STUDY OF PREDICTION POSSIBILITY FOR THE SELECTED MECHANICAL PROPERTY OF DRAWN WIRE

Silvie ADAMCOVÁ, Darja NOSKIEVIČOVÁ

VSB - Technical University of Ostrava, Ostrava, Czech Republic, EU, darja.noskievicova@vsb.cz

Abstract

This study deals with finding the predictive equation to determine the values of the selected mechanical property for drawn wire. The initial prerequisite is to improve (in this case decrease) the values of selected mechanical property through carried out drawing opdwell ation. Based on the data obtained for the values of this mechanical property before and after drawing, the measure of dependence has been examined and tested between variables using correlation analysis. Furthermore a suitable function for description of found dependency using regression analysis will be determined and tested. The aim of identifying the prediction equation is to guarantee (with the selected probability) values of mechanical property only from the metallographic test results before drawing without performing metallographic tests after drawing.

Keywords: Prediction, correlation analysis, regression analysis, drawn wire

1. INTRODUCTION

In cases where the actual testing and measuring is impossible or difficult to implement, it is necessary to seek other alternatives to prove (guarantee) the values of mechanical properties. One of these options is the prediction using regression analysis based on real data from given process. This requires to have mapped out process, know the laws of statistics and be able to interpret the results. This article will therefore be devoted to the issue of prediction of selected mechanical property in order to guarantee these values without the need for testing after the final step of drawing process.

Regression and correlation analysis of the data currently rank among the most widely used statistical methods. These methods are used in many fields like research, economics, health, industry etc. Their aim is to describe the statistical properties of the relationship between two or more random variables. The main advantages of regression and correlation analysis include the ability to detect and quantify the functional relationships between variables. Regression analysis also allows to test various hypotheses about these relationships. It is also used for predicting the consequences of certain decisions.

2. THEORETICAL ASSUMPTIONS

2.1 Correlation analysis

Correlation analysis is concerned with assessing and quantifying the degree of dependence between two or more random variables. The intensity of these relationships may take the form from the complete independence to a complete (functional) dependence. Relationships between variables are explored graphically and using different measures of dependence, which are called correlation coefficients. Statistical inference of the correlation coefficient is based on the theory of probability for the joint distribution of two or more random variables [1]. For more information on the correlation coefficients see the literature [2] [3].

2.2 Regression analysis

Unlike correlation analysis, the regression analysis is in terms more accurate in shape description of the relationship between variables and demonstrates its suitability for predicting the values of the dependent variable using the values of the independent variable. Generally it therefore analyses the relationship between the dependent variable also called regresand (Y) and one or more independent variables, called regressors (X). The relationship is represented by a mathematical model, which is the equation that binds regresand with regressors and probabilistic assumptions, which should satisfy the relationship. The legitimate use of linear regression model is justified if examined regression relationship approximately meets the following requirements:

- linear character between the variables X and Y;
- for the entire range of considered x the value of the residual standard deviation $\sigma_{y,x}$ is constant;
- the y_i values have a normal distribution for given values of clearly identified x_i and they are stochastically independent;
- ε_i variables (i.e. residues of the regression model) are normally distributed with zero mean value: $\varepsilon_i \sim N(0,\sigma^2)$ [4].

A more detailed analysis of regression analysis and its assumptions can be found in the literature [5].

3. STUDY OF THE PREDICTION POSSIBILITY OF SELECTED MECHANICAL PROPERTY

3.1 Input data

There were examined 80 steel grade C40 pairs of samples with four different dimensions to compile the predictive equation. On these samples the test of selected mechanical property before and after drawing process was performed in an accredited laboratory. Measurements were performed by only one laboratory worker to minimize scattering.

For the purpose of this study the input assumption is dependence of the mechanical property before and after drawing. This dependence will be further examined using correlation analysis. The target value, i.e. the value of the mechanical property after drawing, may also influence the size of material and reduction used during drawing operation. However, this last factor will not be, in this study, included in prediction equation, due to more complex and time-demanding data collection. After completion of data collection, the entire study will be updated even for this factor. The chemical composition of the material in this case is not considered because it is already included in the result of mechanical property value before drawing. For the above mentioned reasons, the test results of given mechanical property before and after drawing, and the material dimension were included into source data.

All the statistical analyzes were performed at a significance level α = 0.05 and processed in the statistical program Statistica 12.

3.2 Exploratory data analysis

There was conducted exploratory data analysis of individual variables to obtain initial information about data position, variations of values and characteristics of distribution. The results of this analysis are summarized in **Table 1**. Data file position is on the real line characterized by the average, median and mode. Variation in values of the chosen mean is characterized by the standard deviation, variation range and coefficient of variation. Distribution of random variable is described by characteristics of skewness and kurtosis.

Table 1 A summary of descriptive statistics for the input data

Descriptive statistics	Variable		
	The value of mechanical property after drawing	The value of mechanical property before drawing	Dimension
Number of values	80	80	80
Average	0.06993	0.08404	14.275
Median	0.07100	0.08600	13.635
Mode	multiple	0.0840000	1
Frequency	6	6	20
Sum	5.594	6.723	1142
Minimum	0.04000	0.04000	11.44
Maximum	0.10000	0.12000	18.39
Range	0.060000	0.080000	6.95
Quartile range	0.022000	0.023000	4.23
Dispersion	0.000173	0.000282	6. 817646
Standard deviation	0.013141	0.016792	2.611062
Coefficient of variation	18.79256	19.98208	18.291153
Skewness	-0.205098	-0.443195	0.644447
Standard Error	0.268909	0.268909	0.268909
Kurtosis	-0.48077	-0.06863	-1.030506
Standard Error	0.531786	0.531786	0.531786

As it is seen in **Table 1**, skewness and kurtosis values of regresand (value of mechanical property after drawing) and two regressors vary in the interval from -2 to 2. This could point to an approximate normal distribution of these variables. Also, all the values of the mechanical property vary in desired interval from 0 to 0.12.

3.3 Normality and homogeneity verification of the input data

There have been used exploratory graphs in combination with testing hypotheses to verify the normality and homogeneity of data. To verify the normality histogram with fitted curve of empirical distribution and p - chart (better known as the Q - Q plot) were drawn, for verification of the homogeneity the box plot was used. Shapiro-Wilk and Kolmogorov-Smirnov tests have been used to test the normality. Grubbs test has been used to test the homogeneity of the data. The analysis results are shown in Figures 1 to 3 for the particular variables.

From the analysis of Figures 1-3, it is clear that the first two variables are derived from a normal distribution with 95% probability. P-value mentioned in both tests is greater than 0.05, therefore the null hypothesis on the data normality is confirmed. Normality is obvious from the exploratory graphs for the first two variables (value of mechanical property after drawing and the value of the mechanical property before drawing). The graphs for the variable "dimension" shows that the data do not come from normal distribution. Data do not show a continuous character. These conclusions are also confirmed by the results of the tests, where p-value is less than 0.05. So for this variable we reject the null hypothesis and accept the alternative one. Homogeneity of the data was confirmed, for all three variables using box plots, and the results of Grubbs test, where the p-value was in all cases greater than 0.05.

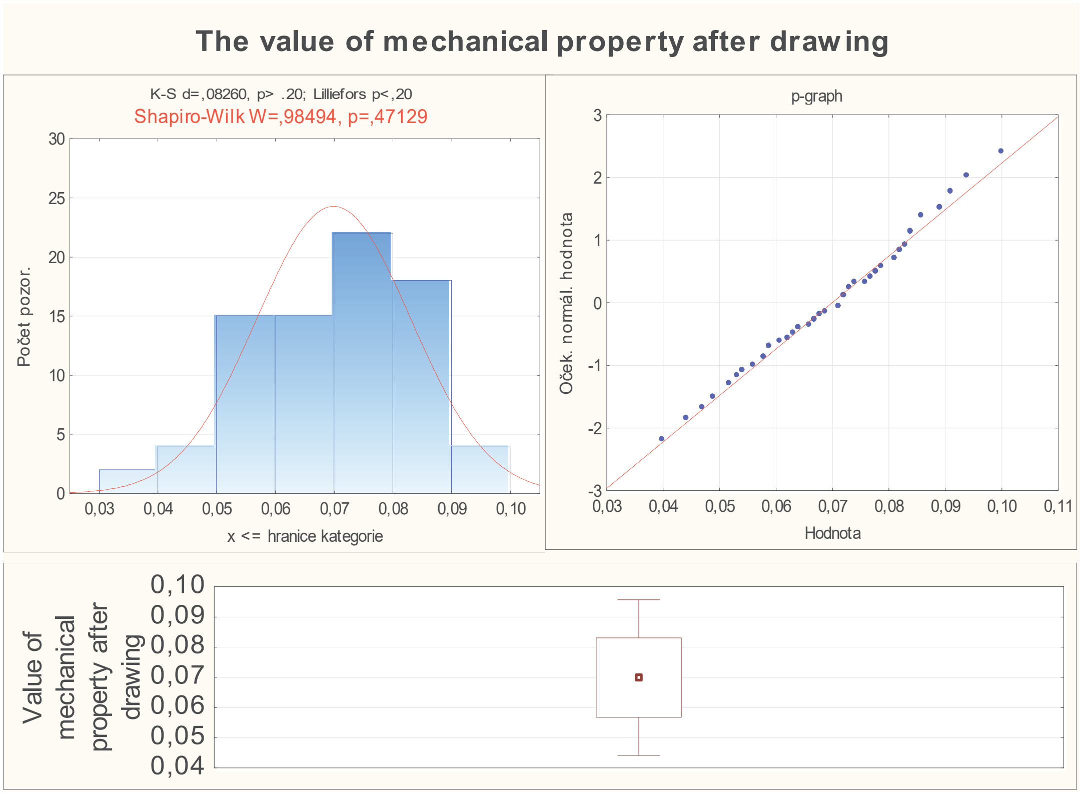

Fig. 1 Normality and homogeneity verification for the variable „Value of mechanical property after drawing"“

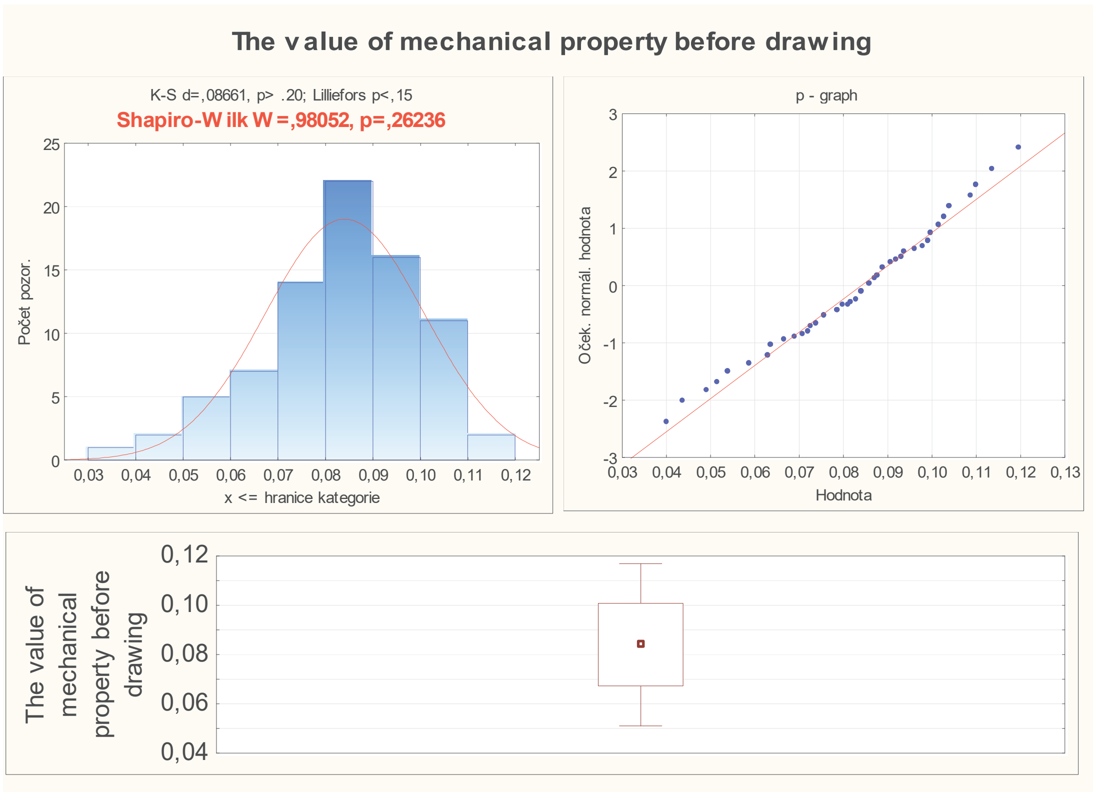

Fig. 2 Normality and homogeneity verification for the variable "Value of mechanical property before drawing"

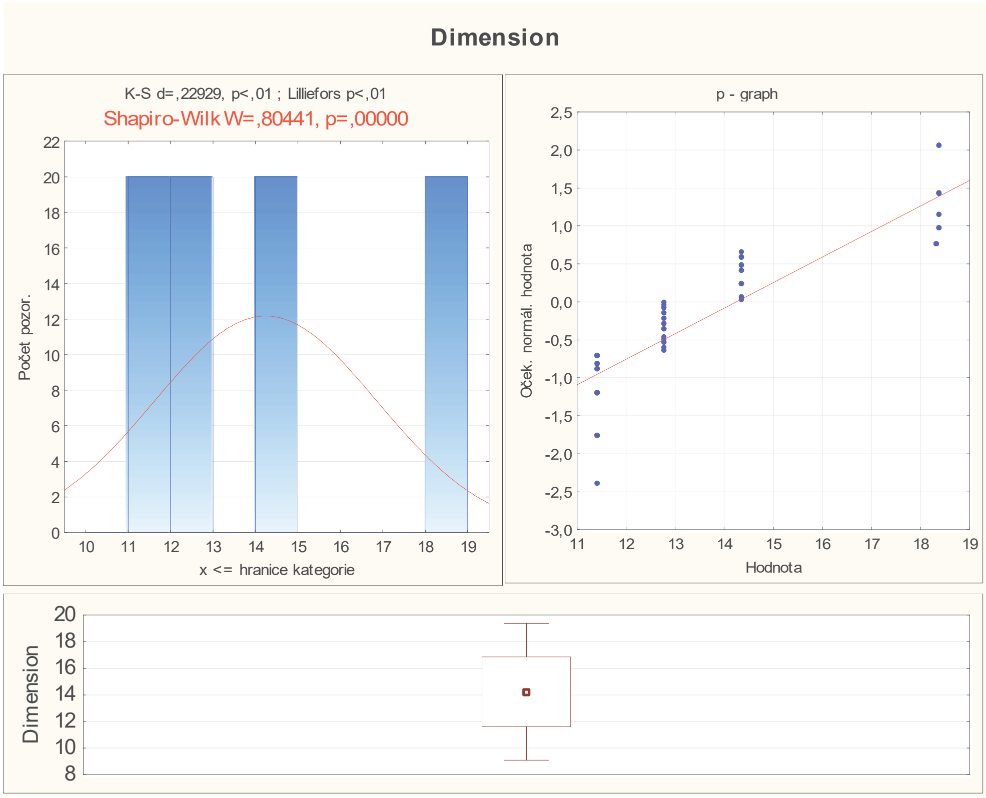

Fig. 3 Normality and homogeneity verification for the variable “Dimension”

On the basis of this analysis, we have verified the input assumptions to build the regression equation. Although for the variable "Dimension" has not been demonstrated normality (which is given by the data character, the analysis includes only 4 values of dimension, namely 4 recurrent values), this variable has been included into the regression equation.

3.4 Correlation between individual variables

For identification dependency between regresand and regressors the correlation analysis of the variables was performed. In particular the correlation coefficient r has been calculated (see **Table 2**) and scatterplots for visual verification of the dependence were constructed (**Fig. 4** for one pair of variables). Red marked correlation coefficient values in **Table 2** represent statistically significant correlations. Values vary from 0.402 till 0.832. All observed dependencies are therefore positive. The strongest correlation is between the variables „Value of mechanical property after drawing“ and „The value of mechanical property before drawing“.

Table 2 The values of correlation coefficient between individual variables

Variable	Coefficient of correlation N=80		
	The value of mechanical property after drawing	The value of mechanical property before drawing	Dimension
The value of mechanical property after drawing	1.00000	0.83220	0.40251
The value of mechanical property before drawing	0.832195	0.00000	0.54607
Dimension	0.40251	0.54607	1.00000

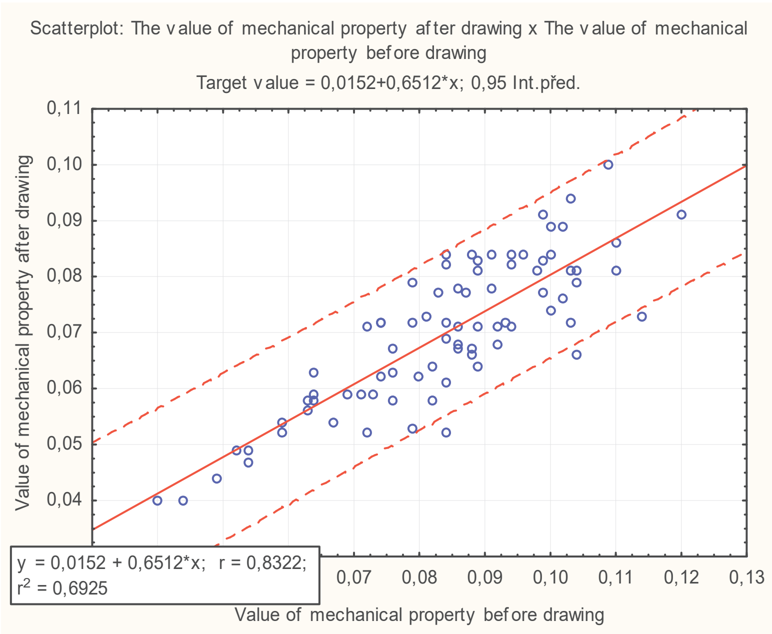

Fig. 4 Correlation between the regresand and regressor „The value of mechanical property before drawing“

4. REGRESSION EQUATION

Based on previous analyses the multiple linear regression (**Table 3**) and analysis of variance ANOVA (**Table 4**) has been performed. Regresand (Y) is the variable "Value of mechanical property after drawing" and regressors (X) are variables „The value of mechanical property before drawing“ and „Dimension“.

Table 3 The result of multiple linear regression

	R= 0.83450017 R2= 0.69639054 F(2.77)=88.308 p<0,0000 Standard error : 0.00733					
N=80	b*	St. error (b*)	b	St. error (b)	t(77)	p-value
Absolute term			0.017821	0.004979	3.57945	0.0006
The value of mechanical property before drawing	0.872596	0.074955	0.682839	0.058655	11.64155	0.00
Dimension	-0.073986	0.074955	-0.000371	0.000376	-0.90706	0.32670

Table 4 ANOVA

	Analysis of variance				
Effect	Sum of squares	sv	Average of squares	F	p-value
Regression	0.009500	2	0.004750	88.30764	0.000000
Residuals	0.004142	77	0.000054		
Sum	0.013642				

From the results (**Table 3** a **4**) it is apparent that regression model as a whole is statistically significant. The coefficient of determination, i.e. the value of R2 describes the proportion of the total variability in dependent variable is explained by given model. In this case, the coefficient of determination equals 0.6963. This means that the target value of given mechanical property is dependent on the given variables by 69%. The remaining 31% is the influence of other factors. Statistically significant regression coefficients are given in **Table 3** and marked in red. This means that into the regression model we include absolute term and a variable „The value of mechanical property before drawing“.

The regression equation thus has the form:

Value of mechanical property after drawing = 0.017821 + 0.682839*Value of mechanical property before drawing + e

In the final step of the analysis assumptions on residues were validated. These assumptions have been discussed in Chapter 2.2. From **Fig. 5** it is apparent that the residues of the model are approximately constantly scattered around zero mean with no obvious systematic dependency or tendencies. The mean value of the residues equals to zero, i.e. E (e) = 0

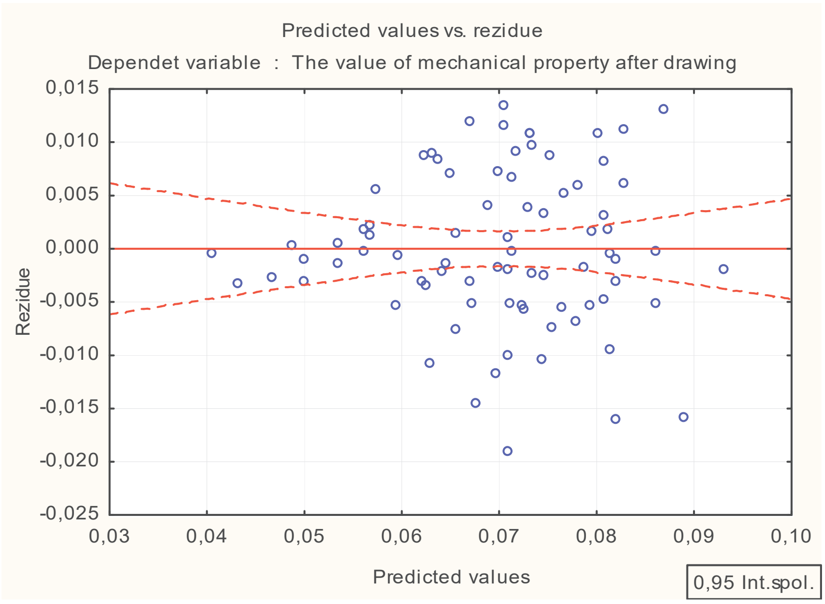

Fig. 5 Graf of scattering residues

Furthermore the normality residue has been verified using tests and exploratory graphs, namely a histogram with fitted curve of normal distribution and p-graph (see **Fig. 6**). P-value of Kolmogorov-Smirnov and Shapiro-Wilk testing is bigger than 0.05 and therefore the residues may be regarded as data coming from a normal distribution. Therefore the prerequisites of regression analysis have been met.

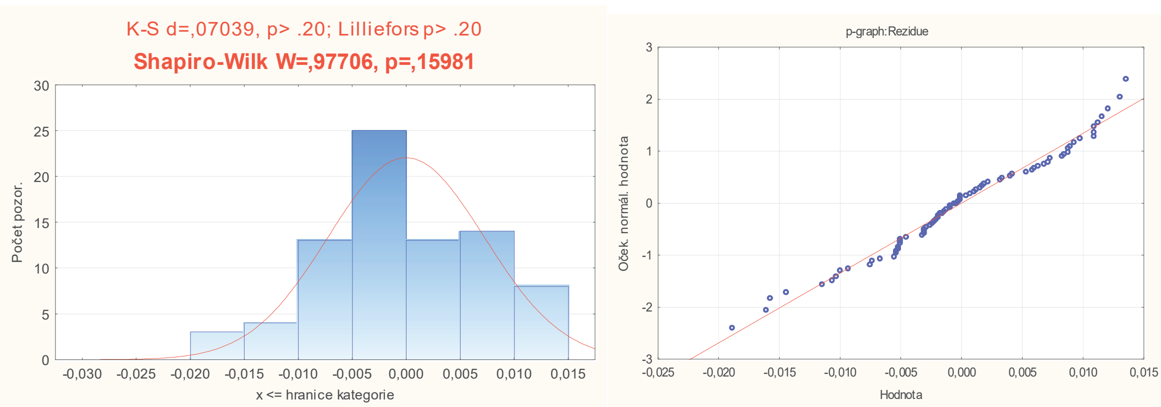

Fig. 6 Histogram and p-graph of residues

CONCLUSION

The aim of this study was to develop a predictive equation for determining the value of selected mechanical property. The correlation and regression analysis has been applied and assumptions validated for the use of these analyses, which were, with one exception met. It was a failure to comply with the normality of the variable „Dimension“. This variable, however, have proved to be statistically insignificant for the proposed predictive model. The resultant predictive equation in the form Value of mechanical property after drawing = 0.017821 + 0.682839*Value of mechanical property before drawing + e will be applied to guarantee the target values of mechanical property and the prediction results will be compared with the real values of mechanical property after drawing. After testing then the predictive equation in the case of obtaining satisfactory results will be used to guarantee the values of the mechanical property after drawing based on the values of mechanical property before drawing.

ACKNOWLEDGEMENTS

This paper was elaborated in the frame of the specific research project SP2014/81, which has been solved at the Faculty of Metallurgy and Materials Engineering, VSB***-TU Ostrava with the support of Ministry of Education, Youth and Sports, Czech Republic.***

REFERENCES

[1] BUDÍKOVÁ, Marie, KRÁLOVÁ, Maria, MAROŠ, Bohumil. Průvodce základními statistickými metodami. 1. vyd. Praha: Grada, 2010, 272 s. ISBN 978-80-247-3243-5.

[2] DAVID M. HIMMELBLAU. *PROCESS ANALYSIS BY STATISTICAL METHODS*. NEW YORK: JOHN WILEY, 1970. ISBN 471-39985-X. KOVÁČ, J., KUPKA, A. Effect of Silicon on the Magnetic Properties of Sheet. *Steel sheets*, 1998, year 49, nr. 11, page 1

[3] MONAHAN, John F. *Numerical methods of statistics*. 2nd ed. New York: Cambridge University Press, c2011, xvi, 447 p. ISBN 978-052-1191-586.

[4] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. 3., přeprac. vyd. Praha: Portál, 2009, 695 s. ISBN 978-80-7367-482-3.

[5] MONTGOMERY, Douglas C. *Introduction to statistical quality control*. 6th ed. Hoboken, N. J.: Wiley, c2009, xiv, 734 p. ISBN 978-047-0169-926.