Use of Harrington’s desirability function in wheat grain quality assessment

A F Melnik, V G Shumetov, B S Kondrashin and M R Mikhaylov

Orel State Agrarian University named after N.V. Parakhin, 69, Generala Rodina St., Orel, 302019, Russia

Email: melnik.anat202@yandex.ru

Abstract. The publication is devoted to the development of a method for assessing the quality of wheat grain based on Harrington's desirability function, which reflects the nonlinearity of the relationship between the quantitative values of quality indicators and the psychological perception of a person. As an empirical base, we used the results of studies on the protein content, gluten content and sedimentation index – Zeleny test of more than 40 varieties of winter wheat, zoned in the Oryol region, the main analysis tool was the factor analysis by the method of principal components. It is substantiated that the set of these wheat grain quality indicators can be described with one slight distortion by one main factor, along with particular desirability functions used to construct a generalized quality indicator with linguistic levels “very good”, “good”, “satisfactory”, “bad” and ”very bad”. The calculation of the values of the main factor of wheat varieties, which are absent in the initial data array, is proposed using an approximate formula that takes into account the indicator that is most closely correlated with it - the protein content. The proposed methodology was tested on the example of assessing the influence of sowing dates on the productivity of winter wheat variety Moscow 39. Based on the results of the studies, it is recommended to use the Harrington’s transformation to evaluate particular indicators of the quality of wheat grain. It is also promising to use the method of principal components for the formation of a generalized function of desirability, when the weights of particular indicators are taken into account “automatically” by using the main component as a normalized variable, which comprehensively reflects the quality of wheat grain.

1. Introduction

An important factor in the formation of quality and food safety is the quality of raw materials. At the same time, the multiplicity and complexity of indicators that form the concept of “quality of raw materials” necessitates the search for a mathematical apparatus for their accounting. The foregoing fully relates to the task of assessing the quality of wheat grain used in the production of bakery products. The interest in this problem is largely caused by the publications of N. M. Derkanosova et al. [1, 2, 3] on the development of a methodology for the integrated assessment of the quality and safety of agricultural products using wheat grain indicators as an example.

2. Problem Statement

While selling grain on the world market, the domestic producers need to focus on grain quality. The main suppliers of cereals and legumes to the world market are the EU countries, as well as Canada and the United States. Due to different climatic conditions of growing, cereals are characterized by different
chemical composition and quality indicators. However, all countries are obliged to have organizations to control the quality and use of grain.

For example, the minimum requirements for the quality of cereals accepted for intervention in the EU member states are defined by EU Commission Regulation No. 824/2000. Cereals are accepted according to the established quality characteristics. In the USA, due to favorable soil-climatic conditions and constant quality of the obtained wheat harvest, it is sufficient to determine the protein content, which is not included in the standard, but it is indicated in the certificate for grain quality on the request of the consumer. There is also no division of wheat grain into food and non-food grain; there are no basic limiting standards. Grain prices are set under the market conditions and depend more on demand and supply for wheat than on its quality.

In addition to the lack of division into food and fodder grains, the American standard of grain classification is based on the biological properties of grain and its condition as a commodity, i.e. grain quality is estimated in two main categories: grain condition (health of grain) and grain purity.

The Russian standard (GOST 9353-2016) provides a commodity classification of wheat grain, which allows to classify grain to a certain class according to its typical composition. The first four classes are classified as food grain and the fifth as non-food (fodder) grain.

One of the most important indicators determining the technological advantages of bakery wheat is the amount and quality of gluten. There is no universal correlation between gluten content and protein content; thus, according to VNIIZ, the ratio between protein and gluten in Russian wheat varies from 0.4 to 2.4, depending on growing regions and agricultural conditions.

Evaluation of the quality of flour produced from soft wheat grain is carried out with the Zeleny test, for which a sedimentation analysis is carried out, or determined on an Infratek 1241 device according to the original method. [4]. On the world market, the sedimentation indicator is of particular importance both for export and import of grain. It is believed that grain with a high index of Greenery guarantees high quality bakery products.

Despite the large number of publications devoted to the analysis of the quality of wheat grain and agricultural products in general, only a few of them are devoted to the use of qualimetric forecasting. According to a number of authors, the use of qualimetric forecasting is one of the effective ways to increase the competitiveness of agricultural products (for example, [5]). There is no clear methodology for conducting qualimetric forecasting in the agricultural sector and, despite the high efficiency of applying qualimetric methods, including qualimetric forecasting in managing the quality of products and services, their use in the agricultural sector is sporadic [5, p. 9].

3. Materials and methods

As the empirical base of the study, we used the results of studies on the grain quality of 46 winter wheat varieties divided into zones in the Oryol region in recent years, presented in Table 1 [6].

Statistical analysis of the data in table 1 was carried out by graphical and analytical tools of the SPSS package version 8.0 [7]. Some graphs were made in the program version 13.0.

An integral element of qualimetric forecasting is availability of information on the rating assessment of various wheat grain quality indicators. This is relevant in connection with the multi-parametric nature of the problem, and a fuzzy approach to the generalized indicator constructing is useful, when instead of the natural (named) values of private quality indicators, they are converted to a dimensionless scale of desirability [8]. The most common is the transformation based on the relationship between the quantitative values of the dimensionless scale and the psychological perception of a person, which appears as a result of observations for real decisions of the experimenter and reflects the objective laws of the systems development - table 2 [8, p.3].

A widespread function that implements such a transformation in the case of quality indicators with a one-sided constraint is Harrington's two-parameter double exponent [9], which is convenient due to the fact that it is defined by only two defining points - the lower and upper limits of the zone "satisfactory". This function varies in the range from 0 to 1, asymptotically approaching to zero at an unacceptable level of the indicator and unity at its ideal level.
Table 1. Grain quality of winter wheat genotypes in 2012

| Variety                  | Code | Protein content,% | Gluten content,% | Sedimentation index - Zeleny test, ml |
|--------------------------|------|-------------------|------------------|--------------------------------------|
| Moskovskaya 39           | 1    | 14.8              | 27.3             | 49.6                                 |
| Nemchinovskaya 57        | 2    | 13.1              | 21.3             | 38.7                                 |
| Moskovskaya 56           | 3    | 13.0              | 22.6             | 35.6                                 |
| Nemchinovskaya 17        | 4    | 14.0              | 23.9             | 46.1                                 |
| Pamyati Fedina           | 5    | 12.4              | 15.9             | 32.9                                 |
| Nemchinovskaya 24        | 6    | 12.4              | 20.2             | 33.9                                 |
| Moskovskaya 40           | 7    | 14.5              | 26.2             | 46.7                                 |
| Eritropermum 902/68      | 8    | 12.8              | 20.5             | 37.0                                 |
| Inna                     | 9    | 11.8              | 18.8             | 32.6                                 |
| Konstantinovskaya        | 10   | 13.0              | 22.2             | 38.0                                 |
| Povolzhskaya 86          | 11   | 13.9              | 25.3             | 44.0                                 |
| Kinelskaya 8             | 12   | 13.4              | 21.8             | 48.1                                 |
| Bezenchukskaya 790       | 13   | 13.3              | 22.8             | 39.6                                 |
| Biryuza                  | 14   | 12.9              | 23.3             | 32.9                                 |
| Bezenchukskaya 380       | 15   | 14.6              | 25.7             | 48.7                                 |
| Svetoch                  | 16   | 14.0              | 24.1             | 44.6                                 |
| Krastal                  | 17   | 13.0              | 22.2             | 38.0                                 |
| Sozvezdiye               | 18   | 13.7              | 25.1             | 41.6                                 |
| Zhemchuzhina Povolzh'ya  | 19   | 12.4              | 21.9             | 32.3                                 |
| Kalach 60                | 20   | 12.7              | 21.6             | 37.9                                 |
| Donchanka                | 21   | 16.4              | 29.4             | 52.6                                 |
| Amazonka                 | 22   | 14.0              | 25.9             | 47.0                                 |
| Lidiya                   | 24   | 14.6              | 25.0             | 50.7                                 |
| Izyuminka                | 25   | 14.4              | 27.2             | 44.5                                 |
| Yermak                   | 27   | 14.0              | 24.0             | 39.0                                 |
| Asket                    | 28   | 14.8              | 27.7             | 50.0                                 |
| Spartak                  | 29   | 14.6              | 26.0             | 47.7                                 |
| Aksinya                  | 30   | 15.9              | 28.0             | 60.7                                 |
| Rostovchanka 7           | 31   | 13.6              | 24.0             | 39.4                                 |
| Dominanta                | 32   | 14.7              | 27.6             | 53.6                                 |
| Arfa                     | 33   | 13.7              | 23.9             | 47.3                                 |
| Kameya                   | 34   | 12.8              | 21.3             | 34.8                                 |
| Gubernator Dona          | 35   | 12.4              | 19.9             | 29.8                                 |
| Missiya                  | 36   | 15.2              | 27.4             | 58.4                                 |
| Oda                      | 38   | 12.8              | 21.3             | 40.8                                 |
| Alyans                   | 39   | 13.5              | 23.6             | 40.2                                 |
| Zolushka                 | 41   | 13.2              | 22.4             | 33.0                                 |
| Ariadna                  | 42   | 14.6              | 25.5             | 52.8                                 |
| Bogdanka                 | 43   | 12.8              | 22.0             | 36.0                                 |
| Korochanka               | 44   | 13.6              | 24.3             | 40.3                                 |
| Sintetik                 | 45   | 14.4              | 23.1             | 49.7                                 |
| Lgovskaya 4              | 46   | 13.4              | 23.4             | 40.3                                 |
| Lgovskaya 8              | 47   | 14.5              | 25.0             | 45.9                                 |
| Skipetr                  | 48   | 13.9              | 23.8             | 45.1                                 |
| Favoritka                | 49   | 12.3              | 20.0             | 31.3                                 |
| Orlovskaya 241           | 50   | 13.7              | 24.4             | 46.5                                 |
Table 2. The relationship between the quantitative values of the dimensionless scale and the psychological perception of a person

| Desirability | Quantitative mark on the desirability scale |
|--------------|--------------------------------------------|
| Very good    | 1.00… 0.80                                 |
| Good         | 0.80… 0.63                                 |
| Satisfactory | 0.63… 0.37                                 |
| Poor         | 0.37… 0.20                                 |
| Very bad     | 0.20… 0.00                                 |

Graphically, it is displayed by an S-shaped curve, the "sensitivity" of which in the areas of desirability close to 0 and 1 is significantly lower than in the middle zone of desirability - the zone "satisfactory".

The research analysts have different approaches to the designation of reference points of the Harrington’s function. Most of them suggest that lower and upper borders of the zone “satisfactory” are to be established by the expertise, but it is also proposed to use the statistical characteristics of real samples for this purpose. So, it is proposed to set these boundaries equal to the minimum and maximum values of the indicators, but in this case the zone “satisfactory” illegally expands according to the principle “satisfies everything that really exists”. In the work [10] a somewhat different approach to the formation of the Harrington’s desirability function with a one-sided restriction is presented, according to which the lower boundary of the zone “satisfactory” equal the arithmetic mean value of the indicator, and the upper boundary is “shifted” towards their large values by one standard deviation.

A statistical approach is also used for the formation of a generalized desirability function. So, in the work [11] it is proposed that a generalized desirability function is to be formed taking into consideration the results of the factor analysis: if partial desirability functions do not correlate or correlate with each other weakly, the multiplicative form of the generalized desirability function is adopted; in the case of their correlation, the additive form is adopted. It is proposed that weights of particular indices in the generalized desirability function be assumed to be proportional to the variance fractions explained by the corresponding main factors.

The cited works are methodological and are of undoubted interest for adapting the statistical approach to the formation of particular and generalized desirability functions for assessing the quality of wheat grains.

4. Results and Discussion

To check the conformity of the distributions of grain quality indicators with the requirements of statistical analysis, Tukey’s diagrams (box plots) were constructed, the analysis of which showed that they do not contain extreme values and are close enough to symmetrical. It should be noted only the presence of a minimum gluten content of wheat grains in the variety of Pamyati Fedina. Additional verification of the distribution of indicators by the Kolmogorov-Smirnov test (table 3) did not reveal statistically significant deviations from the normal, which allows applying of parametric analysis methods to empirical data.

Thus, according to the calculation of the variation coefficient, turned out that the analyzed varieties of winter wheat showed the greatest variability in the “Zeleny test” (sedimentation index), the least in index of protein content, and the informative value has the following sequence: sedimentation, gluten content, protein content (variation coefficients 17.5%; 11.4% and 7.2%, respectively).

In this case, we have three positive quality indicators with one- sided constraint, i.e. large values of these indicators are desirable. Although there is a point of view that increasing of protein content of more than 17–19%, the quality of bread deteriorates, there is no such high protein content in the empirical data sample, and assigning this indicator to positive with one-sided constraint is correct.

All three indicators are quite strongly correlated with each other. Namely, as a result of factor analysis by the method of principal components, it turned out that the array of empirical data can be represented...
with good accuracy by one main factor that very strongly correlates with protein and gluten content and Zeleny test (correlation coefficients are 0.981; 0.948 and 0.947, respectively).

### Table 3. Kolmogorov-Smirnov test of winter wheat grain quality indicators

|                 | Protein content, % | Gluten content, % | Sedimentation index - Zeleny test, ml |
|-----------------|-------------------|------------------|--------------------------------------|
| **N**           | 46                | 46               | 46                                   |
| **Normal**      |                   |                  |                                      |
| **Parameters**  | Std. Deviation    |                  |                                      |
| **Most Extreme**| Absolute          |                  |                                      |
| **Differences** | Positive          |                  |                                      |
| **Kolmogorov-Smirnov Z** | 0.564 | 0.408 | 0.649 |
| **Asymp. Sig. (2-tailed)** | 0.908 | 0.996 | 0.794 |

The building of the desirability function is illustrated by the example of protein content. In accordance with the foregoing, it would be useful to rank varieties by the magnitude of the desirability function of the analyzed indicator values obtained as a result of the psychophysical transformation. Harrington’s desirability function for protein content is positive with one-sided constraint, and, according to [10], has the following form:

\[ d = d(z) = \exp(-\exp(-z)) \]

where \( z \) – the dimensionless values of the indicator

\[ z = (x - x_0)/\text{RMSE}(x). \]

In formula (2) \( x \) – the value of the indicator, \( x_0 \) – the lower boundary of the region “satisfactory” in the original scale, standard deviation, \( \text{RMSE}(x) \) – the root-mean-square error (standard error).

To build the desirability function (1) it is enough to establish the lower boundary of the initial positive indicator \( x_0 \), in this case equal to the arithmetic mean value \( \bar{x}_{avg} \), and its standard error for the analyzed sample. The analyzed sample, including all 46 varieties of wheat, is homogeneous with the numerical characteristics \( \bar{x}_{avg} = 13.7 \) and \( \text{RMSE}(x_i) = 0.98 \), and formula (1) takes the form:

\[ d_i = \exp\{-\exp\left[-\frac{(x_i - 13.7)}{0.98}\right]\}. \]

The diagram of the desirability function of the protein content is shown in Figure 1. It is obvious that within the desirability zone “satisfactory” it is almost linear, and with larger and smaller values of the indicator its steepness decreases.

In figure 1 it is clear that, in accordance with the gradation of the scale of desirability, the varieties Donchanka, Aksinya and Missiya (codes 21, 30 and 36) are characterized by a rating of “very good”, varieties Moskovskaya 39, Asket, Dominanta, Bezenchukskaya 380, Lidiya, Spartak and Ariadna (codes 1, 28, 32, 15, 24, 29 and 42) - by a rating of “good”, varieties Constellation, Harp, Orlovskaya 241, Rostovchanka 7, Korochanka, Alliance, Kinelskaya 8, Lgovskaya 4 and Bezenchukskaya 790 (codes 18, 33, 50, 31, 44, 39, 12, 46 and 13) - the rating is “bad” and the varieties Cinderella, Nemchinovskaya 57, Moskovskaya 56, Konstantinovskaya, Krustal, Biryuza, Erythrospermum 902/68, Kameya, Oda, Bogdanka, Kalach 60, PamyatiFedina, Nemchinovskaya 24, ZhemchuzhinaPovolzh'ya, Gubernator Dona, Favoritka and Inna (codes 41, 2, 3, 10, 17, 14, 8, 34, 38, 43, 20, 5, 6, 19, 15, 45, 49 and 9) - the rating is “very poor (listed in descending order of desirability). The grade “satisfactory” received the varieties Moskovskaya 40, Lgovskaya 8, Izyuminka, Sintetik, Nemchinovskaya 17, Svetoch, Amazonka, Ermak, Povolzhskaya 86 and Skipetr (codes 7, 47, 25, 4, 16, 22, 27, 11 and 48).
Similarly, the particular desirability functions of two other quality indicators – gluten content $x_2$ and sedimentation index – Zeleny test $x_3$ – are calculated by the formulas:

$$d_2 = \exp\{-\exp\left(-\frac{(x_2 - 23.7)}{2.71}\right)\}; \quad (4)$$

$$d_3 = \exp\{-\exp\left(-\frac{(x_3 - 42.5)}{7.45}\right)\}. \quad (5)$$

Graphs of these partial desirability functions are shown in Figure 2.

The generalized desirability function of wheat grain quality $D$ can be defined as the average of the partial desirability functions:

$$D = \frac{d_1 + d_2 + d_3}{3}, \quad (6)$$

but then the results of factor analysis, according to which particular quality indicators are included in the main factor with different weights, will not be taken into account. In this regard, it is
proposed to calculate the generalized function of the desirability of quality by the formula:

\[ D = \exp[-\exp(-f)], \quad (7) \]

where \( f \) – the marker of the main factor.

Comparing the diagram of the proposed variant of the generalized desirability function shown in Figure 3 b with the graph of the particular desirability function of the protein content (Figure 3 a), it is noticeable that their difference is observed only in details.

![Diagram of the proposed variant of the generalized desirability function](image.png)

**Figure 3.** Generalized function of the desirability of wheat grain quality: a - with equal weights of the indicators’ significance; b - taking into account the results of factor analysis

To understand the essence of the results, let’s turn to the theory of factor analysis by the method of principal components analysis (PCA). According to the PCA theory, the components distinguished in the process of factor analysis are linear combinations of initial indicators, and they are ranked by the value of the explained variance. In the case of a very strong correlation between all three initial indicators that we observe, we can restrict ourselves to the first main component, which accounts for 91.9% of the total variance. This main factor correlates most strongly with the protein content (correlation coefficient 0.981), and the correlation coefficients with gluten content and sedimentation index – Zeleny test – 0.948 and 0.947 respectively.

Quantitatively, the main factor is characterized by the so-called “markers” of the factor, the distribution of which obeys the normal law of distribution with the parameters \( f_{avg} = 0 \) and standard deviation \( \text{RMSE}(f) = 1 \), i.e. as a result of factor analysis, we obtain a normalized variable that can play the role of the dimensionless variable \( z \) in formula (1). According to the data used in our factor analysis, the values of the labels of the main factor for wheat varieties are obtained automatically, and for other varieties or other growing conditions, they can be calculated by the formula:

\[ f = -9.922 + 0.362 x_1 + 0.127 x_2 + 0.046 x_3, \quad (8) \]

where \( x_1, x_2 \) and \( x_3 \) – protein content, gluten content and sedimentation index – Zeleny test, respectively.

This formula has the following quality characteristics: determination coefficient 0.963, Fisher criterion 1147.8, and standard deviation of approximation 0.143:

\[ f = -11.480 + 0.707 x_1 + 0.042 x_3. \quad (9) \]

From the comparison of the beta regression coefficients of formula (9), it follows that the protein content makes the largest contribution to the result (0.695 versus 0.315 for the sedimentation index –
Zeleny test), and an alternative formula for the approximate calculation of the marks of the main factor will be:

\[ f = -13.672 + 0.999 x_i \]  \hspace{1cm} (10)

This formula has the following quality characteristics: determination coefficient 0.963, Fisher criterion 1147.8, and standard deviation of approximation 0.194.

Let’s test the proposed methodology for assessing the quality of grain zoned in the Oryol region of Moskovskaya 39 variety using the example of the effect of sowing dates on the productivity of winter wheat [12]. The results of the studies are presented in columns 2 and 3 of table 4 (average in 2013-2015); column 4 shows the values of the labels of the main factor calculated by the approximate formula (10), and column 5 shows the values of the function of desirability of wheat grain quality.

Table 4. The effect of sowing dates on the productivity of winter wheat Moskovskaya 39

| Sowing dates | Yield, t/ha | Protein content, % | Marker of the main factor | Desirability function |
|--------------|------------|--------------------|--------------------------|-----------------------|
| 5.09         | 5.49       | 13.3               | -0.39                    | 0.230                 |
| 15.09        | 5.82       | 14.5               | 0.81                     | 0.642                 |
| 25.09        | 5.63       | 14.0               | 0.31                     | 0.482                 |
| 5.10         | 5.40       | 13.6               | -0.09                    | 0.336                 |
| 15.10        | 4.73       | 12.8               | -0.88                    | 0.089                 |

Taking into consideration the gradations of Harrington’s desirability function, wheat grain quality is rated as “good” and “satisfactory” at sowing dates of September 15 and September 25; earlier and later sowing periods lead to a deterioration in the quality of wheat.

5. Conclusion

Based on the results of statistical studies, it is advisable to use the Harrington’s desirability function in assessing particular indicators of wheat grain quality. It is also promising to use factor analysis by the method of principal components to form a generalized desirability function when weights of particular indicators are carried out “automatically” by using the principal component as a normalized variable, which comprehensively reflects the quality of wheat grain.

References

[1] Derkanosova N M, Ponomareva I N, Shurshikova G V and Vasilenko O A 2018 Application of fuzzy set theory for integral assessment of agricultural products quality Journal of Physics: Conference Series 1015 (3) № 032026
[2] Derkanosova N M, Shurshikova G V and Gins V K 2013 Forecast of Positive Reaction of the Custumers on New Types of Baked Goods Food Technology № 2-3 15-19
[3] Shurshikova G V, Kotarev V I, Derkanosova N M, Vasilenko O A and Zolotareva N I 2015 Methods for Complex Evaluation of the Quality Level of Agricultural Products (By the Example of Wheat Grain) Food Processing: Techniques and Technology № 2 143-148
[4] 2014 Sedimentation index – Zeleny test and its significanse for assessing grain quality, available at: https://www.zol.ru/n/19b30
[5] Dunchenko N I and Yankovskaya V S 2012 Qualimetric Prognostics Use in Agricultural Sector RSAU - MTAA Bulletin 5 9-17
[6] Melnik A F, Amelin A V, Mazalov V I and Nikolaev A N 2015 The Influence of the Variety on the Yield and Quality of Winter Wheat under the Conditions of the Oryol Region Grain Economy of Russia 4 1-5
[7] SPSS Base 8.0 for Windows. Application guide. Translation–Copyright 1998 (SPSS-Rus’)
[8] Liubushin N P and Birkach G E 2014 Harrington’s Desirability Generalized Function in Multiple Parameter Economic Tasks Economic Analysis: Theory and Practice № 18 (369) 2-10
[9] Harrington E C 1965 The Desirability Function Industrial Quality Control 21(10) 494-498
[10] Barbashova E V, Chekulina T A and Shumetov V G 2015 Statistical Approach to Formation of Desirability Function in the Problems of Economic-Mathematical Modelling OrelGIET Bulletin. 2 (32) 94-100
[11] Kotelnikov S A 2012 Method of Forming of the Generalized Indicator of the Reproduction of the Human Scientific Potential of the Russian Innovation System Modern Management Technology 10 (22)
[12] Melnik A F 2017 Sowing Dates in Conditions of Climate Aridisation as a Factor in controlling the Productivity of Winter Wheat RJOAS 11(71) 426-430