UDC 615.1:004.94
DOI: 10.15587/1729-4061.2022.253547

SETTING THE EQUATION OF REGRESSION TO DETERMINE THE TECHNOLOGICAL FACTORS INFLUENCE ON THE CONTENT OF FLAVONOIDS IN THE EXTRACT

Olga Kutova, Rita Sahaidak-Nikitiuk, Inna Kovalevska, Nataliya Demchenko

The aim *of the article is to establish a regression equation that determines the influence of technological factors on the content of flavonoids in the extract for further use in optimizing the technology for obtaining an extract based on horse chestnut. The task of the research was to identify a mathematical model to describe the effect of technological parameters of extraction on the target quality indicators of a given dosage form, in particular, on the quantitative content of flavonoids.*

Materials and methods. *The proposed approach is based on mathematical processing of experimental results obtained according to plan 2^3 using the computer program Mathcad 14 and MS Excel. To establish a mathematical description, an analysis of the separate influence of technological factors on the target indicator was carried out and the possibility of forming a geometric mean function was determined using the corresponding linear regression equations.*

Results*. The general problem of identification was solved, when it was necessary to reveal both the mechanism of influence of technological factors on the value of the target indicator, and to give a quantitative assessment of the unknown parameters of the regression equation. Based on the results of experimental observations, an adequate mathematical model was established in the form of a linear multiple regression equation with the interaction of factors.*

Conclusions. *The obtained mathematical description makes it possible to analyze the influence of technological factors on the quantitative content of the complex of flavonoids in the herbal extract in the range of the investigated factor space, and also to optimize the technological parameters of extraction*

Keywords: *experimental data processing, quantitative factors, multiple regression, identification, geometric mean function*

How to cite:
Kutova, O., Sahaidak-Nikitiuk, R., Kovalevska, I., Demchenko, N. (2022). Setting the equation of regression to determine the technological factors influence on the content of flavonoids in the extract. ScienceRise: Pharmaceutical Science, 1 (35), 52–57. doi: http://doi.org/10.15587/1729-4061.2022.253547

© The Author(s) 2022
This is an open access article under the Creative Commons CC BY license

1. Introduction

The preparation of extracts from plant materials is widely used in the pharmaceutical industry. This process has a number of specific features that complicate the regulation and optimization of its technological parameters. To analyze the results of scientific experiments and solve optimization problems mathematical modelling is widely used, which includes obtaining regression equations [1, 2]. The existing methods for constructing mathematical models allow the use of various modifications of the planned experiment [3], which cannot always be effectively applied for traditional mathematical processing. This is especially true for multifactorial models that include more than two factors and require a large number of experimental observations [4]. Also, in order to establish equations that describe the change in any pharmaco-technological indicator, sometimes one has to deal with the values of the studied parameters unevenly located in the factor space. For this reason, the usual methods of establishing dependencies are unsuitable, and there is a need to develop such methods that would be effective in these conditions [5].

The aim of the article is to obtaining a regression equation that determines the influence of technological factors on the content of flavonoids in the extract for further use in optimizing a technology for obtaining an extract based on horse chestnut.

2. Planning (methodology) of research

When substantiating the technological parameters of the extraction of flavonoids from horse chestnut [6, 7], the method of mathematical planning of a multifactorial experiment was used. As a response function, the yield of flavonoids (y1) was chosen, expressed as a percentage relative to their amount in the feedstock.

The main method of this study is regression analysis. This method is quite effective from the point of view of mathematical statistics and allows you to present the available information in a convenient form.

The choice of the experimental design was determined by the goal set – to describe the nature of the influence of the technological parameters of the extraction on the specific pharmacopoeial characteristics of the extract. Static planning of the experiment of type 2^3 made it possible to

simultaneously vary three quantitative factors and evaluate their linear effects and the effects of their interactions on the content of flavonoids in the extract. To eliminate the nonlinear effects of the influence of technological factors, which requires an increase in the number of experiments, it was proposed to use the geometric mean function. The proposed methodological approach in this study has shown sufficient efficiency in quantitative assessment of the main effects.

3. Materials and methods

To obtain the extract, dried raw materials were used - horse chestnut fruits. As a solvent – a water-ethanol mixture, which allows the extraction of water-soluble phenolic pigments. The prepared raw material was extracted in a laboratory batch extractor. The extraction efficiency was controlled by changing the content of the total flavonoids in the raw material.

The concentration of the extractant (*x1, g/cm*^{3}), the extraction time (*x2, s*) and the ratio of the raw material to the extractant (*x3*), according to the experimental conditions, considering a priori data and practical experience of the researchers, must satisfy the corresponding ranges of values: $40 \le x1 \le 70$; $24 \le x2 \le 48$; $0{,}1 \le x3 \le 0{,}2$.

The considered quantitative factors in this experiment are independent and combinations of factor levels in the study are not tied to each other. It should also be noted that different units of measurement matter. At the same time, the parameters *x1, x2, x3* must ensure the content of flavonoids (*y1=f(x1,x2,x3)*) not less than 0,11 %, since according to experimental data, a further increase in the yield of flavonoids does not occur.

An experimental design containing three factors at two levels: maximum and minimum. This allows you to analyze the minimum possible number of observations to establish a mathematical description in the form of a linear relationship, both without the interaction of factors, and taking it into account. Tab. 1 shows the technological parameters of the extraction process and the content of flavonoids, expressed as a percentage relative to their amount in the feedstock (y1, %).

The processing of the experimental results was carried out using the Mathcad 14 tools (regression analysis [8, 9], graphing [10]) and MS Excel (correlation analysis) [11, 12].

Table 1

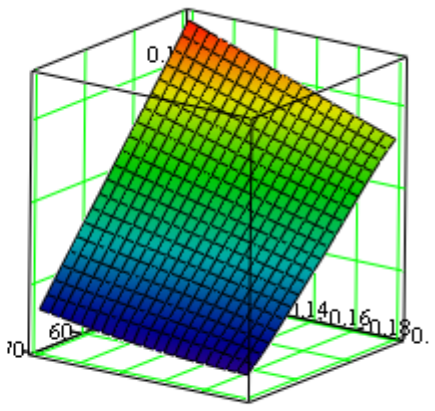
Matrix of complete factorial experiment 2^3

No.	*x1*	*x2*	*x3*	*Y1*
1	70	48	0.2	0.049
2	40	48	0.2	0.092
3	70	24	0.2	0.049
4	40	24	0.2	0.092
5	70	48	0.1	0.054
6	70	24	0.1	0.054
7	40	48	0.1	0.11
8	40	24	0.1	0.11

4. Results

The presence of three independent variables in this experiment excludes the visualization of the experimental response surface for a preliminary assessment of the type of the mathematical model. In this case, graphic interpretation is possible only for two arbitrary variables, provided that the third variable is fixed at a certain level. However, the correlation analysis showed a complete lack of correlation between the content of flavonoids in the extract (y1) and an increase in the extraction time (x2) for more than 24 hours, as well as the absence of a statistical relationship between the concentration of ethanol (x1) and the ratio of the extractant to raw material (x3) in the given ranges of values. Therefore, it is possible to proceed to the mathematical processing of experimental data, considering only two factors - x1 and x3. To avoid the nonlinearity error within the factor space, a full-factor experiment of type 23 was additionally implemented based on the existing plan 3^2 [14, 15]. To establish the regression model, along with the analytical method of establishing, the graphic method was also used (Fig. 1).

$$y1 := \begin{pmatrix} 0.049 \\ 0.092 \\ 0.076 \\ 0.054 \\ 0.11 \\ 0.082 \\ 0.07 \\ 0.1 \\ 0.05 \end{pmatrix} \quad x1 := \begin{pmatrix} 70 \\ 40 \\ 55 \\ 70 \\ 40 \\ 55 \\ 55 \\ 40 \\ 70 \end{pmatrix} \quad x3 := \begin{pmatrix} 0.2 \\ 0.2 \\ 0.15 \\ 0.1 \\ 0.1 \\ 0.1 \\ 0.2 \\ 0.15 \\ 0.15 \end{pmatrix}$$



(x1, x3, y1)

Fig. 1. Graphic interpretation of the experimental data

According to the experimental data, we can assume a mathematical model of a linear form with the interaction of two factors:

$$y1(x1, x3) = a_0 + a_1x1 + a_2x3 + a_3x1x3. \quad (1)$$

To determine the coefficients of equation (1) by the method of least squares (LSM), the Mathcad 14 tools were used. The calculation results for the obtained model:

$$y1(x1, x3) = 0.22 - 0.0023x1 - \\ -0.353x3 + 0.004333x1x3 \quad (2)$$

confirm its adequacy. The average relative error in determining the content of flavonoids in the experimental points is 0.33 %, and in 7 out of 9 calculations it is absent. The statistical significance of the equation and its coefficients was evaluated using MS Excel.

The regression equation is characterized by a high coefficient of determination (R^2=0.999), its adequacy was verified using the Fisher criterion (significance F=1.68·10^{-8}). The coefficients, the absolute value of which is greater than their confidence intervals, are recognized as statistically significant. Intervals estimate for coefficients with 95 % reliability:
–0.00249≤x1≤–0.00212, (P-value – 0.0000005),
–0.42≤x3≤–0.29, (P-value – 0.0000344),
0.00317≤x1x3≤0.0055, (P-value is 0.00021).

Interval estimates for mean values at base points:
0.0475≤y1(70;0.2)≤0.0506, 0.0906≤y1(40;0.2)≤0.0935, 0.0757≤y1(55 ;0.15)≤0.077, 0.0525≤y1(70;0.1)≤0.0556, 0.1086≤y1(40;0.1)≤0.1115, 0.0811≤y1(55; 0.1)≤0.083, 0.0696≤y1(40;0.15)≤0.0715 [16, 17].

Analysis of the plots of residuals as functions of variables indicates the absence of a systematic error in the resulting regression equation (Fig. 2).

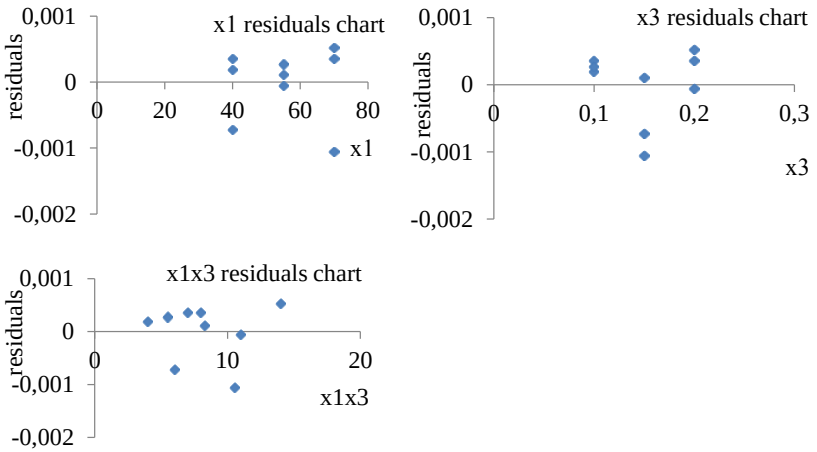


Fig. 2. Plot of residuals as functions of variables

This relationship is empirical, it can be used for extraction times ranging from 24 to 48 hours. Since equation (2) contains only two variables – factors *x1* and *x3*, its graphical interpretation is possible, the analysis of which gives a complete coincidence of the experimental and theoretical response surfaces (Fig. 3).

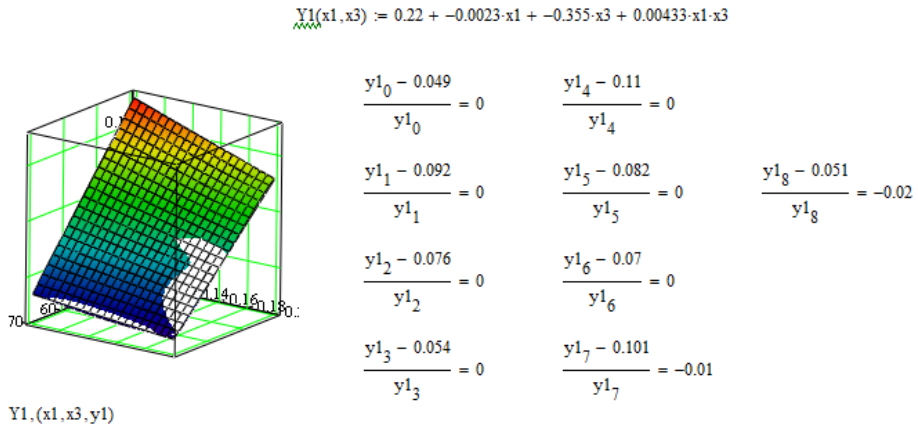
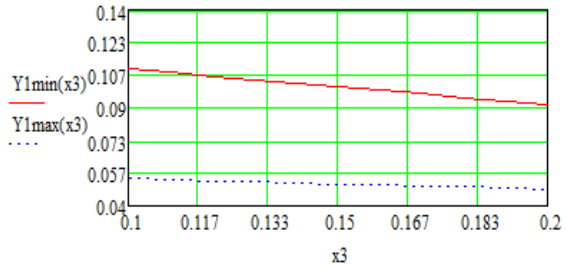


Fig. 3. Comparison of the graphical interpretation of the experimental and theoretical response surfaces and determination of the relative calculation error using equation (2)

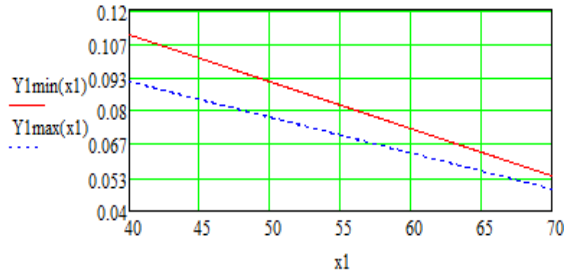
An analysis of equation (2), based on the values of the coefficients and signs in front of them, showed that the concentration of ethanol (*x*1) has the greatest influence on the content of flavonoids in the extract. As the concentration increases from 40 to 70 %, the value of the target decreases. The effect of the x1 factor is an order of magnitude stronger than the *x3* factor, as well as the interaction of the *x*1*x*3 factors. With an increase in the

ratio of extractant to raw material from 0.1 to 0.2, the content of flavonoids decreases. Depending on the level, for example, of the x1 factor, the *x3* factor will manifest itself differently.

The influence of the factors under study can be considered in more detail using the corresponding single-factor graphical dependencies, provided that the second factor is fixed at one of the levels (Fig. 4).



a



b

Fig. 4. Influence of individual factors on the content of flavonoids in the extract: *a* – ethanol concentration; *b* – ratio of extractant to raw material

As could be seen from Fig. 4, the most promising situation is when the x1 factor is fixed at the lower level. At the same time, the content of flavonoids does not decrease by more than 0.09 %. The extractant/raw material ratio equal to 0.1 and the ethanol concentration of 40 % will provide the maximum possible value y1(40;0.1)=0.11.

Using the obtained regression equation, it is possible to determine the values of the factors x1 and x3, which will provide not only a given number of flavonoids, but also their content in the acceptable range. This area will be between the lines of equal values that define the required boundaries of the area of the allowable amount of flavonoids in the extract, for example, 0,1≤y1≤0,11 (Fig. 5, region 1).

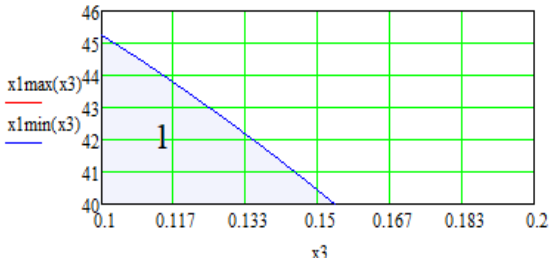


Fig. 5. Range of values of factors x1 and x2 to fulfil the condition 0.1≤y1≤0.11

The set of x1 and x2 values that are included in the generated area will provide the indicated conditions 0.1≤y1≤0.11, for example: y1(41;0.1)=0.108, y1(41;0.12)=0.103.

The authors propose another way to establish the regression equation. It was shown in [18] that in the case when the function of one variable near the boundaries of the multifactorial space does not change its character, it remains monotonically increasing (decreasing), it is possible to obtain a productive mathematical description [11] based on one-factor dependencies (Fig. 6).

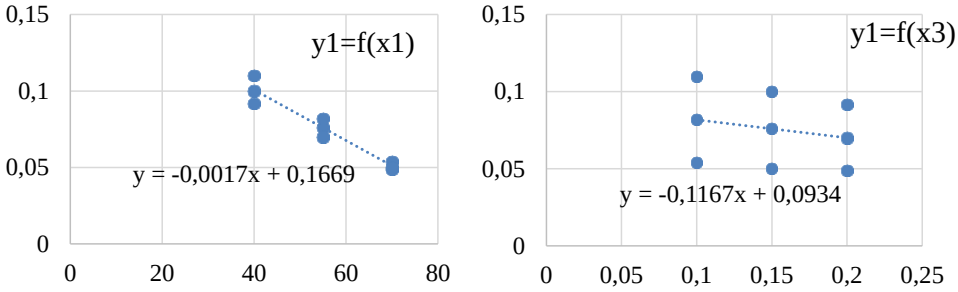


Fig. 6. One-factor graphic dependencies

The main effects of the factors in this study are not "masked" and correspond to the effects at the boundaries of the factor intervals. Let us form the geometric mean function using the obtained linear dependences y(x1), y(x3) (Fig. 6):

$$Y1(x1,x3)=\left((0.167-0.0017x1)(0.0934-0.1167x3)\right)^{1/2}. \quad (3)$$

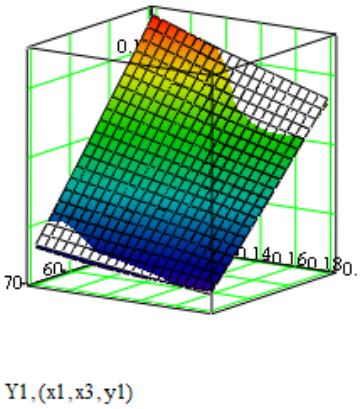
As a result of correcting the function with a polynomial of the 2nd degree LSM, correction coefficients of calculations are established. The required equation takes the form:

$$Y1(x1,x2,x3)=0.028-0.662\times$$
$$\times\left((0.167-0.0017x1)(0.0934-0.1167x3)\right)^{1/2}+$$
$$+17.384(0.167-0.0017x1)(0.0934-0.1167x3).$$

Calculations based on the obtained dependence (4) and its graphic interpretation are shown in Fig. 7. The effectiveness of the regression equation (4) is indicated by a relative error of 1.12 % at the base points and excellent agreement between the experimental and theoretical response surfaces.

The proposed method for establishing the regression equation could be used not only in the case of a full-factor experiment of the n^2 type.

The processing could include experimental points located in an arbitrary order convenient for the researcher.



$$\frac{y1_0 - 0.048}{y1_0} = 0.02 \qquad \frac{y1_4 - 0.109}{y1_4} = 9.091 \times 10^{-3}$$

$$\frac{y1_1 - 0.093}{y1_1} = -0.011 \qquad \frac{y1_5 - 0.081}{y1_5} = 0.012 \qquad \frac{y1_8 - 0.051}{y1_8} = -0.02$$

$$\frac{y1_2 - 0.076}{y1_2} = 0 \qquad \frac{y1_6 - 0.07}{y1_6} = 0$$

$$\frac{y1_3 - 0.055}{y1_3} = -0.019 \qquad \frac{y1_7 - 0.101}{y1_7} = -0.01$$

Fig. 7. Analysis of equation (4)

5. Discussion

The resulting regression equation shows that the completeness of the extraction of flavonoids by a water-ethanol mixture is significantly affected by the concentration of ethyl alcohol and the extractant/raw material ratio. The greater the mass of the extractant relative to the mass of the raw material, the less its concentration increases when the same amount of substance is extracted.

In this regard, the extraction rate will be faster, and the flavonoid yield will increase. At the same time, the mechanism of transfer of the extracted raw material is complicated by the swelling of the raw material, as a result of which it is possible to change the quantitative ratio of the phases during the process, as well as the physical properties of the raw material. The duration of extraction over 24 hours does not affect the yield of flavonoids.

The resulting regression equation could be used as one of the criterion equations for solving the problem of multi-criteria optimization in the course of developing a technology for obtaining an extract based on horse chestnut [19].

The regression equation was obtained in two ways, which showed similar results: by the classical method of regression analysis – LSM and using one-factor dependencies that reflect the main effects of factors on the target indicator. The advantage of the second method is that it could be used under the specified conditions without planning an experiment according to an orthogonal plan [20].

Study limitations. When developing the extract technology, it should be taken into account that the obtained regression equation could be used under certain conditions: the extraction time is at least 24 hours, the concentration of ethanol is from 40 to 70 %, the raw material/extractant mass ratio is from 0.1 to 0.2.

The proposed method for obtaining a two-factor regression equation using the geometric mean function has shown its effectiveness only in the case when the main effects of individual variables coincide with their effects on the boundaries of the factor intervals.

Prospects for further research. The subject of further research is the establishment of regression equations for such target quality indicators of the extract as the content of escin, the amount of dry residue, the content of ethyl alcohol, the density of the extract in order to optimize the technology of the extract based on horse chestnut.

6. Conclusions

A regression equation has been established that characterizes the influence of technological factors of extraction on one of the target quality indicators of an extract based on horse chestnut - the content of flavonoids. The technological parameters of extraction were optimized to ensure the maximum yield of extracted substances: ethanol concentration 40 %, extraction duration 24 hours at a ratio of solvent to raw material of 0.1. The obtained results will be used in the development of a technological scheme for the production of the extract.

Conflict of interest.

The authors declare that they have no conflicts of interest.

Financing

The study was performed without financial support.

References

1. Kenny, H. A., Hart, P. C., Kordylewicz, K., Lal, M., Shen, M., Kara, B. et.al. (2021). The Natural Product β-Escin Targets Cancer and Stromal Cells of the Tumor Microenvironment to Inhibit Ovarian Cancer Metastasis. Cancers, 13 (16), 3931. doi: http://doi.org/10.3390/cancers13163931

2. Orhurhu, V., Chu, R., Xie, K., Kamanyi, G. N., Salisu, B., Salisu-Orhurhu, M. et. al. (2021). Management of Lower Extremity Pain from Chronic Venous Insufficiency: A Comprehensive Review. Cardiology and Therapy, 10 (1), 111–140. doi: http://doi.org/10.1007/s40119-021-00213-x

3. Waltenberger, B., Mocan, A., Šmejkal, K., Heiss, E., Atanasov, A. (2016). Natural Products to Counteract the Epidemic of Cardiovascular and Metabolic Disorders. Molecules, 21 (6), 807. doi: https://doi.org/10.3390/molecules21060807

4. Smyth, R. M., Aflaifel, N., Bamigboye, A. A. (2015). Interventions for varicose veins and leg oedema in pregnancy. Cochrane Database of Systematic Reviews. doi: http://doi.org/10.1002/14651858.cd001066.pub3

5. Yamuna, M., Elakkiya, A. (2017). Mathematical Models in Drug Discovery, Development and Treatment of Various Diseases-A Case Study. Research Journal of Pharmacy and Technology, 10 (12), 4397–4401. doi: https://doi.org/10.5958/0974-360x.2017.00810.1

6. Fytopoulos, A., Ramachandran, R., Pardalos, P. M. (Eds.) (2022). Optimization of Pharmaceutical Processes. Springer International Publishing, VII, 435.

7. Rothlauf, F. (2011). Optimization Methods. Design of Modern Heuristics, 45–102. doi: http://doi.org/10.1007/978-3-540-72962-4_3

8. Stute, W., Thies, S., Zhu, L.-X. (1998). Model checks for regression: an innovation process approach. The Annals of Statistics, 26 (5), 1916–1934. doi: https://doi.org/10.1214/aos/1024691363

9. Hroshovyi, T. A., Martseniuk, V. P., Kucherenko, L. Y., Vronska, L. V., Hyrieieva, S. M. (2008). Matematychne planuvannia eksperymentu pry provedenyy naukovykh doslidzhen v farmatsii. Ternopil: Ukrmedknyha, 377.

10. Draguljic D., Dean A., Voss D. Design and Analysis of Experiments. 2-e ed. Springer International Publishing AG, 2017. 840 p. doi.org/10.1007/978-3-319-52250-0

11. Dean, A., Voss, D., Draguljić, D. (2017). Design and Analysis of Experiments. Springer Texts in Statistics, 840. doi: http://doi.org/10.1007/978-3-319-52250-0

12. Cherneva, K., Furrer, R., Tarigan, B. (2021). SampleSizeR: calculate sample sizes within completely randomized design. Available at: http://shiny.math.uzh.ch/git/reinhard.furrer/SampleSizeR/

13. Lawson, J. (2018). Design and Analysis of Experiments with R. CRC Press. Chapman and Hall/CRC, 620.

14. Lazic, S. E. (2016). Experimental Design for Laboratorist Biologists. 1-e ed. Cambridge University Press, 412.

15. Ferreira, A. P., Tobyn, M. (2014). Multivariate analysis in the pharmaceutical industry: enabling process understanding and improvement in the PAT and QbD era. Pharmaceutical Development and Technology, 20(5), 513–527. doi: https://doi.org/10.3109/10837450.2014.898656

16. Alexopoulos, E. C. (2010). Introduction to Multivariate Regression Analysis, 14 (Suppl 1), 23–28.

17. Zhuravskii, A. A., Zelenskii, A. I., Sinyaeva, A. V., Kutovaya, O. V. (2018). Novyi metod statisticheskoi obrabotki rezul'tatov nauchnyh issledovanii. Naukovo-praktychni zasady zaghaljno-inzhenernoji pidghotovky fakhivciv farmaciji: materialy I Mizhnar. nauk.-prakt. internet-konf. Kharkiv: NPhU, 96–101.

18. Ruban, E. A, Kutsenko, S. A., Kovalevskaya, I. V., Kutova, O. V. (2014). Optimizatsiya parametrov ekstragirovaniya lekarstvennogo rastitel'nogo syr'ya dlya polucheniya slozhnoy nastoyki «Venoton». Vestnik farmatsii. Belorussiya, 2, 37–41.

19. Brent, M. (2009). Essential Mathcad for engineering, science, and math ISE. 2nd ed. Amsterdam: Elsevier; Burlington, MA: Academic Press, 501.

20. Carter Hill, R., Genevieve, B. (2018). Using Excel for Principles of Econometrics, 5th Ed.

21. Kutovaya, O., Kovalevska, I., Sahaidak-Nikitiuk, R. (2021). Determination of regression equations based on the generalized separate influence of factors in pharmaceutical research. Norwegian Journal of development of the International Science, 57 (1), 39–43.

Received date 26.07.2021
Accepted date 15.02.2022
Published date 28.02.2022

Olga Kutova, PhD, Associate Professor, Department of Pharmaceutical Technology of Drugs, National University of Pharmacy, 53 Pushkinska str., Kharkiv, Ukraine, 61002

Rita Sahaidak-Nikitiuk, Doctor of Pharmaceutical Sciences, Professor, Department of Management and Economics of Enterprise, National University of Pharmacy, 53 Pushkinska str., Kharkiv, Ukraine, 61002

Inna Kovalevska, Doctor of Pharmacy Sciences, Associate Professor, Department of Industrial Technolodgy of Drugs, National University of Pharmacy, 53 Pushkinska str., Kharkiv, Ukraine, 61002

Nataliya Demchenko, PhD, Associate Professor, Department of Management and Administration, National University of Pharmacy, 53 Pushkinska str., Kharkiv, Ukraine, 61002

***Corresponding author:** Rita Sahaidak-Nikitiuk, e-mail: sagaidak_rita@ukr.net*