9 | 109.0 | 112.0 |
| Sergipe | 114.5 | 115.9 | 115.8 | 115.4 | 112.0 | 114.7 |
| Bahia | 107.5 | 110.0 | 109.7 | 111.2 | 109.3 | 109.6 |
| **Southeast** | **168.1** | **171.9** | **165.7** | **162.3** | **158.2** | **165.2** |
| São Paulo | 208.3 | 206.6 | 194.7 | 191.3 | 191.5 | 198.5 |
| Minas Gerais | 126.3 | 129.5 | 125.5 | 123.8 | 123.6 | 125.7 |
| Espírito Santo | 162.6 | 167.4 | 163.7 | 159.2 | 154.0 | 161.4 |
| Rio de Janeiro | 175.1 | 184.1 | 178.7 | 175.1 | 163.7 | 175.3 |
| **South** | **184.3** | **193.3** | **180.4** | **180.3** | **180.4** | **183.7** |
| Paraná | 169.8 | 172.9 | 161.0 | 157.1 | 160.4 | 164.3 |
| Santa Catarina | 201.8 | 212.9 | 200.2 | 201.7 | 202.5 | 203.8 |
| Rio Grande do Sul | 181.3 | 194.1 | 180.0 | 182.1 | 178.1 | 183.1 |
| **Midwest** | **183.4** | **190.6** | **187.9** | **184.2** | **185.5** | **186.3** |
| Mato Grosso do Sul | 168.5 | 180.7 | 177.3 | 173.6 | 179.5 | 175.9 |
| Mato Grosso | 194.6 | 204.9 | 211.5 | 206.1 | 214.3 | 206.3 |
| Goiás | 148.9 | 152.9 | 148.5 | 150.7 | 151.5 | 150.5 |
| Distrito Federal | 221.5 | 223.9 | 214.2 | 206.5 | 196.5 | 212.5 |

Source: 2018 Statistical Yearbook of Electric Energy/EPE (2018). Adapted by the authors

the highest tariff adjustment of electricity, accumulating a high of 44.04% (Figure 12).

The main explanation for this phenomenon is based on the specific conditions of the northern states of the country. According to ANEEL (2019), in this region the high dispersion of consumers, the presence of many isolated regions with high-cost local generation and the number of losses above the national average impact on energy costs, causing tariffs to rise.

Combining high tariffs with high consumption, residential consumers in the northern states are the ones with the highest electricity costs in the period analyzed, and they also have the least PV facilities in their homes (Figure 13).

It is essential to highlight that the results obtained in this research refer exclusively to the analyzed period and do not intend to portray trends.

Also, this study has some limitations regarding the amount and availability of data. The small sample size (n = 27) stratified by the FU; the use of old data, such as the HDI (2010); and unavailability of data, such as monthly electricity expenditure per household; may end up affecting the results of the correlation analysis.

**Figure 12:** Average range of energy tariff 2014-2017. In%, for residential consumers
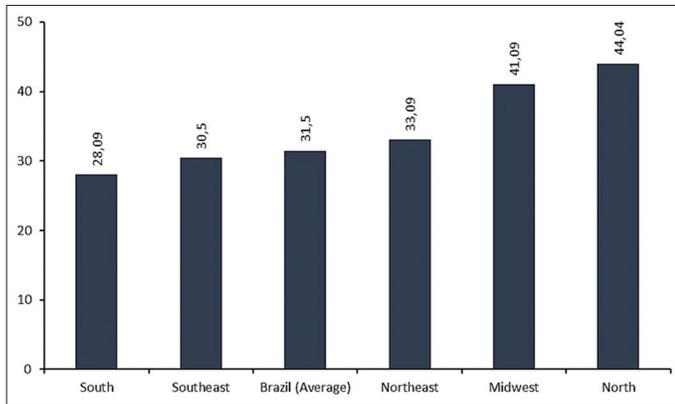


Source: Abrace (2018)

**Figure 13:** Average monthly energy expenditure and quantity of UFV per federative unit
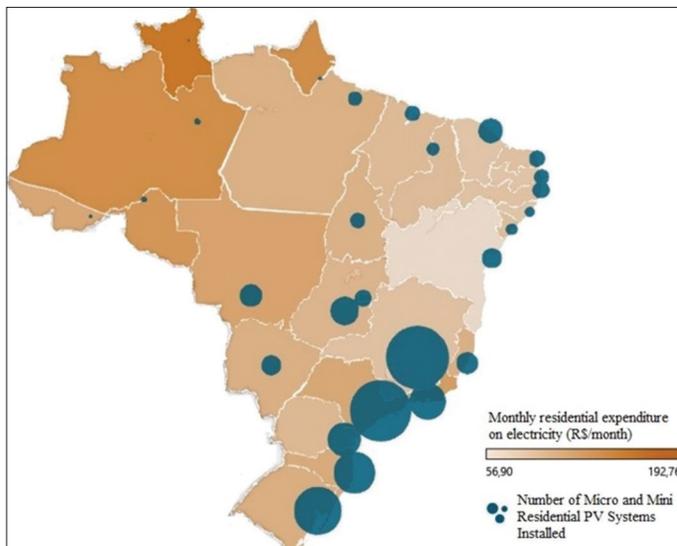


Source: EPE/ANEEL (2019). Elaborated by the authors

Future works may analyze the trend of insertion of PV systems in Brazil, and may also consider other variables in the analysis, such as the family income of owners of PV systems; the average size of installed systems; specific lines of financing; relationship with distributors; among others. Besides, data can be stratified by the municipality, which increases the sample size and reliability of the statistical analysis.

## 6. CONCLUSION

This article aimed to relate the insertion level of distributed generation PV with socioeconomic and geographic factors of each Brazilian state. Initially, we present the scenario of the last 7 years of the distributed generation PV in Brazil, and how the process of growth of this technology was uneven among the regions of the country.

Next, we established five hypotheses about the dynamics of this growth and tested these hypotheses through the statistical analysis

of Pearson's correlation coefficient. From this test, we refuted two hypotheses. First, it was found that the solar index is not related to the level of insertion of PV systems in the country, demystifying the discourse that abundance in solar irradiation is a fundamental factor in developing the PV energy sector. Other factors, such as public policies of socioeconomic development and appropriate incentives, are more effective in the process of expanding this technology than the high rate of solar resources (countries such as Germany and Japan are examples of this). Second, we found that places that spend the most on electricity are the places with the least PV facilities. This last result points to a characteristic process of uneven development in the country, once PV technology - which enables the reduction in the value of the electricity bill - ends up benefiting those states that already have moderate tariffs, thus widening inequality between regions.

It is essential to highlight that the trajectory of PV distributed generation in Brazil is still in its early stages and to analyze such recent events is inevitably uncertain and changing. However, this kind of analysis allows us to look at some almost imperceptible questions that, if not considered and adequately addressed, could lead to severe problems in the future.

Finally, we conclude that the level of technological insertion regarding residential PV distributed generation is related to the socioeconomic factors of each region of the country and that this new model needs guidelines that guide its development to maximize its economic and social potential, without promoting distortions that further intensify the existing inequality in the country.

## REFERENCES

Abrace. (2018), Levantamento da Associação Brasileira de Grandes Consumidores Industriais de Energia e de Consumidores Livres Feito a Pedido do G1. Available from: https://www.g1.globo.com/economia/noticia/conta-de-luz-acumula-alta-media-de-315-entre-2014-e-2017-diz-estudo.ghtml. [Last accessed on 2019 Jul 14].

Aguiar, A.C., Chagas, M., Cohen, C., Mendonça, R. (2007), O Papel Das Tarifas de Energia Elétrica na Queda da Desigualdade de Renda no Brasil. EconPapers. Available from: https://www.EconPapers.repec.org/RePEc: anp:en2007:149.

Almeida, M.S. (2011), Elaboração de Projeto, TCC, Dissertação e Tese: Uma Abordagem Simples, Prática e Objetiva. 1st ed. São Paulo: Atlas.

ANEEL, Agência Nacional de Energia Elétrica. (2018), Banco de Informação de Geração. Available from: https://www.app.powerbi.com/view?r=eyjrijoizjm4njm0owytn 2iwzs00yjviltllmjitn2e5mzbkn2 zlmzvkiiwidci6ijqwzdzmowi4lwvjytct ndzhmi05mmq0lwvhngu5yzaxn zblmsisimmiojr9. [Last accessed on 2018 Oct 23].

ANEEL, Agência Nacional de Energia Elétrica. (2019), Comissão de Minas e Energia: Câmara dos Deputados Requerimento. Brasília: Agência Nacional de Energia Elétrica.

Batistello, P., Balzan, K.L., Pereira, A.T. (2019), BIM no ensino das competências em arquitetura e urbanismo: Transformação curricular. PARC Pesquisa em Arquitetura e Construção, 10, e019019.

Correa, S.M. (2003), Probabilidade e Estatística. 2nd ed. Belo Horizonte, MG, Brasil: PUC Minas Virtual.

Creswell, J.W. (2009), Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, California: Sage.

Creswell, J.W. (2012), Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks: Sage.

Creswell, J.W., Plano, C.V. (2011), Designing and Conducting Mixed Methods Research. Los Angeles: SAGE Publications.

Dancey, C., Reidy, J. (2006), Estatística Sem Matemática Para Psicologia: Usando SPSS Para Windows. Vol. 3. Porto Alegre: Artmed.

EPE. (2014), Inserção da Geração Fotovoltaica Distribuída no Brasil Condicionantes e Impactos (Nota técnica DEA 19/14). Rio de Janeiro: Recursos Energéticos.

EPE. (2018), 2018 Statistical Yearbook of Electricity 2017 Baseline Year. Rio de Janeiro: Eastern Peripheral Expressway. Available from: http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-160/topico-168/Anuario2018vf.pdf.

Filho, D.B., da Silva Júnior, J.A. (2009), Desvendando os mistérios do coeficiente de correlação de pearson (r). Revista Política Hoje, 18, 115-146.

Fonseca, J.J. (2002), Metodologia da Pesquisa Científica. Fortaleza: UEC.

Garcez, C.G. (2017), Distributed eletricity generation in Brazil: An analysis of policy context, design and impact. Utilities Policy, 49, 104-115.

Gil, A.C. (2008), Métodos e Técnicas de Pesquisa Social. 6th ed. São Paulo: Atlas.

Greener. (2019), Estudo Estratégico: Mercado Fotovoltaico de Geração Distribuída Semestre. Available from https://www.greener.com.br/pesquisas-de-mercado. [Last accessed on 2019 May 11].

Guimarães, P.R. (2017), Análise de Correlação e Medidas de Associação. Universidade Federal do Paraná, Curitiba. Available from: https://www.docs.ufpr.br/~jomarc/correlacao.pdf.

IBGE. (2010), Cidades IBGE. Available from: https://www.cidades.ibge.gov.br/brasil/ac/panorama. [Last accessed on 2018 Nov 11].

INPE. (2017), Atlas Brasileiro de Energia Solar. São Paulo: São José dos Campos.

Kowaltowski, D.C., Neves, L.D. (2016), A compreensão do projeto como tópico de pesquisa. PARC Pesquisa Em Arquitetura E Construção, 12, 58-60.

Lange, W.J. (2012), Metodologia de Mapeamento da Área Potencial de Telhados de Edificações Residenciais no Brasil Para Fins de Aproveitamento Energético Fotovoltaico. Rio de Janeiro: EPE/Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

MDIC, SDCI, DECOI. (2018), Relatório Final Grupo de Trabalho Solar Fotovoltaico (05/12/2017a 05/03/2018). Ministério do Desenvolvimento Indústria e Comércio Exterior Secretaria de Desenvolvimento e Competitividade Industrial Departamento de Competitividade Industrial. Brasília: MDIC.

Miranda, R.F. (2013), Análise da Inserção de Geração Distribuída de Energia Solar Fotovoltaica no Setor Residencial Brasileiro. (Dissertação de Mestrado). Rio de Janeiro: Universidade Federal do Rio de Janeiro.

MME, EPE. (2018), Plano Decenal de Expansão de Energia 2027. Brasília: MME.

Moore, D.S. (2003), The Basic Practice of Statistics. 3rd ed. New York: W.H. Freeman and Company.

Mukaka, M.M. (2012), Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. Malawi Medical Journal, 24, 69-71.

Pino, F.A. (2014), A questão da não normalidade: Uma revisão. Rev de Economia Agrícola, São Paulo, 61(2), 17-33.

Rodrigues, S.C. (2012), Modelo de Regressão Linear e suas Aplicações (Dissertação de Mestrado). Covilhã. Portugal: Universidade da Beira Interior.

Taylor, R. (1990), Interpretation of the correlation coefficient: A basic review. The Journal of Defence Modelling and Simulation, 1, 35-39.