UDC 636.4.033
IRSTI 68.39.19

DOI 10.52578/2305-9397-2025-2-10-19

Muratov A. A., researcher, the main author, https://orcid.org/0000-0003-4008-1100,
LLP Kazakh Research Institute for Livestock and Fodder Production, st. Zhandosova 51, Almaty, Republic of Kazakhstan, adilmuratov97@gmal.com
Daugaliyeva A.T., Candidate of Veterinary Sciences, Associate Professor, https://orcid.org/0000-0002-7703-7798
LLP Kazakh Research Institute for Livestock and Fodder Production, st. Zhandosova 51, Almaty, Republic of Kazakhstan, aida1979@bk.ru
Daugaliyeva S.T., https://orcid.org/0000-0002-8826-3942
LLP Scientific Production Center of Microbiology and Virology ‘, Bogenbai Batyr str., 105, Almaty, Republic of Kazakhstan, saule.daugalieva@mail.ru

THE MAIN PROBLEMS OF USING BLUP FOR ASSESSING PIG NUMBERS IN THE REPUBLIC OF KAZAKHSTAN

ANNOTATION

This paper initiates the implementation of the BLUP methodology for assessing the breeding value of pigs in the conditions of domestic pig farming in Kazakhstan. The initial data analysis revealed several critical problems that hinder the correct implementation of statistical calculations. The main issues are the high percentage of missing values and abnormal data in key parameters, such as litter weight, number of piglets, and measurement dates. These shortcomings lead to the impossibility of using a significant part of the records to calculate the complex index using the BLUP method. The paper proposes a comprehensive interdisciplinary approach, including identifying influencing factors based on international experience and regional characteristics, improving the data collection and control system in the information and analytical system (IAS), and developing specialized software. The study results demonstrate that even with correct management of identification data, significant problems remain that require further refinement of the algorithms for cleaning and quality control of the initial data, which must be considered in the subsequent practical implementation of the methodology. Applying the developed method will improve the accuracy of assessing the breeding value of pigs and, in the long term, improve the efficiency of breeding work in the industry.

Key words: *BLUP, breeding value, pig breeding, data analysis, information and analytical system, genetic selection, statistical analysis, data evaluation.*

Introduction. The development of domestic pig farming today requires the introduction of modern methods for accurately assessing the breeding value of animals. The BLUP (Best Linear Unbiased Prediction) method has long established itself in foreign practice as an effective tool for predicting genetic value [1 - 4]. However, its application in Kazakhstan faces several specific problems. One of the key problems is the insufficient completeness and quality of the initial data used to calculate complex breeding indices. Analysis of the provided data showed that 76% to 100% of the records in the central columns, such as nest weight, measurement dates, and number of piglets, are missing or contain abnormal values, significantly complicating the implementation of correct statistical calculations.

The study's relevance is due to the need to adapt international experience to domestic pig farming conditions. This paper proposes an interdisciplinary approach, which includes:

1. Determination of influencing factors considering international methodological approaches and the specifics of regional conditions;

2. Improving the data collection and processing system in the information and analytical system, which requires the development of new software modules for input data control;

The implementation of the proposed measures will not only improve the accuracy of estimates but also create a basis for further software development. The next stage of the study involves integrating genetic information to account for family ties, which will be the next step in improving the assessment methodology. Thus, this study is an essential step toward modernizing the selection system in domestic pig breeding.

This article thoroughly examines the problems associated with missing values and abnormal data in key columns, emphasizing the importance of identifying their cause and understanding potential ways to improve data quality.

Materials and methods of research. The BLUP methodology [5-8] has not yet been applied in domestic pig breeding, so it is impossible to use the Kazakh explicitly works to lay out the working environment. On the other hand, information obtained directly from the farm indicates several traits that are currently the most accessible in the country in terms of their measurement and control methods.

It can be assumed that these traits are the minimum data set in all BLUP variants. However, traditional selection, even for a minimum set of traits, is far from evident since these traits have different biological origins.

The only way to quantitatively evaluate an animal is to assess the BLUP since this method allows us to isolate the genetic contribution to traits' productivity from the totality of other factors' influences [9-10].

To develop a working environment (filling the main table of initial data) for analyzing domestic pig populations in BLUP, information on breeding value is collected according to many criteria. Each indicator is considered separately (fertility, milk yield, lifetime fat thickness, average daily gain, feed conversion, etc.). The necessary data collection should be specified considering national and regional pig breeding characteristics.

The first category of available features consists of indicators of reproductive characteristics of animals (Table 1).

Table 1 – Indicators of reproductive characteristics of animals

Live weight of nest at birth; kg	Multiple pregnancies; heads	Milk yield: kg	Nest weight at weaning: kg
The weight of a litter is the sum of the live weight of all piglets. The weight of a piglet at birth varies from 0.8 kg (the average norm for most breeds is 1.2-1.3 kg) to 1.5 kg.	Prolificacy is determined by the number of live piglets born (including weak, stillborn, mummies)	Milk production of sows is conventionally determined by the live weight of suckling piglets of the entire litter at 21 days. The actual milk production of a sow is calculated based on the fact that approximately 3 kg of the mother's milk is spent to obtain 1 kg of piglet weight gain. Therefore, the weight of suckling piglets at 21 days of age is subtracted from their total birth weight, and the difference is multiplied by 3. Currently, weaning is carried out on farms for 21 to 35 days. In connection with this, the program must make adjustments. For example, a piglet at birth weighed 1.2 kg. Weaning was carried out at 28 days; the live weight was 4.2 kg. To determine milk production, 4.2 - 1.2 = 3 kg 3/28 = 0.108 g 21 - 28 = 7 days. 7 x 0.108 = 0.756 g.	The litter weight at weaning on a given day is determined by adjusting the actual litter weight on the 21st to the 45th day after farrowing (depending on the technology used) using correction factors.

This set of features coincides with the studies on using BLUP methods in pig breeding presented by Suslina et al. [11]. Another set of features is formed based on fattening characteristics (Table 2).

Table 2 – Indicators of Fattening Qualities of Animals

Average daily gain; g.	Precocity	Feed conversion	Body length, cm.	Thickness of bacon, mm.
It is determined by dividing the total increase in the animal's live weight during the fattening period (from when it is put on fattening until it is taken off) by the number of days in this period. C/c = V1 – V2/T where V1 is the initial live weight, kg; V2 is the final live weight, kg; T is the days of fattening, days.	Precocity is the age at which the pigs reach 100 kg of weight. This indicator characterizes the growth energy during pig fattening; the growth intensity is assessed in this case. At the moment, the farm re-weighs at the end of each month. The live weight of young animals can range from 92 to 108 kg. In this regard, this indicator must be converted to 100 kg. For example, a piglet was born on 01.01.2024 with a live weight of 1.2 kg, and weaning was carried out at 28 days (29.01.2024) with a live weight of 4.2 kg. Re-weighing was carried out on 01.07.2024, with a live weight of 108 kg. 108 kg - 4.2 = 103.8 kg / 152 days (from weaning to re-weighing, days) = 0.683 kg 100 - 103.8 = 3.8 kg / 0.683 kg = 5.56 days. (This coefficient should be rounded for 6 days.) 152 days (from weaning to re-weighing days) - 6 days = 146 days (pig maturity)	An indicator of the economic efficiency of feeding and animal husbandry technology. Conversion is the ratio of feed consumed per unit of output, for example, per 1 kg of gain. CCF = total amount of feed / total weight gain	The body's length is measured along the midline of the back from the occipital ridge to the root of the tail using a steel measuring tape with a division value of 1 cm.	The thickness of the fat is determined on live pigs using an ultrasound device to examine the meat qualities of pigs, with an error of no more than 1 mm at the level of the 6-7 vertebrae

The livestock analysis was done using the IAS data for 2021 and 2023, broken down by breed. The primary analytical tool was regression and correlation analysis, which identified patterns in processing an extensive array of data.

The next step in developing the working environment consists of calculating the BLUP score for each trait using an animal model that shows its deviation from the average values in the population being assessed. These specific breeding value values are multiplied by the economic weight of the traits. It is determined based on the actual contribution of the trait to overall profitability and its corresponding weight in selection. The result is a general BLUP index [12]. A more systematic approach description is also presented in [13]. In terms of the animal model, the general concept is as follows:

$$y_v = \sum_{i=1}^{m_v} F_{vi},$$

where v is the number of the trait (in pig farming, there are several for one animal),
F_{vi} is the i-th factor influencing the -th trait,
m_v is the number of factors influencing the -th trait.

For each trait in a herd of n animals (i.e., with a fixed v), y_v is a vector whose components are equal to the value of the given trait for individual animals:

$$y_v = [y_{v1}, y_{v2}, \cdots, y_{vn}].$$

Calculations and statistical analysis were performed in the NEL software environment developed by LLP Kazakh Research Institute for Livestock and Fodder Production. This software platform has several releases designed for different types of livestock. Within each release, several versions were developed due to the desire to improve computing and user characteristics and to increase the number of functions performed, which was caused by the expansion of the list of tasks to be solved. A brief description of all releases is given in Table 3. To perform work on this task, a release (NEL.meat.pg) was developed, designed to work with pig livestock.

Table 3 – Configuration of the NEL software platform for calculations on pig livestock

Release Title	Description
NEL	The basic platform contains functions standard to all releases: • coding of plan matrices • high-performance matrix multiplication • compiling levels of influencing factors of livestock • converting the list of livestock to the pedigree format with the allocation of the genealogical hierarchy • identifying pedigree errors • forming an inverted matrix of relationships • solving the MME equation • primary verification of the initial data • statistical processing of the results • forming the final report performing iterations to select the result according to the maximum likelihood and continuous convergence criteria.
NEL.meat.pg	Special functions and BLUP for pig herds

The basic NEL module contains functions common to all applications (herds). The main ones are the functions of coding plan matrices and the multiplication of coded plan matrices. This function significantly saves RAM and computation time compared to direct execution of these operations using linear algebra formulas. The module was developed due to the need to work with extensive data from domestic herds. Problems with memory and computation time became noticeable, starting with herds exceeding 30,000 animals.

General-purpose functions also include constructing an animal's genealogical line in order from older to younger. This is necessary to ensure calculations in BLUP, a function of ranking factors, and a module for solving the MME equation.

All the functions used to ensure synchronization of animal record numbers in the lists:

• input list of animals;
• list with reliable records;
• list of animals admitted to the calculation;
• list of animals ordered from older to younger.

Along with general functions, specialized functions are required to work with animal species. These functions are constantly being developed as new tasks arise. Specialized functions are grouped into releases by animal species. The data analysis below was performed using the NEL.meat.pg software release.

Results and their discussion. As a result of the work, many missing values were identified (Table 4).

Table 4 – Proportion of missing values in key columns

Indicators	Missing values	Percentage of omissions (%)
Litter weight	55 501	76,20
Age, days/live weight, kg	72 839	100
Actual weighing date on day 21	72 839	100
Date of reaching 100 kg	72 839	100
Fat thickness, cm	72 836	99,99
Measurement date	72 836	99,99
Torse length, cm	72 836	99,99
Number of healthy piglets at farrowing	55 501	76,20
Number of piglets in the litter on day 21	70 916	97,36
Total animals	72 839	

The table shows that most key columns are missing between 76% and 100% of the data. This means the information is simply unavailable for many parameters, making their analysis impossible. The situation description is supplemented by the following illustration (Fig. 1).

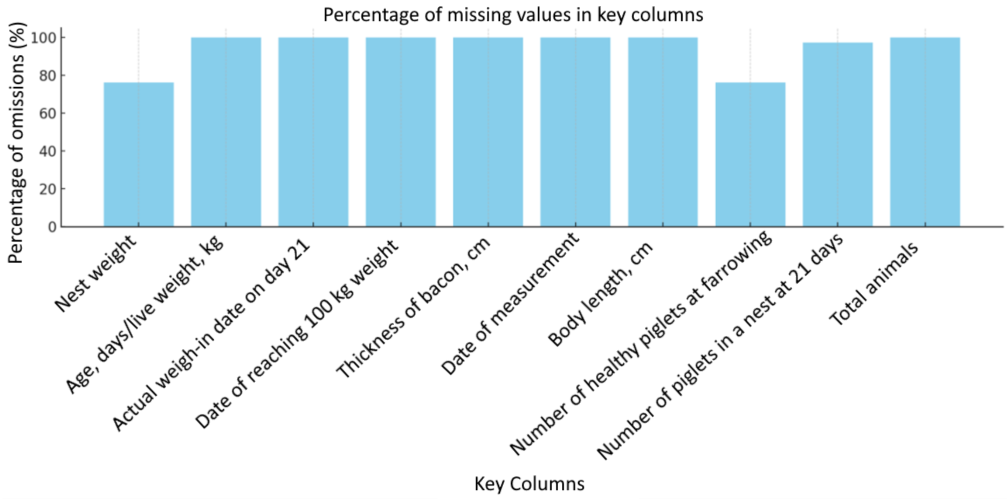

Figure 1 – Percentage of missing records for key positions

The only parameters in which no violations were noted are animal numbers, which are given by the requirements of the Animal Identification Rules in the Republic of Kazakhstan [14].

Also, certain program adjustments were required in cases where data was not missing but recorded in different Excel formats and related to one factor. Without special modifications, such data would have caused the program to stop.

Under these conditions, there were two options for testing the program. The first was to create an artificial population of animals, where a random number generator sets the values of the features and random errors, and an arbitrary decision sets the distributions by farms and other fixed factors. This approach was helpful at the initial stages of program development.

In this situation, another path was possible. Some animals, ~5,000 individuals, had correct records for two features. The program was tested on this subset. The parameters of the test subset of animals are shown in Figure 2.

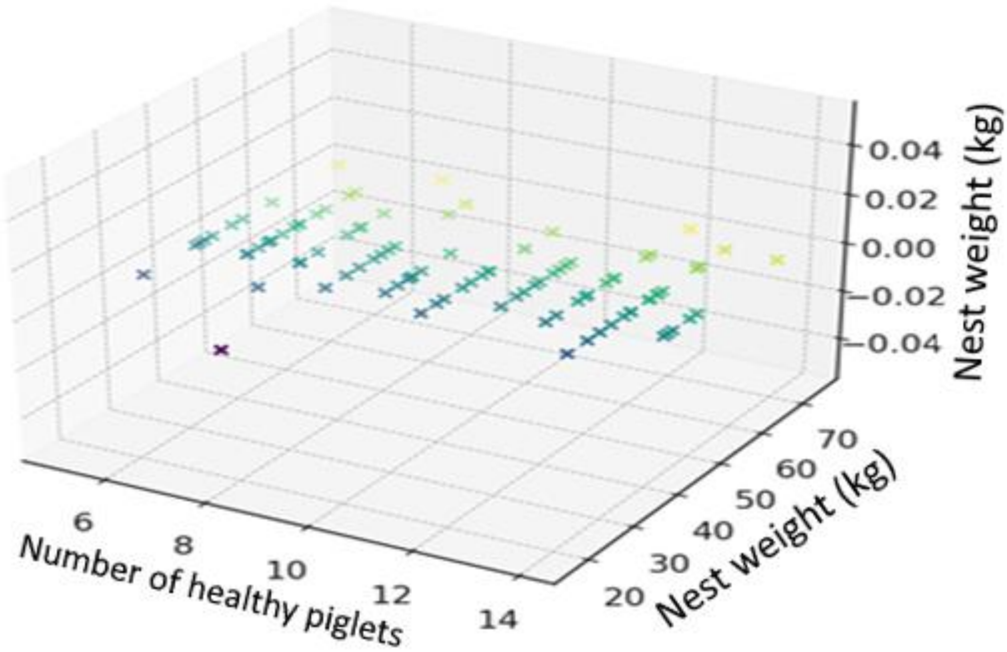

Figure 2 – Distribution of litter and live piglet weights

Abnormal values in the available data.

Litter weight:

• Average: 1.09 kg

• Range: 0 kg to 4 008 465 kg

Abnormal values: Values in millions of kilograms (e.g., 1 301 586 kg and 4 008 465 kg) are unrealistic and indicate serious data entry errors. In addition, values below 1 kg raise doubts about the accuracy of the measurements.

Number of healthy piglets at farrowing:

• Average: 10.7 piglets

• Range: 0 to 54 piglets

Abnormal values: Values above 20 piglets require verification as they exceed the average biological capacity of the sow.

Any statistical conclusions will be unreliable with so many gaps and abnormal values. Using such data can lead to distorted results and incorrect conclusions, which is unacceptable in scientific research.

In addition, the lack of critical data, such as dates of measurements and weighing, makes it impossible to track the dynamics of animal growth and development, which is a key factor in assessing their breeding value.

On the other hand, the correct maintenance of animal numbers can be noted, which allows us to hope for the correct construction of pedigrees and the resulting kinship matrices, necessary elements in BLUP calculations [15-20].

Conclusion. The analysis of the provided raw data for calculating boar breeding indices revealed serious problems: many missing values and anomalous data in key columns. These shortcomings make it difficult to conduct accurate statistical analyses and can significantly distort the results. At the same time, the lack of data on the date of weighing the nest "on the 21st day" leads to the fact that out of 72 839 records, no one can be used to calculate the complex index in BLUP.

Data quality is the foundation for conducting reliable research and making informed decisions. The problems in the current data set make its use risky and can lead to erroneous conclusions. Measures are recommended to improve the data collection, entry, and control processes.

Massive omissions and homogeneity of errors indicate systemic problems:

• Data collection: perhaps the procedures for collecting information were not clearly defined, which led to inconsistency.

• Data entry: Errors could occur when manually entering data or importing data from other sources

Possible causes of problems:

1. Lack of data collection standards: Employees could collect information differently without clear instructions, resulting in incompleteness and inconsistencies.

2. Insufficient staff training: Employees may not understand the importance of accurate and complete data collection, affecting the quality of the information entered.

3. Technical limitations: The hardware or software used could be outdated or unreliable, contributing to errors.

4. Neglect of quality control: Without regular data checks and audits, errors could accumulate and go undetected for a long time.

Suggestions for improving data quality

1. Develop standard data collection procedures

• Create clear instructions: Define what data should be collected and how to ensure its completeness and accuracy.

• Staff training: Conduct training and workshops for all staff involved in data collection and entry, emphasizing the importance of quality information.

2. Implement data quality control systems

• Regular checks: organize periodic data audits to identify and correct errors and omissions.

• Use software tools: implement systems that automatically check data for correctness and warn of possible inconsistencies.

3. Updating technical infrastructure

• Updating software: Switch to modern data management systems that ensure reliable storage and processing of information.

• Automation of data collection: Use sensors and other devices to automatically and accurately collect the necessary information.

4. Recollection of critical data

• Focus on key parameters: prioritize collecting and checking data on the most critical indicators required for analysis.

• Use of alternative sources: refer to other databases or archives to supplement and clarify information if necessary.

5. Analysis and correction of abnormal values

• Identification of outliers: apply statistical methods to detect abnormal values and assess their impact on the overall analysis.

• Verify sources: Check suspect data against original records or logs to confirm or correct the information.

Gratitude. This research is funded by the Ministry of Agriculture of the Republic of Kazakhstan (PTF BR24892783 "Development of integrated management and development of genetic resources of pigs in Kazakhstan using information, molecular genetics and zootechnical methods"). Jafarikia M, Sullivan B. 2014. Genomic tools for improvinghealth and production performance of Canadian pigs.Manuscript n. 506. Proceedings of the 10th WorldCongress Genet Appl Livest Prod, August 2014, VancouverBC, Canada

REFERENCES

1 Bobecek B., Rehacek P., Matusek V. Analysis of reproduction and production parameters in maternal and gene pool populations of boars using the M BLUP AM method in Slovakia [Text] // Biotechnology in Animal Husbandry. 2006; 22(1–2):89–96.

2 Jafarikia M., Sullivan B. Genomics Tools for Improving Health and Production Performance of Canadian Pigs [Text] // Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. 2014; 506.

3 Ruckli A.K., Hortenhuber S.J., Ferrari P., Guy J., Helmerichs J., Hoste R., Hubbard C., Kasperczyk N., Leeb C., Malak Rawlikowska A., et al. Integrative Sustainability Analysis of European Pig Farms: Development of a Multi-Criteria Assessment Tool [Text] // Sustainability. 2022; 14(10):5988. https://doi.org/10.3390/su14105988.

4 Thekkoot D.M., Kemp R.A., Boddicker N.J., Mwansa P., Dekkers J.M.C., Plastow G. Single-step genomic evaluation accuracies for growth and reproductive traits in Yorkshire pigs with genotypes from

two different SNP panels [Text] // Proceedings of the 11th World Congress on Genetics Applied to Livestock Production. 2018; 11 (374).

5 Kuznetsov V.M. BLUP Animal Model – method for genetic evaluation of dairy cattle [Electronic resource] // Kirov: North-East Zonal Agricultural Research Institute named after N.V. Rudnitsky; Available from: https://vm-kuznetsov.ru/files/etude/06_blup_am.pdf (accessed: 12.03.2025).

6 Kuznetsov V.M. BLUP Animal Model for genetic evaluation of pigs: basic model [Electronic resource] // Proceedings of the Conference at Don State Agrarian University, April 22–23, 2010. Kirov: North-East Zonal Agricultural Research Institute; Available from: https://vm-kuznetsov.ru/files/etude/16_blup_am_pig.pdf (accessed: 12.03.2025).

7 Khainatsky V.Yu. BLUP Method for Genetic Evaluation of Beef Sires [Electronic resource] // Animal Husbandry and Fodder Production. 2021; 104(1):20–27. https://doi.org/10.33284/2658-3135-104-1-20 (accessed: 12.03.2025).

8 Wu B., Xiong H., Zhuo L., Xiao Y., Yan J., Yang W. Multi-view BLUP: a promising solution for post-omics data integrative prediction [Electronic resource] // bioRxiv; 2024. https://doi.org/10.1101/2024.11.14.623556 (accessed: 12.03.2025).

9 Henderson C.R. Best Linear Unbiased Estimation and Prediction under a Selection Model [Text] // Biometrics. 1975; 31(2):423–447. https://doi.org/10.2307/2529430.

10 Legarra A., Aguilar I., Misztal I. A relationship matrix including full pedigree and genomic information [Text] // Journal of Dairy Science. 2009; 92(9):4656–4663. https://doi.org/10.3168/jds.2009-2323.

11 Suslina E.N., Novikov A.A., Pavlova S.V., Bashmakova N.V., Fedin G.I., Alekseeva S.I. Genetic evaluation of pigs using the BLUP method [Text] // Proceedings of Timiryazev Agricultural Academy. 2019; 6:150–161. https://doi.org/10.34677/0021-342x-2019-6-150-161.

12 Pig selection – evaluation of pigs breeding value using BLUP [Electronic resource] // Breeding and Genetic Center “Znamensky”. Available from: https://nsgc.ru/kontent/78-ocenka-plemennoj-cennosti-svinej-metodom-blup (accessed: 10.09.2024).

13 Tatiana A.S. Pig breeding selection methods [Electronic resource] // Agroserver. LLC Matrix. Available from: https://agroserver.ru/articles/6336.htm (accessed: 11.09.2024).

14 Animal Identification Rules [Electronic resource] // Approved by Order of the Minister of Agriculture of the Republic of Kazakhstan dated January 30, 2015 №7-1/68. Available from: https://adilet.zan.kz/rus/docs/V1500011127.

15 Sharko F.S., Khatib A., Prokhortchouk E.B. Genomic evaluation of breeding value in Holstein cows using milk production and fertility traits [Text] // Acta Naturae. 2022; 14(1):110–121. Available from: https://actanaturae.ru/2075-8251/article/download/11648/pdf_1 (accessed: 12.03.2025).

16 Sergeev A.I., Golovin A.I. Linear model construction for predicting the breeding value of animals [Text] // Bulletin of KrasGAU. 2021; (7):123–130. Available from: https://www.kgau.ru/vestnik/2021_7/content/22.pdf (accessed: 12.03.2025).

17 Ivanov P.P., Petrov S.S. Predictive models for assessing animal breeding value [Text] // AIC: Economics, Management. 2020; (6):45–50. Available from: https://apknet.ru/prognozirovanie-modeley-rascheta-plem/ (accessed: 12.03.2025).

18 Kabanov A., Melnikova E., Nikitin S., et al. Application of Weighted Single-Step GBLUP Method for Evaluation of Pigs on Meat Productivity and Reproduction Traits [Text] // Animals. 2022; 12(13):1693. https://doi.org/10.3390/ani12131693.

19 Dodd G.R., Gray K., Huang Y., Fragomeni B. Single-Step GBLUP and GWAS Analyses Suggest Implementation of an Unweighted Two-Trait Approach for Pig Selection under Heat Stress Conditions [Text] // Animals. 2022; 12(3):388. https://doi.org/10.3390/ani12030388.

20 Krupa E., Krupova Z., Zákova E., et al. Genetic parameters assessment of reproductive traits in boars using genomic BLUP [Text] // Genes. 2023; 14(11):2003. https://doi.org/10.3390/genes14112003.

ТҮЙІН

Бұл жұмыс Қазақстандағы отандық шошқа шаруашылығы жағдайында шошқалардың асыл тұқымдық құндылығын бағалаудың BLUP әдістемесін енгізу үдерісін бастайды. Бастапқы мәліметтерді талдау статистикалық есептерді дұрыс жүргізуге кедергі келтіретін бірқатар маңызды мәселелерді анықтады. Негізгі мәселелер - жетіспейтін мәндердің жоғары пайызы және қоқыс салмағы, торайлардың саны және өлшеу күндері сияқты негізгі параметрлерде аномальді деректердің болуы. Бұл кемшіліктер BLUP әдісі арқылы күрделі индексті есептеу үшін жазбалардың айтарлықтай бөлігін пайдалануды мүмкін емес етеді. Жұмыс халықаралық тәжірибеге де, аймақтық сипаттамаларға да негізделген әсер етуші факторларды анықтауды, ақпараттық-аналитикалық жүйеде (IAS) деректерді жинау және басқару жүйесін жетілдіруді және мамандандырылған бағдарламалық қамтамасыз етуді әзірлеуді қамтитын кешенді пәнаралық тәсілді ұсынады. Зерттеу нәтижелері сәйкестендіру деректерін дұрыс басқарғанның өзінде бастапқы мәліметтерді тазалау және сапасын бақылау алгоритмдерін одан әрі нақтылауды талап ететін маңызды мәселелер қалып отырғанын көрсетеді, олар әдістемені кейіннен практикалық енгізу кезінде ескерілуі тиіс. Әзірленген әдістемені қолдану шошқалардың асыл тұқымдық құндылығын бағалаудың дәлдігін арттыруға және ұзақ мерзімді перспективада саладағы асылдандыру жұмыстарының тиімділігін арттыруға мүмкіндік береді.

РЕЗЮМЕ

В данной работе инициируется процесс внедрения методики BLUP для оценки племенной ценности свиней в условиях отечественного свиноводства Казахстана. Проведённый анализ исходных данных выявил ряд критических проблем, препятствующих корректному проведению статистических расчётов. Основными проблемами являются высокий процент пропущенных значений и наличие аномальных данных в ключевых параметрах, таких как масса гнезда, количество поросят и даты измерений. Эти недостатки приводят к невозможности использования значительной части записей для расчёта комплексного индекса методом BLUP. В работе предлагается комплексный междисциплинарный подход, включающий определение влияющих факторов на основе как международного опыта, так и региональных особенностей, совершенствование системы сбора и контроля данных в информационно-аналитической системе (ИАС), а также разработку специализированного программного обеспечения. Результаты исследования демонстрируют, что даже при корректном ведении идентификационных данных остаются существенные проблемы, требующие дальнейшей доработки алгоритмов очистки и контроля качества исходных данных, что необходимо учитывать при последующей практической реализации методики. Применение разработанной методики позволит повысить точность оценки племенной ценности свиней и, в перспективе, улучшить эффективность селекционной работы в отрасли.