# CORRELATION ANALYSIS OF THE RELATIONSHIP BETWEEN THE TIME VARIABLE OF OUTPUT AND THE VALUE OF FIXED ASSETS

Batir Mardonov*
Tashkent institute of Textile and Light Industry, Tashkent, Uzbekistan;
batsam@list.ru

Raya Karlibayeva
Tashkent State University of Economics, Tashkent, Uzbekistan;
k.raya3005@mail.ru

Gulnoza Yuldosheva
The Tashkent State University of Economics, Tashkent, Uzbekistan;
gulnozopa0705@gmail.com

Amina Aymuxammedova
The Tashkent State University of Economics, Tashkent, Uzbekistan;
aymukhammedovaa@mail.ru

## ABSTRACT

The main goal of studying economical processes is to identify patterns that are of great importance in practice. Typically, these laws are influenced by many factors, and assessing their level of impact is the most important task. In practice, the factors that arise are often of a random nature, and for their assessment, it is necessary to use special sections of mathematics - probability theory and mathematical statistics. Mathematical statistics is a special branch of mathematics The main goal of studying economical processes is to identify patterns that are of great importance in practice. Typically, these laws are influenced by many factors, and assessing their level of impact is the most important task. In practice, the factors that arise are often of a random nature, and for their assessment, it is necessary to use special sections of mathematics - probability theory and mathematical statistics. Mathematical statistics is a special branch of mathematics that deals with the collection, processing, systematization of observation results and the creation and study of mathematical methods of interpretation in order to determine statistical patterns. While probability theory abstractly studies reality, mathematical statistics, examining the results of inferences from a set of directly observable random events (processes), studies laws in them. The main goal of studying economical processes is to identify patterns that are of great importance in practice. Typically, these laws are influenced by many factors, and assessing their level of impact is the most important task. In practice, the factors that arise are often of a random nature, and for their assessment, it is necessary to use special sections that deals with the collection, processing, systematization of observation results and the creation and study of mathematical methods of interpretation in order to determine statistical patterns. While probability theory studies reality in an abstract way, mathematical statistics, examining the results of inferences from a set of directly observable random events (processes), studies laws in them.

## CCS CONCEPTS

**• mathematical statistics**; **• mathematical model**; **• functional relationship**; **• dispersion**; **• correlation analysis**;

**ACM Reference Format:**
Batir Mardonov, Raya Karlibayeva, Gulnoza Yuldosheva, and Amina Aymuxammedova. 2021. CORRELATION ANALYSIS OF THE RELATIONSHIP BETWEEN THE TIME VARIABLE OF OUTPUT AND THE VALUE OF FIXED ASSETS. In *The 5th International Conference on Future Networks & Distributed Systems (ICFNDS 2021), December 15, 16, 2021, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3508072.3508200


*ICFNDS 2021, December 15, 16, 2021, Dubai, United Arab Emirates*

ACM ISBN 978-1-4503-8734-7/21/12…$15.00
https://doi.org/10.1145/3508072.3508200

## 1 INTRODUCTION

The main task of statistics is (after analysis of variance if the level of influence of external factors on the observation results is determined), find their quantitative relationship between these factors and study their main properties. When studying economical processes, it often happens that one single value of one variable corresponds to several values of another variable [Alyavdin L.A. et al., 1969, Fedoseeva V.V. 2003, Kramer N.Sh. 2010]. If it is known that these values are distributed according to a certain law, then this relationship is called statistical (or stochastic, probabilistic). In some cases, the arithmetic mean of the second variable is compared with the values of the selected variable. This relationship is called correlation. Thus, if several values of the second variable correspond to one single value of the first variable, then this relationship is statistical, if the mean value of the second variable is compared, then it is a correlation, and if some exact value of the second variable is compared, then this relationship will be functional. The statistical relationship is usually expressed as follows: $M_x$(Y)=$\varphi(x)$ or $M_y(X) = \psi(y)$, here $Y(y)$ and $X(x)$ sets of random variables that obey a certain distribution law, $M_x(Y)$, $M_y(X)$ - arithmetic mean, respectively, value $x$ variable $X$ corresponds to the mean of the second variable $M_x(Y)$, and vice versa, if the first variable $Y$, the mean value corresponds to its value $M_y(X)$ variable $X$. The functions presented here are regression functions $\varphi(x)$ and $\psi(y)$. Based on

**Table 1: The relationship between the daily output of Y products (in tons) and the value of the main production x funds (in millions of sums) for a total of 50 similar enterprises**

| The value of OPF bln. UZS X | Mid-intervals $y_i/x_i$ | Daily production (ton)Y | | | | | | | | All$N_i$ | Group average (ton) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Delta l_1$ | $\Delta l_2$ | $\Delta l_3$ | — | $\Delta l_i$ | — | $\Delta l_{k-1}$ | $\Delta l_k$ | | $\bar{y}_i = \frac{1}{N_i}\sum_{j=1}^{k} y_j N_{ij} N_i = \sum_{j=1}^{k} N_{ij}, N = \sum_{i=1}^{m} N_i$ |
| | | $y_1$ | $y_2$ | $y_3$ | — | $y_i$ | — | $y_{k-1}$ | $y_k$ | | |
| $\Delta s_1$ | $x_1$ | $N_{11}$ | $N_{12}$ | $N_{13}$ | – | $N_{1j}$ | — | $N_{1k-1}$ | $N_{1k}$ | $N_1$ | $\bar{y}_1$ |
| $\Delta s_2$ | $x_2$ | $N_{21}$ | $N_{22}$ | $N_{23}$ | — | $N_{2j}$ | — | $N_{2k-1}$ | $N_{2k}$ | $N_2$ | $\bar{y}_2$ |
| $\Delta s_3$ | $x_3$ | $N_{31}$ | $N_{32}$ | $N_{33}$ | — | $N_{3j}$ | — | $N_{3k-1}$ | $N_{3k}$ | $N_3$ | $\bar{y}_3$ |
| ——- | —— | —— | —— | —— | — | – | — | —- | —— | —— | —- |
| $\Delta s_i$ | $x_i$ | $N_{i1}$ | $N_{i2}$ | $N_{i2}$ | — | $N_{ij}$ | — | $N_{ik-1}$ | $N_{ik}$ | $N_i$ | $\bar{y}_i$ |
| ——- | —— | —— | —— | —— | — | —— | — | —- | —— | —- | —- |
| $\Delta s_{m-1}$ | $x_{m-1}$ | $N_{m-11}$ | $N_{m-12}$ | $N_{m-13}$ | — | $N_{m-1j}$ | — | $N_{m-1k-1}$ | $N_{m-1k}$ | $N_m$ | $\bar{y}_m$ |
| $\Delta s_m$ | $x_m$ | $N_{m1}$ | $N_{m2}$ | $N_{m3}$ | — | $N_{mj}$ | — | $N_{mk-1}$ | $N_{mk}$ | $N_m$ | $\bar{y}_m$ |
| | Total $S_j$ | $S_1$ | $S_2$ | $S_3$ | — | $S_j$ | — | $S_{k-1}$ | $S_k$ | $N$ | |
| Group average (billion UZS) $\bar{x}_j == \frac{1}{S_j}\sum_{i=1}^{m} x_i N_{ij}$ $S_j = \sum_{i=1}^{m} N_{ij} N = \sum_{j=1}^{k} S_j$ | | $\bar{x}_1$ | $\bar{x}_2$ | $\bar{x}_3$ | — | $\bar{x}_j$ | — | $\bar{x}_{k-1}$ | $\bar{x}_k$ | | - |

the foregoing, statistical processing between the variables is carried out using regression relationships based on correlation analysis.

Correlation analysis determines whether there is a relationship (known) between variables and estimates its density. On the other hand, regression analysis consists of determining the form of the relationship and studying it [Kramer N.Sh. 1997, Kramer N.Sh. 2007]. In this regard, it is advisable to present statistical data in tabular form.

The remainder of the article proceeds as follows. Section 2 provides the theoretical studies. Section 3 introduces the experimental research. Section 4 presents analysis of results, and Section 5 concludes.

## 2 THEORETICAL STUDIES

Consider the following problem. Let us establish a regression relationship (within a day) between the prepared product of production and the production fund of randomly selected similar companies. Factories start in $t = t_{min}$ hours and last $k$ hours. We put the results of observations in Table 1. The first column (row by row) indicates the cost of production assets (for each row $\Delta s_i = s_i - s_{i+1}$ ( $i = 1..m$) million sum 1 column), in the second column, the average number of production assets (midpoint of the interval) giving, for each row $\Delta x_i = (s_{i+1} + s_i)/2$ billion sums, and the third column consists of parts (each in rows) that show the average amount of products $y_i$ ($i = 1..k$), produced in factories in the allotted time, and the number of factories involved $N_{ij}$ ( $i = 1..m$, $j = 1..k$) indicated in additional (lower)3÷$k$ columns. The main 4th column shows the total number of factories involved $N_i$. In the 5th column, for each row, the average value between groups, calculated using the formula

$$\bar{y}_i = \frac{1}{N_i}\sum\nolimits_{j=1}^{k} N_{ij} y_j \; (i = 1..m).$$

In the $m + 2$ -th line, the average number of the production fund between the groups is calculated

$$\bar{x}_j = \frac{1}{S_j}\sum\nolimits_{i=1}^{m} x_i N_{ij}.$$

In the table $x_j$ and $y_i$ the midpoints of the corresponding intervals are denoted $S_j$ and $N_i$- respectively their frequency. Using the formula $y_x = b_1 x + b_0$ find formulas for calculating the unknown parameters of the linear regression equation. For this purpose, we will apply the method of least squares, according to which the unknown parameters $b_1$and $b_0$ are chosen so that the sum of the squared deviations of the empirical group means $Y$, calculated by the formula, from the values found by the regression equation, was the minimum:

$$S = S\left(b_0, b_1\right) = \sum_{i=1}^{m} (b_0 + b_1 x_i - \bar{y}_i)^2 N_i \rightarrow min$$

By the shape of the broken line, we can assume the presence of a linear correlation in between the two considered $S = S(b_0, b_1)$, equate to zero its partial derivatives,

$$\frac{\partial S}{\partial b_0} = 0 \; and \; \frac{\partial S}{\partial b_1} = 0$$

hence, after transformations, we obtain a system of normal equations for determining the parameters of linear regression:

$$b_0 \sum_{i=1}^{m} N_i + b_1 \sum_{i=1}^{m} x_i N_i = \sum_{i=1}^{m} \bar{y}_i N_i$$

$$b_0 \sum_{i=1}^{m} x_i N_i + b_1 \sum_{i=1}^{m} x_i^2 N_i = \sum_{i=1}^{m} x_i \bar{y}_i N_i$$

**Table 2: The calculation results**

| | $y_{min}$=2 ton | | | | | $y_{min}$ = 2, 5 *ton* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_0$(hour) | 2 | 4 | 6 | 8 | 9 | 2 | 4 | 6 | 8 | 9 |
| $b_{xy}$ | 0.437 | 0.769 | 0.961 | 0.939 | 0.865 | 0.366 | 0.741 | 0.939 | 0.915 | 0.841 |
| $b_{yx}$ | 0.824 | 0.688 | 0.732 | 0.82 | 0.909 | 0.705 | 0.679 | 0.744 | 0.840 | 0.936 |
| $r$ | 0.600 | 0.727 | 0.838 | 0.877 | 0.887 | 0.508 | 0.709 | 0.836 | 0.877 | 0.887 |
| $r_c$ | 5.205 | 7.345 | 10.67 | 12.67 | 12.03 | 4.089 | 6.981 | 10.55 | 12.65 | 13.30 |

The solution to this system can be represented as:

$$b_0 = \bar{y} - b_1\bar{x}, b_1 = b_{xy} = (\bar{x}\bar{y} - \bar{y}\bar{x})/s_x^2$$

Where,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{m} x_i N_i,\ \bar{y} = \frac{1}{N}\sum_{j=1}^{k} y_j S_j,\ \bar{x}\bar{y} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{k} x_i y_j N_i S_j,\ \overline{x^2} = \frac{1}{N}\sum_{i=1}^{m} x_i^2 N_i,\ N = \sum_{i=1}^{l}\sum_{j=1}^{m} N_{ij}.\ s_x^2 = \overline{x^2} - \bar{x}^2$$

Arguing similarly and setting the regression equation in the form $x_y = b_{yx}x + c_0$. Where,

$$b_{yx} = \frac{\bar{x}\bar{y} - \bar{y}\bar{x}}{s_y^2},\ \ s_y^2 = \overline{y^2} - \bar{y}^2.$$

Variable

$$r = b_{xy}\frac{s_x}{s_y} = (\bar{x}\bar{y} - \bar{x}\cdot\bar{y})/s_x s_y = \sqrt{b_{yx}b_{xy}}$$

is a measure of the density of a linear relationship and is called the sample coefficient or simply the correlation coefficient if $r = 0$ there is no linear relationship between $X$ and $Y$sets and. The correlation coefficient satisfies the inequality. To confirm the presence (or absence) of correlation, statistics are calculated.

$$t_c = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

And its value is compared with the Student's t-distribution$t_{1-\alpha;k}$, where $\alpha$ - significance level at$k = N - 2$, the value of which is determined from a known table $t_{1-\alpha;k}$. If the inequality holds$t > t_{1-\alpha,k}$, then the presence of a correlation between $X$ and $Y$sets is confirmed, otherwise, there will be no correlation between them.

## 3 EXPERIMENTAL RESEARCH

All this can be seen with a specific example:

Calculate the correlation coefficient between the value of fixed assets X and the daily output of products Y, where the daily output is produced within 10 hours, the total amount of products produced per day is 20 tons. The total fund is 38 billion UZS. and is sold in parts for each day for 2 billion within 10 days. Thus, if we accept the initial amount of 20 billion UZS, then on the $i$ - day the $X = 20 + 2(i-1)$ $(i = 1 \ldots 10)$ billion UZS. is allocated from the fund. It is assumed that the release of products within 10 hours is subject to the law in time $y = y_{min} + \frac{y_0 - y_{min}}{t_0 - t_{min}}(t - t_{min})$ at $t \le t_0$ , $y = y_0 + \frac{y_0 - y_{min}}{t_0 - t_{min}}(t - t_0)$ at $t \ge t_0$ , where $t_0$ the time of the change in the rate of production with the value $y_0$. With a known value $y_{max}$ there is a connection between $t_0$ and $y_0$: $y_0 = \frac{a_q t_0}{b_0}$ , $a_0 = y_{max} - y_{min}\frac{t_{max} - t_0}{t_{min} - t_0}$ , $b_0 = 1 - \frac{1}{t_{min} - t_0}$ .

The calculations were carried out to calculate the coefficient of variation $r$ $t_{min}$. In calculations it is accepted: $t_{max}$, $y_{\max} = 20$T, $m = 10$, $k = 10$, $N_{11} = N_{12} = 1$, $N_{1j} = 0$, $(j = 3\ldots10)$, $N_{21} = 2$, $N_{22} = N_{23} = 1$, $N_{2j} = 0$ $(j = 4\ldots10)$, $N_{31} = 0$, $N_{32} = N_{33} = 2$, $N_{34} = 3$, $N_{35} = 1$, $N_{3j} = 0$ $(j = 6\ldots10)$, $N_{41} = 0$, $N_{42} = 1$, $N_{43} = 2$, $N_{44} = 4$, $N_{45} = 2$, $N_{46} = 1$, $N_{4j} = 0(j = 7\ldots10)$, $N_{51} = 0$, $N_{52} = 1$, $N_{53} = 1$, $N_{54} = 2$, $N_{55} = N_{56} = N_{57} = 1$, $N_{5k} = 0(j = 8\ldots10)$,$N_{6i} = 0(i = 1..3)$,$N_{64} = N_{65} = N_{68} = 2$, $N_{67} = N_{68} = 1$, $N_{6j} = 0$ $(j = 9\ldots10)$, $N_{7i} = 0(i = 1..4)$, $N_{75} = N_{76} = N_{77} = 1$, $N_{78} = 2$, $N_{7j} = 0(j = 9\ldots10)$, $N_{8i} = 0(i = 1..6)$,$N_{87} = N_{88} = 1$,$N_{8j} = 0(j = 9\ldots10)$, $N_{9i} = 0(i = 1..6)$, $N_{97} = 2.$, $N_{98} = 1$, $N_{9j} = 0(j = 9\ldots10)$, $N_{10i} = 0(i = 1..8)$, $N_{10,9} = N_{10,10} = 1$.

## 4 ANALYSIS OF RESULTS

The calculation results for various values $t_0$ and $y_{min}$ are presented in Table 2.

From the analysis of the tabular data, it follows that, along with the fulfillment of the Student's test, the correlation coefficient $r$ monotonically increases, the regression coefficient $b_{xy}$ first monotonically increases and then decreases, and the coefficient $b_{yx}$ monotonically increases. This circumstance indicates the presence of nonlinearity in the regression equation.

## 5 CONCLUSION

A general scheme for statistical processing of the correlation table data has been developed, taking into account the variability of the table elements, using them in the future to compile a nonlinear regression. The possibility is indicated by selecting the elements of the table to develop the most profitable option for the implementation of the value of fixed assets in production. The possibility of taking into account the variability in time of production for the correlation coefficient is proposed. In the case of a linear increase in output over time, it also increases the coefficient.

## REFERENCES

[1] Alyavdin L.A. *et al.* Planning and analysis of a research experiment as applied to light industry. M. Leg. ind. 1969.
[2] V.V. Fedoseeva. Economic and mathematical methods and applied models. M. 2003.
[3] Kramer N.Sh. Higher mathematics for economists: a textbook for university students studying economics. M. 2010. https://doi.org/10.15863/tas.2019.09.77.28
[4] Tabaei, Bahman P., and William H. Herman. "A multivariate logistic regression equation to screen for diabetes: development and validation." Diabetes Care 25.11 (2002): 1999-2003.
[5] Корреляционныйанализвэкономическихисследованиях. А. С. Пан, Е. В. РадковскаяУральскийгосударственныйэкономическийуниверситет(Екатеринбург)

[6] Ashrafi, M. Z., & Ng, S. K. (2009). Privacy-preserving e-payments using one-time payment details. Computer Standards & Interfaces, 31(2), 321–328. https://doi.org/10.1016/j.csi.2008.04.001
[7] Divgi, Dattaprasad R. "Calculation of the tetrachoric correlation coefficient." Psychometrika 44.2 (1979): 169-172. [8] https://www.datuapstrade.lv/rus/spss/section_15/.
[8] Cash https://www.statmethods.ru/statistics-metody/korrelyatsionnyj-analiz/.
[9] https://works.doklad.ru/view/qUqTfq7xYgM.html.
[10] Boaz, Mona, and Shmuel Smetana. "Regression equation predicts dietary phosphorus intake from estimate of dietary protein intake." Journal of the American Dietetic Association 96.12 (1996): 1268-1270.
[11] Weisstein, Eric W. "Correlation coefficient." https://mathworld. wolfram. com/ (2006).
[12] Eric W. "Correlation coefficient." https://mathworld. wolfram. com/ (2006).
[13] Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. Bmj. 1998 May 9;316(7142):1455-60.
[14] Filliben, James J. "The probability plot correlation coefficient test for normality." Technometrics 17.1 (1975): 111-117.
[15] Fisher, Ronald A. "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." Biometrika 10.4 (1915): 507-521.
[16] Adler, Jeremy, and Ingela Parmryd. "Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient." Cytometry Part A 77.8 (2010): 733-742.
[17] Faller, Alan J. "An average correlation coefficient." Journal of Applied Meteorology 20, no. 2 (1981): 203-205.