Biometric methods in studying the gene pool Indian pea

S.A. Zaitsev[1]*, A.Yu.* Lyovkina[1]*, D.P.* Volkov[1]*, O.B.* Kameneva[1]*, V.V.* Bychkova[1], and *L.A.* Gudova[2]

[1]Russian Research Institute for Sorghum and Maize «Rossorgo», 410050, Saratov, st. 1st Institutskiy proezd, 4, Russia
[2]Federal State Budgetary Scientific Organization «Federal Center of Agriculture Research of the South-East Region», 410010, Saratov, Tulaikov Street, 7, Russia

Abstract. The article presents the results of using biometric methods for assessing the gene pool (using the example of the indian pea). The variation of traits of the model population of indian pea, as well as the correlation coefficients and factor loadings on the variables, is considered. The model population includes 25 samples of different ecological and geographical origin. The nursery was founded according to the scheme of a one-factor complex. Statistical indicators of the sample, coefficients of variation of signs were revealed: >20% - fat content in seeds, fat gather, fiber gather, 10-20% - height of attachment of the lower bean, number of twigs of the 1st order, number of beans per plant, number of seeds in a bean , number of seeds per plant, seed yield, pod width, gather of protein, ash, BEV per unit area, <10% - stem length, weight of 1000 seeds, plant productivity, pod length, content of protein, fiber, ash, BEV in seeds analysis of the experimental data, a matrix of correlation coefficients was calculated, including indicators of vegetative traits and the content of nutrients in the grain. In order to optimize the interpretation of the matrix of correlation coefficients, a factor analysis was performed using the method of principal components. Signs that determine the first, second and third hypothetical factors (stem length, lower bean attachment height, content of protein, fat, fiber, BEV, ash, number of seeds in a bean, weight of 1000 seeds, plant productivity, seed yield, gather per hectare of protein, fiber, ash, BEV, the number of seeds from one plant) determine 63,72% of the accumulated dispersion.

1 Introduction

In agricultural production, located in arid and semi-arid regions, the crop of the indian pea (*Lathyrus sativus L.*), cultivated as a food, fodder and green manure plant, has gained a certain distribution [1]. A high amount of protein and fiber was found in green plants, hay, and seeds, which can be used as a source of additional protein in the diet of humans and animals [2, 3]. Tolerance to soil salinity, a small amount of moisture, low damage by pests and diseases, together with a fairly high yield of seeds and green mass, determine the economic importance of the indian pea. The world market demand for this crop and the potential for its use open up wide opportunities for the resumption of indian pea production

[4]. However, the State Register of Breeding Achievements approved for use in the Russian Federation includes a limited number of varieties [5]. Therefore, studies related to the selection of new varieties and the identification of opportunities to make the most of the potential of plants and growing conditions are relevant.

Factor analysis allows the researcher to determine the significant values of the parameters from the insignificant ones, determine the information content of the data, the relationship, and quantitatively characterize the diagnostic value of the trait. Analysis of breeding material using factor analysis can be divided into stages: compilation of a data matrix, calculation of a matrix of correlation coefficients, calculation and evaluation of factor loads, biological interpretation of common factors [6, 7]. At the same time, these methods of component analysis are used in genetic-breeding studies of various crops: rye [8], corn [9, 10], peas [11], sorghum [12], wheat [13, 14],

To analyze the experiment using a linear model of factor analysis allows the presence of certain conditions: changes in variables are determined by the influence of general and specific factors; the studied sample obeys the multidimensional normal distribution law; specific factors are not correlated either with each other or with general factors; the number of common factors determining the phenomenon under study should be significantly less than the number of analyzed features; the correlation matrix of the observed features should be stable with respect to the change in the sample; there is no autocorrelation in the original observations; the initial data sample should be representative [15, 16, 17, 18].

When analyzing the experimental material, it is taken into account that the contribution of each factor to the total variance is more than 5%, the factor load is considered significant if its absolute value exceeds 0,4. The selection of features that determine each hypothetical factor is also carried out according to the absolute value of the correlation coefficients with the factor. Sometimes the following value is taken as the lower limit of tolerance in the group of these signs: r = 0,8x r max. [19, 20].

2 Methods

Field studies were carried out at the experimental site of the Russian Research Institute for Sorghum and Maize «Rossorgo» in 2015-2018. The hydrothermal coefficient in the period may-september was: 2015 – 0,9; 2016 – 0,48; 2017 – 1,05; 2018 – 0,69. The model population includes 25 samples of the working collection of different ecological and geographical origin: Racheika, Mramornya, Zhemchuzhina, k-7 Sanchinak, k-12, k-17, k-21, k-25, k-30 Gesse Cultive , k-34, k-240, k-278, k-292, k-403, k-703, k-748, k-780 Almortas Flamenca, k-805, k-809, k-834, k-850 , k-924, k-1116, k-1170. The repetition of the experiment is three times. The accounting area of the plot is 10 m^2. Agrotechnics in the experiment - zonal, developed at the Russian Research Institute for Sorghum and Maize «Rossorgo». The study was based on the method of state variety testing of agricultural crops [21]. The content of protein (GOST 10846-91), fat (GOST 13496.15-97), fiber (GOST 13496.2-91), and ash (GOST 26226-95) was determined in seeds and green mass. Mathematical data processing was carried out in accordance with the guidelines [22]. Statistical parameters were used: mean value (x), mean error (sx), variance (s2), standard deviation (s), coefficient of variation (V), minimum feature value (min), and maximum feature value (max).

Interpretation of the results of factor analysis was based on the following principles: highly interconnected indicators are grouped within one group; indicators combined within one group and showing a weak relationship, as well as those included in another factor and insignificant correlations, were not included in the analysis. The analysis includes hypothetical factors with a load of more than 5%. The selection of features that determine

each hypothetical factor was carried out according to the absolute value of the correlation coefficients with the factor.

3 Results

Analysis of the elements of the yield structure of the model population of the indian pea made it possible to identify the statistical indicators of the sample (table 1). Highly variable traits (seed fat content, fat gather, fiber gather) make up 14% in the database. However, the revealed high coefficient of variation (26,6%) of the fat content in the seeds, due to low absolute indicators (0,4-1,3%), does not allow us to focus on the clustering of varieties for this trait. The following slightly varying 8 traits were noted (V less than 10%): stem length, seed size, plant productivity, pod length, content of protein, fiber, ash, BEV in seeds. Greater diversity (V to 20%) between genotypes was noted according to the following traits: lower bean attachment height, branching, number of beans per plant, number of seeds per bean, number of seeds per plant, seed yield, bean width, yield per hectare of protein , ash, BEV.

Table 1. Characteristics of the variability of economically valuable traits of the model population of the Indian pea, average 2015-2018.

Feature	Parameter *					Lim	
	x	s^2	s	V, %	sx	min	max
Stem length, sm	59,6	29,8	5,5	6,7	0,8	50,1	78,8
Lower bean attachment height, sm	16,8	4,6	2,1	13,1	0,4	13,6	23,2
Number of branches, pcs.	7,7	1,7	1,3	17,4	0,3	5,9	10,3
Number of beans per plant**,** pcs.	30,9	20,9	4,6	15,1	0,9	24,1	42,7
Number of seeds in bean, pcs.	1,5	0,02	0,2	10,0	0,03	1,3	1,8
Number seeds per plant, pcs	48,8	54,4	7,4	15,4	1,5	33,8	61,0
Weight of 1,000 seeds, g	218,0	395,6	19,9	9,3	4,1	176,8	266,7
Seeds weight from 1 plant, g	10,5	4,0	2,0	0,4	19,6	7,0	15,7
Seeds yield, t/ha	1,56	0,08	0,3	18,1	0,06	1,05	2,11
bean length, mm	37,2	5,7	2,4	4,0	0,3	34,8	40,8
bean width, mm	13,0	2,5	1,6	12,4	0,3	9,3	15,8
Content in seeds, %							
protein	29,3	1,2	1,1	3,8	0,23	27,1	31,2
fat	0,8	0,05	0,21	26,6	0,04	0,4	1,3
fiber	6,2	0,32	0,57	9,4	0,12	4,9	7,2
ash	3,5	0,01	0,12	3,5	0,02	3,3	3,8
BEV	60,5	3,2	1,8	2,1	0,25	58,0	62,5
Yield per hectare, kg							
protein	395,6	5289,7	72,7	18,8	14,8	249,0	521,7
fat	10,5	8,0	2,8	27,5	0,58	4,6	16,0
fiber	82,2	298,8	17,3	21,5	3,5	44,7	121,1
ash	46,9	64,9	8,1	17,5	1,6	32,0	62,3
BEV	810,9	19948,1	141,2	17,8	28,8	562,9	1118,1

Note: * x - average value, sx - average error, s^2 - variance, s - standard deviation, V - coefficient of variation, %, min - minimum value, max - maximum value.

4 Discussion

When analyzing the database of the experiment, a matrix of correlation coefficients (210 coefficients) was calculated, including 21 indicators of vegetative traits and nutrient content in the seeds of 25 samples of the indian pea (table 2). At the same time, 42 correlation coefficients were found to be significant at the 5% level. Correlation coefficients (>0,59) of seed yield with the following traits were determined: the number of beans per plant, the number of seeds per plant, the weight of 1000 seeds, plant productivity, yield per hectare: protein, fiber, ash, BEV. The length of the stem significantly correlates with the length of the stem from the soil to the point of attachment of the lower bean. No significant correlation was found between the number of seeds in a pod and the studied economically valuable traits of the model population of the indian pea.

Table 2. Matrix of correlation coefficients of the model population of the indian pea, average 2015 – 2018.

			Feature ***				
	1	2	3	4	5	6	7
1	1,00						
2	0,46*	1,00					
3	-0,07	0,15	1,00				
4	0,36	-0,21	0,05	1,00			
5	0,09	-0,12	0,05	-0,25	1,00		
6	0,39	-0,34	0,05	0,82**	0,31	1,00	
7	-0,02	-0,16	0,23	0,27	-0,18	0,16	1,00
8	0,29	-0,34	0,15	0,77**	0,16	0,86**	0,64**
9	0,31	-0,32	0,17	0,75**	0,23	0,88**	0,59**
10	0,23	0,29	0,42*	0,22	-0,11	0,13	-0,23
11	-0,08	-0,14	0,41*	0,31	-0,35	0,13	0,27
12	0,34	0,16	-0,14	0,18	0,20	0,23	-0,15
13	0,07	0,06	-0,04	-0,13	-0,35	-0,27	-0,15
14	0,06	0,25	0,31	0,17	-0,09	0,12	0,13
15	-0,03	0,21	-0,17	-0,24	-0,11	-0,33	-0,18
16	-0,25	0,06	-0,03	-0,23	-0,20	-0,33	0,06
17	0,37	-0,27	0,15	0,74**	0,26	0,88**	0,54**
18	0,33	-0,11	0,11	0,38	-0,15	0,32	0,23
19	0,29	-0,16	0,28	0,72**	0,14	0,80**	0,56**
20	0,31	-0,30	0,12	0,73**	0,22	0,84**	0,58**
21	0,27	-0,33	0,17	0,73**	0,21	0,85**	0,62**

Table 2 continuation

			Feature ***				
	8	9	10	11	12	13	14
8	1,00						
9	0,99**	1,00					
10	-0,01	0,01	1,00				
11	0,22	0,20	0,47*	1,00			
12	0,09	0,14	0,09	-0,22	1,00		
13	-0,31	-0,31	0,29	0,30	0,13	1,00	
14	0,15	0,13	-0,04	-0,02	-0,15	-0,15	1,00
15	-0,34	-0,31	-0,21	-0,26	0,01	0,05	-0,13
16	-0,21	-0,24	-0,08	0,05	-0,67**	-0,14	0,09
17	0,96**	0,98**	0,03	0,15	0,32	-0,27	0,08
18	0,35	0,36	0,33	0,43*	0,28	0,73**	-0,01
19	0,91**	0,90**	0,01	0,16	0,01	-0,31	0,53**
20	0,96**	0,98**	-0,05	0,13	0,16	-0,33	0,09
21	0,99**	0,99**	-0,00	0,22	0,04	-0,34	0,14

Table 2 continuation

	15	16	17	28	19	20	21
	Feature ***						
15	1,00						
16	-0,23	1,00					
17	-0,30	-0,35	1,00				
18	-0,23	-0,30	0,40	1,00			
19	-0,29	-0,16	0,86**	0,30	1,00		
20	-0,12	-0,28	0,97**	0,32	0,87**	1,00	
21	-0,35	-0,11	0,96**	0,32	0,90**	0,96**	1,00

Note: * Significant at the 5% level, ** Significant at the 1% level
***1. Stem length, 2. Lower bean attachment height, 3. Number of branches, 4. Number of beans per plant, 5. Number of seeds in bean, 6. Number seeds per plant, 7. Weight of 1,000 seeds, 8. Seeds weight from 1 plant, 9. Seeds yield, 10. bean length, 11. bean width, 12. Content of protein, 13. Content of fat, 14. Content of fiber, 15. Content of ash, 16. Content of BEV, 17. gather of protein, 18. gather of fat, 19. gather of fiber, 20. gather of ash, 21. gather of BEV.

The use of multivariate statistics methods helps researchers more objectively evaluate the source material by a set of indicators. In order to optimize the interpretation of the matrix of correlation coefficients, a factor analysis was performed using the method of principal components (table 3). When calculating the influence of features on the components, it was found that the largest contribution (dispersion – 39,28%) to the first hypothetical factor is made by the following parameters: the height of attachment of the lower bean, fat content, BEV, ash. In the second factor (dispersion – 12,88%), the largest contribution is made by: the number of seeds in bean, the weight of 1000 seeds, the productivity of the plant, the yield of seeds, the content of fiber, ash, BEV, the gather per hectare of fiber, ash, BEV. The effect of such traits as: stem length, stem length from the soil to the point of attachment of the first bean, the number of seeds in the bean, the number of seeds per plant, content of protein, of ash, gather of protein, of ash to the greatest extent determine the dispersion of the third factor. The fourth hypothetical factor is formed by the contribution of the stem length, height of attachment of the lower bean, plant branching, number of seeds in the bean, bean length, fiber content in seeds, BEV, gather of fiber per hectare. The effects of stem length, lower bean attachment height, number of beans per plant, weight of 1000 seeds, fat content, fiber content form the variance of the fifth factor. The sixth hypothetical factor consists of the total effects of the studied traits: stem length, lower bean attachment height, number of beans per plant, number of seeds per plant, BEV content. Subsequent calculated factors add less than 5% to the accumulated variance and are therefore excluded from the discussion of their effects.

Table 3. Factor loadings of variables on the main components of the model population of the indian pea.

Feature	**Factor**					
	Z1	**Z2**	**Z3**	**Z4**	**Z5**	**Z6**
Stem length, sm	-0,35	-0,34	0,45	0,41	0,21	0,33
Lower bean attachment height, sm	0,30	-0,28	0,15	0,76	0,25	0,03
Number of branches, pcs.	-0,20	-0,24	-0,43	0,44	-0,35	-0,54
Number of beans per plant, pcs.	-0,81	-0,21	-0,02	-0,01	0,16	0,30
Number of seeds in bean, pcs.	-0,18	0,38	0,45	0,11	-0,61	-0,18
Number seeds per plant, pcs	-0,90	0,01	0,20	0,00	-0,19	0,21
Weight of 1,000 seeds, g	-0,56	0,13	-0,42	-0,12	0,38	-0,33
Seeds weight from 1 plant, g	-0,99	0,09	-0,06	-0,06	0,04	0,01
Seeds yield, t/ha	-0,99	0,09	0,01	-0,04	0,00	-0,02
bean length, mm	-0,07	-0,71	-0,11	0,29	-0,42	0,12

bean width, mm	-0,25	-0,55	-0,56	-0,17	-0,15	-0,08
Content in seeds, %						
protein	-0,18	-0,30	0,76	0,00	-0,03	-0,15
fat	0,26	-0,79	-0,01	-0,33	0,16	-0,05
fiber	-0,19	0,07	-0,29	0,63	0,23	-0,09
ash	0,35	0,10	0,32	0,01	0,48	-0,37
BEV	0,28	0,29	-0,66	0,09	0,01	0,46
Yield per hectare, kg						
protein	-0,98	0,02	0,14	-0,03	-0,02	-0,05
fat	-0,42	-0,75	0,02	-0,28	0,10	-0,06
fiber	-0,92	0,10	-0,12	0,24	0,11	-0,06
ash	-0,96	0,13	0,08	-0,05	0,10	-0,08
BEV	-0,98	0,14	-0,09	-0,04	-0,01	0,03
Dispersion	8,25	2,70	2,43	1,75	1,35	1,09
Dispersion, %	39,28	12,88	11,56	8,33	6,43	5,21
Accumulated variance, %	39,28	52,15	63,72	72,05	78,474	83,680

The processes of scientific research are based on the classification of the objects under study. When studying the issue of classification of breeding material, the use of cluster analysis is of particular interest. Unlike other methods of data analysis, this method allows you to group objects not by one parameter, but immediately by a set of features, that is, to collect complete information about the sample and arrange the objects into relatively homogeneous groups. When viewing the values of the fusion coefficient by the minimum of the Euclidean distances of 25 samples of the model population of the indian pea, a significant jump was identified, which leads to the conclusion that it is admissible to group the samples into 8 clusters at the 17th iteration step (the euclidean distance is 26,42) (figure 1). The samples that make up the clusters are characterized by a similar set of features within the clusters (table 4). As a result of this approach to the experimental material, the average value of features for clusters was determined and the parameters by which clusters differ were identified. Clusters are represented by the following forms: 1st: k-12 Stepnaya 12, k-30, k-7, k-25, Zhemchuzhina, k-924, Racheika, k-703, k-809; 2nd: k-240 Stepnaya 240, k-927, k-748; 3rd: Mramornya; 4th: k-21 Stepnaya 21; 5th: k-834, k-850, k-17, k-248, k-403, k-292, k-1116; 6th: k-34, k-780; 7th: k-1170; 8th: k-805.
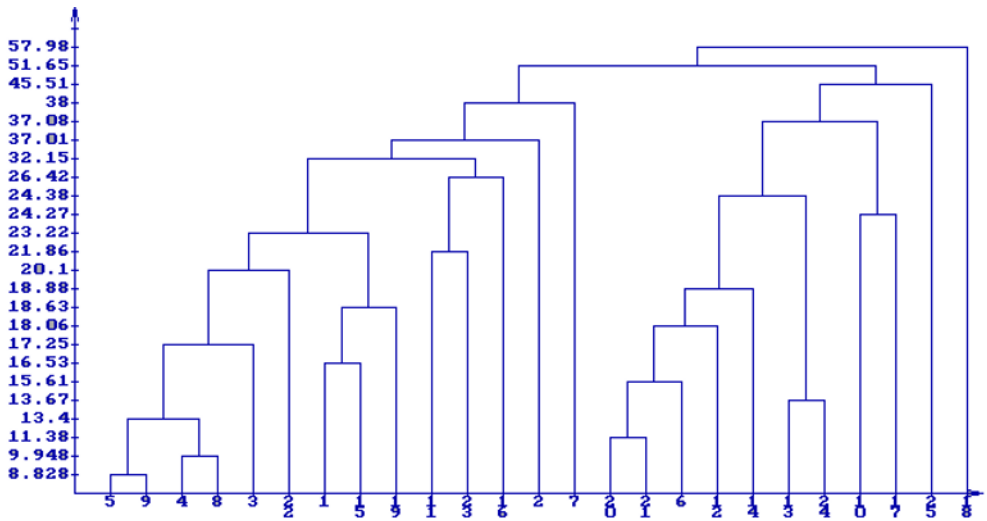
Fig. 1. Dendrogram of similarities-differences of samples of the indian pea according to a set of features, 2015–2018.

The first cluster includes 9 samples and is characterized by significant differences in the height of attachment of the lower bean - from clusters 3, 7, 8; according to the number of seeds per plant from clusters 4, 5, 6, 7; by plant productivity and seed yield, protein the gather - from clusters 4, 5, 6, 7, 8; according to the content of BEV - from clusters 4, 5, 6, 8. Samples of the second cluster occupy an intermediate position between clusters 1 and 3 in terms of plant productivity, content and the gather of BEV.

The seventh cluster is represented by the longest stemmed form with a high attachment of the lower bean.

There were no differences between clusters in plant branching, number of seeds in a bean, weight of 1000 seeds, content of protein, fiber, and ash.

Table 4. Average values of economically valuable traits by clusters, 2015–2018.

Klusters	Feature *								
	1	2	3	4	5	6	7	8	9
1	52,4a	15,4b	26,3a	41,1bc	9,3b	2,18b	363,4b	59,7ab	730,2bc
2	55,6a	16,9b	26,3a	42,3b	8,5ab	1,30b	313,8b	61,7bc	682,3ab
3	58,9a	20,2c	28,0a	39,4b	8,4ab	1,26ab	312,8b	67,2d	719,4bc
4	55,7a	17,5bc	25,1a	33,8a	7,0a	1,05a	249,0a	62,5c	562,9a
5	61,9a	16,2ab	33,6bc	56,6ef	11,9de	1,80de	456,3e	60,2ab	923,9de
6	59,2a	14,7ab	41,3d	58,5f	14,6f	2,10f	513,8f	60,2ab	1069,4f
7	78,8b	23,2d	35,2c	52,1de	11,1cde	1,66cde	43,9cde	58,6a	834,9cde
8	55,3a	13,6a	27,9a	48,9cd	12,0e	1,79e	456,0de	60,8bc	930,8e
F	5,25*	4,61*	12,65*	18,02*	15,45*	16,83*	18,69*	8,6*	12,23*
HCP$_{0.05}$	6,72	2,76	4,02	5,57	1,62	0,22	54,0	1,85	125,9

Note: *1. Stem length, 2. Lower bean attachment height, 3. Number of beans per plant, 4. Number seeds per plant, 5. Seeds weight from 1 plant, 6. Seeds yield, 7. Yield of protein per hectare, 8. Content of BEV in seeds, 9. Yield of BEV per hectare.
** values followed by the same letter do not differ significantly by Duncan's test

5 Conclusion

The revealed coefficients of variation indicate significant differences between the studied forms in terms of seed yield and nutrient collection, which indicates the possibility of using the studied material in breeding to increase these parameters. The traits that determine the first, second and third hypothetical factors determine 63,72% of the accumulated dispersion, which should certainly be taken into account when breeding new varieties of sowing rank. The use of the clustering method makes it possible to group the breeding material of the rank according to a set of economically useful indicators. Obviously, the breeding nursery should include forms belonging to the first, fifth, sixth clusters as initial material: k-12 Stepnaya 12, k-30, k-7, k-25, Zhemchuzhina, k-924, Racheika, k -703, k-809; 2nd: k-240 Stepnaya 240, k-927, k-748, k-834, k-850, k-17, k-248, k-403, k-292, k-1116, k-34, k-780, differing in the height of attachment of the lower bean, plant productivity and seed yield.

References

1. R.J. Hillocks, M.N. Maruthi, Euphytica, **186** (2012)

2. L.I. Zaytseva, V.I. Zhuzhukin, S.A. Zaytsev, Fodder Production, **11,** 24 (2013)

3. V.I. Zhuzhukin, V.S. Gorbunov, S.A. Zaitsev, D.P. Volkov, Grain Economy of Russia, **6**, 48 (2017)

4. M.A. Vishnyakova, T.G. Aleksandrova, S.V.Bulyntsev, T.V. Buravtseva, M.O. Burlyaeva, G.P. Egorova, E.V. Semenova, I.V. Seferova, I.I. Yankov, Agricultural Biology, **1** (2016)

5. State Register of Breeding Achievements Admitted to Use "Plant Varieties", 1 (2021)

6. S.P. Martynov, Molecular and Applied Genetics, Minsk, **9**, 49 (2009)

7. I.Ya. Liepa, *Mathematical Methods in Biological Research* (Riga, 1980)

8. V.M. Bebyakin, T.B. Kulevatova, N.M. Velikanova, Storage and processing of farm products, **12**, 29 (2004)

9. L.A. Gudova, S.A. Zaitsev, V.I. Zhuzhukin, L.G. Kurasova, A.V. Lekarev, Agricultural Scientific Journal, **7** (2021)

10. G.V. Matveeva, L.Yu. Novikova, V.B. Korneev, Corn and Sorghum, **4**, 25 (2010)

11. R.V.Belyaeva, S.N.Agarkova, Legumes and groat crops, **3**, 11 (2018)

12. V.I. Zhuzhukin, A.Yu. Garshin, D.S. Semin, Agroecology, **1**, 12 (2015)

13. V.M. Bebyakin, G.V. Piskunlva, V.A. Matveeva, Grain Economy, **8**, 17 (2003)

14. V.G. Zakharov, V.V. Syukov, O.D. Yakovleva, Vavilov Journal of Genetics and Breeding, **3** (2014)

15. R.I. Ibyatov, F.Sh. Shaikhutdinov, A.A. Valiev Grain Economy of Russia, **2**, 17 (2017)

16. E.M.Makalonov, *Genetics of quantitative traits of agricultural plants*, (Moscow, 1978)

17. S.P. Martynov, T.V. Dobrotvorskaya, Genetic, **8** (1996)

18. V.M. Zhukovskaya, I.B. Muchnik, Statistic (Moscow, 1976)

19. A.M. Dubrov, Moscow, Statistic (1978)

20. V.I. Zhuzhukin, S.A. Zaitsev, D.P. Volkov, Agrarian science, **6**, 48 (2018)

21. Methodology of state variety testing of agricultural crops. Release the first. General part, Moscow (2019)

22. B.A. Dospehov, *Field experience methodology*, (Moscow, 2012)