

Modeling the Quality of New Wheat Varieties in Kazakhstan

Dauletbakovbakytkan (Corresponding Author), Professor
Almaty Technological University, Doctor of Economic Sciences

E-mail: dauletbakovb@mail.ru

Dauletbakov Galym

Takeda Pharmaceuticals, Financial Director

E-mail: galym.dauletbakov@gmail.com

Received: October 30, 2019 Accepted: November 20, 2019 Published: November 2019

doi:10.5296/jfi.v3i1.15912 URL: <http://dx.doi.org/10.5296/jfi.v3i1.15912>

Abstract

The work investigates and demonstrates the analytical capabilities of statistical methods in the development of new varieties of wheat used for the production of pasta. By using examples of Kostanay, Pavlodar, Akmola, North-Kazakhstan region (NKR) and Aktobe regions, for identifying significant factors, there were constructed regression equations due to initial factors, on the main components, according to the panel data. Based on the analysis of various regression models, the strengths of the relationship between the number indicators of the grain quality there was found soft and hard wheat varieties (Mereke 70, Tselina 50, Ertis 97, Astana 2, Astana, Aktobe39, Baiterek), taking into account the peculiarities of grain production in this area.

Keywords: wheat flour, new wheat varieties, pasta properties, pasta products, quality

1. Introduction

Currently, it is important to study the nature of the relationship between the different indicators of grain quality, significantly affecting the consumer, nutritional value of the final product. Many scientific institutions develop mathematical models that show the quantitative dependence of the achieved quality parameters on the conditions of cultivation.

It is important to identify patterns of changes in interdependent indicators of grain quality. Selection of groups of indicators with stable close correlative connections allows us to

present an overall picture of the perspective dynamics of qualitative characteristics and creates a potential opportunity to influence them through comprehensive optimization of factors that significantly affect the physical and biochemical properties of the grain, at the lowest and only reasonable cost and with high efficiency.

The nature and closeness of the relationship between the different indicators of grain quality is important to know when developing new varieties of wheat used for pasta production. This will allow purposefully influence the formation of the whole complex of useful properties of grain and can serve as a basis for the prediction of new wheat varieties used for pasta production. It is known that between the individual indicators of grain quality there is a close correlation. However, despite the large number of works, the data on the relationship between the individual indicators of wheat grain quality are contradictory and do not cover the whole complex of indicators (Dauletbaev, Iztayev, Aryngazin, Omarov, & Syrlybayeva, 2007).

In recent years, Kazakhstan (Dauletbaev, Sultangaliyeva, & Kuanyshovna, 2016; Dauletbaev, Sultangaliyeva, & Dnimova, 2018; Kazakhstan: non-waste processing technology of grain crops; Modern aspects of scientific and technological support for the processing of agricultural raw materials and waste, 2015) and neighboring countries (Torzhinskaya et al., 1965; Serduykov, 1970; Kazakova, 1975; GOST R 52189-2003, 2004; Butkovskiy & Kasatov, 2003) have paid much attention to scientific research on the quality of flour, depending on its use in the food sector of the economy.

The differentiated approach is applied abroad to the quality of flour depending on its use (Landi, 1987; Tiefenbacher & Dobrovics, 2000; Roy Chung, 2004). For example, in the UK, the following indicators are used to assess the technological advantages of wheat grain: the fall number, as measured by the Hagberg-Perten method, the test for protein content, much attention is paid to such an indicator as wheat variety (GOST R 52189-2003, 2004).

Currently, in the Republic Kazakhstan, the main control of the baking properties of grain is carried out at the final stage of its processing, when considerable funds and efforts have been expended in obtaining the finished product. So we have only several opportunities to influence its quality. Therefore the transition from the quality control of the finished product to the preliminary control at the stage of its production is very actual and important today. Throughout the entire production cycle – *«grain indicator is - flour»*, that allows to considerably optimize the cost of research and improve the quality of the final product, as well as to regulate the properties of intermediate products in the right direction (Dauletbaev, Sultangaliyeva, & Kuanyshovna, 2016).

In the framework of such a technological cycle of the introduction of aggregated complex indicators, it has not been put anywhere so far. In this regard, the actual task is to simulate the quality of new wheat varieties used for the production of pasta.

This statement of the problem can be solved on the basis of factor analysis, which allows for known numerical values of the considered indicators to determine the number of hidden factors behind them, each of which receives a certain substantive characteristic due to its relations with the considered indicators. By distributing of grain quality indicators to factors

we can significantly reduce the number of variables to be considered in the grain problems.

The purpose of factor analysis as a branch of mathematical statistics is to develop models and methods that allow analyzing and interpreting arrays of experimental or observed data regardless of their physical nature (Roy Chung, 2004; Lawley & Maxwell, 1967; Kharman, 1972).

The array of the studied information is a rectangular matrix X_{ij} , strings which describe n objects observations ($i=1,2,\dots,n$), and columns describe the set of features that describe these observations ($j=1,2,3,\dots,m$).

There are many excellent books on factor analysis. More detailed description is given by the authors - Lawley and Maxwell (1967), Kharman (1972), Zhukovskaya and MuchnikI (1976).

2. Objects and Methods of Research

For experimental studies used samples of grain of soft wheat (Mereke 70, Tselina 50, Ertis 97, Astana 2, Astana, Aktobe39, Baiterek) of Kostanay, Pavlodar, Akmola, North-Kazakhstan region (NKR), Aktoberegion and determined the physical and biochemical properties of grain, characterizing their quality.

The moisture content of the grain was determined by the standard method GOST 13586.5-93 by drying milled grain weights in an electrical cabinet СЭИИ-3М (SESH-3M) at a temperature of 130°C for 40 minutes followed by cooling and determining the difference in mass resulting from dehydration. Humidity expressed as a percentage.

To determine the mass of 1000 grains, the sample after mixing the weed and grain impurities was mixed and distributed evenly in the form of a square, which was divided diagonally into four triangles and samples of 500 whole grains were counted out from every two opposite triangles). The mass of both samples was folded and received a mass of 1000 grains. The difference between the masses of the two samples should not exceed 5% of their average value (GOST 10820-89).

The quantity and quality of gluten grain was determined by GOST 13586.1-68. The amount of gluten was determined in a dough obtained from 25 g of milled grain and 14 ml of tap water, 20 minutes after kneading. The quality of gluten was determined in a 4-gram sample after 15 minutes of binning in water at a temperature of $+18^{\circ}\text{C}$. The elastic properties of gluten were determined on the device Gluten Deformation Meter – IDK-1. The results of measuring the elastic properties of gluten expressed in conventional units of the scale of the device, and depending on this gluten was attributed to the appropriate quality group.

The number of drops was determined on the Hagberg-Perten device and expressed in seconds. The method for determining the number of falls is an international method for determining α -amylase in wheat and rye grains, in flour. The method is based on the rapid gelatinization of a suspension of flour or coarse flour in a boiling water bath, followed by a change in the liquefying effect of α -amylase on starch. The values of the incidence number are inversely proportional to the amount of α -amylase in the sample (Ustimenko, 2007).

The hard grain measurements are made on UK infrared analyzer (PacificScientific 4250), pre-calibrated from the device SKCS 4100 (SingleKernelcharacteristicsystem) (Abugalieva, Srokbayev, Dracheva, & Savin, 1997; Abugalieva, Dracheva, Fursov, & Esimbekova, 1998).

The mass fraction of protein in grain was measured by the Kjeldahl method (Abugalieva, Srokbayev, Dracheva, & Savin, 1997), which is based on the burning of organic components of the sample products in the Kjeldal flask in the presence of sulfuric acid. The nitrogen released in this process was determined by titration and the protein content was calculated from its amount.

3. Results and Its Discussion

In order to develop new wheat varieties used for the production of pasta it is necessary to consider following factors: assessment of the potential properties of the grain; analysis of technological schemes, parameters and choice of the minimum number of parameters, in a complex form, taking into account influence of the physicochemical structural, mechanical and biological characters of the processed grain.

Based on analysis of the presented data we conclude that physical (nature, mass 1000 grains, glassiness, hard grain size) and biochemical (quantity and quality of gluten, protein, ash content, number of falling) indicators of wheat samples studied vary comprehensively. The data obtained for soft wheat varieties are presented in the Table 1.

The procedure of factor analysis is carried out using the software package “STATISTICA 10” using the principal component method (Borovikov, 2001).

4. Correlation Matrix

The matrix of pair correlations between indicators is calculated on the basis of the matrix of sequence numbers (Figure 1).

The matrix of correlation coefficients (Figure 1) is a quantitative expression of the degree of connection between each pair of the considered indicators. Analysis of the matrix of pair correlation coefficients shows that the yield of flour, % is most closely related to the variables of Nature, g/l, and Glassiness,%. The structure of relationships between groups of indicators is revealed by factor analysis.

The Calculation of correlation matrix is the first stage of factor analysis.

After clicking Ok you can go to the next dialog box. Define Method of Factor Extraction (Figure 2).

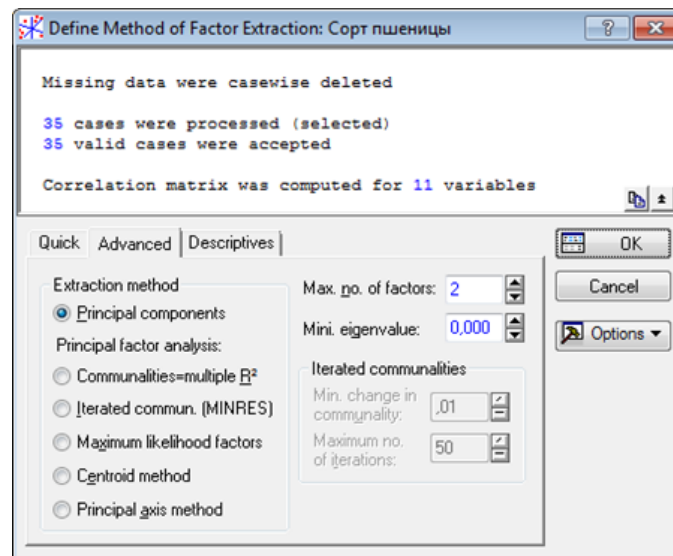


Figure 2. Method of Factor Extraction

Choose an option Principal components and click the OK button.

The system will quickly calculate, and a window will appear on the screen Factor Analysis Results (Figure 3).

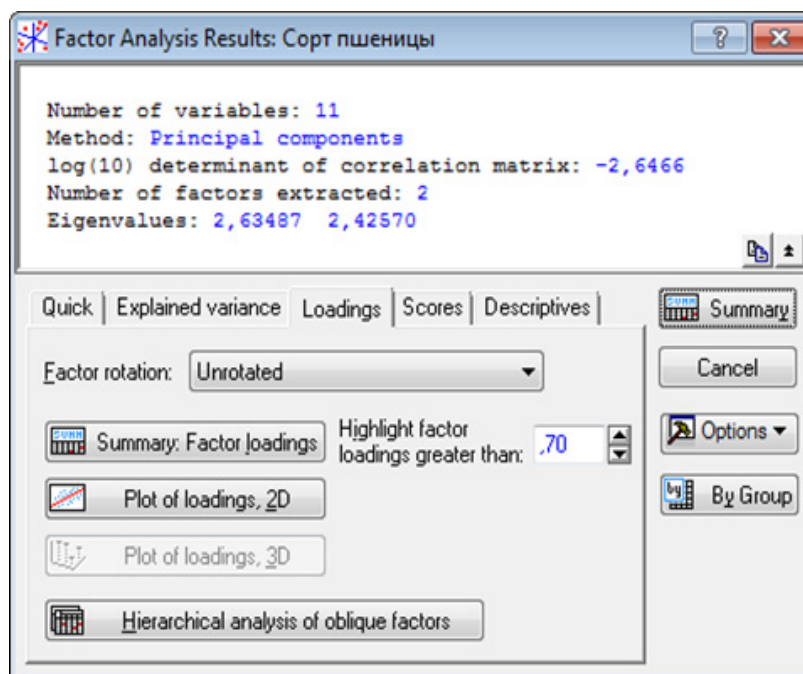


Figure 3. Factor Analysis Results

In the upper part of the Factor Analysis Results window an informational message is given:

Number of variables – 11;

Analysis Method (метод анализа) – Principal components;

log(10) determination of correlation matrix -2,6466;

Number of Factor extraction – 2;

Eigenvalues – 2,63487; 2,42570.

At the bottom of the window there are subsections that allow you to view the results of the analysis numerically and graphically.

Plot of loadings, 2D and Plot of loadings, 3D (Load graphs) – these options will build graphs of factor loads projected on the plane of any two selected factors (Figure 4) and in the projection into space of three selected factors (which requires the presence of at least three selected factors).

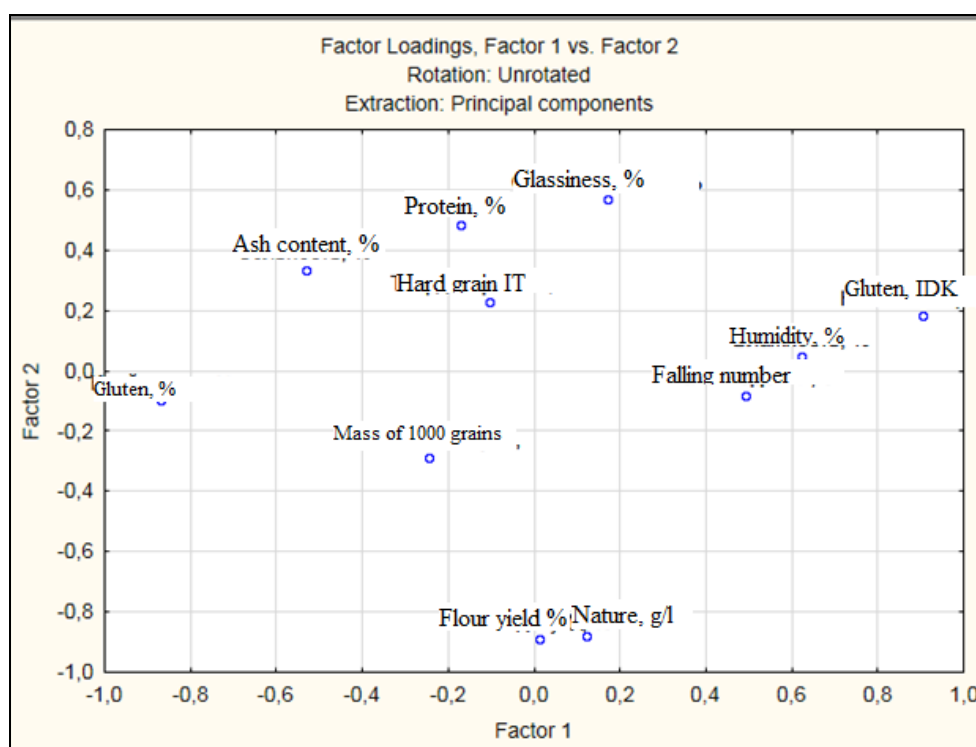


Figure 4.

Summary. Factor loadings. This option causes a table with current factor loads. (Figure 5), that is, the factors calculated for this method of rotation, which is specified to the right of the corresponding button.

Variable	Factor Loadings (Unrotated) (Сорт пшеницы) Extraction: Principal components (Marked loadings are >,700000)	
	Factor 1	Factor 2
Humidity, %	0,623774	0,045762
Nature, g/l	0,123160	-0,879611
Mass of 1000 grains	-0,243864	-0,286723
Glassiness, %	0,172705	0,568210
Hard grain, IT	-0,102817	0,229772
Ash content, %	-0,531817	0,336338
Gluten, %	-0,869099	-0,097295
Gluten, IDK	0,905918	0,181363
Protein, %	-0,172798	0,485416
Falling number, %	0,491818	-0,081616
Flour yield, %	0,011907	-0,891204
Expl. Var	2,634866	2,425695
Prp. Totl	0,239533	0,220518

Figure 5. Table with current factor loads

Summary. Factor loadings. This option causes a table with current factor loads. (Figure 5), that is, the factors calculated for this method of rotation, which is specified to the right of the corresponding button.

In this table, columns correspond to factors, and rows correspond to variables, and for each factor the load of each source variable is indicated, showing the relative magnitude of the projection of the variable on the factor coordinate axis. Factor loadings can be interpreted as correlations between the corresponding variables and factors — the higher the modulus load, the greater is the closeness of the factor to the original variable. They represent the most important information for the interpretation of the factors obtained. In the generated table, to facilitate the interpretation, factor loads will be highlighted in absolute value greater than 0,8.

Apparently, the first factor is more correlated with variables than the second. They are difficult to interpret, the question arises, what is the point of giving the second factor. In this case, it is advisable to resort to the rotation of the axes, hoping to get a solution that can be interpreted in the subject area.

Click on the menu Factor rotation (Figure 6).

The simple structure we can receive by rotation, in which most observations are located near the axes of coordinates. It is impossible to obtain a simple structure with a random configuration of observations

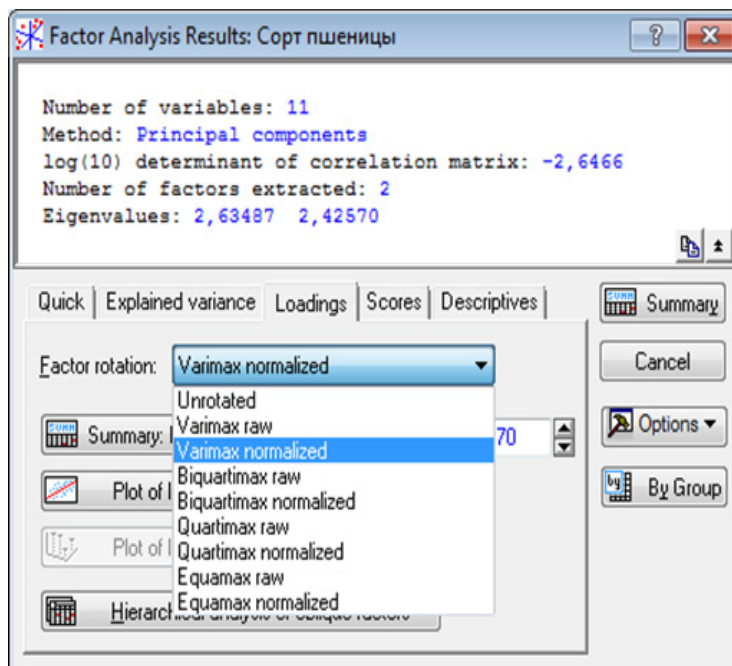


Figure 6. Menu Factor rotation

We can select different axis rotations in this drop-down menu. The window offers several possibilities to evaluate and find the desired turn by the following methods: Varimax; Biquartimax; Quartimax; Equamax.

The additional term in method name – normalized – indicates that the factor loadings in the procedure are normalized, i.e. divided by the square root of the corresponding dispersion. The term “raw” indicates that rotating loadings are not normalized.

Press the button Varimax normalized.

The system will produce the rotation of the factors by the method of Varimax normalized, and the window Factor Analysis Results (appears again on the monitor).

Press again in this window the button Plot of Loadings 2D. We will again see the loadings schedule (Figure 7).

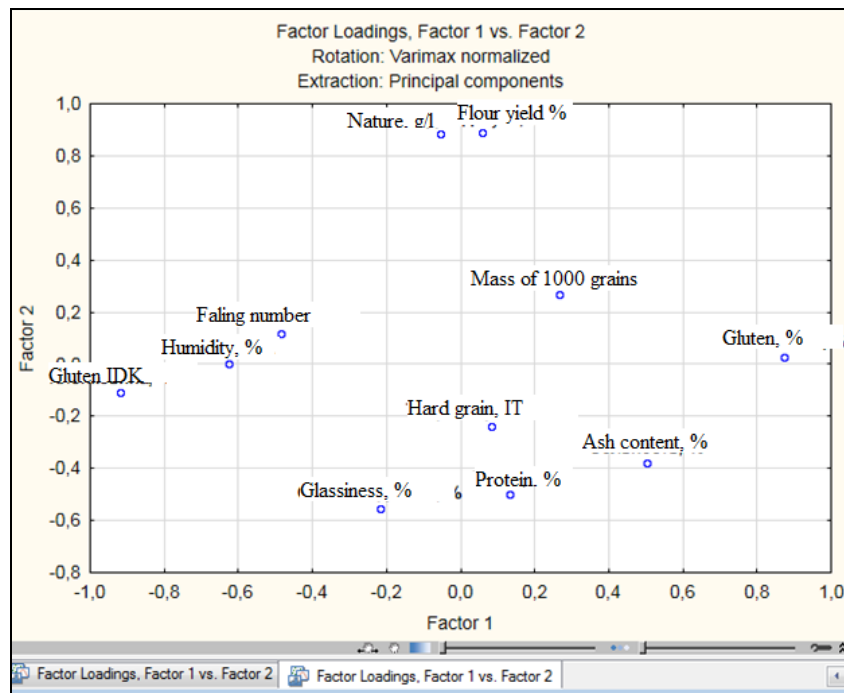


Figure 7. Loadings Schedule

In this drop-down menu, we can select different axis rotations. The window offers several possibilities to evaluate and find the desired turn by the following methods: Varimax; Biquartimax; Quartimax; Equamax.

The additional term in method name – normalized – indicates that the factor loadings in the procedure are normalized, i.e. divided by the square root of the corresponding dispersion. The term “raw” indicates that rotating loadings are not normalized.

Press the button Varimax normalized.

The system will produce the rotation of the factors by the method of Varimax normalized, and the window Factor Analysis Results (appears again on the monitor).

Press again in this window the button Plot of Loadings 2D. We will again see the loadings schedule (Figure 7).

Of course, this chart is slightly different from the previous one. We can consider the load more numerically by initiating the button Factor loadings. Click on the button Summary. Factor loadings, you open the window (Figure 8):

Variable	Factor Loadings (Varimax normalized) (Copr nu Extraction: Principal components (Marked loadings are >,700000)	
	Factor 1	Factor 2
Humidity, %	-0,625440	0,003445
hNature, g/l	-0,053590	0,886573
hMass of 1000 grains	0,265662	0,266653
hGassiness, %	-0,216864	-0,552865
hHard grain IT	0,084425	-0,237148
hAsh content %	0,503713	-0,377128
hGluten %	0,874059	0,028632
hGluten IDK	-0,917377	-0,109544
hProtein %	0,134081	-0,497504
hFallinf number %	-0,483875	0,120048
hFlour yield %	0,058230	0,889380
Expl.Var	2,633572	2,426990
Prp.Totl	0,239416	0,220635

Figure 8.

Now the solution can be interpreted. Factors are more often interpreted by loadings. The first factor is most closely related to Gluten, %.

Second factor is – Flour yield, % and Nature, g/l.

Thus, we made the classification of variables into two groups. The question arises: how many factors should be limited in practice?

To do this, there is a criterion in the STATISTICA software package - Scree plot. In the window Factor Analysis Results press the button Scree plot and then we get the following graph of eigenvalues (Figure 9):

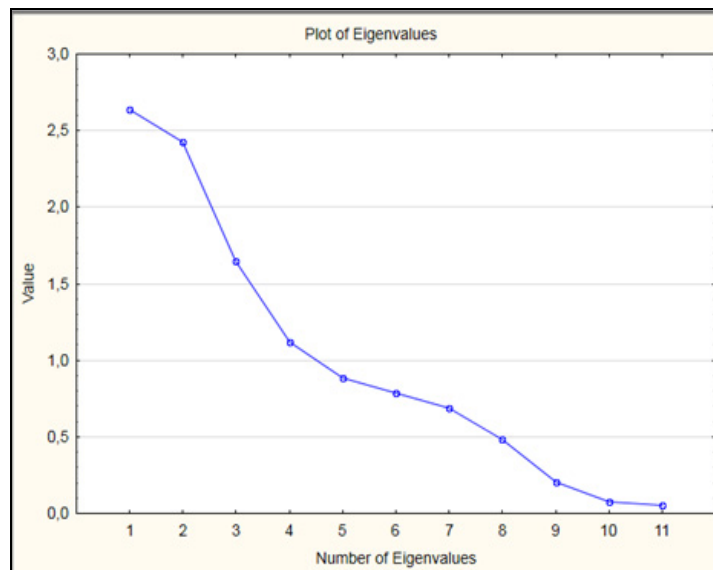


Figure 9.

In points with coordinates 1, 2, 3, 4 shedding slows down most significantly, therefore, theoretically it can be limited to four factors (Figure 10):

Variable	Factor Loadings (Varimax normalized) (Сорт пшеницы) Extraction: Principal components (Marked loadings are >.700000)			
	Factor 1	Factor 2	Factor 3	Factor 4
Humidity, %	-0.676105	0.072295	0.008738	0.218330
Nature g/l	0.029417	0.942391	0.041993	-0.137940
Mass of 1000 grains	-0.005076	-0.065435	0.928951	0.072487
Glassiness, %	-0.018460	-0.219885	-0.841936	0.158146
Hard grain IT	0.056517	0.070837	-0.163439	0.822125
Ash content %	0.397329	-0.557260	0.290541	-0.020981
Gluten %	0.891933	0.044244	0.062259	0.115927
Gluten IDK	-0.916849	-0.121855	-0.137390	-0.169125
Protein %	-0.006047	-0.408328	0.143730	0.657214
Falling number	-0.526364	0.089302	0.095962	-0.019697
Flour yield %	0.055830	0.876440	0.293890	-0.025340
Expl.Var	2.535784	2.221175	1.823764	1.248291
Prp.Totl	0.230526	0.201925	0.165797	0.113481

Figure 10.

The solution found can be interpreted. Factors are more often interpreted by loadings. The first factor is most closely related to Gluten, %.

Second factor – Flour yield, % and Nature, g/l.

Third factor – Mass of 1000 grains and and Glassiness, %.

Fourth factor – Hard grain.

Thus, we classified the variables into four groups.

Again, press in this window the button Plot of Loadings 2D. We will again see the loading schedule (Figure 11).

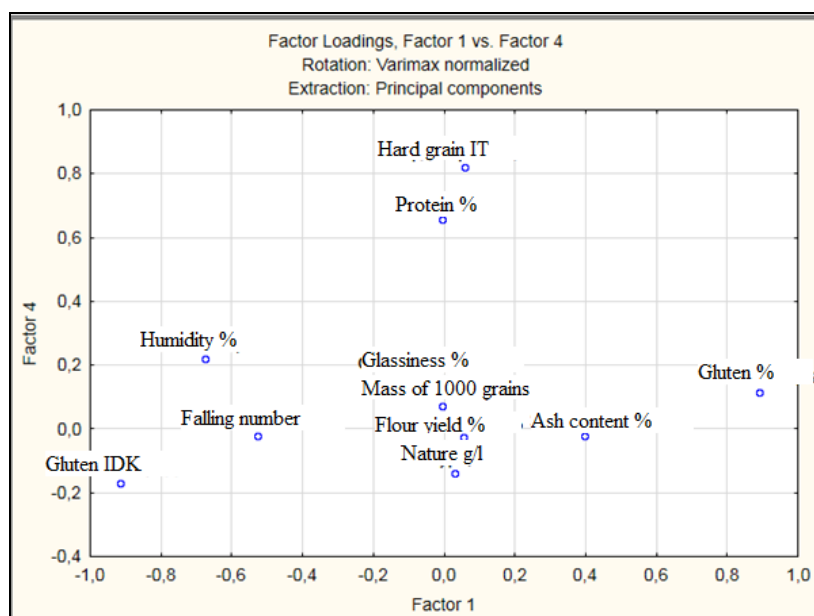


Figure 11.

Comparison of the results confirms the conclusion about the objective nature of the revealed factor structure: for all four methods of calculating the matrix of factor loadings turned out to be very close, which confirms the proposed interpretation using the method of principal components.

Correlation Matrix Transformation (Figure 1) into the Matrix of Factor Loadings (Figure 5, Figure 8 and Figure 10) was implemented by us in several ways: using the centroid method, the minimum residual method, maximum likelihood method and principal components method.

The line Expl.Var (Figure 5) shows the dispersion attributable to one or another factor. The line Prp.Totl shows the proportion of dispersion attributable to the first, second and third factors. The contribution of each factor to the total variability of 24% and 22%, respectively.

The basis for determining the factors was a consistent series of eigen values of the original correlation matrix. The Factor Loading Matrix presented in Figure 5 is obtained using the principal component method.. The contribution of each of the factors to the total dispersion was compared with a certain threshold; on this basis, factorization is discontinued after the extraction of four factors (Figure 5).

Figures 4, 7 and 11 show the results of the rotation of the coordinate system for wheat.

The abscissa axis shows the load on the first factor, the ordinate axis - on the second. In the space of factors, quality indicators are depicted by vectors, the squares of the modules of which are equal to the factor dispersions.

According to the basic principle of factor analysis, vectors should form beams. The scatter of vectors relative to the center of the beam characterizes the uncertainty when replacing a beam of feature vectors by the factor. Figure 11 clearly shows the presence of four beams. The first of these includes indicators: the amount of gluten. Second - Nature, g/l and Flour yield, %. Third – Mass of 1000 grains and Glassiness, %. Fourth – Hard grain.

First factor, therefore, can be interpreted as a factor of the grain indicator, the second is the flour indicator, the third and fourth are the grain indicator.

Indicators as Humidity,%, Ash content,%, Protein,% and Falling number are the independent feature that does not depend on the indicators that form the beams f1, f2, f3 и f4. It estimates grain from the flour-grinding point of view (outputs of varietal flour).

Thus, to assess the quality of grain, it is necessary to take into account specific indicators. In this regard, we are faced with the task of developing the basics of an integrated system for assessing the quality of grain, taking into account the intended purpose of grain and grain products, using modern analysis methods that increase the reliability of ensuring the adequacy of consumer properties of raw materials to the requirements of the food processing industry.

In the course of our work, at the production stage using factor analysis, an indicator was developed for assessing the quality of grain throughout the entire production cycle:

«grain indicator is flour», allows to significantly reduce the cost of research and predict the quality of the final product, as well as adjust the properties of intermediate products in the right direction;

identified significant indicators of grain quality;

compiled examples of classification indicators for assessing the quality of grain.

Our derived system of factors: «grain indicator is flour», shows that each of these parameters finds its maximum expression by means of different factors.

5. Conclusions

- 1). The factor analysis method gives possibility to move from a large number of indicators of grain quality to a limited number of complex factors which determine the relationship between these indicators and their technological value.
- 2). The obvious benefit of the method is the possibility to identify relationships between the studied grain quality indicators that are not revealed by the methods of correlation analysis and regression equations.

- 3). The system of four factors «grain indicator is flour» gives technological advantage of grain-wheat, identically described in the factor space.
- 4). The ash content index characterizes only one side of the milling production (flour yield) and not uniquely expressed through the factors, since it and is not directly related to the macro-value of grain and flour.

References

- Abugaliev, A. J., Dracheva, L. M., Fursov, O. V., & Esimbekova, M. A. (1998) Common wheat hardness in Kazakhstan: NIRS, varieties and grain quality breeding strategy. *16th JCC conference "Cereal Science - its contribution to health and well being"*, May 9-12, 1998, Vienna, Austria, p.56.
- Abugaliev, A. J., Srokbayev, S., Dracheva, L., & Savin, V. N. (1997). Spring wheat in Kazakhstan. Current status and Future Directions. - Proceedings of the Kazakhstan. - CIMMYT Conference. - Shortandy, Akmola, September 22-24, 1997. - P. 49-56.
- Borovikov, V. P. (2001). *STATISTICA. The art of analyzing data on a computer*. For professionals. Saint-Petersburg: Piter.
- Butkovskiy, V. A., & Kasatov, D. A. (2003). Milling industry of Great Britain. *Bread products*, 12, 14-16.
- Dauletbakov, B., Iztayev, A. I., Aryngazin, K. Sh., Omarov, T. E., & Syrlybayeva, L. M. (2007). Grain quality modeling using factor analysis. *Food technology and service*, 2, 3-7.
- Dauletbakov, B., Sultangaliyeva, L. S., & Dnimova, Z. (2018). Forecasting Methods of Agrometeorological Conditions in the Northern Zone of the Republic of Kazakhstan. *Agricultural Sciences*, 9(9). Retrieved from <https://www.scirp.org/Journal/PaperInformation.aspx?PaperID=87645>
- Dauletbakov, B., Sultangaliyeva, L. S., & Kuanyshovna, P. K. (2016). Improvement of the methodology for grain quality assessment. *Journal of Theoretical and Applied Information Technology*, 92(1).
- GOST R 52189-2003. (2004). *Wheat flour: General specifications. - Introd. 2003-29-12 (III, p.8)*. M.: Standards publishing.
- Kazakhstan: non-waste processing technology of grain crops*. <https://www.newsland.com/news/detail/id/693657/>
- Kazakova, I. E. (1975). Modeling of technological quality of grain by factor analysis. University news USSR, Food technology, № 3, 88, 1975.
- Kharman, G. (1972). *Modern factor analysis* (Transl. from Eng.). *Statistics*, 1972.
- Landi, A. (1987). Garatteristiche ottimali del grano duro e tenero per I prodotti destinati aii'alimentazione umana. *Informatore agrario*, 43(36), 29-31.
- Lawley D., & Maxwell A. (1967). Factor analysis as a statistical method (Transl. from Eng.).

Mir, 1967.

Modern aspects of scientific and technological support for the processing of agricultural raw materials and waste. *Collection of reports of the international scientific-practical conference*, 9-10 october 2014 - Astana, 2014.

Roy Chung, K. H. (2004). Weak flour for strong cookies / Roy Chung Kah Hee // Flour - Food for Life. *материалы 2nd International Muhlenchemie Symposium*, 17/18 June 2004, Hamburg.

Serduykov, P. I. (1970). *Flour-milling signs of wheat quality and their significance for flour milling*. CNIITEI Ministry of gas USSR, M.,

Tiefenbacher, K., & Dobrovics, M. (2000). Mehle tur die Waffeiherstellung. Ver-offentl. Arbeitsgemensch. *Getreideferschung e.v.*, 285, 65-73.

Torzhinskaya, L. R., Romenskiy, N. V., Atanas, L. G., & Umleva, N. G. Red. Of USSR universities. *Food technology*, 2(12), 1965.

Ustimenko, T. V. (2007). *Workshop: assessing the quality of grain and grain products. Methodical instructions* (p. 176). Saint-Petersburg: GIORD.

Zhukovskaya, V. M., & Muchnik, I. B. (1976). Factor analysis in socio-economic research. *Statistics*, 1976.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).