RECENT J. (2023), 71:221-227
https://doi.org/10.31926/RECENT.2023.71.221

Research on Correlation Analysis of Stainless Steel's Experimental Results

Ioan MILOSAN
Transilvania University of Brasov, Romania, milosan@unitbv.ro

Abstract
The analysis of industrial processes mainly uses fundamental statistical methods, which allow the drawing of conclusions from the observed values on the distribution of the frequency of various parameters, their interaction, as well as the verification of the validity of certain hypotheses. When studying the dependence between the parameters of a process, there should be determined whether the parameters initiating the process are independent of each other or influence each other (depend on each other), thus to determine the correlation between the parameters and to establish the type of the connection between parameters. In this paper, the Correlation Analysis is used to define and characterize the relationship between the process performance, Y, microhardness [HV_{100}] and variables X_1, carbon percentage, C [%], respectively X_2, chromium percentage, Cr [%], of the specific chemical composition of six stainless steels, used in industry. Based on the values obtained for the 3 simple correlation coefficients (r_{yx1}=0.98; r_{yx2}=0.85; r_{x1x2}=0.87) and for the multiple correlation coefficient ($r_{y,x1x2}$ =0.98), after applying the Student and Fischer Criteria, there is accepted the alternative hypothesis (H_1), asserting with a probability of 95% the existence of a correlation between the dependent variable Y and independent variables X_1 and X_2, an thus there can be stated that X_1 and X_2 are the technological parameters of the studied process, highly influencing it.

Keywords
correlation analysis, simple and multiple correlation, Pearson correlation coefficient (PCC), microhardness

1. Introduction

The correlation analysis can be used in experimental research to solve the problems within the preliminary experiment. It is used to estimate the statistical connections existing between different factors and the state variable(s). On this occasion, based on statistical criteria, there are selected the highly influencing factors and interactions, and they are considered in carrying out the basic experiment, being introduced as variables in the mathematical model of the process.

Correlation, in general, defines the interdependence (connection) between the observed variables of a process and, in particular, measures the degree of association between the variables. The intensity of the connection between the state variables can be measured using the correlation coefficient, also referred to as the "the Pearson correlation coefficient (r)", which measures the linear correlation between two sets of data [1, 2].

Being the most frequently used, the Person correlation coefficient (r) (linear correlation coefficient) refers to the degree and direction of simultaneous variation of the values of one variable in relation to the other, according to a linear model [3-5].

This paper uses the correlation analysis based on the Pearson correlation coefficient in order to determine whether the two analyzed factors (the carbon and chromium contents of the chemical composition of six stainless steels) influence or not the HV_{100} microhardness of these steels in the raw cast state, i.e. we analyze whether the two studied factors are technological parameters or not.

2. Research Objectives

The main objectives of this paper are to define and characterize the connection between the process performance, Y, microhardness [HV_{100}] and variables, the technological factors X_1, carbon percentage [%], respectively X_2, chromium percentage [%], of the specific chemical composition of six stainless steels.

In this case, there was calculated the simple correlation coefficient r_{xy} as well as the multiple correlation coefficient $r_{y.x1x2}$ establishing the three characteristics of the correlation:

1. direction: positive (+) or negative (-);

2. degree of association: between -1 and 1;
3. shape: linear or non-linear.

3. Simple and Multiple Correlation

3.1. Simple Correlation

In terms of the correlation between two variables $Y(y_1, y_2, ..., y_n)$ and $X(x_1, x_1, ..., x_n)$, the r_{xy} quantity is referred to as the simple correlation coefficient [4, 6] . In practice, the r_{xy} (the Pearson correlation coefficient, PCC) coefficient is calculated based on the experimental data of an "n" volume sample, using the expression [6]:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\,(S_x \cdot S_y)} \tag{1}$$

where: x_i = values of the x-variable in a sample; $\bar{x}$ = mean of the values of the x-variable; y_i = values of the y-variable in a sample; $\overline{y_i}$ = mean of the values of the y-variable; S_x = standard deviation of x_i, S_y = standard deviation of y_i ; n = sample quantity.

The Pearson correlation coefficient (r), PCC, takes values between r = –1 (perfectly negative correlation) and r = +1 (perfectly positive correlation). The absence of any connection (correlation) between the variables is shown by r = 0.

According to its classification [4], the values of the Pearson correlation coefficient (r) are as follows:

$r \subset [-0.10, +0.10]$, negligible correlation;
$r \subset [0.10, +0.39] \cup [-0.10, -0.39]$, weak correlation;
$r \subset [0.40, +0.69] \cup [-0.40, -0.69]$, moderate correlation;
$r \subset [0.70, +0.89] \cup [-0.70, -0.89]$, strong correlation;
$r \subset [0.90, +1.00] \cup [-0.90, -1.00]$, very strong correlation.

The plus or minus symbol of the linear relationship (the direction of the relationship) between two variables is interpreted as follows:

- if $r > 0$, the two variables have a direct proportional relationship, i.e. if the value of one of them increases, the value of the other also increases;
- if $r < 0$, the two variables have a reverse proportional relationship, i.e. if the value of one of them increases, the value of the other decreases.

The simple correlation coefficient is tested using the *Student Criterion* [1, 2], with the following hypotheses:

H_0 (null hypothesis): the X and Y variables are not related, the correlation coefficient, r, is not significant, and thus there is no relationship between the variables or there is only a random relationship;
H_1 (alternative hypothesis): the X and Y variables are related, the correlation coefficient, r, is significant with a probability of 95% (an assumed risk of 5%), and thus there is a significant relationship between the two variables, or, in other words, the correlation coefficient is statistically significant.

The *Student Criterion* [1, 2] is calculated according to the expression (2):

$$t_c = \frac{|r_{yx}|\sqrt{\nu}}{\sqrt{1 - r_{yx}^2}} \tag{2}$$

where: r_{yx} is the simple correlation coefficient, Pearson (PCC); ν is the number of degrees of freedom, $\nu = n - 2$; n is the number of experimental determinations.

The *p* result of the test is the probability of making an error if rejecting the H_0 hypothesis of the test, which is a number between 0 and 1. If p is lower than the selected α significance threshold (α = 0.05 is the most used value), the H_0 hypothesis is rejected and the H_1 hypothesis is accepted.

The *p* values are interpreted as follows: $p < 0.05$, the statistical connection is significant (95% confidence); $p < 0.01$ the statistical connection is significant (99% confidence); $p < 0.001$, the statistical connection is highly significant (99.9% confidence); $p > 0.05$, the statistical relationship is insignificant.

If $t_c > t_T$, there is accepted the hypothesis of a correlation connection between the two variables, and the correlation connection is absent otherwise.

There is defined t_T [1, 2] the tabular value of the criterion Student ($t_T = t_{0.05;\,\nu=n-2}$); 0.05 is the value of α (statistical coefficient of the confidence level used).

3.2. Multiple Correlation

In terms of the correlation between a (Y) process state variable and the (X_1, X_2, ..., X_k) factors acting on it, this is a multiple correlation and the $r_{y \cdot x1x2,\dots xk}$, quantity is referred to as multiple correlation coefficient [1, 2, 7-10].

In this case, as well, the calculations are carried out in practice based on the experimental data of an "n" volume sample and [1, 2] are determined according to the expression (3):

$$r_{y \cdot x1x2\dots xk} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}} \tag{3}$$

where: $\bar{y}$ is the arithmetic mean of the experiment values y_i; $\tilde{y}$ is the calculated value of the Y variable (at point i) using the regression equation.

In the particular case when one Y dependent variable and two X_1, X_2 independent variables are considered, the multiple correlation coefficient can be determined at sample level, based on the simple correlation coefficients between paired variables using the expression [1, 2]:

$$r_{y \cdot x1x2} = \sqrt{\frac{r_{yx1}^2 + r_{yx2}^2 - 2r_{yx1}r_{yx2}r_{x1x2}}{1 - r_{x1x2}^2}} \tag{4}$$

The significance of the multiple correlation coefficient is tested using the Fischer's criterion [1, 2] with the following hypotheses:

H0 (null hypothesis): the independent variables X_1, X_2, ..., X_k do not influence the state variable Y or, in other words, the X_1, X_2, ..., X_k factors are not technological parameters;

H1 (alternative hypothesis): there is accepted, with a 95% probability (an assumed risk of 5%), the existence of a correlation between the Y dependent variable and the X_1, X_2, ..., X_k group of independent variables.

Fischer's criterion [1, 2] is calculated according to the expression (4):

$$F_c = \frac{n-k-1}{k} \cdot \frac{r_{y \cdot x1x2\dots\dots xk}^2}{1 - r_{y \cdot x1x2\dots xk}^2} \tag{4}$$

where: *n* is the number of determinations; *k* is the number of independent variables.

If $F_c > F_T$, there is accepted with a 95% probability (1-α) the hypothesis of the existence of a correlation between the Y dependent variable and the X_1 , X_2,...,X_k group of independent variables. If $F_c \leq F_T$, there is considered that the independent variables X_1, X_2, ..., X_k do not influence the state variable Y or, in other words, the X_1, X_2, ..., X_k factors are not technological parameters; There is defined F_T the tabular value of the *Fischer Criterion*, $F_T = F_{(0.05;\, \nu1=k;\, \nu2=n-k-1)}$, where ν_1 and ν_1 are the number of degrees of freedom for the *Fischer Criterion* [1, 2].

4. Experimental Procedure

The correlation analysis assesses the connection between the Y process performance, the microhardness [HV_{100}] and the process variables, the technological factors X_1, the percentage of carbon [%], respectively X_2, the percentage of chromium [%], in the specific chemical composition of six industrial stainless steels, for which the simple and multiple correlation coefficients were determined, according to the data in Table 1. The microhardness values [HV_{100}] were determined on six cylindrical specimens, Ø10 mm in diameter and 10 mm thick, and three parallel determinations were carried out for each specimen.

The analysis of the data presented in Table 1 reveals as follows:

- the 1.4848 and 1.4837 steel grades have a 100% austenitic structure, with an average hardness of 195.67 HV_{100} (the 1.4848 steel), respectively 203 HV_{100} (the 1.48437 steel);
- the 1.4312 and 1.4408 steel grades have a mainly austenitic structure with 5% ferrite, and an average hardness of 188.33 HV_{100} (the 1.4312 steel), respectively 189 HV_{100} (the 1.4408 steel);
- the 1.4136 and 1.4776 steel grades have a mainly austenitic structure with 10% ferrite, and an average hardness of 203 HV_{100} (the 1.4136 steel), respectively 208.33 HV_{100} (the 1.4408 steel).

Additionally, simple and complex chromium carbides can also be found in all steel grades.

Table 1. Data used in the correlation analysis

No.	Steel grade	C [%] (X1)	Cr [%] (X2)	Process performance, hardness [HV_{100}]				Raw cast structure**
				Y_1	Y_2	Y_3	(Y*)	
1	EN 1.4312	0.13	17.50	186	190	191	182.33	A+5% Fe
2	EN 1.4408	0.18	16.85	204	199	222	189.00	A+5% Fe
3	EN 1.4848	0.40	22.10	204	192	191	195.67	A
4	EN 1.4136	0.45	22.60	193	171	183	203.00	A+10% Fe
5	EN 1.4837	0.50	25.00	200	201	208	203.00	A
6	EN 1.4776	0.64	22.50	200	201	208	208.33	A+10% Fe

Y* is the process performance obtained as the average of the 3 determinations (Y_1, Y_2, Y_3);
** Determined using the Schaeffler diagram; A = austenite; Fe = ferrite

The microhardness was measured using the FM 700 Microhardness Tester, applying the Vickers method with a 100 gf load. The tests were carried out at the C08 Research Centre – Advanced Metal, Ceramic and Composite Materials and Technologies – within the Research and Development Institute of the Transilvania University of Brașov.

5. Calculation of the Pearson Correlation Coefficient (PCC)

5.1. Calculation of the Simple Pearson Correlation Coefficient

The Pearson coefficients specific to the simple correlation are calculated with the expression (1) using the SPSS 23.0 software, obtaining the results shown in Figures 1, 2, 3 and Table 2.

Descriptive Statistics

	Mean	Std. Deviation	N
C[%]	0.3833	0.19480	6
Cr[%]	21.0917	3.20756	6
Hardness [HV_{100}]	196.8883	9.81093	6

Fig. 1. Descriptive statistic (Mean and Std. Deviation) of the data used in specific calculations

Correlations

		Hardness [HV_{100}]
C[%]	Pearson Correlation	0.976**
	Sig. (2-tailed)	0.001
	N	6
Cr[%]	Pearson Correlation	0.850*
	Sig. (2-tailed)	0.032
	N	6

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Fig. 2. The Pearson correlation coefficient (r_{yx1}, r_{yx2})

Correlations

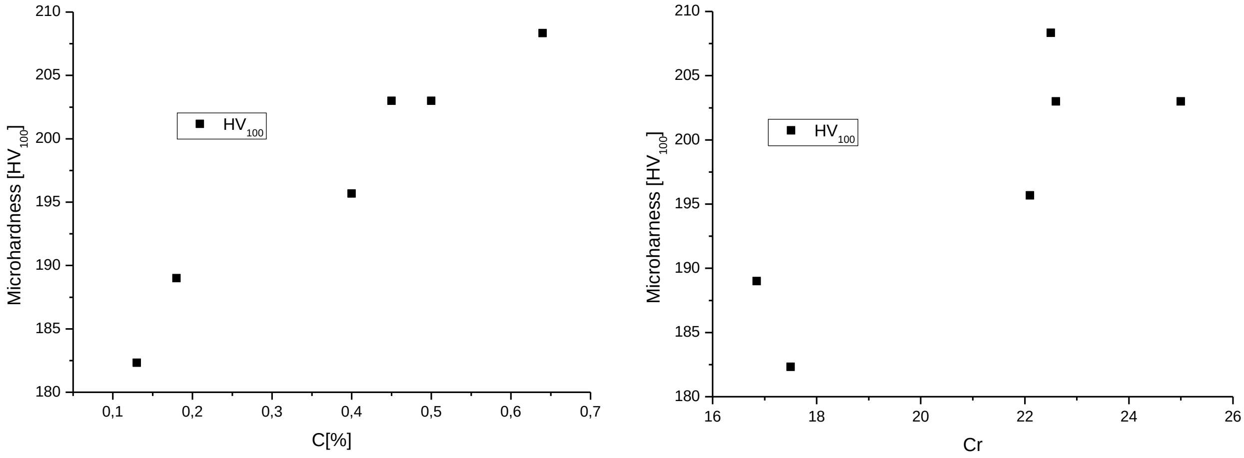
		Cr[%]
C[%]	Pearson Correlation	0.867*
	Sig. (2-tailed)	0.026
	N	6

* Correlation is significant at the 0.05 level (2-tailed)

Fig. 3. The Pearson correlation coefficient r_{x1x2} for pairs of variables used in this study

An important step in analyzing correlations between two quantitative variables is the shape of the scatterplot [1, 11]. The scatterplot, shown in Figure 4 is plotted to identify the existence of a dependency relationship between the analyzed variables, as well as the shape and direction of the dependence relationship.

The data cumulated from Figures 2, 3 and 4 is shown in Table 2.



a) b)

Fig. 4. Scatterplot: a) Y = f(X_1); b) Y = f(X_2); a) The connection between the percentage of carbon, C[%], and the microhardness of steels [HV_{100}]; b) The connection between the percentage of chromium, Cr [%], and the microhardness of steels [HV_{100}]

Table 2. Data cumulated: CPP, strength and direction of Simple Correlation

No.	Variable1	Variable2	CCP (r) value	p-value	Strength	Direction	Conclusions
1	Process performance, (Y)	C [%], (X_1)	r_{yx1} = 0.98	0.001	Strong	Positive	Very strong positive correlation
2	Process performance, (Y)	Cr[%], (X_2)	r_{yx2} = 0.85	0.032	Strong	Positive	Strong positive correlation
3	C[%], (X_1)	Cr[%], (X_2)	r_{x1x2}=0.87	0.026	Strong	Positive	Strong positive correlation

By analyzing the data presented in Figures 2, 3 and 4 and Table 2, there can be interpreted as follows:

- ✓ When one variable changes, the other variable changes in the same direction.
- ✓ In the case of the research presented, the following interpretations can be made:
  - C [%] & Hardness [HV]: as the percentage of carbon C [%] increases, the hardness [HV] also increases;
  - Cr [%] & Hardness [HV]: as the percentage of chromium Cr [%] increases, the hardness [HV] also increases;
- ✓ There was a strong, positive correlation between C [%] and microhardness [HV_{100}], which was statistically significant (r_{yx1}=0.98; n=6; p=0.001).
- ✓ There was a strong, positive correlation between Cr [%] and microhardness [HV_{100}], which was statistically significant (r_{yx2}=0.85; n=6; p=0.032).
- ✓ There was a strong, positive correlation between C [%] and Cr [%], which was statistically significant (r_{x1x2}=0.87; n=6; p=0.026).

The significance of the simple correlation coefficient (presented in Table 2) is tested using the *Student Criterion*. Using the expression (2) [1, 2] and taking into account the number of degrees of freedom, ν = 4 (ν = n-2; in this case, ν = 6-2 = 4), the following results are obtained:

$$t_{c\ ryx1} = \frac{|r_{yx}|\sqrt{\nu}}{\sqrt{1-r_{yx}^2}} = 9.849;\ t_{c\ ryx2} = \frac{|r_{yx}|\sqrt{\nu}}{\sqrt{1-r_{yx}^2}} = 3.226;\ t_{c\ rx1x2} = \frac{|r_{yx}|\sqrt{\nu}}{\sqrt{1-r_{yx}^2}} = 3.529.$$

The tabular value of the *Student Criterion* ($t_T = t_{0.05;\ \nu=n-2}$); the value of $t_T = t_{0.05;\ \nu=4}$ is chosen from the tables [1, 2] as being t_T =2.776 and this value is compared to the previously calculated values, resulting in the following data: $t_{c\ ryx1} > t_T$ (9.849>2.776); $t_{c\ ryx2} > t_T$ (3.226>2.776); $t_{c\ rx1x2} > t_T$ (3.529>2.776).

Since all the calculated values specific to the Student criterion are greater than the tabular value of this criterion, there is accepted the alternative hypothesis (H_1), thus accepting with a probability of 95%

the existence of a correlation between the Y dependent variable (process performance, HV_{100}) and the group of independent variables X_1 (C[%]) and X_2 (Cr[%]); according to [1, 2], it follows that the three simple correlation coefficients, r_{yx1}, r_{yx2} and r_{x1x2}, are significant.

5.2. Calculation of the Multiple Pearson Correlation Coefficient

The Pearson coefficients specific to the multiple correlation are calculated using the expression (4), and there is obtained the value of the multiple correlation coefficient $r_{y \cdot x1x2}$:

$$r_{y\cdot x1x2} = \sqrt{\frac{(0.98)^2 + (0.85)^2 - 2\,(0.98 \cdot 0.85 \cdot 0.87)}{1-(0.87)^2}} = \sqrt{0.960} = 0.98$$

The obtained value of the multiple correlation coefficient shows a strong positive correlation between C [%], Cr [%] and microhardness [HV_{100}], ($r_{y\cdot x1x2} = 0.98$; n = 6).

The significance of the multiple correlation coefficient ($r_{y\cdot x1x2} = 0.98$) is tested using the Fischer criterion. By using the expression (4), knowing that n = 6 and k = 2 (the number of independent variables: X_1 and X_2), the following result is obtained:

$$F_c = \frac{n-k-1}{k} \cdot \frac{r^2_{y\cdot x1x2\ldots xk}}{1-r^2_{y\cdot x1x2\ldots xk}} = \frac{6-2-1}{2} \cdot \frac{(0.98)^2}{1-(0.98)^2} = \frac{3}{2} \cdot \frac{0.9604}{0.0396} = 36.379$$

The tabular value of the Fisher criterion, $F_T = F_{(0.05;\nu1=k;\,\nu2=n-k-1)} = F_{(0.05;\,2;\,3)} = 9.55$ [1, 2].

Since $F_c > F_T$ (36.379> 9.55), there is accepted the alternative hypothesis (*H1*), specifying that:

- there is accepted with a probability of 95% the existence of a correlation between the dependent variable Y (process performance, HV_{100}) and the group of independent variables X_1 (C [%]) and X_2 (Cr [%]), and thus the X_1 and X_2 variables are correlated with the process performance Y;
- the X_1 (C [%]) and X_2 (Cr [%]) factors are the technological parameters of the studied process, having a significant influence thereon.

6. Conclusion

The analysis of all data taken into account leads to the following conclusions:

a) The following results were obtained from the Simple Pearson Correlation Coefficient calculation:

- ✓ $r_{yx1} = 0.98$, and it can be stated that there was a very strong, positive correlation between C [%] and the microhardness [HV_{100}], which was statistically significant;
- ✓ $r_{yx2} = 0.85$, and it can be stated that there was a strong, positive correlation between Cr [%] and the microhardness [HV_{100}], which was statistically significant;
- ✓ $r_{x1x2} = 0.87$, and it can be stated that there was a strong, positive correlation between C [%] and Cr [%], which was statistically significant;
- ✓ The testing of the simple correlation coefficient using the Student Criterion resulted in the following values: $t_{c\,ryx1} = 9.849$; $t_{c\,ryx2} = 3.226$; $t_{c\,rx1x2} = 3.529$; all three of which are greater than the tabular value of the criterion, $t_T = 2.776$;
- ✓ Based on the results obtained, there is accepted the alternative hypothesis (*H_1*), stating as follows:
  - There is accepted with a probability of 95% the existence of a correlation between the dependent variable Y (process performance, HV_{100}) and the group of independent variables X_1 (C[%]) and X_2 (Cr[%]);
  - The three simple correlation coefficients, r_{yx1}, r_{yx2} and r_{x1x2}, are significant.

b) When calculating the Multiple Pearson Correlation Coefficient, $r_{y\cdot x1x2}$, the following aspects are noted:

- ✓ The value obtained for the coefficient was $r_{y\cdot x1x2} = 0.98$, showing a very strong positive correlation between C [%], Cr [%] and the microhardness [HV_{100}].
- ✓ When testing the significance of the multiple correlation coefficient using the Fischer Criterion, there is observed that the calculated value is greater than the tabular value of this criterion, $F_c > F_T$ (36,379> 9.55);
- ✓ Based on the data obtained, there is accepted the alternative hypothesis (*H_1*), ascertaining with a probability of 95% the existence of a correlation between the dependent variable Y (process performance, HV_{100}) and the group of independent variables X_1 (C[%]) and X_2 (Cr[%]), and there can be stated as follows:

- the X_1 and X_2 variables are correlated with the process performance Y;
- the X_1 (C [%]) and X_2 (Cr [%]) factors are the technological parameters of the studied process, having a significant influence thereon.

References

1.Taloi D., Bratu C., Florian E., Berceanu E. (1983): *Optimizarea proceselor metalurgice* (*Optimization of metallurgical processes*). Editura Didactiă şi Pedagogică, Bucureşti, pp. 76-96

2. Taloi D. (1987): *Optimizarea proceselor tehnologice. Aplicaţii în metalurgie* (*Optimization of technological processes. Applications in metallurgy*). Editura Academiei Române, Bucureşti, pp. 162-184

3. Senthilnathan S. (2019): *Usefulness of Correlation Analysis*. https://dx.doi.org/10.2139/ssrn.3416918

4. Schober, P., Boer C., Schwarte L. A. (2018): *Correlation Coefficients: Appropriate Use and Interpretation*. Anesthesia & Analgesia, ISSN 1526-7598, Vol. 126, p. 1763-1768

5. Raju N. S., Brand P. A. (2003): *Determining the significance of correlations corrected for unreliability and range restriction*. Applied Psychological Measurement, ISSN 1552-3497, Vol. 27, p. 52–71

6. Makowski D., Ben-Shachar M. S., Patil I., Lüdecke D. (2020): *Methods and Algorithms for Correlation Analysis in R*. Journal of Open Source Software, ISSN 2475-9066, Vol. 5, is. 51, https://doi.org/10.21105/joss.02306

7. Sherry A., Henson R.K. (2005): *Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer.* Journal of Personality Assessment, ISSN 1532-7752, Vol. 84, is. 1, pp. 37-48, https://doi.org/10.1207/s15327752jpa8401_09

8. Algina J., Olejnik S. (2003): *Sample Size Tables for Correlation Analysis with Applications in Partial Correlation and Multiple Regression Analysis*. Multivariate Behavioral Research, ISSN 0027-3171, Vol. 38, is. 3, pp. 309-323, https://doi.org/10.1207/S15327906MBR3803_02

9. Tsai M.Y. (2015): *Comparison of concordance correlation coefficient via variance components, generalized estimating equations and weighted approaches with model selection.* Computational Statistics & Data Analysis, eISSN 1872-7352, Vol. 82, pp. 47-58, https://doi.org/10.1016/j.csda.2014.08.005

10. Wang Y., Yao G., He Y. (2015): *Research on Correlation analysis of industry electricity quantity*. Proceedings of International Conference on Information Technology and Management Innovation (ICITMI 2015), ISBN 978-9-4625-2112-4, ISSN 2352-538X, pp. 906-912, DOI:10.2991/icitmi-15.2015.153

11. Miot H.A. (2018): *Correlation analysis in clinical and experimental studies*. Jornal Vascular Brasileiro, ISSN 1677-7301, Vol. 17, is. 4, pp. 275–279, doi: 10.1590/1677-5449.174118