

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН



ҚазҰТУ ХАБАРШЫСЫ _____

_____ **ВЕСТНИК КазНУ**

VESTNIK KazNRTU _____

№6 (130)

Главный редактор
И. К. Бейсембетов – ректор

Зам. главного редактора
Б.К. Кенжалиев – проректор по науке

Отв. секретарь
Н.Ф. Федосенко

Редакционная коллегия:

С.Б. Абдыгаппарова, Б.С. Ахметов, З.С. Абишева- акад. НАНРК, Л.Б. Атымтаева, Ж.Ж. Байгунчеков- акад. НАНРК, А.Б. Байбатша, А.О. Байконурова, В.И. Волчихин (Россия), К. Дребеншted (Германия), Г.Ж. Жолтаев, Р.М. Искаков, С.Е. Кудайбергенов, С.Е. Кумеков, В.А. Луганов, С.С. Набойченко – член-корр. РАН, И.Г. Милев (Германия), С. Пежовник (Словения), Б.Р. Ракишев – акад. НАН РК, М.Б. Панфилов (Франция), Н.Т. Сайлаубеков, А.Р. Сейткулов, Фатхи Хабаши (Канада), Бражендра Мишра (США), Корби Андерсон (США), В.А. Гольцев (Россия), В. Ю. Коровин (Украина), М.Г. Мустафин (Россия), Фан Хуаан (Швеция), Х.П. Цинке (Германия), Т.А. Чепуштанова, Г.Ж. Елигбаева, Б.У. Куспангалиев

Учредитель:

Казахский национальный исследовательский технический университет
имени К.И. Сатпаева

Регистрация:

Министерство культуры, информации и общественного согласия
Республики Казахстан № 951 – Ж “25” 11. 1999 г.

Основан в августе 1994 г. Выходит 6 раз в год

Адрес редакции:

г. Алматы, ул. Сатпаева, 22,
каб. 616, тел. 292-63-46
Nina. Fedorovna. 52 @ mail.ru

[15] Ермуханова Н.Б. Научные основы радиационной безопасности в нефтегазовых объектах//Актуальные научные исследования в современном мире. 2018. № 5-8 (37). С. 21-26.

[16] Цомбуева Б.В., Горяшкина З.В., Щербакова Л.Ф. Метод очистки почвы от нефтяного загрязнения с помощью природных сорбентов//Вестник Волгоградского государственного университета. Серия 11: Естественные науки. - Волгоград, 2017. - Т.7. № 2. - С.19-25.

[17] Абдибаттаева М.М., Нурымова Р.Д., Жиенбаева Л.Б., Бекетова А.К. Оценка эффективности использования рисовой шелухи в качестве нефтесорбента//Научный журнал: Исследования, результаты. - Алматы, 2014. - С.112-118.

[18] D. L.Bish, D.W.Ming.Natural Zeolites: Occurrence, properties, applications/D.L. Bish,D.W.Ming(Eds.). - Mineral.Geochem,2001.-Vol.45.-P.619-654.

Ермуханова Н.Б., Танжарыков П.А., Керимбекова З.М.

Оценка техногенного воздействия радионуклидов на почву нефтегазовых месторождений

Резюме. Рассматриваются проблемы загрязнения почв радионуклидами, такие как проблема очистки и дегазации почвенного загрязнения и ранее использованных технологических трубопроводов. На нефтяных и газовых месторождениях загрязнителями почвы с радионуклидами являются химические реагенты, буровые растворы, буровые и нефтяные шламы, сырая нефть и загрязненная нефтью почва. Основными дозообразующими радиоактивными источниками в почвах, загрязненных нефтяными отходами, являются радиевые, ториевые и радоновые изотопы. Накипь устаревшего оборудования состоит в основном из солей радиоактивных изотопов. Эти соли в почве усиливают техногенный эффект. Радиоактивные источники в загрязненных почвах в крупнейших нефтедобывающих зонах Казахстана вызвали ухудшение состояния окружающей среды. Поэтому основной проблемой является прежде всего очистка почвы от радионуклидов и, во-вторых, улучшение методов дезактивации ранее использованных технологических трубопроводов С этой целью был проведен и проанализирован мониторинг загрязнения почв в месторождение Кумколь.

Ключевые слова. загрязнение почв радионуклидами, химические реагенты, буровые растворы, пластовые воды, изотопы радий, торий и радон, осадочных отложений в технологическом оборудовании.

Ermukhanova N.B., Tanzharykov P.A., Kerimbekova Z.M.

Assessment of the technogenic impact of radionuclides on the soil of oil and gas deposits

Summary. The article deals with the problems of soil contamination with radionuclides, such as the problem of cleaning and degassing of soil pollution and previously used technological pipelines. In oil and gas fields, soil pollutants with radionuclides are chemical reagents, drilling fluids, drilling and oil sludge, crude oil and oil-contaminated soil. The main dose-generating radioactive sources in soils contaminated with oil waste are radium, thorium and radon isotopes. The scale of obsolete equipment consists mainly of salts of radioactive isotopes. These salts in the soil enhance the technogenic effect. Radioactive sources in contaminated soils in the largest oil-producing areas of Kazakhstan have caused environmental degradation. Therefore, the main problem is first of all cleaning the soil from radionuclides and, secondly, improving the methods of decontamination of previously used technological pipelines for this purpose, monitoring of soil contamination in the Kumkol field was carried out and analyzed.

Key words: soil contamination with radionuclides, chemical reagents, drilling muds, formation water, isotopes of radium, thorium and radon, sediment in technological equipment.

UDK 004.89;004.93;004:85

D.K. Darkenbayev

(Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan
e-mail: dauren.kadyrovich@gmail.com)

NUMERICAL SOLUTION OF THE REGRESSION MODEL FOR ANALYSIS AND PROCESSING OF BIG DATA

Abstract: In this article, numerical methods for solving the regression model for processing large amounts of data are considered. Due to the increase in the population on the globe, the world data volume is more than doubled every two years. There were great difficulties in data processing and storage. One of the most important problems today is the processing of large amounts of data, data security and rapid analysis, as well as the determination of their structure, monitoring of growth rates and speed of processing. The article presents a regression analysis of external factors affecting the growth of the data volume, a numerical solution of the regression model for processing large amounts of data. I found the regression coefficient, the free term and the coefficient of determination of this problem using the method of least squares. Defined the direction of the data and plotted the function. I was convinced that the method of least squares is the

most optimal and effective method for analyzing and constructing a regression model. Accordingly, these methods are a reliable statistical tool for testing and evaluating linear relationships.

Key words: Model, regression analysis, data, information, processing, function, variables, characteristics of relationships.

Introduction

Currently, many companies compete with each other by increasing the volume of processing big data, as the consumers of their software grow every year. This leads to the fact that software development companies must improve the processing of big data. Today the processing of large data and ensuring the safety and storage of them in a safe place is very important for everyone, the solution of such a task must be fast and reliable [1]. Usually big data sets are a process of continuous accumulation of various types of unstructured data. It describes a set of data that exponentially grows, which are big, rough and unstructured for analysis by relational database methods. Terabytes or petabytes - the exact amount is not as important as understanding where the data ends and how they can be used. The information value of big data sets, examples of tasks that can be solved by analyzing big data streams:

- predict the amount of data, conduct regression analysis, improve the work of technical support services, websites and traffic;
- development of predictive models
- detection of fraud in real time
- Risk analysis
- construction of situational premises;
- interactive analytical processing, etc.

Regression analysis is one example that includes methods and analyzes of processing big amounts of data. This problem is solved by using an adequate regression model using the existing and new technology algorithms. The regression model is an equation in which the variable explained is represented as a function of the explanatory variables. For example, the model of demand for some goods, depending on its price and the income of buyers. By the form of the function, linear and nonlinear regression models are distinguished. The methods of evaluation and analysis of linear regression models are most often studied in detail and in the econometric analysis[2,3].

The regression equation

The main goal of the regression analysis is to determine the relationship between some characteristic Y of the observed phenomenon or object and the variables x_1, x_2, x_n , which determine, explain the changes in Y . The variable Y is called the dependent variable (response), the variables x_1, x_2, x_n are called factors (regressors). The establishment of the form of dependence, the selection of the regression model (equations) and the evaluation of its parameters are the tasks of regression analysis[4].

Regression analysis studies models of the form $Y = \varphi(X) + \varepsilon$, where Y is the resultant characteristic (response, random dependent variable); X - factor (nonrandom non-dependent variable); ε is a random variable characterizing the deviation of factor X from the regression line (residual variable). The regression equation is written as: $y_x = \varphi(x, b_0, b_1, \dots, b_p)$, where x is the value of X ; $y_x = M_x(Y)$, $b_0, b_1, \dots, b_p$ are the parameters of the regression function φ . Thus, the problem of regression analysis consists in determining the function and its parameters and the subsequent statistical study of the equation. Depending on the type of the chosen equation, linear and nonlinear regression are distinguished (in the latter case, further refinement is possible: quadratic, exponential, logarithmic, etc.). Depending on the number of interrelated characteristics, pair and multiple regression are distinguished. If the relationship between the two characteristics (result and factor) is studied, then the regression is called a pair if there are multiple (multifactorial) regressions between three or more signs.

At the first stage of the regression analysis, the observational data or the experiment is represented graphically.

The dependence between the variables X and Y is represented by points on the coordinate plane (x, y) and connect them by a broken line. This broken line is called the empirical line of Y regression by X . By the form of the empirical regression line, the assumption is made about the form (form) of the dependence of the variable Y on X . In this case it is logical to assume a linear dependence. If the form of the function φ

in the regression equation is chosen, then the least-squares (OLS) method is used to estimate the unknown parameters $b_0, b_1, \dots, b_p$. According to the method, the unknown parameters of the function are chosen so that the sum of the squares of the deviations of the experimental (empirical) values of y_i from their calculated (theoretical) values is minimal, i.e.

$$S = \sum_{i=1}^n (y_{i\text{exp}} - y_i^p)^2 = \sum_{i=1}^n (y_{i\text{exp}} - \varphi(x_i, b_0, b_1, \dots, b_p))^2 \rightarrow \min$$

where y_i^p is the value calculated from the regression equation; $y_i - y_i^p = \varepsilon$ - deviation (error, balance); n is the number of pairs of source data[5].

Pairwise linear regression model

Let's consider a pair linear regression model of the interrelation of two variables, for which the regression function $\varphi(x)$ is linear. We denote by $y(x)$ the conditional mean of Y in the general population for a fixed value of x with the variable X . Then the regression equation will have the form:

$y_x = ax + b$, where a is the regression coefficient (linear slope of the linear regression line). The regression coefficient shows how many variables on average Y varies with the change in the variable X by one unit. With the help of the least squares method, formulas are obtained for calculating the parameters of linear regression:

Table 1. Formulas for calculating linear regression parameters

Free member b	The regression coefficient a	Coefficient of determination
$b = \frac{\overline{y} \cdot \overline{x^2} - \overline{x} \cdot \overline{xy}}{\overline{x^2} - (\overline{x})^2}$	$a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2}$	$R^2 = \frac{\sum (y_i^p - \overline{y})^2}{\sum (y_i - \overline{y})^2}$
Testing the hypothesis of the significance of the regression equation		
$H_0 : R^2 = 0$	$H_1 : R^2 > 0$	$F_{obs} = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p}$
$F_{cur}(\alpha; k_1; k_2), k_1 = p, k_2 = n - p - 1, \text{ Annex 7 (for linear regression } p = 1)$		

The direction of communication between variables is determined based on the sign of the regression coefficient. If the sign with a regression coefficient is positive, the relationship between the dependent variable and the independent variable will be positive. If the sign for the coefficient of regression is negative, the relationship between the dependent variable and the independent variable is negative (inverse)[6].

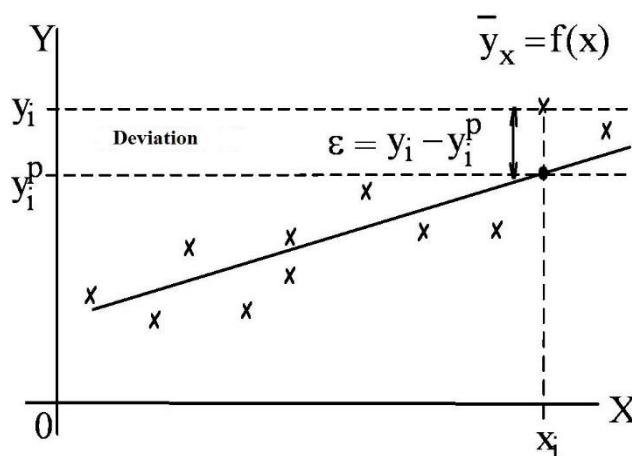


Figure 1 - The concept of deviation (linear regression)

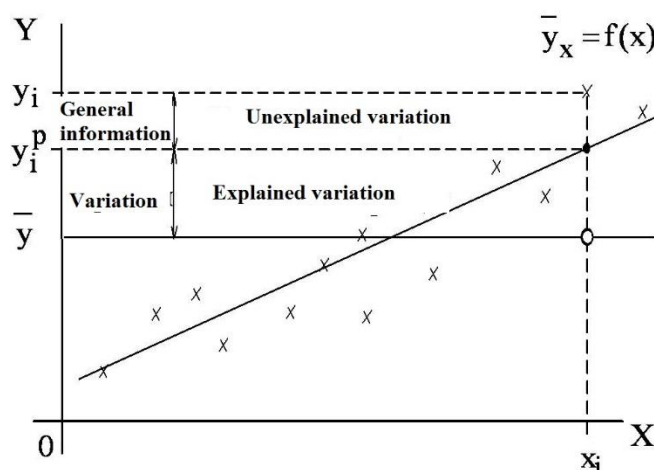


Figure 2 - Graphical interpretation of the determination coefficient (linear regression)

To analyze the overall quality of the regression equation, the determination coefficient R_2 , also called the square of the multiple correlation coefficient, is used. The coefficient of determination (measure of certainty) is always within the interval $[0;1]$. If the value of R_2 is close to one, this means that the model constructed explains almost all the variability of the corresponding variables. Conversely, the R_2 value is close to zero, which means poor quality of the constructed model. The coefficient of determination R_2 shows, by how many percentages $(R^2) \cdot 100\%$, the regression function found describes the relationship between the initial values of Y and X . Fig. 3 shows $(y_i^p - \bar{y})$ - the variation explained by the regression model and $(y_i - \bar{y})$ is the total variation. Accordingly, the quantity $(1 - R^2) \cdot 100\%$ shows how many percent of the variation of the parameter Y are due to factors not included in the regression model. With a high value of the determination coefficient $R^2 \geq 75\%$, we can make a prediction $y^* = f(x^*)$ for a particular value of x^* within the range of the original data. For forecasts of values that are not within the range of the initial data, the validity of the obtained model can not be guaranteed. This is due to the fact that the influence of new factors that the model does not take into account may appear. The significance of the regression equation is estimated using the Fisher criterion (see Table 1). Under the condition that the null hypothesis is valid, the criterion has a Fisher distribution with the number of degrees of freedom $k_1 = p, k_2 = n - p - 1$ (for paired linear regression $p = 1$). If the null hypothesis deviates, then the regression equation is considered statistically significant. If the null hypothesis does not deviate, then the statistical insignificance or unreliability of the regression equation is recognized[7].

Formulation of the problem.

In the mechanical shop, the structure of the cost of production and the share of purchased components are analyzed. It was noted that the cost of components depends on the time of their delivery. As the most important factor affecting the delivery time, the distance traveled is selected. Regression analysis of supply data:

Distance, miles	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Time, min	16	13	19	18	12	11	8	14	9	16

For regression analysis:

1. to plot the initial data, to approximately determine the nature of the dependence;
2. Select the form of the regression function and determine the numerical coefficients of the model by the method of least squares and the direction of the connection;
3. Evaluate the strength of regression dependence using the coefficient of determination;

4. Evaluate the significance of the regression equation;

5. Make a forecast (or a conclusion about the impossibility of forecasting) according to the adopted model for a distance of 2 miles.

1. The constructed points are not exactly on the line: apart from the distance at the time of delivery, traffic jams, time of day, road works, weather, driver's qualification, mode of transport influence. But these points are collected along a straight line, so we can assume a linear positive relationship between the parameters.

2. Let us calculate the sums necessary for calculating the coefficients of the linear regression equation and the determination coefficient R_2 :

№	x_i	y_i	x_i^2	$x_i y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	12,25	56,00	15,22	2,63	5,76
2	2,4	13	5,76	31,20	12,30	1,70	0,36
3	4,9	19	24,01	93,10	18,95	28,59	29,16
4	4,2	18	17,64	75,60	17,09	12,15	19,36
5	3,0	12	9,00	36,00	13,89	0,08	2,56
6	1,3	11	1,69	14,30	9,37	17,88	6,76
7	1,0	8	1,00	8,00	8,57	25,27	31,36
8	3,0	14	9,00	42,00	13,89	0,09	0,16
9	1,5	9	2,25	13,50	9,90	13,67	21,16
10	4,1	16	16,81	65,60	16,82	10,36	5,76
Σ	28,9	136	99,41	435,30	—	112,42	122,40

$$\bar{x} = \frac{\sum n_i x_i}{n} = 2,89; \bar{y} = \frac{\sum n_i y_i}{n} = 13,6; b = \frac{13,6 \cdot 9,941 - 2,89 \cdot 43,53}{9,941 - 2,89^2} = 5,91; a = \frac{43,53 - 2,89 \cdot 13,6}{9,941 - 2,89^2} = 2,66.$$

The desired regression dependence has the form: $y^p = 2,66x + 5,91$. We determine the direction of the relationship between the variables: the sign of the regression coefficient is positive, therefore, the connection is also positive, which is confirmed by the graphical assumption[8].

3. Let's calculate the determination coefficient: $R^2 = \frac{112,41}{122,40} = 0,92$ or 92%. Thus, the linear model explains

92% of the variation in the delivery time, which means the correct choice of the factor (distance). 8% of the time variation is not explained, which is due to other factors affecting the delivery time, but not included in the linear regression model.

4. Let's check the significance of the regression equation:

$$F_{\text{observations}} = \frac{0,92^2}{1 - 0,92^2} \cdot \frac{10 - 1 - 1}{1} = 44,1$$

As $F_{observations} = 44,1 > F_{curve}(0,05;1;10-1-1) = 5,32$ - regression equation (linear model) is statistically significant.

5. Solve the problem of forecasting. Since the determination coefficient R_2 has a sufficiently high value and the distance of 2 miles for which it is necessary to make a prediction is within the range of the initial data, it is possible to make the prediction:

$$y^*(x = 2 \text{ miles}) = 2,66 \cdot 2 + 5,913 = 11,2 \text{ min.}$$

Regression analysis is conveniently carried out using the capabilities of various mathematical programs. The "Regression" mode of operation serves for calculating the parameters of the linear regression equation and checking its adequacy for the process under study. In the dialog box, you need to fill in the following parameters:

- The input interval Y is a range of data by the resultant characteristic, it must consist of one column.
- The input interval X is the range of cells containing the values of factors (independent variables).

Number of input ranges (*columns*) ≤ 16 .

- A check mark is set if the first line of the range contains a heading.
- The Reliability level checkbox is activated if you enter a reliability level in the field next to it, which is different from the default level 95%.

The constant is zero. This check box must be set if the regression line passes through the origin ($b = 0$).

- A check box in the Residual and Residuals group is set if you want to include the corresponding columns or graphs in the output range.

Conclusion of the results

Regression statistics	
Multiple R	0,958275757
R-square	0,918292427
The normalized R-square	0,90807898
Standard error	1,11809028
Observations	10

	Coefficients	Standard error	t-statistics	P-Value
Y-crossing	5,913462144	0,884389599	6,686489927	0,00015485
Variable X	2.65970168	0,280497238	9,482095791	1,26072E-05

Consider the results of the regression analysis presented in the table. The value of the R -square, also called a measure of certainty, characterizes the quality of the regression line obtained. This quality is expressed by the degree of correspondence between the initial data and the regression model (calculated data). In our example, the measure of certainty is 0,91829, which indicates a very good fit of the regression line to the original data and coincides with the coefficient of determination R_2 , calculated by the formula. Multiple R - coefficient of multiple correlation R - expresses the degree of dependence of the independent variables X and the dependent variable Y and is equal to the square root of the determination coefficient. In a simple linear regression analysis, the multiple coefficient R is equal to the linear correlation coefficient ($r = 0,958$). Coefficients of the linear model: the Y -intersection derives the value of the free term b , and the variable x_1 - the regression coefficient a . Then the linear regression equation: $y = 2,659x + 5,9135$ (which agrees well with the calculation results). Next, we check the significance of the regression coefficients: a and b . Comparing the values of the columns coefficients and standard error in the table in pairs, we see that the absolute values of the coefficients are greater than their standard errors. In addition, these coefficients are significant, which can be judged from the values of the P -value, which are less than a given level of significance $\alpha = 0,05$ [9].

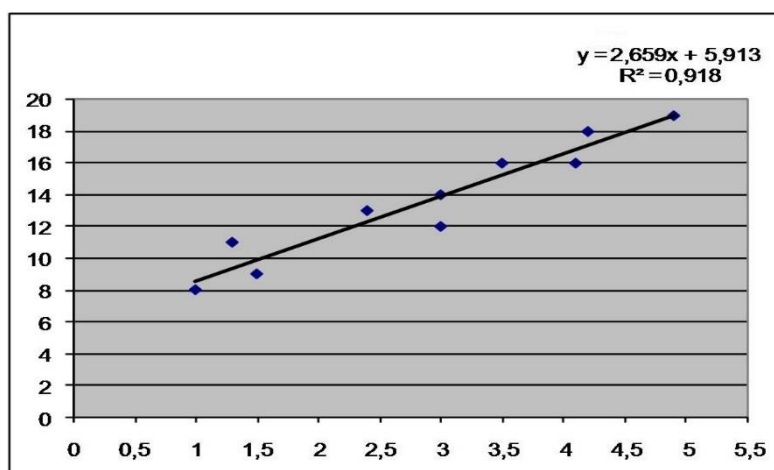


Figure 3. Initial data and regression line

Observation	Predicted Y	Remains	Standard errors
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
5	13,89256718	-1,892567185	-1,795356486
6	9,371074328	1,628925672	1,545256778
7	8,573163824	-0,573163824	-0,543723571
8	13,89256718	0,107432815	0,101914586
9	9,903014664	-0,903014664	-0,8566318
10	16,81823903	-0,818239033	-0,776210624

The table shows the results of the withdrawal of residuals. With this part of the report, we can see the deviations of each point from the constructed regression line. The greatest absolute value of the remainder in this case is 1,89256, the smallest is 0,055399. For better interpretation of these data, a plot of the initial data and a constructed regression line are constructed. As can be seen from the construction, the regression line is well "adjusted" to the values of the initial data, and the deviations are random[10].

Conclusion

The creation of a regression model is an iterative process. This process is aimed at finding effective independent variables, to explain the dependent variables that we are trying to model or understand by starting a regression tool to determine which quantities are effective predictors. Then, step-by-step delete or add variables until you will not find the best fitting regression model. As the process of creating a model is often research, it should never become a simple "fit" of data. Building process the regression model should take into account theoretical aspects, opinion of experts in this field and common sense.

REFERENCES

- [1] D.Caire, R.Kossmann. Credit scoring: Is It Right for Your Bank?, Bannock Consulting, 2003;
- [2] C.Ong, J.Huang, G.Tzeng, Building credit scoring models using genetic programming, Expert Systems with Applications 29, 2005-41-47p.
- [3] K.Nurlybayeva, G.Balakayeva. Algorithmic Scoring Models, Applied Mathematical Sciences, Vol.7, no.12, 2013-571-586p.
- [4] G.A.Timofeeva, A.V.Martynenko. Ekonometrika. - E.: UrGUPS, 2016 –
- [5] 9-12p.
- [6] Glen J.Myatt. Making Sense of Data. A Practical Guide to Exploratory Data Analysis and Data Mining. Published by John Wiley & Sons, Inc. Hoboken, New Jersey, 2007 – 162-164p.
- [7] John O.Rawlings, Sastry G.Pantula, David A.Dickey. Applied Regression Analysis: A research Tool, Second Edition. Springer Verlag, New York, Berlin, Heidelberg SPIN 10660129. 1998-24-25p.

[8] Daniel Zelterman. Applied Linear Models with SAS. Published in the USA by Cambridge University Press, New York. 2010-62-64p.

[9] 8.Rao V.Dukkipati Numerical Methods. Copyright 2010, New Age IntKinect for Windows SDK, <https://msdn.microsoft.com/en-us/library/hh855347.aspx> (date of review: 10.03.2015).

[10] Steven C.Chapra. Second Edition Applied Numerical Methods with Matlab for Engineers and Scientists. Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2012 by The McGraw-Hill Companies, Inc. All rights reserved. Previous editions © 2008 and 2005-320p.

[11] N.R.Draper, H.Smith. Applied Regression Analysis, Second Edition. Published by John Wiley & Sons New York, Chichester, Toronto, Singapore. 1981-29-31p.

Даркенбаев Д. К.

Үлкен көлемді деректерді талдаудағы және өңдеудегі регрессиялық модельдің сандық шешімі

Түйіндеме. Бұл мақалада үлкен көлемді деректерді өңдеудегі регрессиялық модельдің сандық шешімі қарастырылды. Жер шарында адам санының көбеюіне байланысты, дүние жүзілік ақпарат көлемі екі жыл сайын екі еседен артық өсіп отыр. Деректерді өңдеумен сақтауға қатысты қиындықтарда аз емес. Қазіргі таңдағы өзекті мәселелердің бірі деректерді өңдеу, жедел талдау жүргізу, деректердің қауіпсіздігін қамтамасыз ету, сонымен қатар деректердің құрылымын анықтау, өсу үдерісіне мониторинг жасау өңдеу жылдамдығын арттыру болып тұр. Мақалада ақпарат көлемінің артуына тікелей әсер ететін сыртқы факторларға талдау жасалып, үлкен көлемді деректерді өңдеудегі регрессиялық модельдің сандық шешімі табылды. Қойылған есептің регрессия коэффициентін, бос мүшесін және детерминация коэффициентін кіші квадраттар әдісін қолданып таптым. Алынған деректердің байланысу бағыты анықталып сәйкесінше функция сұлбасы алынды. Кіші квадраттар әдісі регрессиялық талдаудың ең тиімді әрі нәтижелі әдісі екеніне көз жеткіздім. Сәйкесінше бұл әдістер сызықтық қатынастарды тестілеу және бағалау үшін сенімді статистикалық құрал болып табылады.

Түйінді сөздер: Модель, регрессиялық талдау, деректер, ақпарат

Даркенбаев Д. К.

Численное решение регрессионной модели для анализа и обработки больших данных

Резюме: Рассмотрены численные методы решения регрессионной модели для обработки больших объемов данных. Из-за увеличения численности населения на земном шаре, мировой объем данных увеличивается более чем в два раза каждые два года. Возникли большие трудности обработки данных и их хранения. Одной из наиболее важных проблем является обработка больших объемов данных, безопасность данных и быстрый анализ, а также соответственно определение их структуры, мониторинг темпов роста и увеличение скорости обработки. В статье приведен регрессионный анализ внешних факторов, влияющих на рост объема данных, численное решение регрессионной модели при обработке больших объемов данных. Найден коэффициент регрессии, свободный член и коэффициент детерминации данной задачи, используя метод наименьших квадратов. Определено направление полученных данных и построен график функции. Следовательно, наименьших квадратов является самым оптимальным и эффективным методом для анализа и построения регрессионной модели. Соответственно, эти методы являются надежным статистическим инструментом для тестирования и оценки линейных отношений.

Ключевые слова: Модель, регрессионный анализ, данные, информация, обработка, функция, переменные, характеристики связей.

УДК 004.9

O.Z. Sembiev, Zh.S. Kemelbekova, A.H. Mahatova

(M. Auezov South Kazakhstan State University,

Shymkent, The Republic of Kazakhstan.

E-mail: ordabai@mail.ru)

LOAD BALANCING IN THE CIRCUIT SWITCHING MODE ON THE NODES OF AN ASYNCHRONOUS NETWORK

Abstract. The problem of the quality of servicing the computer network is very important, since the Internet technology is going on in the country. Very different Internet technologies are developing in Kazakhstan. One of them is a broadband network with the integration of services based on asynchronous data transfer technology. Therefore, we investigated load distributions in the circuit switching mode for each node, in the absence of free nodes, the adjacent load distribution along the bypass paths.

Key words: asynchronous transfer mode, channel switching, integrated group paths, traffic, virtual connections.

CONTENTS

<i>Omirsirikov M.Sh., Isayeva L.D., Asubayeva S.K., Kembayev M.K., Togizov K.S.</i>	
THE ROLE OF THREE-DIMENSIONAL MODELS OF DEPOSIT AND THERMODYNAMIC CONDITIONS OF ITS FORMATION AT SELECTING AND EVALUATING RESOURCES OF PERSPECTIVE SITES.....	3
<i>Zhassymbekov B., Kembayev M.</i>	
GEOLOGICAL AND STRUCTURAL FEATURES OF THE LOCATION AND CONDITIONS FOR THE FORMATION OF URANIUM ORE AT THE TORTKUDUK DEPOSIT	10
<i>Janaleyeva K.M., Mazhitova G.Z., Kabiev Y.S.</i>	
METHODOLOGICAL APPROACHES AND PRINCIPLES FOR IDENTIFICATION AND MAPPING OF AGROLANDSCAPES OF THE REPUBLIC OF KAZAKHSTAN.....	14
<i>Ozgeldinova Zh., Ussalinov Y., Mukayev Zh.</i>	
LANDSCAPE DIVERSITY EVALUATION OF THE TERRITORY OF PAVLODAR REGION.....	22

Technical sciences

<i>Abdeli D.Zh., Yskak A.S., Bae W., Novriansyah A.</i>	
MODELING OF WATER SHUT-OFF TECHNOLOGY IN THE BOTTOM HOLE OF OIL WELLS USING LIQUID GLASS.....	29
<i>Zhanikulov N., Dzhanmuldaeva Zh., Taimasov B., Ayazbekov M., Masatbayev A.</i>	
RESEARCHES FOR A WALL CERAMIC BRICK OUTPUT USING COAL PRODUCTION WASTES OF LINGER MINES.....	35
<i>Bekyashev A. 3, Kazagachev V. N.</i>	
LABORATORY RESEARCHES USING ARDUINO FOR THE STUDENTS IN INNOVATION TECHNOLOGY PARK.....	40
<i>Sagyntayeva A.K., Adyrbaykyzy R.</i>	
THE CULTIVATION TECHNOLOGY OF OYSTER MUSHROOM.....	44
<i>Kelamanov B.S., Zhumagaliev E.U., Samuratov E.K., Akuov A.M., Abilberikova A.A.</i>	
THERMODYNAMIC-DIAGRAM ANALYSIS OF MN-FE-SI-C SYSTEM.....	47
<i>Sakibayeva S.A., Suigenbayeva A.Zh., Abdurazova P.A., Serikov A. S., Konysbek Zh. K.</i>	
STUDY OF THE EFFECT OF MODIFIED ZEOLITE FILLER ON THE RUBBER PROPERTIES.....	53
<i>Sekerbayev K. S., Taurbayev Ye. T., Chirvony V. S., Martinez-Pastor J. P., Timoshenko V. Yu., Taurbayev T. I.</i>	
MODELING OF PHOTOLUMINESCENCE TRANSIENTS IN THIN FILMS OF METAL-HALIDE PEROVSKITES: A ROLE OF SURFACE NONRADIATIVE RECOMBINATION.....	58
<i>Baikenzheyeva A.S.</i>	
ANALYSIS AND SAFETY PROTECTION OF EMPLOYEES BY INDIVIDUAL MEANS OF PROTECTION.....	64
<i>Satpayev D.A., Poltavtseva V.P., Degtyaryova A.S., Partyka J.</i>	
Ti-Cr-N COATINGS STRUCTURE AND HARDNESS ALTERATION UNDER THE IMPACT OF XENON IONS AND ANNEALING.....	69
<i>Rakhimbekov S.M.</i>	
INNOVATIVE COURSE OF MINING TECHNOLOGY DEVELOPMENT.....	73
<i>Muradov A.D., Kyrykbayeva A.A., Taipova B.G.</i>	
EFFECTS OF ELECTRON IRRADIATION AMOUNT AND TEMPERATURE ON THE DEFORMATION OF POLYIMIDE FILMS.....	79
<i>Omarova A.B., Tulemisova Zh.K., Bekenova M.T., Kurbanova G.V., Zetbek G.S.</i>	
STUDY OF LACTIC ACID MICROORGANISMS SEPARATED FROM FERMENTED MILK PRODUCTS OF KAZAKHSTAN	84
<i>Bakhtaev Sh.A., Tergeusizova A.S., Toigozhinova A.Zh.</i>	
METHODS DEVELOPMENT FOR LINEAR PARAMETERS MEASURING OF MOVING NONCONDUCTING STRINGS.....	90
<i>Zaurbek A., Dzhuruntayev D. Z.</i>	
FUNCTIONAL MODELING OF THE LINEAR FEEDBACK SHIFT REGISTER ON THE FPLD IN A CADSYSTEM QUARTUS II USING THE VERILOG LANGUAGE.....	97
<i>Nurzhanov Ch., Naizabayeva L.</i>	
A MODEL OF THE GROWTH KINETICS OF A NEW PHYTO-REMEDY PLANT MISCANTHUS X GIGANTEUS ON THE HEAVY METALS CONTAMINATED SOIL.....	105
<i>Jomartova Sh., Ziyatbekova G., Issimov N., Mazakova A., Bolatkhanova Zh., Sametova A.</i>	
BIOMETRIC BIOMETRIC IDENTIFICATION OF THE MAN FACE ON THE BASIS OF MULTICRITERIA OPTIMIZATION.....	110

<i>Duissenov H.J., Koshinbaeva M.J., Kozhabekova P.A., Tastemir D.B., Makhanova Z.A.</i>	
METHODS OF SOFTWARE DEVELOPMENT USING CRYPTOGRAPHIC CONVERTING ALGORITHMS.....	115
<i>Bektursunova A.K., Botabayev N.E., Aripbayeva A.E.</i>	
OBTAINING TEXTILE MATERIALS WITH ANTIBACTERIAL PROPERTIES.....	121
<i>Ermukhanova N.B., Tanzharykov P.A., Kerimbekova Z.M.</i>	
ASSESSMENT OF THE TECHNOGENIC IMPACT OF RADIONUCLIDES ON THE SOIL OF OIL AND GAS DEPOSITS.....	125
<i>Darkenbayev D.K.</i>	
NUMERICAL SOLUTION OF THE REGRESSION MODEL FOR ANALYSIS AND PROCESSING OF BIG DATA.....	132
<i>Sembiyev O.Z., Kemelbekova Zh.S., Mahatova A.H.</i>	
LOAD BALANCING IN THE CIRCUIT SWITCHING MODE ON THE NODES OF AN ASYNCHRONOUS NETWORK.....	139
<i>Akasheva Zh.K., Assilbekov B.K., Kudaikulov A.A.</i>	
STIMDESIGN SOFTWARE MODULE FOR AUTOMATIC SELECTION OF A SUITABLE METHOD FOR OIL FLOW STIMULATION TO THE WELL.....	146
<i>Boranbayev A.S., Boranbayev S.N., Nurusheva A.M., Yersakhanov K.B.</i>	
APPROACHES, METHODS AND TOOLS TO ENSURE THE RELIABILITY OF THE INFORMATION SYSTEM FUNCTIONING.....	153
<i>Boranbayev S.N., Boranbayev A.S., Nurusheva A.M., Slabkevich V.N.</i>	
ANALYSIS OF RISK METHODS IN INFORMATION SYSTEMS.....	159
<i>Syrmanova K., Kovaleva A., Hinrichsen K.-O., Kaldybekova Zh., Beloborodov B.</i>	
DETERMINATION OF ULTRASOUND INFLUENCE ON UNDESIRABLE COMPONENTS OF USED MOTOR OIL.....	165
<i>Syrmanova K., Kovaleva A., Hinrichsen K.-O., Sakibayeva S., Beloborodov B.</i>	
ADSORPTION TREATMENT OF USED MOTOR OIL BY VERMICULITE SORBENT.....	171
<i>Omarbekova A. S., Zakirova A. B., Nurgazinova G.</i>	
SET QUESTIONS AND ANSWERS WITHIN THE SCOPE OF ALGEBRA KNOWLEDGE BASE.....	177
<i>Yelemesov K., Bortebaev S., Sydykbekova S., Utyanov A.</i>	
COMPOSITE APPLICATIONS APPLICATION TO PRODUCE PUMP CASES	181
<i>Genbach A.A., Bondartsev D.Yu.</i>	
MODELING OF THERMAL STRESSES THAT DEplete POROUS COATINGS OF HEAT EXCHANGE SURFACES BOILER-TURBINE EQUIPMENT.....	185
<i>Abdrakhmanov R., Saparkhojayev N., Rustamov E., Mukasheva A.</i>	
ALGORITHM BASES OF FRACTAL KNOWLEDGE BASES DESIGNING IN INTELLECTUAL SYSTEMS.....	192
<i>Saparkhojayev N., Mukasheva A.</i>	
INFORMATION SYSTEM DEVELOPMENT OF INFORMATION RESOURCES DESIGN AND USE FOR EVALUATION OF PARAMETERS AND RECOMMENDATIONS BASED ON BIG DATA TECHNOLOGY TOOLS: OPERATING MONGODB.....	198
<i>Nadirov K. S., Otarbayev N.Sh., Sadyrbaeva A.S., Turebekova A.M., Aikenova S.Zh., Baibotayeva S.E.</i>	
IMPROVING THE DEMULSIFIER REAGENTS EFFICIENCY TO PARAFFIN OIL DEHYDRATION.....	204
<i>Golubev V.G., Sadyrbayeva A.S., Turebekova A.M., Baibotayeva S.E.</i>	
SOME REGULARITIES OF THE PROCESS STUDY OF OIL DISPLACEMENT FROM THE OIL RESERVOIR MODEL BY THE METHOD OF ALKALINE EXPOSURE.....	209
<i>Golubev V.G., Sadyrbaeva A.S., Besbaeva N.A., Kasimova Zh.Zh. Aikenova S.Zh.</i>	
THE INFLUENCE OF AQUEOUS ALKALI AND POLYACRYLAMIDE ON THE WETTING ANGLE OF THE OIL CORE SURFACE	213
<i>Artykbayev D.Zh</i>	
THE CHANGE IN THE SOIL STRENGTH PARAMETERS IN THE SLOPE BODY UNDER DYNAMIC INFLUENCES.....	217
<i>Tusupova B.Kh., Kulbayev T.T., Sarman M.A.</i>	
AN OVERVIEW OF CONSTRUCTION AND DEMOLITION WASTE MANAGEMENT ISSUES IN KAZAKHSTAN: RECYCLING OF DEMOLITION WASTES AS A TRANSITION TO GREEN CONSTRUCTION.....	221
<i>Batayev N.A., Shuvatov T.T., Kuzyrgaliyev A.R.</i>	
MODELING OF GAS TURBINE COMPRESSOR UNIT OPERATING MODES.....	227

<i>Skakov M.K., Mukhamedova N.M., Wieleba W.</i>	
MECHANICAL PROPERTIES OF SILICON CARBIDE CERAMICS	235
<i>Smakanov B., Uvaliyeva I.</i>	
DIAGNOSTICS OF MECHANICAL VEHICLE DEFECTS BASED ON A PROBABILITY LOGIC METHOD.....	240
<i>Manap K.R., Nurmat M., Serikkyzy M.S.</i>	
QUALITY ASSESSMENT OF RED DRY WINE IN THE REPUBLIC OF KAZAKHSTAN.....	248
<i>Pazylova D.T., Shevko V.M., Tleuov A.S., Tleuova S.T.</i>	
THERMODYNAMIC MODEL OF LEAD, ZINC, COPPER COMPOUNDS INTERACTION WITH SODIUM AND CALCIUM CHLORIDES CONTAINED IN A STILL WASTE LIQUID.....	253
<i>Kaldybayev R.T., Kaldybaeva G.Yu., Temirshikov K.M., Mahmudova M.A.</i>	
INVESTIGATION OF THE POSSIBILITY OF STABILIZING THE DECOMPOSITION OF HYDROGEN PEROXIDE DURING THE BLEACHING OF COTTON CELLULOSE.....	261
<i>Aripbayeva A.E., Kaldybayev R.T., Stepanov S.G., Kaldybayeva G.Yu., Mirzamuratova R.Sh.</i>	
ANALYSIS OF TECHNICAL REQUIREMENTS FOR THE DESIGN OF FIRE PRESSURE HANDS.....	266
<i>Kenzhegulov A.K., Mamayeva A.A., Panichkin A.V.</i>	
INVESTIGATION OF CALCIUM-PHOSPHATE COATINGS OBTAINED BY HF MAGNETRON SPUTTERING.....	271
<i>Kaldybayev R.T., Kaldybayeva G.Yu., Tolesh Z.</i>	
INVESTIGATION OF QUALITY INDICATORS BY BICOMPONENT AND COTTON STRETCH NEEDS.....	276
<i>Zhumanbayeva N., Balbekova B.</i>	
WELDED JOINTS FEATURES AND STEEL PROPERTIES OF MARTENSITIC AND PEARLITIC CLASSES.....	280
<i>Omarova A.B., Tulemisova Zh.K., Kaumenova A.E., Suleimenova Zh.M.</i>	
DEVELOPMENT OF REGULATORY TECHNICAL DOCUMENTS FOR THE PRODUCTION OF PROBIOTIC PRODUCTS.....	284
<i>Nugman E.Z., Abekova A.Zh., Mustafa A.K., Tulebaev N.S., Bekbosynova B.A</i>	
SIZE DETECTION OF SPINDLE CONNECTIONS OF NEW RADIAL SHEAR SET.....	288
<i>Baikenzheeva A.S.</i>	
APPROPRIATE STATE AND WAYS TO INCREASE SERVICE EFFICIENCY AND SAFETY OF THE RAILWAY TRANSPORT ROAD WAY OF THE REPUBLIC OF KAZAKHSTAN.....	295
<i>Bolatbek M.A., Mussiraliyeva Sh.Zh.</i>	
DETECTION OF EXTREMIST TEXTS USING MACHINE LEARNING METHODS.....	299
<i>Azanov N.P., Manas Zh.B.</i>	
THE SECURITY PROBLEMS AND CRYPTOGRAPHIC STRENGTH OF ELLIPTIC CURVES.....	304
<i>Kambatyrov M.B., Nazarbek U.B.</i>	
MARKET ANALYSIS OF SUPPLEMENT FEED IN THE REPUBLIC OF KAZAKHSTAN.....	309
<i>Burumbayeva G. R., Dyussova R. M., Seitenova G. Zh., Ivashkina E. N.</i>	
UNIT RENOVATION OF THE CATALYTIC CRACKING PROCESS.....	314
<i>Dyussova R. M., Burumbayeva G. R., Seitenova G. Zh., Ivanchina E. D.</i>	
THE EFFICIENCY DETERMINATION OF VARIOUS ISOMERIZATION PROCESS TECHNOLOGIES BY METHOD OF MATHEMATICAL MODELING.....	320
<i>Mostovoi A.S., Nurtazina A.S., Kadykova Yu.A.</i>	
FIRE-PROOF EPOXY COMPOSITES WITH HIGHER COMPLEX OF OPERATIONAL PROPERTIES..	327
<i>Kassekeyeva A.B.</i>	
DIFFICULTIES TO CONSTRUCT IR THESAURUS.....	333
<i>Savinkin V. V., Ratushnaya T. Yu., Kyznetsova V. N.</i>	
RESEARCH OF MECHANICAL PARAMETERS OF TURBINE BLADES OF CHP RESTORED BY PLASMA ENERGY SOURCE.....	337
<i>Skakov M.K., Vityuk G.A., Vityuk V.A., Kotov V.M., Zhanbolatov O.B.</i>	
IMPLEMENTATION OF SET ENERGY RELEASE PROFILE IN THE HETEROGENEOUS FUEL ELEMENT UNDER THE TESTS IN THE IMPULSE GRAPHITE REACTOR.....	345
<i>Skakov M.K., Irkimbekov R.A., Vurim A.D., Vityuk G.A.</i>	
APPLICATION OF SPATIAL KINETICS MODEL AT IGR REACTOR EXPERIMENTS.....	357
<i>Abduraimova K., Zakirova A.B., Omarbekova A.S.</i>	
DEVELOPING A MODEL OF SECURE DATA MIGRATION TO THE CLOUD ON THE BASE OF A COMBINATION OF CRYPTOGRAPHIC ALGORITHMS.....	351
<i>Uskenbayeva R., Satybaldiyeva R., Bektemysova G., Moldagulova A., Aitim A.</i>	
PECULIARITIES OF MODELING BUSINESS PROCESSES IN THE UNIVERSITY USING THE EXAMPLE OF IITU.....	364

Physical and mathematical sciences

<i>Yerbay M.K., Muradov A.D.</i> EFFECT OF ELECTRON IRRADIATION AND FILLER POLYETHYLENE TEREPHTHALATE ON THE PHYSICAL AND CHEMICAL PROPERTIES OF POLYIMIDE.....	371
<i>Muradov A., Berzhanova Zh.M.</i> THE EFFECT OF GAMMA IRRADIATION FILLER CONCENTRATION AND DOSE ON THE MECHANICAL PROPERTIES OF THE «POLYIMIDE – GLASSY» SYSTEM.....	375
<i>Shaikhova G.N., Tuleugaliyeva G.G., Syzdykova A.M.</i> CONSERVATION LAWS FOR NONLINEAR INTEGRABLE EQUATIONS OF HIROTA TYPE.....	380
<i>Zhumabekov M.E., Mukhtarova M.N., Nurgaliyeva K.E., Maulyenova N.U., Danlybayeva A.K.</i> THE EFFECTIVE USE OF INDIVIDUAL SOLAR POWER STATIONS.....	387
<i>Korunbayeva L.A., Shokobaeva G.T., Myrzakhmetova U.A.</i> KINETICS OF OXIDATION OF PIPE STEEL 17G1S AT VARIOUS TEMPERATURES.....	392
<i>Ajan B., Inkarbekov M.</i> MODAL DISCONTINUOUS GALERKIN METHOD FOR LARGE EDDY SIMULATION OF TURBULENT FLOWS.....	399
<i>Orazbayev B.B., Kamenov N.M., Zhenisbekuly Zh., Kassenova L.G.</i> DEVELOPMENT OF MATHEMATICAL MODELS OF THE MAIN OIL PIPELINE ASSEMBLIES IN A FUZZY ENVIRONMENT.....	405
<i>Nurgabyl D.N., Kauharova R.Sh.</i> GENERAL BOUNDARY VALUE PROBLEM FOR A SYSTEM OF DIFFERENTIAL EQUATIONS WITH SMALL PARAMETERS FOR DERIVATIVES.....	412
<i>Nurgabyl D.N., Zhurimbay G.A.</i> MAKING RESULT OF A DISTURBED BOUNDARY VALUE PROBLEM WITH UNBOUNDED BOUNDARY CONDITIONS.....	416
<i>Sarmanbetov S.A., Sagatbayeva A., Tukan A.</i> LINEAR CLASSIFICATOR OF NEURAL NETWORKS ON THE BASIS OF PROBABILITY THEORIES.....	420
<i>Uskenbayeva R.K., Kuandykov A.A., Beisembekova R.N., Karimzhan N.B., Kozhamzharova D.Kh.</i> DEFINITIONS AND SPECIFICS SOLUTION OF ONE CLASS TASKS OF CONSTRUCTING OBJECT BY ROBOT UNIT.....	425
<i>Mustafin M.A., Makashov N.K.</i> SEVERAL TASKS OF COMPUTATIONAL GEOMETRY.....	432
<i>Tukenova L. M.</i> APPROACH DECISION FOR ONE LOCAL PROBLEM OF THE OCEAN.....	432

Chemical and metallurgical sciences

<i>Besterekov U., Ermekov S.R.</i> CHANGES IN THE CHEMICAL COMPOSITION OF WATER-SALT SYSTEMS. THE INITIAL FIGURATIVE POINTS IN THE UNSATURATED SOLUTION AREA	441
<i>Ainabekov I., Burkitbayeva B., Avchukir Kh.</i> BENZYL CHLORIDE MONOACETYL DIACRYL TRI-ETANOL AMINE CORROSION INHIBITOR OF STAINLESS STEEL IN PHOSPHORIC MEDIUM.....	448
<i>Kenzhaliev B.K., Berkinbayeva A.N., Surkova T.Yu., Dosymbayeva Z.D., Chukmanova M.T.</i> STUDY OF URANIUM SORPTION BY NATURAL SORBENT.....	454
<i>Asylhanov J. S., Markina D. B., Kishibaev K. K., Nechipurenko S. V., Atshabarov A. A., Tokaev R. R., Efremov S. A., Kalugin S. N.</i> OBTAINING AND RESEARCH OF THE GROWTH STIMULATOR ON THE BASIS OF 2,2- DIMETHYL-4- (5'-AMINO-1'2'4'-TRIAZOLE-3'-THIOPROPYNYL) - TETRAHYDROPYRAN -4- OLEFINS.....	460
<i>Kambarova G.A., Abdurazova P.A., Makulbaeva A.M., Raiymbekov Y.B.</i> INVESTIGATION OF THE PURIFICATION PROCESS BY SUCCESSIVE RECRYSTALLIZATION OF THE AMMONIUM DIHYDROGEN PHOSPHATE.....	467
<i>Avchukir Kh., Burkitbayeva B.D.</i> INDIUM ELECTRODEPOSITION KINETICS ON A TITANIUM ELECTRODE FROM CHLORIDE SOLUTIONS.....	469

Besterekov U., Yermekov S.R.

CHANGES IN THE CHEMICAL COMPOSITION OF WATER-SALT SYSTEMS IN THE SATURATED SOLUTION RANGE.....	476
---	-----

Military sciences

Adautov A.U., Martikyan A.S., Muhamedjanov B.S.

THE MAIN CATEGORIES OF BORDER SCIENCE AND SET UP ITS CONCEPTUAL FRAMEWORK.....	482
---	-----

Salii S.M.

BORDER POLICY OF THE REPUBLIC OF KAZAKHSTAN AT THE PRESENT STAGE AND WAYS TO IMPROVE IT.....	486
---	-----

Rysbayeva G.P.

THE ANALYSIS AND CALCULATION OF CONTOUR OF ELASTIC STRESSES CONCENTRATION TAKING INTO ACCOUNT THE TECTONIC FORCES OF MILITARY UNDERGROUND CONSTRUCTION	491
--	-----

Редакторы:

Н.Ф. Федосенко

Верстка на компьютере:

Л. Касжанова

Подписано в печать 25.11.2018 г.

Формат 60x84 ¹/₈. Усл. п.л 32,0.

Тираж 500 экз. Заказ № 219.

Адрес редакции:

ул. Сатпаева, 22, КазННТУ каб. 616, тел. 292-63-46 ,Nina.Fedorovna. 52 @ mail.ru

Департамент маркетинга и коммуникаций КазННТУ

Казахского национального исследовательского технического университета имени К.И. Сатпаева