Research Article
Open Access

Application of correlation-regressional analysis for modeling the influence of the component composition of drinking water on the features of life activity tests

Abstract

The scientific and practical significance of multifactor correlation analysis is shown and substantiated. This achieves a more objective and comprehensive assessment of the degree of negative impact on the environment. To assess the quality of drinking water, blood cells of test-organisms were examined. The prospects of using hematological parameters of test organisms in bio testing are shown. The method consists in determining the action of toxicants on specially selected organisms under standard conditions with the registration of changes at cellular levels. As an optimal set for determining some structural and functional changes in the cell of blood due to toxic effects, a factor indicators are lymphocytes, monocytes, segmental neutrophils, rod-shaped neutrophils, basophils, eosinophil's.

Keywords: correlation-regression analysis, drinking water, blood cells

Volume 6 Issue 2 - 2020

Maya Vergolyas,[1] Vitaliy Khromyshev,[2] Elena Khromysheva,[2] Igor Khaliman[3]

[1]Department of Fundamental Disciplines with a course of pharmacology, International Academy of Ecology and Medicine, Ukraine
[2]Department of Organic and Biological Chemistry, Bogdan Khmelnitsky Melitopol State Pedagogical University, Ukraine
[3]Department of Environmental Safety and Environmental Management, Bogdan Khmelnitsky Melitopol State Pedagogical University, Ukraine

Correspondence: Maya Vergolyas, Department of Fundamental Disciplines with a course of pharmacology, International Academy of Ecology and Medicine, Kiev, Ukraine, Fax (8044)5608965, Tel (8050)5410100, Email vergoyas@meta.ua

Received: July 26, 2020 | **Published:** August 12, 2020

Introduction

The problem of studying the relationship between environmental indicators and the current environmental factors is one of the most important problems in the analysis of the negative man-made impact on the environment and public health. Any environmental policy involves the regulation of environmental variables and should be based on knowledge of how these variables affect certain performance indicators, the improvement of which is the ultimate goal of the decision-maker.[1] In practice, not all environmental phenomena and processes can be studied using the method of deterministic factor analysis, because in most cases they cannot be reduced to mathematical dependencies, where the values of the factor corresponds to a single value of the performance indicator.[2] Stochastic dependencies, which differ in approximation and uncertainty, are more common in environmental studies. They are manifested only on average by a significant number of objects and observations. In stochastic dependences, each value of the factor indicator can correspond to several values of the performance indicator. This is due to the fact that all the factors on which the performance indicator depends are complex and interrelated. Depending on how optimally the various factors are combined, the degree of influence of each of them on the value of the performance indicator will be equal.[2,3] The relationship between the factors and the performance indicator will be manifested if you take a large number of observations of the studied objects and compare their values. Then, in accordance with the law of large numbers, the influence of other factors on the performance indicator is smoothed out and neutralized. This makes it possible to establish the strength of the connection and the relationship between the phenomena being studied.

A correlation (stochastic) relationship is an incomplete, probabilistic relationship between indicators that manifests itself only in a mass of observations. There are paired and multiple correlations. Pairwise correlation is the relationship between two indicators, one of which is factorial and the other is effective.

Environmental phenomena and processes depend on many factors. As a rule, each factor separately does not define the phenomenon completely. Only a set of factors in their relationship can give a more or less complete picture of the nature of the phenomenon being studied. Multiple correlation arises from the interaction of several factors with the performance indicator.[3] The use of correlation analysis allows you to solve the following problems:

1. To determine the change in the performance indicator under the influence of one or more factors (in absolute terms), i.e. to determine how many units will change the value of the performance indicator when the factor changes by one.
2. To establish the relative degree of dependence of the performance indicator on each factor.

The study of correlations is of great importance in the analysis of environmental processes. This is manifested in the fact that the factor analysis is significantly deepened, the place and role of each factor in the formation of the level of research indicators is established, knowledge about the studied phenomena is deepened, the patterns of their development are determined. As a result, global environmental projects and current environmental measures are more accurately substantiated. Against the background of the research, the result of environmental activities of enterprises and organizations is more objectively assessed and the internal reserves for improving the environmental situation in the studied areas are more fully determined.[3,4] Multifactor correlation analysis consists of several stages:

I. At the first stage the factors influencing the studied indicator are defined and the most essential for the correlation analysis are selected.

II. In the second stage, the source information required for correlation analysis is collected and evaluated.

III. In the third stage, the nature is studied and the relationship between factors and performance is modeled, i.e. the mathematical equation that most accurately expresses the essence of the studied dependence is selected and substantiated.

IV. At the fourth stage the calculation of the main indicators of correlation analysis is carried out.

V. At the fifth stage the statistical estimation of results of the correlation analysis and their practical application is given.

Materials and methods

To select the factors that will be included in the multiple correlation in the future, a pairwise regression analysis is performed, which establishes a relationship between the performance indicator and each of the factors. Substantiation of the connection equation is performed by comparing parallel series, grouping data and line graphs. The placement of points on the graph will show the relationship between the studied indicators: rectilinear or curvilinear. The simplest equation, which characterizes the rectilinear relationship between two indicators, is the equation of the line:

$$Y_x = a + bx$$

where x is the factor index; Y - performance indicator; a i b - regression parameters to be found.

This equation describes a relationship between two traits in which there is a steady increase or decrease in the value of the performance as the factor changes by a certain amount. Coefficient a is a constant value of the performance indicator, which is not related to the change of this factor and determines the basic level of action of other factors. Parameter b shows the change in the performance indicator with increasing or decreasing the value of the factor per unit of measurement. Substituting the corresponding value of x into the regression equation, we can determine the aligned (theoretical) value of the performance indicator Y.[1–4] Thus, it is possible to determine the degree of dependence between the phenomena studied, but regression analysis does not answer the question: whether this relationship is close, and whether this factor has a significant or secondary effect on the value of the performance indicator. To measure the closeness of the relationship between factor and performance indicators, a correlation coefficient is determined. In the case of a rectilinear form of relationship between the studied indicators, the correlation coefficient is calculated by the formula:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}} = \frac{\overline{xy} - \sum x \sum y}{\sqrt{\left(\sum x^2 n - \left(\sum x\right)^2\right)\left(\sum y^2 n - \left(\sum y\right)^2\right)}}$$

The correlation coefficient can take values from 0 to ±1. The closer its value is to 1, the closer the relationship between the phenomena being studied and vice versa. The multifactor model includes those of the factors whose correlation coefficients are quite high, and the relationship between factor and performance is straightforward. One of the conditions of correlation analysis is the homogeneity of the studied information with respect to its distribution near the average level. The criteria for the homogeneity of information are the standard deviation and the coefficient of variation, which are calculated for each factor and performance indicator. The standard deviation shows the absolute deviation of individual values from the arithmetic mean. It is determined by the formula:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

The coefficient of variation characterizes the relative degree of deviation of individual values from the arithmetic mean. The formula is used for its calculation:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

The higher the coefficient of variation, the more scattered the data around the average value of this indicator and the less uniformity (homogeneity) of the studied objects. The variability of the variation series is considered insignificant if the coefficient of variation does not exceed 10%, medium - at 10-20%, significant - more than 20%, but less than 33%. The value of the coefficient of variation, exceeding 33% indicates the heterogeneity of information and the impossibility of its use in this form in further calculations. After selecting factors and evaluating the source information, an important task in multifactor correlation analysis is to model the relationship between factor and performance indicators, ie to select the appropriate equation that best describes the dependencies being studied. The same techniques are used to substantiate it as to establish the existence of a connection. If the relationship of all factor indicators with the results is rectilinear, then to record these dependencies, you can use a linear function:

$$Y_x = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n$$

In this model, the coefficients b_i show how many units the performance indicator will change with the change of the factor by one in absolute terms. The solution of the problem of multifactor correlation analysis is performed on a PC using standard programs using the extensive capabilities of Microsoft Excel. The connection (regression) equation is usually calculated step by step. First, one factor is taken into account, which has the greatest impact on the performance indicator, then the second, third, etc. At each step, calculate the relationship equation, multiple correlation and determination coefficient, Fisher's criterion, and so on. Their size at each step is compared with the previous one. The greater the value of the coefficients of multiple correlation and determination, and the greater the Fisher's calculation criterion of the corresponding tabular value, the more accurately the equation of the relationship describes the relationships between the studied indicators. If the addition of the following factors does not improve the estimated indicators of the relationship, they should be discarded, i.e. stop at the level where these indicators are most important. The coefficients of the equation show the quantitative impact of each factor on the performance of others.[1–4] In order to ensure the reliability of the equation of communication and the legitimacy of its use for practical purposes, it is necessary to give a statistical assessment of the reliability of communication indicators.

Fisher's criterion (F-ratio) is used for this purpose. Fisher's criterion is calculated as follows:

$$F = \frac{\sigma^2_{âîñïîä}}{\sigma^2_{îñò}} \quad \sigma^2_{âîñïîä} = \frac{\sum\left(Y_{x_i} - \overline{Y}_x\right)^2}{m-1}$$

$$\sigma^2_{îñò} = \frac{\sum\left(Y_i - Y_{x_i}\right)^2}{n-m}$$

where Y_{x_i} the individual value of the performance indicator, calculated by the equation; $\overline{Y}_x$ the average value of the performance indicator, calculated from the original data; Y_i the actual individual value of the performance indicator; m is the number of parameters in the communication equation, taking into account the free member of the equation; n is the number of observations. The actual value of the F-ratio is compared with the table and a conclusion is made about the reliability of the connection. If Ffact <Ftable, then the hypothesis of no relationship between performance and factor indicators is rejected and the model is considered adequate. If the correlation analysis is performed correctly, the obtained equation can be used for practical purposes:[4]

1) Assessment of the results of the impact of harmful factors on the environment.
2) Calculation of the influence of factors on the growth of the performance indicator.
3) Calculation of reserves to improve the level of the studied indicator.
4) Planning and forecasting its value.

Thus, multifactor correlation analysis has important scientific and practical significance. It allows us to study the patterns of change of the performance indicator depending on the behavior of various factors, to determine their impact on the value of the performance indicator, to establish which of them are primary and which are secondary. This achieves a more objective and comprehensive assessment of the degree of negative environmental impacts.[4,5]

Results and discussion

Correlation-regression analysis of the influence of the component composition of drinking water on the vital signs of test organisms, on the example of fish. Consider a multifactor correlation model, where the effective indicators are nitrates, mg/dm^3,[6] bicarbonates, mg / dm^3, calcium, mg / dm^3, chlorides, mg / dm^3, dry organic matter residue, mg/dm^3. Factor indicators are lymphocytes, monocytes, segmental neutrophils, rod-shaped neutrophils, basophils, eosinophils.[7,8] Correlation analysis revealed a high relationship between influencing factors and indicators. The results of the analysis are presented in Table 1. Thus, the bulk of the active factors have a strong influence on the number of formed elements in the blood of fish, namely: lymphocytes, monocytes, basophils, eosinophils (where the correlation coefficient significantly exceeds 0.55). The composition of drinking water has a weaker effect on segmental neutrophils (here the maximum value of the correlation coefficient is 0.327) and rod-shaped neutrophils (here the maximum value of the correlation coefficient is 0.518) the effect of nitrates on monocytes is insignificant (here the value of the correlation coefficient is 0.389).[7,8] The above indicates the correct choice of current factors to build the model. Paired linear models show a high density of fit of the experimental data to the linearly aligned curves. This is evidenced by the examples of graphical images of the developed linear models, which are presented in Figure1-Figure 3. As can be seen from Tables 2 & Table 3, adequate linear multifactor models have been constructed, which have high values of grouping density between actual and calculated data (all coefficients of determination and correlation are close to and greater than 0.9).

Table 1 Analysis of correlations

Estimation of density correlation connections	Lymphocytes	Monocytes	Segmental neutrophils	Stick core neutrophils	Basophils	Eosinophils
Chlorides, mg / dm^3	-0,791	0,674	-0,219	0,518	0,810	0,881
Nitrates, mg / dm^3	-0,583	0,389	-0,085	0,109	0,650	0,737
Bicarbonate, mg / dm^3	-0,882	0,715	0,050	0,476	0,837	0,947
Calcium, mg / dm^3	-0,672	0,554	-0,327	0,320	0,769	0,791
Dry residue, mg / dm^3	-0,767	0,609	-0,123	0,352	0,797	0,880
Characteristics of correlation	Strong feedback	Strong direct connection	Weak feedback	Weak direct connection	Strong direct connection	Strong direct connection

Table 2 Designation of parameters of models

Marking	Factors	Marking	Performance indicators
X1	Chlorides, mg/dm^3	Y1	Lymphocytes
X2	Nitrates, mg/dm^3	Y2	Monocytes
X3	Bicarbonates, mg/dm^3	Y3	Segmental neutrophils
X4	Calcium, mg/dm^3	Y4	Rod-shaped neutrophils
X5	Dry residue, mg/dm^3	Y5	Basophils
		Y6	Eosinophils

Citation: Vergolyas M, Khromyshev V, Khromysheva E, et al. Application of correlation-regressional analysis for modeling the influence of the component composition of drinking water on the features of life activity tests. *Adv Tissue Eng Regen Med Open Access.* 2020;6(2):37–42. DOI: 10.15406/atroa.2020.06.00114

Table 3 Five-factor models of dependence of shaped elements of peripheral blood of fish on the content of inorganic components and dry residue of organic substances in drinking water

Complete model for lymphocytes: Y1=221,50,55X1+0,42X20,38X30,91X4+0,35X5					
0,34755	-0,91102	-0,37899	0,42411	-0,55027	221,52717
0,45399	0,92958	0,28930	2,45655	1,92588	31,07868
0,93772	11,48401				
3,01109	1	R^2=	0,93772	F=	3,01109
1985,54619	131,88238	Corel. =	0,96836	Ftab. =	6,60789
Complete model for monocytes: Y2=2,15-0,13X1-0,94X2+0,11X3-0,46X40,08X5					
-0,08498	0,45908	0,11521	-0,93534	-0,13168	2,15136
0,22567	0,46208	0,14381	1,22111	0,95732	15,44867
0,88397	5,70850				
1,52374	1,00000	R^2=	0,88397	F=	1,52374
248,27017	32,58698	Corel. =	0,94020	Ftab. =	6,60789
Complete model for segmental neutrophils: Y3=18,45+0,39X1+0,69X2-0,016X3-0,4X4-0,034X5					
0,03371	-0,40085	-0,01658	0,69264	0,38723	18,49649
0,07712	0,15791	0,04915	0,41730	0,32716	5,27946
0,94380	1,95083				
3,35852	1,00000	R^2=	0,94380	F=	3,35852
63,90853	3,80575	Corel. =	0,97149	Ftab. =	6,60789
Complete model for rod-shaped neutrophils: Y4=2,87+0,11X1-0,2X2+0,03X3-0,12X4-0,36X5					
-0,03565	0,12035	0,03299	-0,20092	0,10888	2,87495
0,05922	0,12126	0,03774	0,32045	0,25122	4,05407
0,87124	1,49804				
1,35327	1,00000	R^2=	0,87124	F=	1,35327
15,18446	2,24411	Corel. =	0,93340	Ftab. =	6,60789
Complete model for basophils: Y5=2,84-0,089X1-0,21X2+0,15X3+0,54X4-0,16X5					
-0,15860	0,53835	0,14621	-0,20649	-0,08860	2,83978
0,13865	0,28391	0,08836	0,75027	0,58819	9,49190
0,94946	3,50739				
3,75761	1,00000	R^2=	0,94946	F=	3,75761
231,12678	12,30179	Corel. =	0,97440	Ftab. =	6,60789
Complete model for eosinophils: Y6=-0,05+0,279X1+0,094X2+0,091X3+0,17X4-0,083X5					
-0,08295	0,17319	0,09098	0,09386	0,26664	-0,04998
0,11880	0,24326	0,07571	0,64284	0,50397	8,13277
0,95607	3,00517				
4,35242	1,00000	R^2=	0,95607	F=	4,35242
196,53474	9,03106	Corel. =	0,97779	Ftab. =	6,60789

Citation: Vergolyas M, Khromyshev V, Khromysheva E, et al. Application of correlation-regressional analysis for modeling the influence of the component composition of drinking water on the features of life activity tests. *Adv Tissue Eng Regen Med Open Access.* 2020;6(2):37–42. DOI: 10.15406/atroa.2020.06.00114

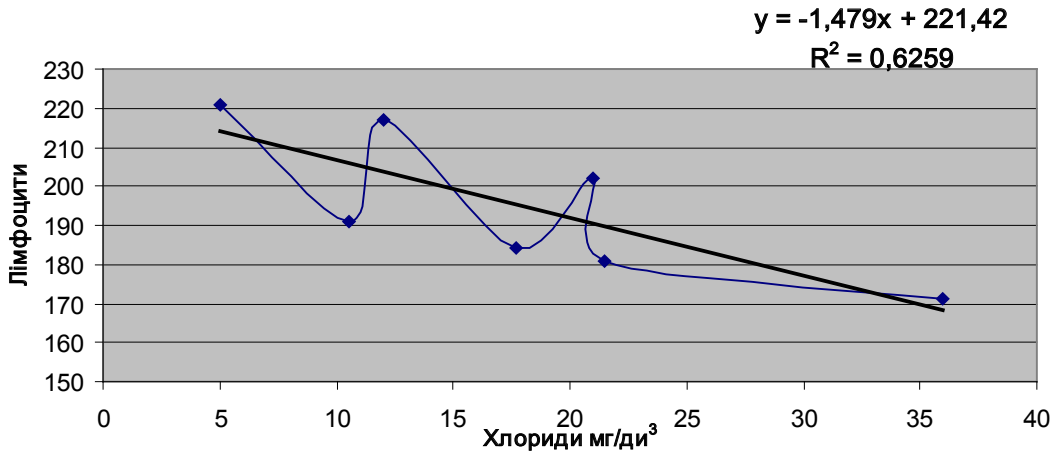

Figure 1 Linear model "Lymphocytes-chlorides".

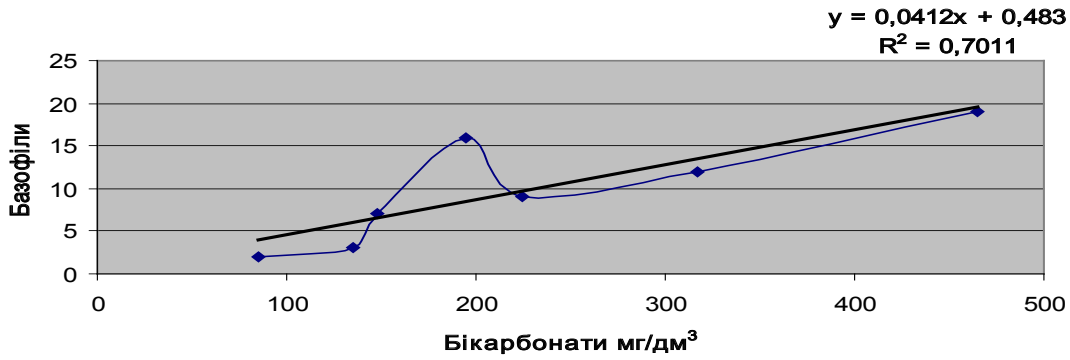

Figure 2 Linear model "Basophils-bicarbonates".

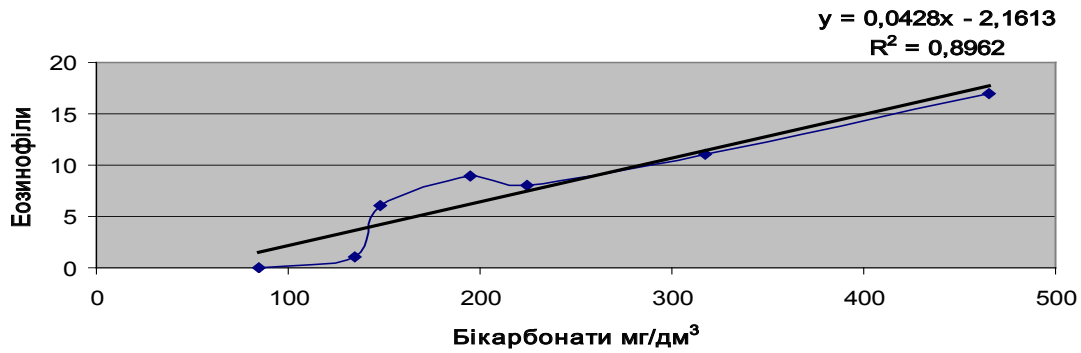

Figure 3 Linear model "Eosinophils-bicarbonate.

Future prospective

These models allow with high accuracy to predict the impact of drinking water parameters on the performance of fish and, thus, to predict the quality of drinking water in a wide range of studied parameters. The obtained patterns allow for a more in-depth factor analysis of the relationships between the complex action of the components of drinking water and the level of life of living organisms.[7,8] To improve the quality of drinking water, it is possible to recommend the introduction of such environmental measures aimed at optimizing the composition of drinking water, which will help to increase the level of protective forces of all plant and animal organisms, including humans.

Conclusion

The following prognostic conclusions can be made from the conducted correlation-regression analysis and modeling of the influence of current factors on the test objects.

a) Increase in the content of chlorides in water will lead to a significant increase in their toxic effects primarily on blood lymphocytes, basophils, eosinophils and monocytes.

b) An increase in the content of nitrates in water will have a negative effect on the content of basophils, eosinophils and lymphocytes in the blood.

Citation: Vergolyas M, Khromyshev V, Khromysheva E, et al. Application of correlation-regressional analysis for modeling the influence of the component composition of drinking water on the features of life activity tests. *Adv Tissue Eng Regen Med Open Access.* 2020;6(2):37–42. DOI: 10.15406/atroa.2020.06.00114

c) Increase in bicarbonate content will have a detrimental effect on eosinophils, lymphocytes, basophils and monocytes.

d) Increased calcium content will affect the concentration of eosinophils, basophils and blood lymphocytes;

e) An increase in the dry residue in the water indicates a deviation from normal concentrations of lymphocytes, eosinophils, basophils and monocytes.

f) It should be noted that deviations from the normal values of most components of drinking water will not significantly affect the segmental and rod-shaped neutrophils.

This dependence indicates the presence and the body of the studied fish inflammatory and infectious processes, as well as indicates the adaptation of fish immunity. Therefore, to improve the quality of drinking water, it is recommended to implement such organizational and environmental measures aimed at optimizing the component composition of drinking water, which will increase the level of protective forces of living organisms, including humans.

Acknowledgments

None.

Conflicts of interest

The author declares that there is no conflict of interest.

References

1. Kuznetsov DI, Mammadov RM. Ecological assessment by the method of biotesting the quality of water bodies in the regions of oil production. *Siberian Ecological Journal*. 2009;3:37–39.
2. Kaza M, Mankiewicz-Boczek J, Izydorczyk K, et al. Toxicity Assessment of Water Samples from Rivers in Central Poland Using a Battery of Microbiotests - a Pilot Study. *Polish J of Environ Stud*. 2007;16(1):81–89.
3. Espigares M, Roman I, Gonzalez Alonso JM, et al. Proposal and application of an ecotoxicity biotest based on *Escherichia coli*. *Journal of Applied Toxicology*. 1990;10(6):443–446.
4. Bulgakov NG, Dubinina VG, Levich AP, et al. A Method of Searching for Correlation Between Hydrobiological Indices and Abiotic Factors (Using Commercial Fish Catches and Productivity as Examples). *Biology Bulletin of the Russian Academy of Science*. 1995;22(2):184–190.
5. Krainyukov OM. Regression analysis of the relationship between the results of measurements of the component composition and determination of levels of water toxicity. *Bulletin of KhNU Ser Ecology*. 2013:68–73.
6. Khromyshev VO, Khromysheva OO, Levina AV, et al. Determination of nitrate ions in food by standard and express methods / Materials of the XI International Research and Practical Conference. *Trends of Modern Science*. 2015;19:56–58.
7. Vergolyas MR. Determination of toxicity of water samples using hematological parameters of fish. Factors of experimental evolution of organisms. Kyiv. 2015;17:299–302.
8. Vergolyas MR. Blood as integrated system of organism. *Science Rise*. 2016;2(1):7–11.