Review Article

Correlation-regression analysis of the relationship between the volume of sales of goods and the use of rotary photo tables for their presentation in online stores

.

ABSTRACT

Aims: The aim of this article was to determine the existence of a relationship between the volume of sales of goods by an online store and the use of rotary photo tables for the presentation of goods, using mathematical methods and models.
Study design: In order to determine the relationship between the volume of sales of goods at the online store and the use of rotary photo tables for the presentation of goods, economic and mathematical modeling was carried out on the base of the data of online store "Ukrasa".
Place and Duration of Research: Russian Federation
Methodology: To achieve the goals, the research used such methods as analysis and synthesis, induction and deduction, comparison and grouping, methods of mathematical modeling (correlation-regression analysis).
Results: On the basis of mathematical methods, it was established that there is a relationship between the volume of sales of goods by the online store and the number of goods that are presented with the help of photos and videos made with the use of turntables for photography. As a result of the research, it was established that 84.78% of the total change in the volume of sales of goods is explained by the change in the presentation of goods on the site with the help of rotary photo tables.
Conclusion: The conducted research showed the importance of using rotary photo tables for presenting goods in online store. The economic interpretation of the parameters of the model shows that the increase in the number of goods presented with the help of photos and videos made with the use of rotary tables for photography will increase by one hundred units, then the volume of sales of goods will increase by 6,705 thousand hryvnias.

Keywords: Internet trade, rotary photo tables, product photography, scope of implementation, correlation, regression, influence.

1. INTRODUCTION

Product photography in modern conditions is gaining more and more popularity. Many stores go online and need high-quality photos that will help to show the products from all sides and in the most attractive light. It is advisable to use rotary photo tables for product photographs for this purpose. Under such conditions, the question of determining the relationship between the volume of sales of goods in the online store and the use of rotary photo tables for the presentation of goods becomes particularly relevant. It is expedient to carry out such research with the help of a mathematical model, which is considered to be the most effective

tool for the research of complex economic problems and allows modeling the main properties of real processes and phenomena.

The purpose of this research was to determine the existence of a relationship between the volume of sales of goods by an online store and the use of rotary photo tables for the presentation of goods, using mathematical methods and models, namely correlation-regression analysis for the assessment of economic processes and trends, as well as the development of a dependence model between the volume of sales of goods by the online store and the use of rotary photo tables for the presentation of goods.

Mathematical modeling actively penetrates into various spheres of human activity and enriches many sciences. In the field of economic processes, mathematical modeling plays a very important role. Since, mathematical models allow to describe essential correlation between economic processes and phenomena, to forecast economic indicators, to develop management strategies of economic objects [3].

The use of mathematical methods significantly expands modeling capabilities, allows formulating new tasks, and improves the quality of management decisions. Since economic-mathematical models simulate the main properties of real processes and phenomena with the help of various mathematical apparatuses, they are one of the most effective tools for researching complex economic problems.

The conducted research of literary sources showed that for today there is no research conducted with aim of determining the influence between the volume of sales of goods in the online store and the use of rotary photo tables for the presentation of goods, which explains the need for this research.

2. MATERIAL AND METHODS / EXPERIMENTAL DETAILS / METHODOLOGY

To achieve the set goals, the research used such methods as analysis and synthesis, induction and deduction, comparison and grouping, as well as methods of mathematical modeling and mathematical statistics. The information base of the research is made up of the works of recognized specialists, as well as data from the online store "Ukrasa".

Correlation-regression analysis is one of the most common and flexible methods of statistical data processing. This method originated in 1795, when the English researcher Francis Galton proposed the theoretical foundations of the regression method, and in 1801 he calculated the flight path of the planet Ceres with its help. He also introduced the term "correlation" (from English word "relation"), but the method was used at first in the 18th century by french paleontologist Georges Cuvier [14]. We can also name the French crystallographer Auguste Bravet, the German physicist Gustav Theodor Fechner, the English economist and statistician Francis Edgeworth, who were first to express their ideas about the quantitative measurement of correlation phenomena from the middle to the end of the 19th century. At different times, such well-known scientists in the field of theoretical statistics as Karl Friedrich Gauss (Germany), Andrian Marie Legendre (France), Karl Pearson (England) and others worked on the theory of analysis [10].

Correlation-regression analysis is based on the construction and analysis of an economic-mathematical model in the form of a regression equation (correlation relationship), which shows the dependence of a characteristic on the factors that determine it.

According to the correlation-regression analysis, the following stages are carried out:

• preliminary analysis (the main directions of the entire analysis are determined, the method of evaluating the performance indicator and the list of the most significant factors are determined);

• collection of information and its primary process;

• a model of correlation-regression dependence is built (one of the most important stages);

• assessment and analysis of the model.

The main tasks of correlation-regression analysis are reduced to the selection of the most important factors that influence the resulting characteristic, measurement of the closeness of the relationship between factors, identification of correlations factors, and assessment of factors that have the maximum impact on the result.

The tasks of regression analysis generally consist in establishing of the form of dependence, determining the regression equation and using it to estimate influence of the main values on the dependent variable, predicting the possible values of the resulting characteristic with the given values of factors of influence [5].
In order to confirm or refute the dependence between the volume of sales of goods by the online store and the use of rotary photo tables for the presentation of goods, a correlation-regression analysis was conducted.

It was expected that there is a linear relationship between the indicator of the volume of sales of goods by the online store and the use of rotary photo tables for the presentation of goods, which is expressed by the equation:

$$y = a + bx \quad (1)$$

where:
y - the value of the resulting characteristic (dependent variable);
x - the value of the factor characteristic (independent variable);
a - the start of the countdown, or the value of y at x=0 (has no economic meaning)
b - regression coefficient [9].
If *b*>0, then the correlation is direct, if *b*<0, then the correlation is inverse, if *b*=0, there is no correlation. The general equation of multiple linear regression has the form:

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \quad (2)$$

where y is the resulting variable:
$x_1, x_2 \ldots x_n$ – independent variables;
$a_0, a_1 \ldots a_n$ – parameters.
An equation of this type is called a regression equation, its main task is to establish a quantitative relationship between characteristics. The parameters of the equation a and b are determined by the method of least squares, which makes it possible to find such a theoretical regression line, which, in comparison with others, passes closest to the points of the correlation field of the actual data, that is, gives the smallest sum of squares of the deviations of the actual values of the resulting characteristic from the aligned (theoretical) values [1].

The procedure for obtaining a system of normal correlation equations is as follows. To obtain the first equation of the system, it is necessary to multiply all values of the original correlation equation by the coefficient for the first unknown (a) and add the obtained products. Then, to obtain the second equation, it is necessary to multiply all values of the original equation by

the coefficient for the second unknown (b) and also add all the products. The technique of obtaining a system of normal equations remains similar to the construction of a system of equations with a larger number of variables [11]. So, the system of correlation equations can be written in the following form:

$$\begin{cases} \sum y = an + b\sum x, \\ \sum yx = a\sum x + b\sum x^2 \end{cases} (3)$$

The correlation equations are used to calculate the theoretical regression line and the expected values of the dependent variables at the corresponding values of the factor (factors). Studying the correlation relationship, along with solving the regression equation, it became necessary to determine the level of closeness of the correlation between the factors, using a special relative indicator called the correlation coefficient.

3. RESULTS AND DISCUSSION

In recent years, there has been a steady trend towards the reorientation of the trade sphere into an online format.

The global pandemic has changed the attitude of users towards online trading, which has significantly affected the development of the e-commerce market. Mall closures and quarantine restrictions in 2020 forced shoppers to adopt new habits [4].

According to the [6], at the beginning of the pandemic, namely from February to April 2020, there was a decline in purchasing activity. This was due to the fact that many consumers were not yet accustomed to the new conditions of self-isolation and shopping in online stores. However, since April 2020, the situation has leveled off - the share of online purchases started to increase, and by the fourth quarter of 2020, the number of visits to online stores increased by almost 13% compared to the same period in 2019 [2].

It is noteworthy that in 2020 the share of buyers of online stores increased, but not the turnover of trade. That is, users switched to online under the influence of the pandemic, but did not buy more than before. But in 2021, the growth of both traffic and purchasing activity is observed. In 2021, the share of visits to online stores increased by 7%, and the turnover of electronic commerce increased by 34% compared to 2020 [13].

Quarantine restrictions have influenced consumer habits: people continue to buy not only food and beverages, but also electronics, jewelry, clothing, furniture, and more. Online trade has rapidly entered everyday life, thanks to which online stores have significantly improved their positions, as well as gained new customers - people of the older generation and people living in small settlements [19].

In 2021, consumer engagement metrics increased compared to last year — buyers began to spend more time on sites and visit more pages on one resource. At the same time, audience loyalty decreased: in 2020, 73% of users visited more than six sites of the same topic per day, and in 2021, this indicator increased to 76% [10].

It can be said that buyers began to browse the site even more carefully when choosing a product, because they want to study various offers before making a purchase [20].

Under such conditions, competition among online stores is steadily growing, so it is especially important to determine the factors that force the consumer to buy goods in the

particular online store. One of these factors can be the creation of a high-quality image of the product on the website using the rotary tables for product photography.

The product photography is becoming more and more popular every year, and the main reason for this phenomenon is the active development of the Internet business sphere, namely: internet stores. For trading in the "network", it is desirable to have high-quality photos of what needs to be sold. A rotary table for product photography greatly facilitates the process of creating photos and videos, and also improves the quality of images [8]. A rotary table or as it is also called "3D-table" consists of a disk that rotates and a fixed base on which it is fixed. The use of this device allows you to shoot products and to create GIF-animation, as well as obtain three-dimensional images of products. The advantages of such pictures are obvious, as a result, a potential buyer can look at the product he is interested in from all sides, which has a positive effect on the volume of sales.

To build a correlation-regression model, we chose the following factors:
x_1 - Number of product items presented in the store, in total;
x_2 - The number of items of the product presented in the store, with a description and photo;
x_3 - Number of items of the product presented in the store, presented with a description and photo using a rotary table, pcs.
The data for correlation regression analysis are presented in Table 1.

Table 1 . Data for carrying out a correlation-regression analysis of the influence of the relationship between the volume of goods sold and the use of rotary photo tables for their presentation in the online store "Ukrasa"

Year	The volume of sales of goods, thousand UAH.	The number of items of the product presented in the store		
		Total, pcs.	presented with description and photo, pcs.	presented with a description and photo using a rotary table, pcs.
	Y	*X1*	*X2*	*X3*
2015	131805	3980	3215	2890
2016	90615	2950	2770	2060
2017	93270	3100	2880	2236
2018	112154	4080	3200	2560
2019	130832	4400	3125	2670
2020	153781	4170	2990	2760
2021	155337	3900	3220	3098

Source: created by the author

We will analyze the relationship between the volume of sales of goods and the number of goods presented in the online store with only a description, with a description and a photo, as well as with a photo taken using a rotary photo table from 2015 to 2021 using the correlation coefficient. This coefficient characterizes the linear correlation (that is, the relationship that is given by a certain value and direction) of two or more variables.

To calculate the correlation, the Pearson formula is used:

$$r_{xy} = (\frac{1}{n-1}) \sum (\frac{x-\mu_x}{\sigma_x}) * (\frac{y-\mu_y}{\sigma_y}) \quad (4)$$

where:

r_{xy}- correlation coefficient;
n – number of periods;
x_n, y_n– values of variables x and y;
μ_x, μ_y– average arithmetic value of x and y;
σ_x, σ_y– standard deviations x and y.
The value of the correlation coefficient is between -1 and +1, that is, the correlation can be both positive and negative. When the value of the correlation coefficient is -1, there is a perfect negative correlation; when the value of the correlation coefficient is +1, there is a perfect positive correlation. In other cases, there is a positive correlation, a negative correlation, or no correlation between the two variables [12].

To determine the correlation between the selected indicators in the period from 2015 to 2021, we will calculate the value of the correlation indicators. The level of correlation of the factors selected for building the model with the resulting indicator is presented in Table 2.

Table 2. Correlation level of the factors selected for building the model of correlation regression analysis with the volume of sales of goods by the online store "Ukrasa"

Indicator		***Y***	***X 1***	***X 2***	***X 3***
Volume of sales of goods	*Y*	1			
The number of items of the product presented in the store	*X 1*	0.757583287	1		
The number of items of the product presented in the store with a description and photo	*X 2*	0.635940009	0.784897585	1	
The number of product positions presented in the store with photos using rotary photo tables	*X 3*	0.920778827	0.754676827	0.854582238	1

Source: created by the author

The value of the correlation coefficient is always in the range from -1 to +1, the closer the value is to 0, the less dependence of the indicators on each other, and accordingly, the closer the value is to 1, the higher the level of dependence [15]. Note that a correlation coefficient with a positive value means a direct relationship, and with a negative value, an inverse relationship [16]. A high degree of dependence is considered to be the value of the correlation coefficient (modulo) greater than 0.85.

The assessment of the closeness of the connection between the studied features was carried out using the Chaddock scale, the characteristics of the obtained ties are presented in the figure 1.

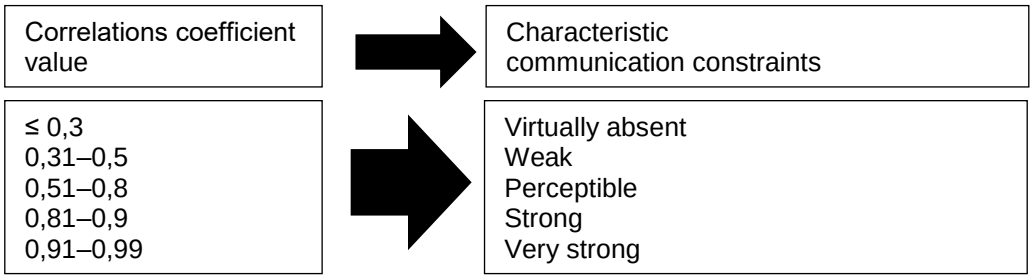

Fig. 1. Chaddock scale

Source: [17]

Based on this, it can be argued that the number of products presented using photos and videos made using rotary tables for photography has a significant impact on the volume of sales in the online store (table 3), since the correlation coefficient for this factor is greater than 0.85. So, to build a correlation-regression model, it is advisable to use the indicator of the number of items of the product presented in the store with photos using rotary photo tables.

Table 3. Characteristics of the closeness of the relationship between the performance indicator and factor characteristics

Indicator	Characteristic of correlation strength
The number of items of the product presented in the store	Notable
The number of items of the product presented in the store with a description and photo	Notable
The volume of foreign economic operations with the countries of the European Union	Strong

Source: created by the author

The significance check of the regression model is carried out using Fisher's F-Criterion, the calculated value of which is defined as the ratio of the variance of the original series of observations of the studied indicator and the unbiased estimate of the variance of the residual sequence for this model.

If the estimated value with $k_1=(m)$ and $k_2=(n-m-1)$ (m is the number of factors in the model) "degrees of freedom" is greater than the tabular value at the specified level of significance, then the model is considered significant. To assess the statistical significance of the paired linear regression, the null hypothesis is put forward that the equation as a whole is statistically insignificant: $R^2=0$ at the level of significance α [18].

Next, the actual value of the F- Criterion is determined by the formula:

$$F = \frac{R^2}{1-R^2} * \frac{n-m-1}{m} (5)$$

where:

m=1 for paired regression.

The tabular value (F tab.) is determined based on the Fisher distribution table for a given level of significance, taking into account that the number of degrees of freedom for the total sum of squares (larger variance) is 1, and the number of degrees of freedom of the residual sum of squares (smaller variance) in linear regression is n-2.

F table. is the maximum possible value of the criterion under the influence of random factors with the given degrees of freedom and significance level α. The significance level α represents the probability of rejecting the correct hypothesis, given that it is true. Usually, α is taken equal to 0.05.

If the actual value of the F-Criterion is less than the table value, then there is no reason to reject the null hypothesis [16]. And if not, then the null hypothesis is rejected and the alternative hypothesis about the statistical significance of the equation as a whole is accepted with probability (1-α).

In our case, the tabular value of the criterion with degrees of freedom k1=1 and k2=5, tabular F = 6.61, and calculated for the studied model F = 27.85.

Since the actual value of F > F tabular, the coefficient of determination is statistically significant (the regression equation estimate found is statistically reliable), and the proposed model is significant.

To confirm the significance of the correlation coefficient of the constructed model, we use the Student's criterion. The formula which is used to calculate it:

$$t = \frac{|r|(n-2)^{1/2}}{(1-r^2)^{1/2}} (6)$$

The calculated value of the Student's t criterion is 5.2781, the table value is 2.571. Since the actual value of the Student's coefficient is greater than the tabular one, the linear correlation coefficient is considered significant, and the relationship between the volume of sales of goods in the online store and the number of goods that are presented with the help of photos and videos made with the use of rotary tables for photography is real.

Based on the data in Table 1, using formula 1, we write down the system of equations, as a result of solving which, we get the regression coefficients: b = 67.0488, a = -51065.01. By substituting them into the equation of linear dependence, we get a regression equation which looks as following:

$$y = 67{,}0488x_1 - 51065{,}0969 (7)$$

where y is the volume of sales of goods;

x_1 – the number of products that are presented with the help of photos and videos made using rotary tables for photography.

The value of the coefficient of determination r^2 for the constructed model is at the level of 0.8478, which means that 84.78% of the variations in the sales of goods are linearly related to the factor included in the model.

So, if the number of goods that are presented with the help of photos and videos made with the use of rotary tables for photography increases by one hundred, then the volume of sales of goods will increase by 6705 thousand hryvnias.

4. CONCLUSION

The conducted research showed the importance of using rotary photo tables for presenting goods in an online store. In the course of the research, it was proved that there is a close relationship between the volume of sales of goods by the online store and the number of goods that were presented with the help of photos and videos made with the use of rotary tables for photography. It was also established that the correlation between the volume of sales of goods by the online store and the number of goods presented without the use of rotary photo tables is weak. This trend is quite clear and is explained by the fact that consumers of goods from online stores make a choice in favor of goods that are most widely and comprehensively presented on the website of the online store. It is the use of rotary tables for product photography that makes it possible to present the product from all possible angles and demonstrate all the features and advantages of the offered product to the potential consumer. Paired linear regression was chosen in the research. Its parameters were estimated by the method of least squares. The statistical significance of the equation was checked using the coefficient of determination and Fisher's criterion. It was established that 84.78% of the total variability of Y in the model is explained by the change in *x*. It was also established that the parameters of the model are statistically significant. The economic interpretation of the parameters of the model shows that the increase in the number of goods presented with the help of photos and videos created with the use of rotary tables for photography will increase by one hundred units, then the volume of sales of goods will increase by 6,705 thousand hryvnias.

REFERENCES

1. Akoglu H. User's guide to correlation coefficients. Turkish journal of emergency medicine. 2018; 18(3): 91–93.
2. Amankwah-Amoah J., Khan Z., Wood G. COVID-19 and business failures: The paradoxes of experience, scale, and scope for theory and practice. European Management Journal. 2021; 39(2): 179–184.
3. Bassey E., Akra U. Canonical Correlation of Multivariate Regression Analysis on Economic Factors in Nigeria. Asian Journal of Probability and Statistics. 2021; 4: 1-7.
4. Cumming D., Johan S., Khan Z., Meyer M. E-Commerce Policy and International Business. Manag Int Rev. 2022; 2: 1-23.
5. Cong H., Hu X., Du Y. Correlation Analysis and Regression Fitting of Multiple Sulfide Aging Characteristic Parameters. IEEE Transactions on Dielectrics and Electrical Insulation. 2022; 2: 1-17.
6. Dos Santos E., Gomes J. Interactions between e-commerce users during the COVID-19 pandemic period: What came and what remained. Work. 2022; 73(s1): 177-187.

7. Duan L. Correlation Analysis between Atmospheric Environment and Public Sentiment Based on Multiple Regression Model. Wireless Communications and Mobile Computing. 2022; 3: 1-8.
8. Grechyshkin Y. Mathematical modeling of the development of the potential of integrated business structures. Management. 2021; 33: 141-148.
9. Jobsta L., Heck D, Moshagena M. A comparison of correlation and regression approaches for multinomial processing tree models. Journal of Mathematical Psychology.2020; 98:123-131.
10. Homapour E., Su L., Caraffini F., Chiclana, F. Regression Analysis of Macroeconomic Conditions and Capital Structures of Publicly Listed British Firms. Mathematics. 2022; 10: 11-19.
11. Hwang S., Kim J. Correlation Between Tractor Variables and Loan Support Limit in South Korea Through Regression Analysis. Journal of Biosystems Engineering. 2022; 47: 101-107.
12. Khatti J., Grover K. A Study of Relationship among Correlation Coefficient, Performance, and Overfitting using Regression Analysis. International Journal of Scientific and Engineering Research. 2022; 13: 1074-1085.
13. Mellos E., Paparrigopoulos T. Substance use during the COVID-19 pandemic: What is really happening? Psychiatriki. 2022; 33(1): 17-20.
14. Mohammed M., Feroskhan M., Sreekanth K., Karunamurthy R., Shivakumar S., Nazaruddin T., Yunus K. Regression-Analysis-Based Empirical Correlations to Design Regenerative Flow Machines. Energies. 2022; 15: 1-23.
15. Yanqing Y. Model-free tests for series correlation in multivariate linear regression.Journal of Statistical Planning and Inference. 2020; 206: 179-195.
16. Papageorgiou S. On correlation coefficients and their interpretation. J Orthod. 2022; 49(3): 359-361.
17. Schober P., Boer C., Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg. 2018; 126(5): 1763-1768.
18. Tabatabai M., Bailey S., Bursac Z., Tabatabai H., Wilus D. An introduction to new robust linear and monotonic correlation coefficients. BMC Bioinformatics. 2021; 22(1): 170-186.
19. Tzeremes, P. Productivity, efficiency and firm's market value: Microeconomic evidence from multinational corporations. Bull. Appl. Econ. 2020; 7: 95-101.
20. Vetter T., Mascha E. Unadjusted Bivariate Two-Group Comparisons: When Simpler is Better. Anesth Analg. 2018; 126(1): 338-342.