DOI: 10.32347/2412-9933.2022.50.102-105
UDC 004.021

Zelenko Eduard
Postgraduate student of the Department of Software for Automated Systems,
https://orcid.org/0000-0002-9939-3830
Cherkasy State Technological University, Cherkasy

DETERMINING THE CORRELATION BETWEEN DATASETS FOR CALCULATION OF THE RETAIL PRICE WHEN USING SOFTWARE AGENTS

Abstract. *The development and support of algorithms for calculating prices in the field of trade is one of the current trends in the development of information systems in recent years, which attracts the attention of specialists of various profiles. Experts in the field of economics see in such algorithms new opportunities for improving automated systems, transforming them into public automated systems of a new generation with advanced means of presenting various digital information resources and accessing them, created taking into account the need for integration and the use of APIs. Within the framework of the outlined problem, the scientific tasks of developing models, methods, algorithms and programs that simulate processes for data processing and price calculation in order to determine their main characteristics for the construction of mathematical software for automated retail price calculation systems are important [12].*
In general, effective pricing contributes to the subordination of production to social needs, and an adequate level of prices contributes to economic growth, provides an effective competitive environment and orients production to innovative content [1]. The above determines the relevance of the development and improvement of algorithms for calculating the retail price of a product in the event, that already known algorithms determine the price inaccurately (for example, when there is insufficient amount of necessary data due to certain factors). The use of the latest ones in such situations quite often increases the risk of selling the product at an irrelevant price, which in turn can lead to a decrease in turnover or loss of funds.

Keywords: information technology; enterprise economy; price premium; retail price; pricing algorithm

Problem statement

With a significant increase in the price premium, the trading company risks losing customers, however, compensating for this risk with low income, the company reduces the risk of its own losses.

Under the conditions of a risk-free price premium, the amount C_0, that the buyer pays to the trading company is as follows:

$$C = \left(\frac{1}{1 - P_H}\right) \times C_0 \qquad (1)$$

The presence of the risk of non-return of the purchase price of the goods, which is measured by the probability P_H, thereby leads to an increase in the amount of the retail price that the buyer must pay, measured by the coefficient K_c.

$$K_c = \frac{C}{C_0} = \frac{1}{(1 - P_H)} \qquad (2)$$

Graphically, the dependence of the coefficient of increase in the retail price of the amount paid by the buyer for the product K_c on the probability of non-return of the purchase price of the product is shown in figure 1.

With the probability values (P_H) of the product purchase price non-return ranging from 0,3 to 0,6 not only the risk itself increases significantly (it becomes unacceptable), but also the amount that the buyer pays for the product (retail price of the product). So already at $P_H = 0{,}5$ the buyer is forced to pay twice the price compared to the risk-free price, and the trading premium increases several times [2, p. 215–216].

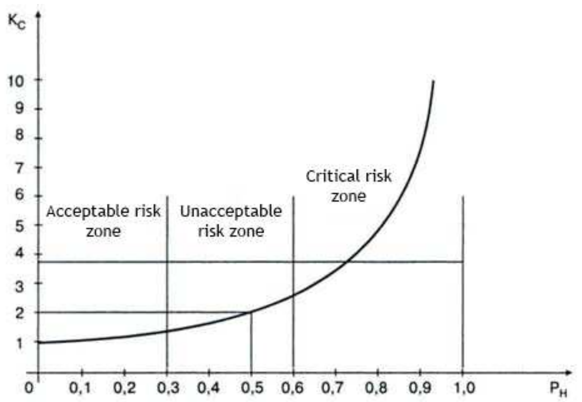

Figure 1 – The dependence curve of the retail price increase on the product purchase price non-return probability

With such a method of supply chain management as dropshipping, the online retailer uses service packages provided by dealers. At the beginning of a startup life cycle, the profitability of cooperation with dealers is determined both by the presence of certain optional features in particular, and by the cost of the provided service packages as a whole. Thus, the use of more expensive packages increases the risk of losing money in case of project failure, and less expensive ones – limits the retailer's ability to use certain functions and/or data. The essence of the problem in the second case is the lack of an opportunity to control the risk of receiving an invalid retail price during the calculation.

Analysis of recent research and publications

Problems related to pricing are considered in scientific research by many scientists. For example, in work [6] the construction and functioning forms of the internal economic mechanism and its regulatory framework at the enterprise are considered. The textbook [10] deals with the principles of setting prices for products and enterprises services of all ownership forms, the peculiarities of the prices classification in modern conditions are determined. The factors that affect the price level are analyzed. The comparison of Ukrainian and foreign price regulation and foreign trade prices is presented.

In the work [7], the main provisions of the theory of pricing, price functions, principles, factors and regularities of pricing are revealed, as well as the classification features and types of prices in the modern economy are highlighted. The textbook [11] highlights the basics of pricing methodology. Particular attention is paid to the theoretical bases of pricing, the definition of policy and the strategy for setting prices, and the formation of prices, taking into account the competitiveness of the product.

The analysis of these and other sources demonstrates the presence of a tendency towards constantly emerging new requirements for pricing, which is due to the development of technologies. Exceptional situations require research to create innovative methods of price calculation.

Aim

The purpose of the article is to form a hypothesis regarding the use of additional incoming data (which is involved when using the dropshipping) based on their correlation with the main incoming data in order to develop software using software agents to improve the accuracy of retail price calculation.

Main material presenting

For the experimental implementation of the algorithm, the dependence of the following important values was calculated (others do not take part in the calculations):

- price (p) – price for one product unit (blue);
- price_4 (p4) – price for four units of the product (red);
- avg_price (avg) – average price (green);
- availability (bal) – balance in stock (orange).

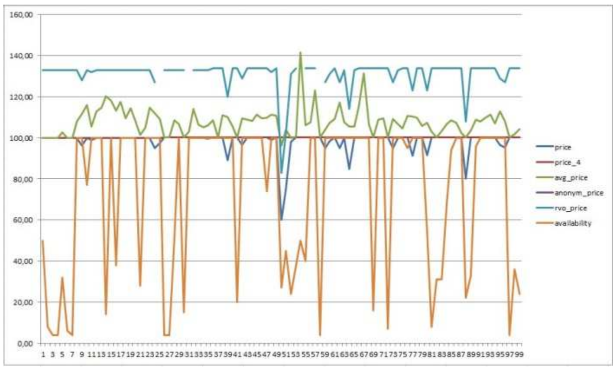

Figure 2 – Graph of dependence of price values and balance for products 101-200 from the list (values of other ranges were also analyzed, for example: 1-100, 60000-60100)

Initial analysis confirmed the existence of a relationship between the values, but not in all cases. It can be argued that the average price value in most cases depends on the total balance value, but with certain variations of the variables, the dependence may disappear, for example, if the value for four units of the product is missing.

For a more accurate definition and display of graphical dependence, it became necessary to build an additional correlation graph that would confirm the presence or absence of dependence.

To begin with, the correlation coefficient was calculated for several pairs of data sets, for example: for the average price and the total balance, for the price of four units of the product and the total balance, etc.

The Pearson correlation coefficient (hereafter PCC) characterizes the existence of a linear relationship between two values.

Let two samples be given $x^m = (x_1, \dots, x_m), y^m = (y_1, \dots, y_m)$;

Then the PCC is calculated using the formula:

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \sum_{i=1}^{m}(y_i - \bar{y})^2}} = \frac{\operatorname{cov}(x, y)}{\sqrt{s_x^2 s_y^2}}, \quad (3)$$

where $\bar{x}, \bar{y}$ – sample averages x^m and y^m, s_x^2, s_y^2 – sample variances, $r_{xy} \in [-1, 1]$.

Pearson correlation coefficient – is a statistical indicator of the dependence of two random variables. The correlation coefficient can take values from -1 to +1, where the value -1 will indicate the absence of correlation, 0 – about zero correlation, and +1 – about full correlation of values. That is, the closer the value of the correlation coefficient is to +1, the stronger the correlation between two random variables.

It's also called the tightness of the linear relations, so $|r_{xy}| = 1 \Rightarrow x, y$ – linearly dependent and $r_{xy} = 0 \Rightarrow x, y$ – linearly independent. [9, p. 238-239; 4, p. 23, 75, 109; 3, p. 9-13; 5, p. 62; 8]

Let's take several sets of data:

p1	p4	avg	max	bal
10	10,01	18,56	42,5	76
10	10	19,86	44,5	70
10	10	10,82	14,2	70
10	10	10	10	50
10		17,25	24,5	2
10,01	10,01	16,97	42,5	100
10,5	16	15,87	16	41
11,58	17,26	22,77	32,83	27
10	13	13,95	15,5	43
12,25	13,53	13,35	13,35	14
14	14,9	14,85	14,9	17
61,5		74,74	78,46	12
75,38		80,07	86,27	12
62,89		71,2	76,13	12
100	400	500		100

Figure 3 – Data sets,
*where **max** – maximum price for one product unit*

Calculating the Pearson correlation coefficient revealed the following:

p1/p4	p1/avg	p1/bal	p4/avg	p4/bal	avg/bal
0,609758	0,788977	-0,02435	1,00	0,473071	0,381679

Figure 4 – Correlation coefficients for the main data sets.

It can be concluded from the figure 4 that the correlation is full for datasets **p4** and **avg**, and almost zero – for datasets **p1** and **bal**.

Graphically, the full correlation in this case can be depicted as follows:

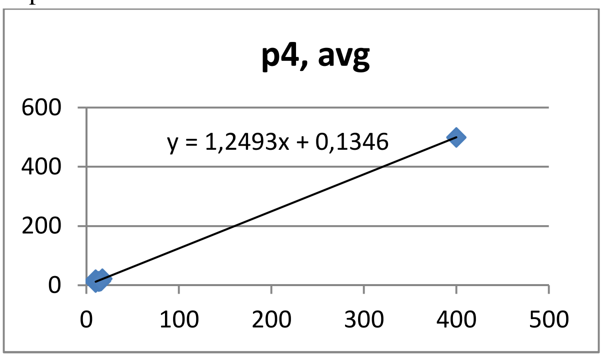

*Figure 5 – Full correlation for datasets **p4** and **avg***

The 400/500 point can be disregarded, as it represents fictitious values for the test of possible situations. Other points are very close to each other, which indicates a fairly close correlation.

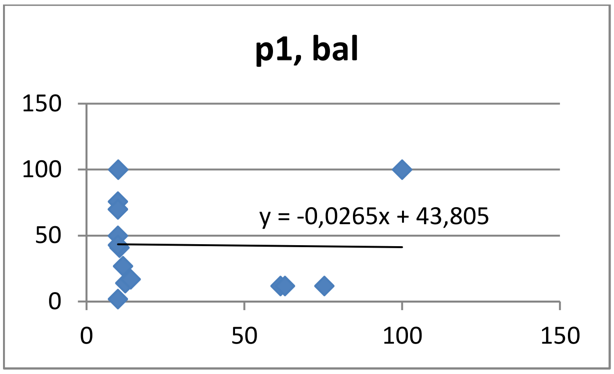

*Figure 6 – Almost zero correlation for datasets **p1** and **bal***

Conclusion

Current research work has identified a correlation between incoming data, which suggests the existence of possibility to use the additional data for retail price calculation accuracy improving with use of software agents in case of the main input data irrelevance (for example, the price and balance of the product for each supplier at the right time).

In this case, the main problem was the risk of calculating a price premium that would go beyond the allowed price range (based on competitors' prices). Therefore, the task of developing an algorithm that would calculate the safe retail price of the product in the absence of the necessary data arose.

To solve the problem in this work, the sets of variation series for prices and product balance were analyzed in order to determine the dependency between the data of these sets using the Pearson correlation coefficient. Coefficients are presented both in writing and graphically.

As a result, the new method of the product retail price calculating will include the use of only those data between which the existence of dependencies was revealed during the research (in this case, the values of the following variables: price_4, avg_price and availability).

The results and conclusions obtained in the course of the current work are not exhaustive, and the work itself may require additional research.

References

1. Andrus, O. I. (2013). Analysis of problems of the modern pricing in Ukraine and ground of ways of increase of his efficiency. *Journal "Efektyvna ekonomika"*, 5.

2. Borovkova, V. A. (2004). Risk management in trade. St. Petersburg, 288.

3. Chen, Peter Y. & Popovich, Paula M. (2002). Correlation: Parametric and Nonparametric Measures. (1st ed., Quantitative Applications in the Social Sciences, series no. 07-139). Thousand Oaks, CA: Sage. ISBN 0-7619-2228-8.

4. Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd ed.). 567. ISBN 0-8058-0283-5.

5. Rodgers, Joseph Lee & Nicewander, Alan W. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42, 1, 59-66. DOI: 10.2307/2685263.

6. Krush, P. V. (2008). Internal economic mechanism of the enterprise. *National Technical University of Ukraine "Kyiv Polytechnic Institute"*, 206.

7. Krush, P. V. (2019). Pricing. *National Technical University of Ukraine*, 291.

8. Pearson's correlation coefficient [Online]. (2022). Available: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient – Accessed on: February 06, 2022.

9. Rahman, Mezbahur & Zhang, Qichao. (2015). Comparison among Pearson correlation coefficient tests. *Far East Journal of Mathematical Sciences (FJMS),* 99, 237-255. DOI: 10.17654/MS099020237.

10. Shkvarchuk, L. O. (2006). Price formation. K.: Kondor, 460. ISBN: 966-351-051-X.

11. Verkhohlyadova, N. I., Il'ina, S. B., Ivannikova, N. A., Slabko, Ya. Ya. and Lysenko Yu. V. (2006). Bases of price formation. K.: Kondor, 252. ISBN 966-351-082-X.

12. Zelenko, E. (2019). Algorithm of determination of risk limit at calculation of retail price. *Scientific-practical conference "Problems of science development in the context of transformations of society",* Poltava, *Publishing House "Molodyy vchenyy", (Technical Sciences)*, pp. 81–83.

The article was received by the editorial board 17.06.2022

Зеленько Едуард Васильович
Аспірант кафедри програмного забезпечення автоматизованих систем, *https://orcid.org/0000-0002-9939-3830*
Черкаський державний технологічний університет, Черкаси

ВИЗНАЧЕННЯ КОРЕЛЯЦІЇ МІЖ МНОЖИНАМИ ДАНИХ ДЛЯ РОЗРАХУНКУ РОЗДРІБНОЇ ЦІНИ ПРИ ВИКОРИСТАННІ ПРОГРАМНИХ АГЕНТІВ

Анотація. *Розробка та підтримка алгоритмів для розрахунку цін у сфері торгівлі – один з актуальних напрямів розвитку інформаційних систем в останні роки, що привертає увагу фахівців різного профілю. Фахівці в галузі економіки бачать у подібних алгоритмах нові можливості для вдосконалення автоматизованих систем, перетворення їх у публічні автоматизовані системи нового покоління з розвиненими засобами подання різноманітних цифрових інформаційних ресурсів і доступу до них, створювані з урахуванням необхідності інтеграції та використання API. У межах окресленої проблеми важливими є наукові задачі розроблення моделей, методів, алгоритмів та програм, які здійснюють моделювання процесів для обробки даних та розрахунку цін з метою визначення їх основних характеристик для побудови математичного програмного забезпечення автоматизованих систем розрахунку роздрібних цін [12]. Загалом ефективне ціноутворення сприяє підпорядкуванню виробництва суспільним потребам, а адекватний рівень цін сприяє економічному зростанню, забезпечує ефективне конкурентне середовище, орієнтує виробництво на інноваційний зміст [1]. Вищезазначене обумовлює актуальність розроблення та вдосконалення алгоритмів для розрахунку роздрібної ціни товару в разі, коли вже відомі алгоритми визначають ціну неточно (наприклад, при недостатній кількості необхідних даних, що зумовлено певними факторами). Використання останніх у подібних ситуаціях доволі часто збільшує ризик реалізувати товар по неактуальній ціні, що своєю чергою може призвести до зменшення товарообігу або втрати грошових коштів.*

Ключові слова**: **інформаційні технології; економіка підприємства; цінова надбавка; роздрібна ціна; алгоритм ціноутворення

Link to the publication

APA Zelenko, Eduard. (2022). Determining the correlation between datasets for calculation of the retail price when using software agents. *Management of Development of Complex Systems*, 50, 39–43, dx.doi.org\10.32347/2412-9933.2022.50.39-43.

ДСТУ Зеленько Е. В. Визначення кореляції між множинами даних для розрахунку роздрібної ціни при використанні програмних агентів. *Управління розвитком складних систем.* Київ, 2022. № 50. С. 102 – 105, dx.doi.org\10.32347/2412-9933.2022.50.102-105.