

ДОСТОВЕРНОСТЬ ДАННЫХ ПЕРВИЧНОГО ЗООТЕХНИЧЕСКОГО УЧЁТА В МОЛОЧНОМ СКОТОВОДСТВЕ

¹**Е.В. Камалдинов**, доктор биологических наук, доцент

¹**А.Ф. Петров**, и.о. заведующего лабораторией прикладной биоинформатики

¹**К.С. Шатохин**, кандидат биологических наук, старший научный сотрудник

¹**К.Н. Нарожных**, кандидат биологических наук

¹**В.Г. Маренков**, кандидат биологических наук

¹**Т.А. Жигулин**, магистрант

¹**О.В. Богданова**, старший преподаватель

²**П.Н. Пальчиков**, директор АО «Новосибирскагроплем»

¹**А.А. Плахова**, доктор биологических наук, доцент

¹*Новосибирский государственный аграрный университет, Новосибирск, Россия*

²*АО «Новосибирскагроплем», с. Ярково, Новосибирской обл., Россия*

E-mail: infoplem@yandex.ru

Ключевые слова: молочное скотоводство, племенное дело, первичный зоотехнический учёт, селекция, разведение, сельское хозяйство.

Реферат. Целью настоящей работы является оценка баз данных первичного зоотехнического учёта в Западной Сибири. Объектом исследований были записи о молочной продуктивности (удой за всю лактацию, содержание жира в молоке, содержание белка в молоке, количество молочного жира, количество молочного белка), длительности лактации (сервис-период, сухостойный период, межотельный период), возраст первого плодотворного осеменения и сведения о происхождении 20000 коров голштинской породы. Оценку достоверности исходных данных осуществляли исходя из предположения о гауссовом распределении при отсутствии значимого влияния человеческого фактора. С этой целью применяли критерий Андерсона–Дарлингга и соответствующую визуализацию с использованием гистограмм и графиков квантиль–квантиль. Список признаков молочной продуктивности был составлен исходя из значений критерия Андерсона–Дарлингга. Установлено, что самые высокие уровни данного критерия соотносились с такими показателями, как молочный жир и белок, тогда как удой практически отсутствовал в перечне признаков молочной продуктивности. Это объясняется тем, что у животных большинства предприятий величина удоя была выше пороговых значений бонитировки. Наряду с использованием статистических критериев, проведен анализ генеалогических деревьев исследуемых племенных предприятий, который позволил выявить неправомерное отнесение нескольких десятков потомков к одной матери. Таким образом, представленный подход может быть использован для выявления выбросов, сопряженных с человеческим фактором, неправильным методическим обеспечением процесса отбора проб и ошибками в работе лабораторий селекционного контроля качества молока, связанными с отбором проб и их доставкой.

RELIABILITY OF PRIMARY ZOOTECNICAL RECORDS IN DAIRY FARMING.

¹**E.V. Kamaldinov**, Doctor of Biological Sciences, Associate Professor

¹**A.F. Petrov**, Acting Head of the Applied Bioinformatics Laboratory

¹**K.S. Shatokhin**, Ph.D. in Biological Sciences, Senior Researcher

¹**K.N. Narozhnykh**, Ph.D. in Biological Sciences

¹**V.G. Marenikov**, Ph.D. in Biological Sciences

¹**T.A. Zhigulin**, Master's student

¹**O.V. Bogdanova**, Senior Lecturer

²**P.N. Palchikov**, Director of the JSC «Novosibirskagroplem»

¹**A.A. Plakhova**, Doctor of Biological Sciences, Associate Professor

¹*Novosibirsk State Agrarian University, Novosibirsk, Russia*

²*JSC Novosibirskagroplem, Yarkovo settlement, Novosibirsk region, Russia*

E-mail: infoplem@yandex.ru

Keywords: dairy cattle breeding, breeding, primary zootechnical accounting, breeding, breeding, agriculture.

Abstract. The purpose of this paper is to evaluate the primary databases of zootechnical records in western Siberia. The object of the study was the records on milk productivity (milk yield during the whole lactation, milk fat content, milk protein content, amount of milk fat, amount of milk protein), duration of lactation (service period, dry period, inter-breeding period), age of the first fruitful insemination and information on the origin of 20,000 Holstein cows. The validity of the raw data was assessed by assuming a Gaussian distribution without significant human influence. For this purpose, the Anderson-Darling test and corresponding visualization using histograms and quantile-quantile plots were applied. The list of traits of milk production was based on the values of the Anderson-Darling criterion. The authors found that the highest levels of this criterion were correlated with milk fat and protein. And the indicator “milk yield” was practically absent from the list of traits of dairy productivity. These results can be explained by the fact that in most enterprises, the value of milk yield was higher than the appraisal threshold values. An analysis of the genealogical trees of the studied breeding enterprises was carried out, along with the use of statistical criteria. This analysis revealed the inappropriate assignment of several dozen of offspring to a single mother. Thus, the presented approach can be used to identify outliers associated with human factors. And it can also be related to improper methodological support of the sampling process and errors in the work of the laboratories of selective milk quality control associated with the sampling and delivery of samples.

Разведение и селекция сельскохозяйственных животных традиционно являются важнейшими составляющими современного животноводства, требующими использования современных подходов и методов. Согласно некоторым исследователям [1], вклад генетических факторов в формирование признаков молочной продуктивности может достигать 30 %. Наряду с признаками отбора важно уделять внимание показателям здоровья и продолжительности хозяйственного использования [2–5].

Племенной учёт в России имеет ряд уникальных особенностей, отличающих его от подходов, применяемых за рубежом. Отечественную селекцию сопровождают выраженная децентрализованность и фрагментированность исходных данных. В условиях ограничения импорта маточного поголовья и племенного материала актуальными представляются вопросы сохранения генетического потенциала по основным признакам молочной продуктивности. Решение обозначенных задач в условиях многолетнего межлинейного кроссирования за рубежом становится малоэффективным при сохранении практики использования классических методов селекции. В странах с развитым молочным скотоводством давно применяются линейные смешанные модели, требующие высокого качества входных данных [6–9]. Важным аспектом является разработка инструментов контроля качества доступного цифрового материала и нивелирование человеческого фактора [10–12]. Дефрагментация исходных данных позволяет значительно повысить уровень контроля используемой информации и эффективность селекции [13–15]. Эффективным методом проверки предположений исследователя и

исключения вероятности принятия решения, основанного на значимом влиянии случайно проявившихся во времени факторов, является повторяемость [16]. Низкое качество исходных данных приводит к снижению точности прогноза изучаемых признаков [6].

Целью данной работы является оценка достоверности первичных данных племенного учёта в молочном скотоводстве Западной Сибири для повышения точности прогноза генетического потенциала молочного скота.

ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЙ

Лабораторией прикладной биоинформатики Новосибирского ГАУ совместно с региональным информационно-селекционным центром (РИСЦ) АО «Новосибирскагроплем» протестированы генеалогические структуры и пригодность данных молочной продуктивности для последующего построения математических моделей и автоматизации процесса подбора родительских пар. Были обработаны материалы первичного зоотехнического учёта за последние 20 лет по маточному поголовью крупного рогатого скота голштинской и черно-пестрой пород Западной Сибири. Оценка генеалогической структуры популяций осуществлялась путём визуализации с использованием специализированных библиотек языка статистического программирования R. В основе проверки консолидированности данных первичного цифрового учёта лежал статистический критерий Андерсона–Дарлингса [17], позволяющий судить о соответствии эмпирических распределений распределению Гаусса.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

Медианы максимальных и минимальных значений признаков, которые были рассчитаны в ходе исследования, могут отражаться на

длительности лактации (табл 1). Полученные результаты отражают региональную особенность ведения баз данных ИАС «Селэкс» и дают общее представление о качестве доступного цифрового материала без предварительного усечения.

Таблица 1

Приемлемые границы некоторых признаков, связанных с лактацией
Acceptable limits for some signs associated with lactation

Параметр	Значения по базам*	Приемлемые границы
Возраст первого плодотворного осеменения, мес	От 6,9±1,04 до 116,4±8,7	13–18
Сухостойный период	От -1,2±0,2 до 373,2±72,4	30–120
Сервис-период	От 13,2±1,8 до 1012,6±106,7	30–120
Межотельный период	От 236,9±14,9 до 1226,5±98,4	350–470

* Лимиты изменчивости представляют собой медианные значения лимитов изменчивости.

* Limits of variability are the median values of the limits of variability.

Очевидными представляются проблемы, связанные с воспроизводительными способностями животных, где, возможно, не последнюю роль играет фактор кормления. Это отражается на максимальном возрасте первого плодотворного осеменения, который превышает приемлемую границу более чем в 6 раз (см. табл. 1). По всей видимости, описываемые проблемы ведут к изменению лимитов других признаков, в особенности сервис-пе-

риода. Полученные результаты представляются первоочередными для региональных информационно-селекционных центров (РИСЦ), поэтому требуют особого внимания и мониторинга.

В базах данных предприятий отобраны показатели молочной продуктивности, которые были проранжированы по уровню критерия Андерсона–Дарлинга за 2021 г. (табл. 2).

Таблица 2

Возможная степень искажения данных первичного зоотехнического учёта по признакам молочной продуктивности

Possible degree of distortion of data for primary zootechnical accounting based on milk production

Признак (за всю лактацию)	Номер хозяйства	n	Mean	SE	Me	Min	Max	Range	IQR	Cv	AD
1	2	3	4	5	6	7	8	9	10	11	12
Белок, кг	1	1788	391,7	5,41	342,9	3,3	1589	1586	128,3	58,4	157,4
Жир, кг	1	1788	467,4	6,47	411,1	3,9	1895	1891	159,2	58,5	152
Жир, %	2	678	3,77	0,005	3,75	3,63	4,85	1,22	0,04	3,3	121
Белок, кг	3	1970	339,3	2,86	327,8	2,4	1767	1764	97	37,4	69,2
Жир, кг	3	1970	412,9	3,44	398,6	2,8	2070	2067	118,3	37	64,9
Жир, кг	4	280	293,8	21,7	216,6	60,4	3114	3053	85,6	123,4	60,9
Белок, кг	5	279	236,3	17,1	177	46,5	2499	2452	69,9	121,1	60,1
Жир, кг	6	830	184,1	1,69	177,2	23,9	530,3	506,4	31,3	26,4	55,1
Белок, кг	7	462	161,6	5,5	163,5	4,2	2456	2452	74,7	73,2	52,6

1	2	3	4	5	6	7	8	9	10	11	12
Жир, кг	7	462	189,6	6,27	185,3	4,2	2776	2772	83,5	71,1	49,4
Белок, кг	8	829	144,5	1,28	139,1	18,9	408,7	389,8	25,3	25,6	49,4
Удой, кг	9	895	4681	40,6	4569	664	13403	12739	823,3	25,9	48,3
Удой, кг	10	471	5076	162,7	5203	132	72128	71996	2298	69,6	45,4
Белок, %	11	1101	3,17	0,001	3,17	3,03	3,4	0,37	0,02	0,9	43,2
Жир, %	12	693	3,77	0,003	3,75	3,59	4,17	0,58	0,05	2	41,4
Белок за 305 дней лактации, %	2	758	3,21	0,002	3,2	3,05	3,56	0,51	0,051	2	41,4
Жир за 305 дней лактации, %	7	586	3,78	0,003	3,75	3,64	4,01	0,37	0,051	1,9	41,2

Представленные данные демонстрируют приоритет по белку и жиру подавляющего большинства предприятий, на которых возможно искажение данных первичного зоотехнического учёта. На это в первую очередь указывают коэффициент вариации и лимиты изменчивости. Максимальный удой за всю лактацию 72128 кг объясняется ошибкой ввода вследствие человеческого фактора, а средний уровень был достаточно низким и мог быть также преднамеренно изменён. На это косвенно указывали низкие значения средних арифметических 9 и 10 хозяйств. Минимальные значения удоя за всю лактацию 664 и 132 кг

могли свидетельствовать о раннем выбытии животного по причине болезни.

Возможные искажения могут быть частично нивелированы предварительным усечением входной информации в границах, значения вариант за пределами которых с большой долей вероятности являются необъективными (табл. 3). Такой подход приведёт к снижению числа выбросов и повышению точности перспективных математических моделей в отдельных случаях до 80% [12]. Для молочных пород скота, отличных от чернопёстрой и голштинской, приведённые границы могут отличаться.

Таблица 3

Предлагаемые границы усечения признаков молочной продуктивности и воспроизводительных качеств молочного скота

Proposed truncation limits for signs of milk productivity and reproductive qualities of dairy cattle

Параметр	Значения по базам*	Приемлемые границы
Возраст первого плодотворного осеменения, мес	От 6,9±1,04 до 116,4±8,7	13–18
Сухостойный период	От -1,2±0,2 до 373,2±72,4	30–120
Сервис-период	От 13,2±1,8 до 1012,6±106,7	30–120
Межотельный период	От 236,9±14,9 до 1226,5±98,4	350–470

Искажение процентного содержания жира и белка в молоке может выражаться заполнением одного «округлённого» значения (рис. 1, Б). При этом зачастую количество вариант непропорционально меньше объёма выборки. При учёте молочной продуктивности наблюдаются многовершинность гистограммы

(см. рис. 1, А) и отклонение распределения квантилей от теоретически ожидаемого (см. рис. 1, Б–Г). Поэтому одной из рекомендаций проверки корректности внесения первичных данных является проверка с помощью гистограмм и графиков распределения квантилей.

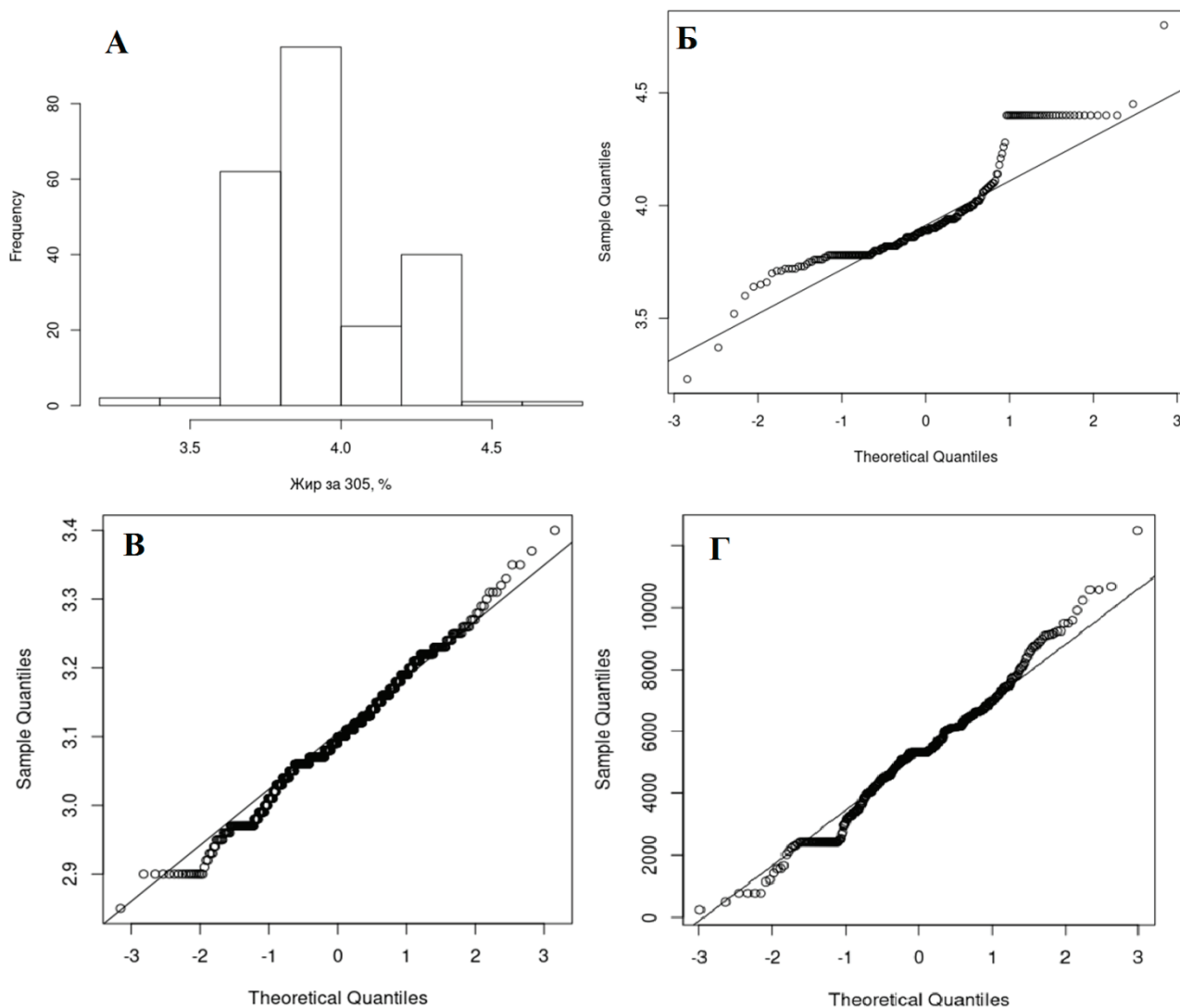


Рис. 1. Гистограмма распределения молочного жира, % (А) и графики квантиль–квантиль молочного жира, % (Б), молочного белка, % (В) и удоя, кг (Г)

Fig. 1. Histogram of milk fat distribution, % (A) and quantile-quantile graphs of milk fat, % (B), milk protein, % (C), and milk yield, kg (D)

Оценка распределения признаков молочной продуктивности на соответствие гауссову распределению (см. табл. 2) и графический подход (см. рис. 1) представляются одним из возможных методов, автоматизирующих поиск возможных ошибок первичного материала. В эмпирических данных маловероятно появление одинаковых значений вариант по сравнению с ожидаемыми уровнями [17]. Это утверждение коррелирует с тем утверждением, что в начале ранжированного списка располагаются признаки, в намеренном изменении которых специалисты хозяйств заинтересованы в большей степени. Причиной тому являются существующие требования при проведении бонитировки, и поэтому удой

здесь встречался не так часто. С другой стороны, несоблюдение требований при отборе для исследования проб в лабораториях селекционного контроля качества молока также может приводить к появлению смещения исходных значений.

Другим источником преднамеренных изменений могут быть генеалогические деревья. Исследование генеалогических структур позволило обнаружить наиболее часто встречаемые варианты ошибок – зацикливания и дублирования. Тем не менее в ряде случаев можно было наблюдать случаи преднамеренной фальсификации этого вида данных, когда к одной матери приписывалось слишком много потомков (рис. 2).

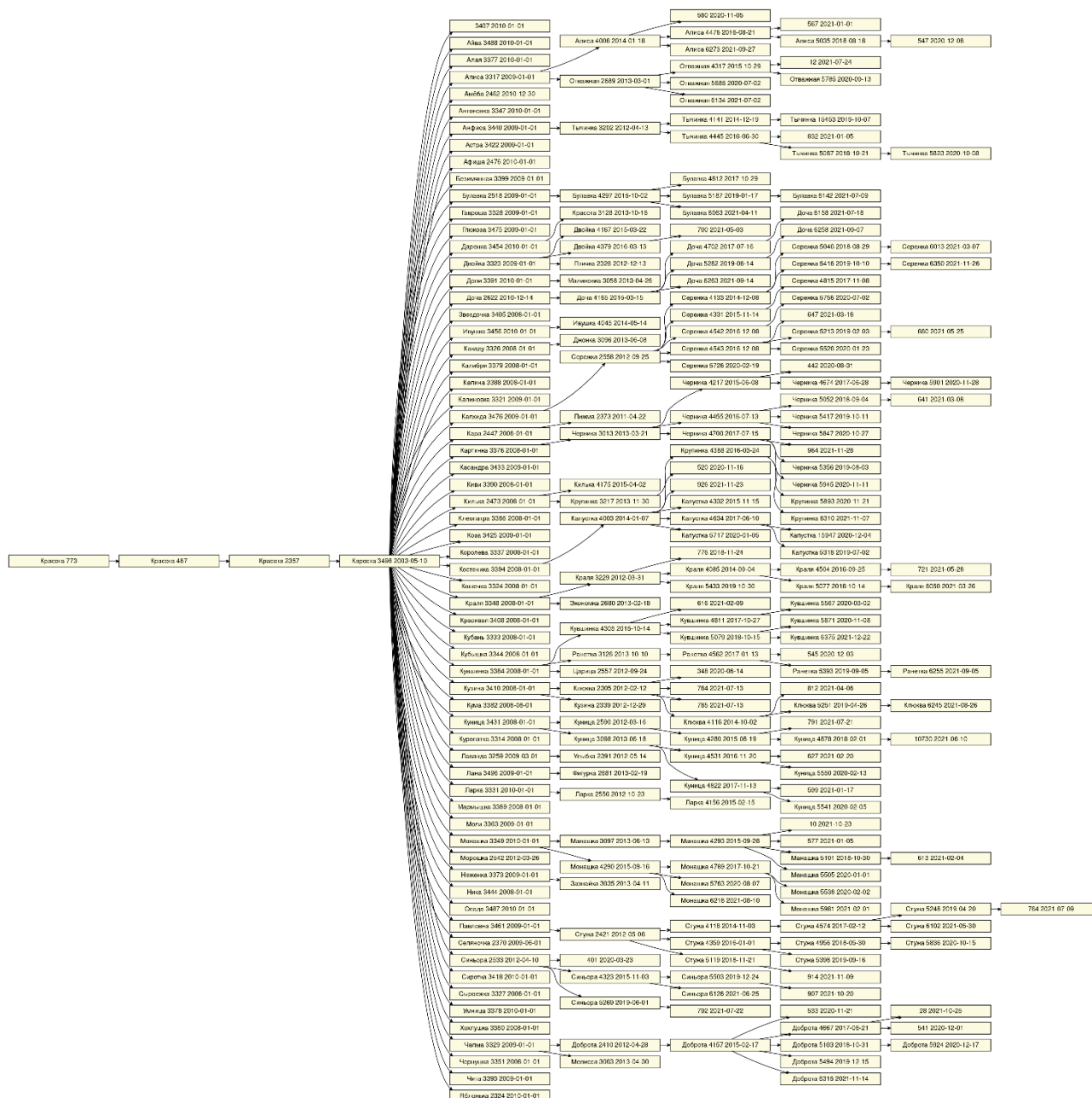


Рис. 2. Пример недостоверного генеалогического дерева семейства

Fig. 2. An example of an invalid family tree of a family

Несложно предположить, что при оценке достоверности происхождения по группам крови, микросателлитному профилю или SNP в таких хозяйствах количество неподтвержденных записей будет велико. Такое явление создаёт дополнительные трудности в приведении первичного электронного учёта к необходимому уровню и может быть связано с недостаточно высокой квалификацией зоотехников-селекционеров и племучётчиков или низким уровнем производственной культуры.

ВЫВОДЫ

1. Предложен подход, позволяющий выявлять факты искажения данных первичного зоотехнического учёта в молочном скотоводстве с использованием критерия Андерсона-Дарлингга и методов визуализации эмпирических распределений.

2. Преднамеренное искажение или ошибки, связанные с функционированием лабораторий селекционного контроля качества молока, чаще всего наблюдаются в отношении признаков, оцениваемых при бонитировке.

3. Автоматизация анализа генеалогической структуры стада позволит своевременно выявлять факты преднамеренного искажения и повышать производственную культуру профильных специалистов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Factors* affecting genetic improvement for milk production of dairy cattle at farm level in Central Thailand / M. Sarakul, S. Koonawootrittriron, T. Suwanasopee, M. A. Elzo, A. Hirunwong, T. Thongprapi // *Proceedings of the 48th Kasetsart University Conference, Animals*, February 3-5, 2010. – Bangkok (Thai), 2010. – P. 150–157.
2. *The future* of phenomics in dairy cattle breeding / J.B. Cole, S.A.E. Eaglen, C. Maltecca [et al.] // *Anim Front.* – 2020. – Vol. 10, N 2. – P. 37–44. – DOI: 10.1093/af/vfaa007.
3. *Hill W.B.* Maintenance of quantitative genetic variation in animal breeding programmes // *Livestock Production Science*. 2000. – Vol. 63. – P. 99–109. – DOI: 10.1016/S0301-6226(99)00115-3.
4. *Tropical* milk production systems and milk quality: a review / E.J. Ramírez-Rivera, J. Rodríguez-Miranda, I.R. Huerta-Mora [et al.] // *Trop Anim Health Prod.* – 2019. – Vol. 51. – P. 1295–1305. – DOI:10.1007/s11250-019-01922-1.
5. *Strategies* to improve the efficiency of beef cattle production / S.A. Terry, J.A. Basarab, L.L. Guan, T.A. McAllister // *Canadian Journal of Animal Science.* – 2020. – Vol. 101, N 1. – P. 1–19. – DOI:10.1139/cjas-2020-0022.
6. *Mrode R.A.* Linear models for the prediction of animal breeding values. – Wallingford: CAB International Publ., 2014. – 360 p.
7. *Use of Random Regression Test-Day Model to Estimate Genetic Parameters of Milk Yield in Holstein Cows* / Y. Fazel, M. Fozi, A. Esmailzadeh [et al.] // *Open Journal of Animal Sciences.* – 2018. – Vol. 8. – P. 27–38. – DOI: 10.4236/ojas.2018.81003.
8. *Inoue K.* Application of Bayesian causal inference and structural equation model to animal breeding // *Anim Sci J.* – 2020. – Vol. 91, N 1. – P. e13359. – DOI: 10.1111/asj.13359. PMID: 32219948; PMCID: PMC7187322.
9. *Bayesian* estimation of genetic variance and response to selection on linear or ratio traits of feed efficiency in dairy cattle / M.S. Islam, J. Jensen, P. Løvendahl [et al.] // *J Dairy Sci.* – 2020. – Vol. 103, N 10. – P. 9150–9166. – DOI: 10.3168/jds.2019-17137.
10. Assessment of the variability of reproductive abilities of a black and white cattle using genealogical data and paratypical factors / E.V. Kamaldinov, O.D. Panferova, O.V. Efremova, V.G. Marenkov, A.F. Petrov, I.N. Ryumkina // *Data in Brief.* – 2021. – Vol. 35. – P. 106842. – DOI: 10.1016/j.dib.2021.106842.
11. *Моделирование* объема туловища по линейным признакам скота ирменского типа / А.Ф. Петров, Е.В. Камалдинов, О.Д. Панферова [и др.] // *Сибирский вестник сельскохозяйственной науки.* – 2020. – № 50 (6). – С. 106–114. – DOI: 10.26898/0370-8799-2020-6-12.
12. *Роль* фиксированных факторов в изменчивости удоя скота ирменского типа в условиях промышленного комплекса / А.Ф. Петров, Е.В. Камалдинов, О.В. Богданова [и др.] // *Вестник НГАУ.* – 2021. – № 4 (61). – С. 137–149. – DOI: 10.31677/2072-6724-2021-61-4-137-149.
13. *Hu Z.-S., Park C.A., Reesy J.M.* Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB // *Nucleic Acids Research.* – 2019. – Vol. 47, N D1. – P. D701–d710. – DOI: 10.1093/nar/gky1084.
14. *Nayeri S., Sargolzaei M., Tulpan D.* A review of traditional and machine learning methods applied to animal breeding // *Anim Health Res Rev.* – 2019. – Vol. 20, N 1. – P. 31–46. – DOI: 10.1017/s1466252319000148.
15. *Data* quality in recommender systems: the impact of completeness of item content data on prediction accuracy of recommender systems / B. Heinrich, M. Hopf, D. Lohninger, [et al.] // *Electron Markets.* – 2021. – Vol. 31. – P. 389–409. – DOI: 10.1007/s12525-019-00366-7.
16. *Лепехина Т.В., Бакай А.В., Бакай Ф.Р.* Изменчивость и повторяемость признаков молочной продуктивности у коров черно-пестрой породы // *Зоотехния.* – 2018. – № 12. – С. 10–11.

17. *Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход* / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов [и др.]. – Новосибирск: Изд-во НГТУ, 2011. – 888 с.

REFERENCES

1. Sarakul M., Koonawootrittriron S., Suwanasopee T., Elzo M.A., Hirunwong A., Thongprapi T., Factors affecting genetic improvement for milk production of dairy cattle at farm level in Central Thailand, *Proceedings of the 48th Kasetsart University Conference, Animals, February 3-5, 2010, Bangkok (Thai), 2010*, pp. 150–157.
2. Cole J.B., Eaglen S.A.E, Maltecca C., Mulder H.A., Pryce J.E., The future of phenomics in dairy cattle breeding, *Anim Front.*, 2020, Vol. 10, No. 2, pp. 37–44, DOI: 10.1093/af/vfaa007.
3. Hill W.B., Maintenance of quantitative genetic variation in animal breeding programmes, *Livestock Production Science*, 2000, Vol. 63, pp. 99–109, DOI: 10.1016/S0301-6226(99)00115-3.
4. Ramírez-Rivera E.J., Rodríguez-Miranda J., Huerta-Mora I.R., Cárdenas-Cágal A., Juárez-Barrientos J.M., *Tropical milk production systems and milk quality: a review. Trop Anim Health Prod.*, 2019, Vol. 51, pp. 1295–1305, DOI:10.1007/s11250-019-01922-1.
5. Terry S.A., Basarab J.A., Guan L.L., McAllister T.A., Strategies to improve the efficiency of beef cattle production, *Canadian Journal of Animal Science*, 2020, Vol. 101, No. 1, pp. 1–19, DOI:10.1139/cjas-2020-0022.
6. Mrode R.A., *Linear models for the prediction of animal breeding values*, Wallingford: CAB International Publ., 2014, 360 p.
7. Fazel Y., Fozi M., Esmailzadeh A., Fazel F., Niazi A., Rahmati S., Qasimi M., Use of Random Regression Test-Day Model to Estimate Genetic Parameters of Milk Yield in Holstein Cows, *Open Journal of Animal Sciences*, 2018, Vol. 8, pp. 27–38, DOI: 10.4236/ojas.2018.81003.
8. Inoue K., Application of Bayesian causal inference and structural equation model to animal breeding, *Anim Sci J.*, 2020, Vol. 91, No. 1, pp. e13359, DOI: 10.1111/asj.13359. PMID: 32219948; PMCID: PMC7187322.
9. Islam M.S., Jensen J., Løvendahl P., Karlskov-Mortensen P., Shirali M., Bayesian estimation of genetic variance and response to selection on linear or ratio traits of feed efficiency in dairy cattle, *J Dairy Sci.*, 2020, Vol. 103, No. 10, pp. 9150–9166, DOI: 10.3168/jds.2019-17137.
10. Kamaldinov E.V., Panferova O.D., Efremova O.V., Marenkov V.G., Petrov A.F., Ryumkina I.N., Assessment of the variability of reproductive abilities of a black and white cattle using genealogical data and paratypical factors, *Data in Brief.*, 2021, Vol. 35, pp. 106842, DOI: 10.1016/j.dib.2021.106842.
11. Petrov A.F., Kamaldinov E.V., Panferova O.D., Efremova O.V., Rogozin V.A., *Sibirskij vestnik sel'skhozjajstvennoj nauki*, 2020, No. 50 (6), pp. 106–114, DOI: 10.26898/0370-8799-2020-6-12. (In Russ.)
12. Petrov A.F., Kamaldinov E.V., Shatokhin K.S., Efremova O.V., Rogozin V.A., *Vestnik NSAU*, 2021, No. 4, pp. 137–149, DOI: 10.31677/2072-6724-2021-61-4-137-149. (In Russ.)
13. Hu Z.-S., Park C.A., Reecy J.M., Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB, *Nucleic Acids Research*, 2019, Vol. 47, No. D1, pp. D701–d710, DOI: 10.1093/nar/gky1084.
14. Nayeri S., Sargolzaei M., Tulpan D., A review of traditional and machine learning methods applied to animal breeding. *Anim Health Res Rev.*, 2019, Vol. 20, No. 1, pp. 31–46, DOI: 10.1017/s1466252319000148.
15. Heinrich B., Hopf M., Lohninger, D., Schiller A., Szubartowicz M., Data quality in recommender systems: the impact of completeness of item content data on prediction accuracy of recommender systems, *Electron Markets*, 2021, Vol. 31, pp. 389–409, DOI: 10.1007/s12525-019-00366-7.
16. Lephina T.V., Bakaj A.V., Bakaj F.R., *Zootehnika*, 2018, No. 12, pp. 10–11. (In Russ.)
17. Lemeshko B.Y., Lemeshko S.B., Postovalov S.N., Chimitova E.V., *Statisticheskij analiz dannyh, modelirovanie i issledovanie veroyatnostnyh zakonornostej. Komp'yuternyj podhod* (Statistical data analysis, modeling and probabilistic patterns research. Computer approach), Novosibirsk state technical university Publ., 888 pp.