

# Evaluation of the influence of sowing time on yield and quality of the wheat grain using Multivariate dispersion analysis

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

Orel State Agrarian University named after N.V. Parakhin, General Rodin St., 69, 302019 Orel, Russia

**Abstract.** The statistical analysis of the results of studies on the effect of sowing time on the yield and quality of winter wheat grains, from the perspective of multivariate variance analysis and multiple comparison of averages, is carried out in the article. It was found that the application of the procedure of a multivariate generalized linear model allows comparing the variants of experiments for all indicators in the aggregate. An important advantage of multiple comparisons over paired Student comparisons is also the ability to isolate homogeneous subgroups of experiments. In this study, post hoc comparisons according to the Tukey criterion were applied, the results of which are presented in tables of homogeneous experiment subgroups and multiple comparisons of indicators' differences. The distribution of experiment variants by zones of desirability of Harrington functions of yield and generalized indicator of grain quality, depending on shift of sowing time is also presented. It is justified that the maxima of yield and quality of winter wheat grains provide a sowing period of 10 days later than previously recommended for the region. The use of multivariate dispersion analysis and multiple comparison of average increases the reliability of statistical conclusions, allows you to distinguish homogeneous versions of experiments in agricultural production.

## 1 Introduction

The choice of winter wheat sowing time under the conditions of climate aridization in the Oryol region is very important. This is due to the increase of the warm autumn period by 2-3 weeks with an average daily temperature of 10-12°C and often with a lack of moisture in the soil, in comparison with the previously recommended winter wheat harvest dates for the zone. Winter wheat sowing at a later date increases the length of the field preparation period, which leads to the increase in yields and affects positively the quality improvement. So, in the work [1] on the example of the wheat variety Moskovskaya 39, it was shown that sowing at a later date ensured an increase in grain-unit, protein and gluten accumulation in the grain by 8-24 gram per liter, 0.7-1.3%, an increase in yield by 0.14-0.33 t/ha respectively, in comparison with the recommended period.

---

\*Corresponding author: [melnik.anat202@yandex.ru](mailto:melnik.anat202@yandex.ru)

A.A. Lyubishchev pointed out the importance of using dispersion analysis to make informed conclusions based on the results of the biological studies. He noted that "one of the great advantages of dispersion analysis is that it makes it possible to verify the correctness of modifications to the applied treatments" [4]. In the cited work, the technique of multiple comparisons of averages was not considered, however, it was noted that even with a significant difference between different experience options, subsequent mathematical processing gives a more valuable conclusion.

In the work [5], on the example of the study on the effect of the probiotic "Olin" on the live weight of broilers, it is shown that a multiple comparison of the averages upon criteriumSheffe test using the procedure of a generalized linear model gives the researcher comprehensive information about the reliability of statistical conclusions. It is noted that if the number of variants in the experiment is not large (up to four), then for their comparison you can rely on the LSR criterion, but the maximum comparison accuracy is achieved for only one pair of averages, and rank criteria should be used to increase the accuracy of the comparison [6].

Taking into account the above mentioned, statistical analysis of the results of the studies on the effect of sowing time on quality indicators of the grains of winter wheat variety Moskovskaya 39 given in the work [6] is performed from the point of view of multiple comparison of average.

## 2 Statistical study methodology

In the cited work, five test options were compared according to the scheme of Table 1.

**Table 1.** Scheme of the experiment

Option	Sowing date
1 <sup>st</sup> (Control)	5 <sup>th</sup> of September
2 <sup>nd</sup> Experimental	15 <sup>th</sup> of September
3 <sup>rd</sup> Experimental	2 <sup>th</sup> of September
4 <sup>th</sup> Experimental	5 <sup>th</sup> of October
5 <sup>th</sup> Experimental	15 <sup>th</sup> of October

To obtain conclusions that are not so tightly connected to the reference sowing date – 5<sup>th</sup> of September - we enter a quantitative variable - a sowing date shift, ranging from 0 to 40 days. The initial data for the simulation are shown in Table 2.

**Table 2.** Empirical simulation basis

Option of the experiment	1	2	3	4	5	HCP <sub>05</sub>	s <sub>x</sub>
Sowing date	5.09	15.09	25.09	5.10	15.10	-	-
Sowing date shift, days	0	10	20	30	40	-	-
Crop yield, dt/ha	54,9	58,2	56,3	54,0	47,3	1,4	0,54
Grain-unit, gram per liter	780	786	805	798	795	6,4	2,47
Protein, %	13,3	14,5	14,0	13,6	12,8	0,33	0,13
Gluten, %	23,7	25,0	23,5	22,1	19,1	0,9	0,35
Sedimentation, cbcm.	45	47	39	37	31	3,3	1,27

Table 2 shows the average yield and quality of wheat grains over three years, while the computer implementation of the dispersion analysis requires "raw" data for the years of the experiment. The values of the smallest significant difference for all LSD<sub>05</sub> indicators are given in the work [6], which allows to calculate standard error of experiment  $s_x$  and to

perform simulation of three times replication of experiments, adding to set of average arrays "averages plus or minus standard deviation".

### 3 Results and Discussion

Based on the results of our simulation of triplicate replication, we have five options of the experiments and data not only on the yield, but also on four indicators of wheat grain quality, and a valid method for detecting the impact on these indicators of the shift in the sowing time is a dispersion analysis according to the multidimensional generalized linear model scheme. In this case, multivariate analysis of variance should be preferred to one-dimensional one, since dependent variables correlate with each other: according to correlation data analysis, it was revealed that most Pearson correlation coefficients are statistically significant at a level at least 0.05, and considering the correlation of dependent variables with each other will allow taking into account all the connections hidden in the numerical data.

Applying the procedure of a multivariate generalized linear model allows to compare the variants of experiments with each other at the same time for all indicators. The calculations carried out in this procedure confirm the importance of the impact of the sowing time shift on yield and grain quality indicators in the aggregate.

An important advantage of multiple comparisons over paired comparisons according to the Student criterion is the possibility of separation homogeneous subgroups. In this study, post hoc comparisons according to the Tukey criterion were applied, the results of which are presented in tables of homogeneous subgroups and multiple comparisons of differences. There are many such tables - five for homogeneous subgroups and five for multiple comparisons of differences; Tables for yields are given below as an example.

**Table 3.** Homogeneous subgroups by yield (Tukey test, level of significance of criterion of difference between subgroups  $p=0,05$ )

Sowing date shift, days	Repeatability	Crop yield, dt/ha (subgroup)			
		1	2	3	4
40	3	47,3			
30	3		54,0		
0	3		54,9	54,9	
20	3			56,3	
10	3				58,2
Level of significance of the subgroup difference criterion		1,000	0,314	0,060	1,000

As it follows from Table 3, the 10-day and 40-day shifts form by grain yields independent subgroups 1 and 4, while the yield in the 10-day shift experience is significantly higher than in the control - 58.2 dt/ha comparing to 54.9 dt/ha, and in the 40-day shift experience - less control (47.3 dt/ha). At the same time, subgroups 2 and 3 include a control experience, and the differences in yield in sowing dates shift the by 20 and 30 days from the control are not statistically discernible.

These findings correlate to the results of multiple differences comparisons - Table 4: 95% confidence intervals for yield differences when shifting the sowing dates by 20 and 30 days include zero, which reflects their statistical insignificance, while the yield differences when shifting the sowing dates by 10 and 40 days significantly differ from zero.

**Table 4.** Posterior pairwise comparisons of yield by Tukey criterion

(J) Shift of sowing dates, days	(J) Shift of sowing dates, days	Average difference (I-J)	Standard error	Value	95% confidence interval	
					Lower bound	Upper bound
0	10	-3,300	0,4409	0,000	-4,751	-1,849
	20	-1,400	0,4409	0,060	-2,851	0,051
	30	0,900	0,4409	0,314	-0,551	2,351
	40	7,600	0,4409	0,000	6,149	9,051

Similarly, the 10, 20, and 40 day shifting options for the protein content form independent subgroups, while the protein content in the 10 and 20 day shifting experiment is markedly higher than the control - 14.5% and 14.0%, respectively, against 13.3% in the control, and in the 40 day shifting experiment - less control (12.8%). At the same time, the difference in protein content when shifting the sowing dates by 30 days from the control is not statistically significant.

Further from the analysis of the composition of homogeneous subgroups it follows, that options for shifting the sowing time by 10, 20 and 40 days in terms of gluten content form independent subgroups, gluten content in the experiment with 10 days shift is higher than control - 25,0%, in comparison to 23.7% in the control, respectively, and in the 30 and 40 day shift, less control (22.1% and 19.1%, respectively). At the same time, the difference in gluten content by the sowing date shift by 20 days from the control is not statistically significant.

It was also determined that options of sowing date shift by 20 days by grain-unit form a single independent subgroup, moreover, the grain-unit in this experience is higher than control - 805 gram per liter against 780 gram per liter in control, and with sowing date shift by 10 days, grain-unit is not statistically different from control. Another subgroup combines experiments with the sowing date shift by 30 and 40 days, but the grain-unit in it is only slightly higher than control.

The last of the wheat grain quality indicators recorded in the experiment under discussion is sedimentation. It follows from table 5 that the option of the sowing date shift by 10 days forms by sedimentation with control a common subgroup 3, the indicator of 47 cb cm in this experiment is not statistically different from the control level of 45 cb cm. Subgroup 2 combines experiments with the sowing date shift by 20 and 30 days, the sedimentation in it is slightly lower than control - 39 cb cm. and 37 cb cm., respectively. Subgroup 1 with the 40-day sowing date shift is characterized by the minimum sedimentation value of 31 cbcm.

**Table 5.** Homogeneous subgroups by sedimentation (Tukey test, significance level of the criterion of difference between subgroups  $p=0,05$ )

Shift of sowing dates, days	Repeatability	Sedimentation, cbcm (subgroup)		
		1	2	3
40	3	31		
30	3		37	
20	3		39	
0	3			45
10	3			47
Level of significance of the subgroup difference criterion		1,000	0,363	0,363

These findings correlate the results of the multiple difference comparisons shown in Table 6: 95% confidence interval for the sedimentation difference in the 10-day seeding shift includes zero, which reflects its statistical insignificance, while the sedimentation differences in the 20, 30 and 40-day seeding shift differ significantly from zero.

**Table 6.** Posterior pairwise comparisons of sedimentation by Tukey criterion

(I) Shift of sowing dates, days	(J) Shift of sowing dates, days	Average difference (I-J)	Standard error	Value	95% confidence interval	
					Lower bound	Upper bound
0	10	-2,00	1,037	0,363	-5,41	1,41
	20	6,00	1,037	0,001	2,59	9,41
	30	8,00	1,037	0,000	4,59	11,41
	40	14,00	1,037	0,000	10,59	17,41

An important result of the generalized linear model procedure is the evaluation of variance analysis model parameters.

The peculiarity of one-dimensional models of variance analysis is that each variable in them is represented in the form of the sum of the average value, the effect of the experiment variant and error. In mathematical form, a single-factor linear model is presented in the following way:

$$Y_{ij} = \mu_0 + \alpha_i + \varepsilon_{ij} \quad (1)$$

where  $Y_i$  – is the observed value of the output variable  $Y$  at the  $i$ -th level of the factor;  $\mu_0$  – is the estimation of the free coefficient of the model;  $\alpha_i$  – is the estimation of the main effects at the  $i$ -th level;  $\varepsilon_{ij}$  – is a random error.

The interpretation of the variance analysis model parameters is explained by the linear yield model (Table 7).

**Table 7.** Linear Yield Model Parameter Evaluation

Parameter	B	Standard error	t	Significance	95% confidence interval	
					Lower bound	Upper bound
Crop yield, dt/ha						
Free term	47,300	0,312	151,715	0,000	46,605	47,995
[shift=0]	7,600	0,441	17,237	0,000	6,618	8,582
[shift=10]	10,900	0,441	24,722	0,000	9,918	11,882
[shift=20]	9,000	0,441	20,412	0,000	8,018	9,982
[shift=30]	6,700	0,441	15,196	0,000	5,718	7,682
[shift=40]	0	.	.	.	.	.

Table 7 shows Least Squares (LS) estimates of model parameters (1). In it, the constant is  $\mu_{00}=47,3$  dt/ha, and the effect of option 5 is taken as zero, that is, the yield of wheat grain with the sowing date shift by 40 days is 47.3 dt/ha. The effects (yield addition) of experimental variants are counted already from this level; thus, yield according to the experiment version 2 (sowing date shift by 10 days) is characterized by the addition of  $\alpha_2=10,9$ dt/ha, with a 95% confidence interval (9.9... 11.9) dt/ha, which corresponds to yield (57.5... 58.9) dt/ha.

The high quality of the linear yield model reflects the  $p$ -level values of its parameters proximity to zero (at least 0.0005) and the proximity to the unit of the determination coefficient is 0.986. The last one means that the single-factor linear yield model explains 98.6% of the total variance, and only 1.4% of the variance is due to other variability factors.

Now let's appeal to the models of wheat grain quality indicators.

One of the most important indicators determining the technological advantages of bread-making wheat is the indicator of the quantity and quality of gluten. In experiments on the effect of the sowing date shift on grain quality, the following indicators were recorded: grain-unit, protein and gluten content, sedimentation. For all these indicators, linear models turned

out to be adequate - the determination coefficients show high values from  $R^2 = 0,951$  (grain-unit) to  $R^2 = 0,968$  (protein content and sedimentation) and  $R^2 = 0,980$  (gluten content).

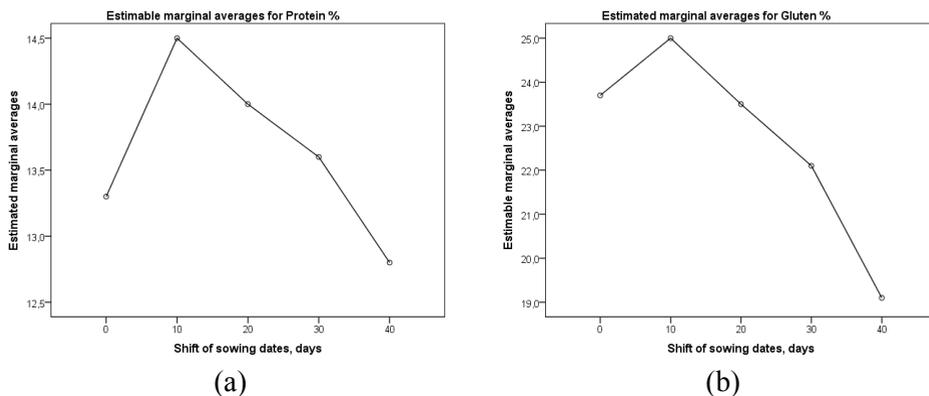
LS-estimates of wheat grain quality indicators model parameters are given in Table 8. In this table, the free member of the models corresponds to the 40-day shift of the seeding time relative to the control, i.e. the effects are counted from this level.

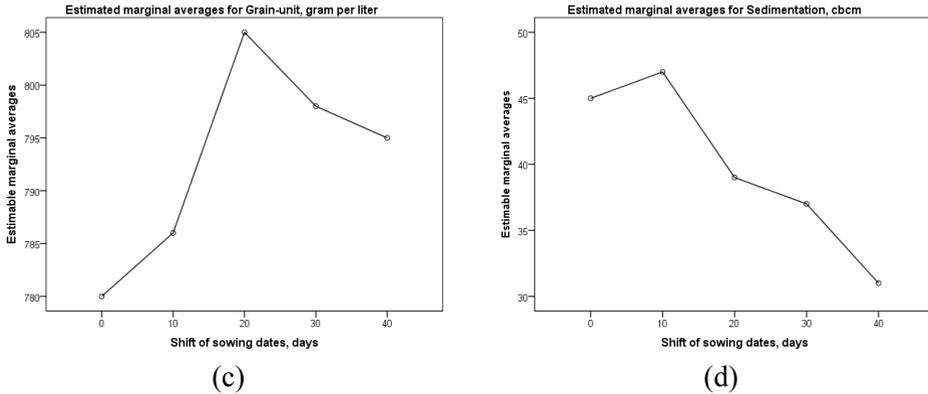
**Table 8.** Estimates of parameters of linear models of wheat grain quality indicators

Parameter	B	Standard error	t	Significance	95% confidence interval	
					Lower bound	Upper bound
Protein content, %						
Free term	12,800	0,075	170,540	0,000	12,633	12,967
[shift =0]	0,500	0,106	4,711	0,001	0,263	0,737
[shift =10]	1,700	0,106	16,016	0,000	1,463	1,937
[shift =20]	1,200	0,106	11,305	0,000	0,963	1,437
[shift =30]	0,800	0,106	7,537	0,000	0,563	1,037
Gluten content, %						
Free term	19,100	0,202	94,520	0,000	18,650	19,550
[shift =0]	4,600	0,286	16,097	0,000	3,963	5,237
[shift =10]	5,900	0,286	20,646	0,000	5,263	6,537
[shift =20]	4,400	0,286	15,397	0,000	3,763	5,037
[shift =30]	3,000	0,286	10,498	0,000	2,363	3,637
Grain-unit, gram per liter						
Free term	795,000	1,426	557,482	0,000	791,823	798,177
[shift =0]	-15,000	2,017	-7,438	0,000	-19,494	-10,506
[shift =10]	-9,000	2,017	-4,463	0,001	-13,494	-4,506
[shift =20]	10,000	2,017	4,958	0,001	5,506	14,494
[shift =30]	3,000	2,017	1,488	0,168	-1,494	7,494
Sedimentation, cbcm.						
Free term	31,000	0,733	42,278	0,000	29,366	32,634
[shift =0]	14,000	1,037	13,501	0,000	11,690	16,310
[shift =10]	16,000	1,037	15,430	0,000	13,690	18,310
[shift =20]	8,000	1,037	7,715	0,000	5,690	10,310
[shift =30]	6,000	1,037	5,786	0,000	3,690	8,310

The last two graphs show the 95% confidence intervals of the parameters. Almost all of them do not include zero, which reflects the statistical significance of the observed effects.

The effect of the sowing time shift on wheat grain quality can be observed on the diagrams of Figure 1 generated in the program environment of the "linear generalized model" procedure.



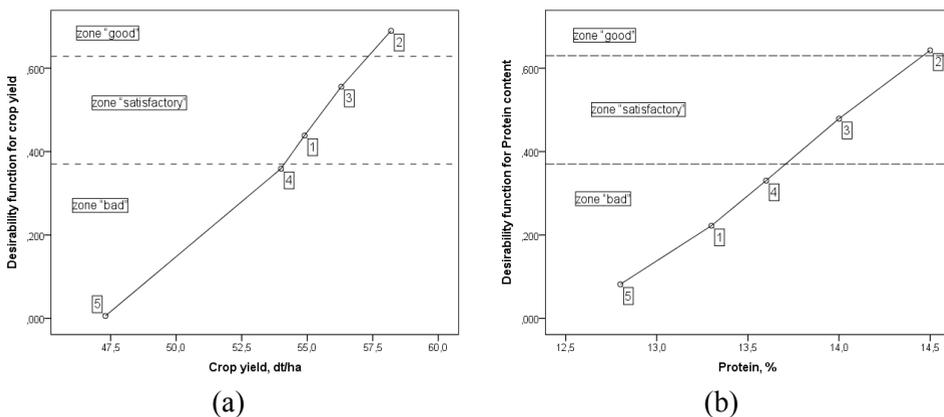


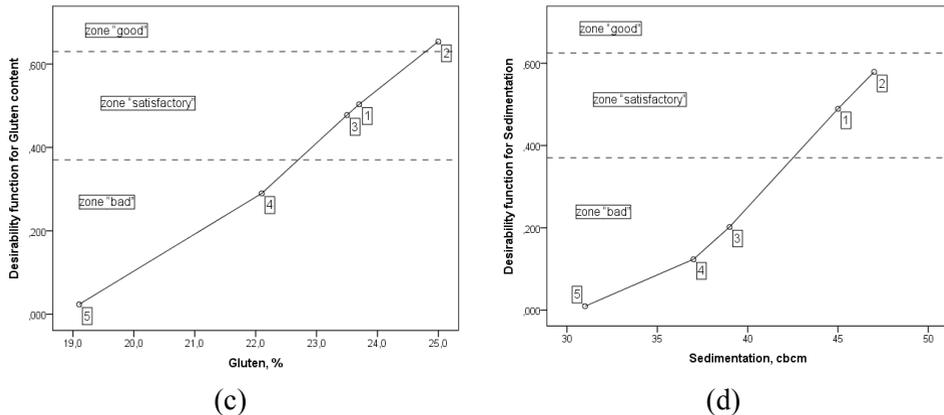
**Fig. 1.** Dependence of wheat grain quality indices on sowing time shift: (a) - protein content; (b) - gluten content; (c)–grain-unit; (d) - sedimentation

These diagrams show that the maxima of all quality indicators, except grain-unit, fall on option 2 with the sowing period shift relative to control by 10 days. Taking into consideration that the maximum yield also falls on this option, the wheat sowing date shift by 10 days compared to the previously recommended one can be considered optimal.

Analysing the results of our correlation analysis, the yield is positively and strongly correlates with wheat grain quality indicators. In addition, the quality indicators are positively and strongly correlated with each other, which is of clear interest to combine them into a kind of generalized quality indicator. It is recommended to use the Harrington desirability function normalized to the interval (0; 1) [8] as such an indicator as it was implemented in our previous work [1]. In this work, Harrington's generalized desirability function was determined from the array of 46 varieties, and then used to assess the quality of Moskovskaya 39 wheat grain depending on the sowing date; at the same time, the assessment "well" and "satisfactorily" corresponded to the sowing date shift by 10-20 days.

We built Harrington's partial desirability functions for all three quality indicators. In addition, we also built a desirability function for yield; according to the method described in [13], the lower reference points for their construction were the average values of the parameters for the sample, the upper ones were apart from them by the values of standard deviations. With this selection of reference points, all functions are of the nature of S-shaped curves - Figure 2.



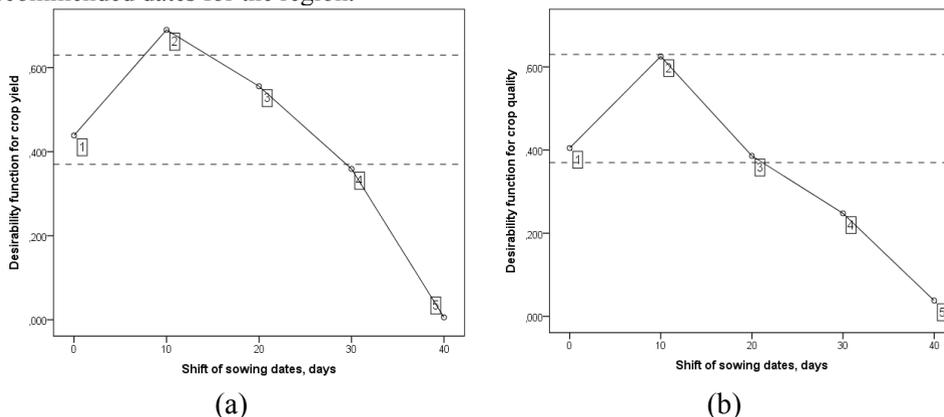


**Fig. 2.** Distribution of experiment options by sowing dates (numbers close to marks) by desirability zones: (a) – yield; (b) – protein content; (c) - gluten content; (d) - sedimentation. The "Satisfactory" zone is highlighted with dashed lines

It follows from Figure 2 that if in terms of yield, gluten content and sedimentation, the results of the control test are evaluated as "satisfactory," then in terms of protein content - as "bad." While shifting the sowing date by 10 days "transfer" estimates of yield and protein content, gluten into a "good" zone.

The small sample size (five options of the experiments) does not allow correct factorial analysis, however, the use of the main components method revealed the tendency that three particular functions of Harrington's desirability - protein and gluten content as well as sedimentation - can be combined into one main factor explaining more than 80% of the total dispersion. This makes it possible to determine the generalized wheat grain quality index as the arithmetic mean of the particular Harrington desirability functions.

Figure 3 shows the distribution of experience variants by crop desirability zones and a generalized grain quality indicator, depending on the sowing date shift. It can be seen that the maxima of both correlations fall on the sowing dates shift by 10 days later than the recommended dates for the region.



**Fig. 3.** Distribution of experiment options (numbers close to marks) by desirability zones: (a) – yield; (b) – generalized grain quality index. The "Satisfactory" zone is highlighted with dashed lines

## 4 Conclusion

1. Based on the results of the statistical studies performed, the procedure of a multidimensional generalized linear model is recommended for modeling the effect of agricultural conditions on the yield and quality indicators of wheat grains.

2. Multiple comparison of averages, based on the methods of dispersion analysis, increases the reliability of statistical conclusions, allows to distinguish homogeneous versions of experiments in agricultural production.

3. It is statistically justified that the maxima of yield and quality of wheat grains provide a sowing period of 10 days later than previously recommended for the region.

## References

1. A.F. Melnik, V.G. Shumetov, B.S. Kondrashin, M.R. Mikhaylov, IOP Conference Series: Earth and Environmental Science, 422 (2008)
2. B.A. Dospekhov, *Methods of Field Experiment*, 351 (Agropromizdat, 1985)
3. S. Glanz, *Medico-Biological Statistics*, 459 (Praktika, 1998)
4. Lyubishev, *Dispersion Analysis in Biology*, 6 (Moscow State University Press, 1986)
5. V.G. Shumetov, A.S. Kolomeichenko, V.S. Buyarov, S.Yu. Metasova, *Bulletin of Agrarian Science*, **4(67)**, 113 (2017)
6. A. F. Melnik, *Bulletin of Agrarian Science*, **1(82)**, 41 (2020)
7. V.G. Shumetov, E.V. Barbashova, *Innovative development of the Russian economy vol .3: Statistical and tool researches of development*, 321 (2017)
8. E.C. Harrington, *Industrial Quality Control*, **21(10)**, 494