The Importance of Data in the Modeling Process of the Realized Net Coal Production Level

Jagoda LINCZOWSKA[1], Barbara KOWAL[2], Marek KĘSEK[3], Paweł PIWOWARCZYK[4]

[1] student; AGH University of Krakow, Poland, email: jagodalin@student.agh.edu.pl
[2] Ph.D., DSc, Eng.; AGH University of Krakow, Poland, email: bkowal@agh.edu.pl, ORCID: 0000-0003-4643-1140
[3] Ph.D., DSc, Eng.; AGH University of Krakow, Poland, email: kesek@agh.edu.pl, ORCID: 0000-0001-6217-8435
[4] Południowy Koncern Węglowy S. A.; Główny inżynier, Kierownik Działu Inwestycji i Przygotowania Produkcji, email: komunikacja@pkw-sa.pl

http://doi.org/10.29227/IM-2025-01-20

Submission date: 29-05-2025 | Review date: 15-06-2025

Abstract

Data analysis and model building are important elements of enterprise management, especially in mining, where the risk of conducting business is high. The article presents the procedure of building an econometric model that presents the realized level of net coal extraction. The procedure of model construction was carried out using one of the data selection methods, namely the graph method. Although the process of selecting variables for the model is correct, it is not sufficient for the correctness of the model itself. The example given shows what problems may arise and answers the question of whether modeling alone is a sufficient process to describe the analyzed phenomenon.

Keywords: *coal mining, data analysis, modeling, data collection, data quality, production process*

Introduction

The hard coal mining process is complex and includes several stages from planning mining operations, through access and preparatory works (drilling shafts and workings enabling access to deposits), reinforcement and liquidation works (equipping previously completed roadway workings with longwall complex equipment and haulage of ore together with power supply installations), operational works (the period of putting the mining wall into operation), ventilation, haulage of ore and mechanical processing of coal (sorting, cleaning and preparation for further sale or transport) [1, 2, 3]. Each stage aims to obtain raw material from seams usually located at great depths. The entire process is carried out in compliance with safety rules and with the use of ventilation systems that ensure appropriate air circulation and gas removal. The exploitation of workings is most often carried out using the longwall method, which involves mining coal using mining combines and transporting it to the surface by conveyor belts.

Coal extraction level modeling is a process of creating analytical tools that allow for forecasting its volume [4]. A key element of this process is data analysis, which enables the identification of patterns, risk assessment, or optimization of activities. The use of modern methods, such as data mining or machine learning, allows not only for precise forecasts but also for effective support of the decision-making process [5]. Thanks to this, mine management and management institutions can make more informed strategic, operational, and environmental decisions, increasing efficiency and reducing the risk associated with mining activities.

The work aims to identify the elements influencing coal extraction and to create an econometric model for the realized net coal extraction level based on data obtained from the mining enterprise, as well as to show the problems that may arise and to answer the question of whether modeling alone is a sufficient process to describe the analyzed phenomenon. The article focuses on a fragment of the coal extraction process, namely, from starting the shearer to transporting coal by belt conveyor. The specific objectives were: to conduct the procedure of building an econometric model and check its correctness.

Data and methodology

Data

The analysis used mining data provided by Południowy Koncern Węglowy S.A. and information obtained from Internet resources (Table 1). The obtained data are monthly. Data from 16 months (May 2023 – August 2024) were analyzed. The data (variables X1-X13) refer to the coal mining process that was carried out in wall number 03 in seam 510 at a depth of 900 m. As mentioned earlier, the data concern only a fragment of the process. Variable X14 is the average coal price (USD/t) in a given month, calculated based on the highest and lowest value in a given month [6].

Methodology

Modelling of the realised net coal production level was carried out in several stages (Fig. 1).

After collecting the data, a two-step selection was carried out. First, it was checked whether the variables have a fundamental impact on the modeling of the studied phenomenon. For this purpose, a preliminary selection was carried out using the classic coefficient of variation:

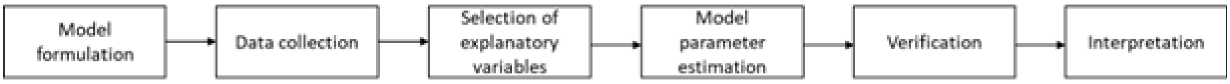
$$V_j = \frac{s_j}{\bar{x}_j}, \tag{1}$$

Where:
s_j – standard deviation of variable X_j,
$\bar{x}_j$ – arithmetic mean of the variable X_j.

Tab. 1. Wykaz zmiennych wziętych do modelowania. Źródło: [7]

Tab. 1. List of variables taken for modeling. Source: [7]

Variables	Description
Y	realized net coal production level (t)
X1	number of shearer starts
X2	number of longwall conveyor starts
X3	number of stage loader starts
X4	number of conveyor systems starts
X5	shearer operating time (min)
X6	longwall conveyor operating time (min)
X7	stage loader operating time (min)
X8	conveyor system operating time (min)
X9	shearer downtime (min)
X10	longwall conveyor downtime (min)
X11	beam stage conveyor downtime (min)
X12	conveyor system downtime (min)
X13	numer of rock mass tremors
X14	average coal market price (USD/t)



Rys. 1. Etapy budowy modelu. Źródło: opracowanie własne

Fig. 1. Model building stages. Source: own study

$$\mathbf{R_0} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}$$

a) b)

Rys. 2. Wektor korelacji R₀ (a) oraz w macierzy współczynników korelacji R (b). Źródło: [9]

Fig. 2. The correlation vector R₀ (a) and the correlation coefficient matrix R (b). Source: [9]

This allowed us to reject variables that were not sufficiently differentiated and that met the dependence $V_j<V^*$, where the value $V^*=30\%$ was assumed [8]. For variables that were characterized by sufficient differentiation, the strength of the linear relationship between them (explanatory variables Xi) and the explained variable Y was determined. To assess this strength, the Pearson correlation coefficient r was used according to the formula:

$$r_i = \frac{\sum_{t=1}^{n}(y_t-\bar{y})(x_{ti}-\bar{x_i})}{\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2\sum_{t=1}^{n}(x_{ti}-\bar{x_i})^2}}\ (i = 1,2, \dots, m) \tag{2}$$

The coefficients are presented in the correlation vector R_0 and the correlation coefficient matrix R (Fig. 2), respectively.

In the second stage of selection, i.e., selection of variables for the model, the graph method was used. The main assumption of this method is to select explanatory variables that are weakly connected to each other and strongly connected to the explained variable. The critical value of the correlation coefficient r* was calculated using the minimax rule:

$$r^* = \min_i \max_j |r_{ij}|, \text{ for } j \neq i. \tag{3}$$

Then, the elements of the R matrix satisfying the dependence $|r_{ij}| \leq r^{\wedge *}$ were removed, considering these connections as insignificant. Based on the new R' matrix, a graph of connections between variables (having non-zero correlation coefficients) was constructed. In the next step, variables were selected for the model, and its verification was carried out.

Results

Data Variation

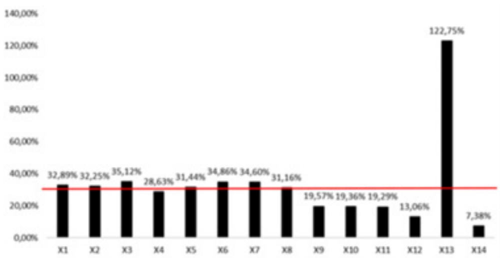
The initial selection allowed us to select variables that are statistically differentiated from 14 explanatory variables. The critical value of the coefficient of variation V* was assumed to be 30%. The level of the coefficient of variation for individual variables is presented in Fig. 3.

From the set of 14 variables considered, 6 variables were eliminated, for which the coefficient of variation obtained values lower than the critical value. The variables rejected in the initial selection were: X4 – number of conveyor system starts, X9 – shearer downtime (min), X10 – longwall conveyor downtime (min), X11 – beam stage conveyor downtime (min), X12 – conveyor system downtime (min), X14 – average coal market price (USD/t). The remaining 8 variables are considered to be suitably differentiated and will be included in further analysis. It can be seen that variable X13 – number of rock mass tremors – is characterized by very high variability.

The strength of linear dependence

For the variables that remained after the initial selection, the correlation vector R_0 and the correlation coefficient matrix R were created (Fig. 4).

All values of the correlation coefficient of the explanatory variables are characterized by a positive and relatively high strength of dependence on the explained variable Y, the realized level of net coal extraction (correlation coefficients above 0.5). For one variable, the number of shearer starts, the correlation coefficient with Y is lower and amounts to 0.49 in rounding. In turn, the highest strength of connection with the variable Y is characterized by the variable X13 – the number of rock tremors. Looking at the strength of the dependence between the explanatory variables themselves, it is positive (except for the variable X13 – the number of rock mass tremors, for which the correlation coefficient with the variable number of shearer starts is approximately -0.06. A negative value of the coefficient indicates an inverse relationship be-



Rys. 3. Współczynniki zmienności dla zmiennych objaśniających. Źródło: opracowanie własne

Fig. 3. Coefficients of variation for explanatory variables. Source: own study

$$R_0 = \begin{vmatrix} 0.50896 \\ 0.488678 \\ 0.590946 \\ 0.57786 \\ 0.639004 \\ 0.631999 \\ 0.672779 \\ 0.683303 \end{vmatrix}$$

a)

$$R = \begin{vmatrix} 1 & 0.778952 & 0.757115 & 0.830721 & 0.803951 & 0.782957 & 0.559248 & -0.0556 \\ 0.778952 & 1 & 0.938603 & 0.861581 & 0.840341 & 0.842352 & 0.663245 & 0.161416 \\ 0.757115 & 0.938603 & 1 & 0.911445 & 0.89979 & 0.902967 & 0.800966 & 0.338202 \\ 0.830721 & 0.861581 & 0.911445 & 1 & 0.990215 & 0.988429 & 0.864661 & 0.210944 \\ 0.803951 & 0.840341 & 0.89979 & 0.990215 & 1 & 0.997878 & 0.87909 & 0.307315 \\ 0.782957 & 0.842352 & 0.902967 & 0.988429 & 0.997878 & 1 & 0.899817 & 0.321147 \\ 0.559248 & 0.663245 & 0.800966 & 0.864661 & 0.87909 & 0.899817 & 1 & 0.466289 \\ -0.0556 & 0.161416 & 0.338202 & 0.210944 & 0.307315 & 0.321147 & 0.466289 & 1 \end{vmatrix}$$

b)

Rys. 4. Wektor korelacji R_0 (a) oraz w macierzy współczynników korelacji R (b) dla zmiennych wystarczająco zróżnicowanych. Źródło: własne

Fig. 4. The correlation vector R_0 (a) and the correlation coefficient matrix R (b) for sufficiently differentiated variables. Source: own study

$$R' = \begin{vmatrix} 1 & 0.778952 & 0.757115 & 0.830721 & 0.803951 & 0.782957 & 0.559248 & 0 \\ 0.778952 & 1 & 0.938603 & 0.861581 & 0.840341 & 0.842352 & 0.663245 & 0 \\ 0.757115 & 0.938603 & 1 & 0.911445 & 0.89979 & 0.902967 & 0.800966 & 0 \\ 0.830721 & 0.861581 & 0.911445 & 1 & 0.990215 & 0.988429 & 0.864661 & 0 \\ 0.803951 & 0.840341 & 0.89979 & 0.990215 & 1 & 0.997878 & 0.87909 & 0 \\ 0.782957 & 0.842352 & 0.902967 & 0.988429 & 0.997878 & 1 & 0.899817 & 0 \\ 0.559248 & 0.663245 & 0.800966 & 0.864661 & 0.87909 & 0.899817 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

Rys. 5. Macierz R'. Źródło: opracowanie własne

Fig. 5. Matrix R'. Source: own study

tween the variables, i.e., in the case of an increase in the number of tremors, the level of extraction decreases.

Selection of variables for the model

To select variables for the model, the graph method was used. The critical value of the correlation coefficient r* calculated according to formula 3 was 0.466, and the correlation coefficients that satisfied the inequality $|r_{ij}| \leq r^*$, were removed from the R matrix, considering them as insignificant connections. The new R' matrix is presented in Fig. 5.

In this way, information was obtained that variable X13 is an isolated variable, i.e., a variable that has no connections with the remaining explanatory variables (zeroed correlation coefficients, Fig. 5). Then, a graph of connections between variables was constructed (Fig. 6).

It can be seen that all variables (except X13) are characterized by the same number of connections. Therefore, it was necessary to select the variable that is most strongly correlated with the explained variable Y. According to the previously determined Pearson r correlation coefficients, included in the correlation vector R0, the variable that showed the strongest connection (not taking into account the isolated variable) with Y is variable X8. Its strength of dependence and connection with Y was 0.673 (Fig. 4a). Therefore, the variables that were selected by the graph method for the model of variable Y are X13 and X8.

Estimation of model parameters

For variables selected for the model using the graph method (Section 3.3), i.e., x8 and X13, the model parameters were estimated. The LINEST function of MS Excel was used for this purpose (Table 2).

Verification

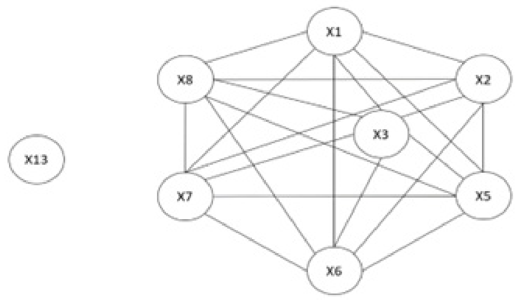
The model verification was based mainly on determining the quality of the model fit to the empirical data, i.e., on the coefficient of determination R2. According to [10], the fit of the regression line to the data was divided into three categories: 0.6-0.8 means satisfactory, 0.8-0.9 good, 0.9-1.0 very good. Table 2 shows that the fit is approximately 63%. This indicates that variables X8 and X13 explain the level of net coal extraction by 63%. Regression analysis was also performed; the statistics are presented in Table 3.

The multiple R for the graph method was approximately 79%. To this extent, the theoretical values of the explained variable are related to the empirical values of this variable. On the other hand, the standard deviation of the residuals in this method is approximately 8,658 tons. Table 4 shows the theoretical and actual values of the explained variable obtained using the graph method.

A normality test was also performed for the residual components. Table 5 presents the results of the test. The obtained values of $e'_{n:n}$ and Fe_n' were rounded to three decimal places.

The calculations obtained in Table 5 indicate that the number of cells (K) is equal to 10. According to the table values for the Hellwig test, the assumed significance level of 0.05 and the given number of observations, the residual components do not have a normal distribution (the number of cells is not in the critical range of 3-8) [8].

Based on the e'n:n values from Table 5, a histogram was generated, which is shown in Fig. 7.



Rys. 6. Graf powiązań pomiędzy zmiennymi objaśniającymi. Źródło: opracowanie własne
Fig. 6. Graph of relationships between explanatory variables. Source: own study

Tab. 2. Estymacja parametrów modelu. Źródło: opracowanie własne
Tab. 2. Estimation of model parameters. Source: own study

a2	a1	b
106.0853	1.536534	31377.81
43.00019	0.649949	7689.357
0.627183	8658.09	#N/D!
R^2		

The histogram for the distribution of residuals of the graph method does not represent the Gaussian distribution in its structure, so it was confirmed that the distribution of residuals is not normal.

Additionally, statistical significance measures are used in model verification. To start the analysis of the significance of parameters, it is necessary to first state the null and alternative hypotheses:

$$H_0: B_j = 0, j = 0,1,\ldots,k. \qquad H_1: B_j \neq 0. \tag{5}$$

This hypothesis is put forward when there is a doubt about the structural parameters and their possible value equal to zero [8].

One of the measures of the significance of the parameters is the p-value, which reflects (for the Student t-test) the critical level of significance. The general level of significance that is adopted is p equal to 0.05. When the p-value is less than or equal to 0.05, the null hypothesis is rejected and the alternative one is selected [11]. Analysis of variance was performed (Table 6).

The measurement of the significance of structural parameters (Table 6) showed that the p value is less than 5% for both variables, so based on the assumptions from formula 5, it can be assumed that the alternative hypothesis is true and the structural parameters of the model are different from zero. The analyses carried out confirmed that the structural parameters of the model are different from zero, which indicates that they are statistically significant.

Interpretation of the model

The regression constant in the equation is 31,377.81. This means that when variable X8 and variable X13 take a value equal to zero, the net coal extraction level will be 31,377.81 tons.

The values One of the remaining regression parameters allows us to state that if the conveyor system operation time (X8) were increased by one minute, the hard coal extraction level would increase by 1.54 tons (in the case when variable X13 is constant). Accordingly, when variable X8 takes a constant value, an increase in the number of rock mass tremors (X13) by one unit will result in an increase in hard coal extraction by 106.09 tons.

This is an Interpretation resulting directly from the obtained model.

Discussion of the results

From a procedural point of view, the model was constructed correctly. However, although variables X8 and X13 entered the model, they do not have a direct impact on variable Y. The substantive value of the model is not consistent with industry and economic knowledge. For this reason, the model contains inaccuracies.

Variable X8 (conveyor system operation time) cannot be treated as a variable directly influencing the volume of extraction, unless it constitutes a bottleneck in the transport system. In turn, the dependence of coal extraction on the number of rock mass tremors (X13) obviously exists, but not in a direct and linear way. Rock mass tremors can affect extraction, for example, by temporarily suspending extraction in areas with a high risk of rock bursts, damaging the structure of excavations, causing additional safety work, reducing the length of the wall, or shortening the working time, restrictions on the intensity of extraction. The discussed relationship between these two variables is quite the opposite. Mining causes stresses in the rock mass, which can lead to the release of energy in the form of tremors. Especially intensive mining or mining of thick seams at great depths can increase the risk of seismicity. It cannot, therefore, be said that more tremors directly affect the amount of extraction, because it also depends on other factors such as local conditions (intensity, location), technology, and risk management. This is certainly an important element of risk management in mining.

In connection with the above, the model should not be considered correct in a situation when variables are incorrectly interpreted by the researcher. As an example, we can show the redundancy of information introduced by variables X5, X6, X7, X9, X10, X11 (Fig. 8), which in fact correspond to one parameter, which is the time of operation or standstill of the longwall complex elements.

The presented correlation coefficients for selected variables cover only 15 periods, because period 16 did not have

Tab. 3. Statystyka regresji dla metody grafowej. Źródło: opracowanie własne

Tab. 3. Regression statistics for the graph method. Source: own study

Regression Statistics for the Graph Method	
Multiple of R	0,791948666
R square	0,627182689
Fitted R-squared	0,56982618
Standard error	8658,090177
Observations	16

Tab. 4. Zestawienie wartości Y teoretycznych i rzeczywistych na podstawie metody grafowej. Źródło: opracowanie własne

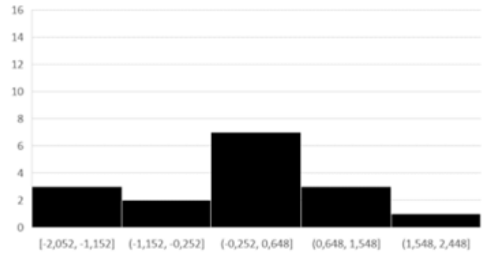
Tab. 4. Comparison of theoretical and actual Y values based on the graph method. Source: own study

Observation	**Expected Y**	**Residual ingredients**	**Real Y**
1	48868,88604	-10553,79832	38315,08771
2	55048,24573	7928,145701	62976,39143
3	54560,78818	6747,798158	61308,58633
4	66125,29033	-1185,526514	64939,76382
5	66402,65045	-9139,887596	57262,76286
6	77785,75616	2616,383519	80402,13968
7	68798,05302	4429,965484	73228,0185
8	53672,30573	2757,299102	56429,60483
9	53113,05083	-9973,083061	43139,96777
10	49805,59166	-16535,66204	33269,92962
11	58008,70605	7091,293947	65100
12	57269,19246	-3269,192463	54000
13	52191,45272	3608,547281	55800
14	51547,06483	2452,935168	54000
15	42287,58854	13512,41146	55800
16	34587,62982	-487,6298209	34100

Tab. 5. Rozkład reszt dla metody grafowej. Źródło: opracowanie własne

Tab. 5. Distribution of residuals for the graph method. Source: own study

Observation	e_t	$e_t' = e_t/s$	$e'_{n:n}$	Fe_n'	**Episode**
1	-10553,79832	-1,309364733	-2,052	0,020	[0;0,0625]
2	7928,145701	0,983611214	-1,309	0,095	[0,0625;0,125]
3	6747,798158	0,83717053	-1,237	0,108	[0,125;0,1875]
4	-1185,526514	-0,147083217	-1,134	0,128	[0,1875;0,25]
5	-9139,887596	-1,133946861	-0,406	0,343	[0,25;0,3125]
6	2616,383519	0,324603541	-0,147	0,442	[0,3125;0,375]
7	4429,965484	0,549606918	-0,060	0,476	[0,375;0,4375]
8	2757,299102	0,342086336	0,304	0,620	[0,4375;0,5]
9	-9973,083061	-1,237317868	0,325	0,627	[0,5;0,5625]
10	-16535,66204	-2,051509045	0,342	0,634	[0,5625;0,625]
11	7091,293947	0,879786587	0,448	0,673	[0,625;0,6875]
12	-3269,192463	-0,405594762	0,550	0,709	[0,6875;0,75]
13	3608,547281	0,447697066	0,837	0,799	[0,75;0,8125]
14	2452,935168	0,304325202	0,880	0,811	[0,8125;0,875]
15	13512,41146	1,676427243	0,984	0,837	[0,875;0,9375]
16	-487,6298209	-0,060498152	1,676	0,953	[0,9375;1]
Standard deviation of residuals [*s*]			8060,243303		



Rys. 7. Rozkład reszt dla metody grafowej. Źródło: opracowanie własne

Fig. 7. Distribution of residuals for the graph method. Source: own study

the same dimension (annual). As can be seen, the negative values of the correlation coefficients between X5, X6, X7 and X9, X10, X11 show the correct direction of the relationship between them, i.e. the shearer is either working or has a downtime (X5 and X9), the same in the case of the wall conveyor (X6 and X10) or the beam stage loader (X7 and X11). However, the values of all mutual correlation coefficients are high, which may indicate the previously mentioned redundancy of information.

Summary

Data analysis plays a key role in mining process modeling, as it allows for a better understanding of the complex relationships between factors that affect mining efficiency and profitability. Data analysis can identify patterns of relationships, which can lead to more realistic and accurate models. Such data analysis supports decision-making, such as planning mining schedules, resource allocation, and cost forecasting, minimizing risk, and improving mine operational efficiency.

The conclusions from the conducted analysis are as follows:

- information was obtained regarding explanatory variables that are significant and irrelevant to the model,
- among the significant variables (according to the model) that have a real impact on the achieved level of extraction were: conveyor system operating time

Tab. 6. Analiza wariancji dla metody grafowej. Źródło: opracowanie własne

Tab. 6. Variance analysis for the graph method. Source: own study

VARIANCE ANALYSIS

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1639402359	819701179.4	10.93481275	0.001639549
Residual	13	974512831.6	74962525.51		
Total	15	2613915190			

	Coefficients	*Standard error*	*t Stat*	*p-Value*	*Bottom 95%*	*Top 95%*	*Bottom 95,0%*	*Top 95,0%*
Cross-Cut	31377.8103	7689.357386	4.080680437	0.001299357	14765.96357	47989.65694	14765.96357	47989.65694
X8	1.53653402	0.649948782	2.364084774	0.034312751	0.132405042	2.940662994	0.132405042	2.940662994
X13	106.085306	43.00018579	2.467089491	0.028289206	13.18905287	198.9815601	13.18905287	198.9815601

Rys. 8. Korelacje między zmiennymi. Źródło: opracowanie własne

Fig. 8. Correlations between variables. Source: own study

	X5	X6	X7	X9	X10	X11
X5	1.000000	0.9794873	0.9739276	-0.8752729	-0.8357481	-0.8535367
X6	0.9794873	1.000000	0.9956199	-0.8509631	-0.8435689	-0.8637057
X7	0.9739276	0.9956199	1.000000	-0.8457283	-0.8393471	-0.8667521
X9	-0.875273	-0.8509631	-0.8457283	1.000000	0.985649	0.982301
X10	-0.835748	-0.8435689	-0.8393471	0.985649	1.000000	0.9959595
X11	-0.853537	-0.8637057	-0.8667521	0.982301	0.9959595	1.000000

(X8) and the number of rock mass tremors (X13),

- the model with the above data achieved a fit of almost 63%, which indicates a poor fit,
- the lack of normality in the distribution of residuals does not disqualify the model, but may indicate a violation of the model assumptions, which may result in the invalidity of the statistical test results,
- the structural parameters of the model are statistically significant but not substantively significant.

As you can see, data analysis and modeling are very important, although not a sufficient process for describing the analyzed phenomenon, as in the example provided. In addition to the mathematical and analytical approach, expert knowledge should be used, especially from people related to the industry or specialty that the modeling concerns. The article showed that the fundamental problem that may arise when creating a model is data and its quality, i.e. the degree to which it is useful for describing the phenomenon being studied (the explained variable). The example shows that this stage of data collection (but also its selection) is crucial for the subsequent stages of model construction. Data selected incorrectly may result in the final form of the model being substantively incorrect. Therefore, the data should be analyzed at the data collection stage to see if they have a direct impact on the analyzed phenomenon.

The inconvenience of the presented research is certainly the relatively small amount of data. However, they show how important and crucial the first stage of modeling is for the final form of the model. In the future, we plan to repeat similar research on a larger amount of data, as well as compare the obtained results with other similar walls or check the selection of variables for the model using other methods.

This paper was supported by the AGH University of Science and Technology [No. 16.16.100.215].

Literatura – References

1. Piechota, S. Technika podziemnej eksploatacji złóż, Część 1: Podstawowe zasady i technologie wybierania kopalin stałych, Biblioteka Szkoły Eksploatacji Podziemnej, Kraków 2003.
2. Brzychczy, E.; Trzcionkowska, A. New possibilities for process analysis in an underground mine [in:] Jonek-Kowalska, I. (ed.) Organization zna Management, Politechnika Śląska, Zeszyty Naukowe Nr 1987 , Gliwice 2017
3. Gawlik, L. Wpływ poziomu wydobycia węgla kamiennego na koszty jego pozyskania w kopalniach, Studia, Rozmprawy, Monografie 148; Wydawnictwo Instytutu Gospodarki Surowcami Mineralnymi i Energią PAN, Kraków 2008
4. Kęsek, M.; Ogrodnik, R. Method for determining the utilization rate of thin-deck shearers based on recorded electromotor loads, Energies 2021, 14(13), 4059; https://doi.org/10.3390/en14134059
5. Kęsek, M. The k-means grouping method as a mean to control the performance of the production process, Journal of the Polish Mineral Engineering Society 2020, vol. 1, no. 1, https://doi.org/10.29227/IM-2020-01-83
6. https://www.wnp.pl/gornictwo/notowania/ceny_wegla/, 05. 11.2024
7. Zmienne, dla których Południowy Koncern Węglowy S.A. udostępnił dane
8. Dziechciarz, J. (pr. zbior.). Ekonometria. Metody, przykłady, zadania. Wyd. 2 poprawione, Wydawnictwo Akademii Ekonomicznej im, Oskara Langego we Wrocławiu, Wrocław 2002
9. Nowak, E. Zarys metod ekonometrii. Zbiór zadań., Wydawnictwo Naukowe PWN, Warszawa 1997
10. Nowakowski, E. W. Podstawy ekonometrii z elementami algebry liniowej, Wszechnica Polska Szkoła Wyższa TWP w Warszawie, 2011
11. Górecki, B.R.: Ekonometria podstawy teorii i praktyki, Wydawnictwo Key Text, Warszawa 2010

Znaczenie jakości danych w procesie modelowania zrealizowanego poziomu wydobycia węgla netto

Analiza danych i budowanie modeli są ważnym elementem zarządzania przedsiębiorstwem, szczególnie wydobywczym, gdzie ryzyko prowadzenia działalności jest duże. W artykule przeprowadzono procedurę budowy modelu ekonometrycznego przedstawiającego zrealizowany poziom wydobycia węgla netto. Przeprowadzono procedurę konstrukcji modelu z wykorzystaniem jednej z metod selekcji danych, a mianowicie metody grafowej. Pomimo, że proces doboru zmiennych do modelu jest prawidłowy, to jednak niewystarczający dla poprawności samego modelu. Przytoczony przykład pokazuje jakie problemy mogą się pojawić oraz odpowiada na pytanie czy samo modelowanie jest wystarczającym procesem do opisu analizowanego zjawiska.

Słowa kluczowe: *wydobycie węgla, analiza danych, modelowanie, gromadzenie danych, jakość danych, proces produkcyjny*