

Оценка качества и информативности показателей признаков при формировании селекционного индекса красных молочных пород крупного рогатого скота

А. П. Шевчук[✉], Т. А. Князева

Всероссийский научно-исследовательский институт племенного дела, п. Лесные Поляны, Московская область, Россия

[✉]E-mail: vniiplem.kras@outlook.com

Аннотация. Надежность селекционного индекса определяется качеством исходных данных племенного учета и обоснованностью выбора признаков, включаемых в его структуру. **Цель работы** – оценить качество и информативность показателей хозяйственно полезных признаков красных молочных пород крупного рогатого скота для их последующего отбора к включению в комплексный селекционный индекс. **Методология и методы.** Материалом послужил массив из 43 987 записей по 18 061 корове 16 племенных хозяйств (ИАС «Селэкс. Молочный скот», годы рождения 2011–2024, лактации с 1 по 5). По 17 признакам молочной продуктивности, воспроизводства и здоровья вымени проведен описательный статистический анализ, выявлены выбросы по биологическому и статистическому (границы Тьюки) критериям, осуществлено логарифмическое преобразование для стабилизации дисперсии и корреляционный анализ по Пирсону; обработка выполнена в среде R. **Научная новизна.** Впервые для популяции красных пород проведена комплексная проверка структуры и качества первичных данных племенного учета как самостоятельный этап, предшествующий конструированию индекса: с количественной оценкой доли биологически недостоверных записей по каждому признаку и обоснованием отбора показателей по критериям информативности и отсутствия избыточности. **Результаты.** Показатели признаков воспроизводства и содержания соматических клеток в молоке характеризуются выраженной правосторонней асимметрией (коэффициент до 2,48) и высокой долей выбросов, которые для возраста 1-го осеменения достигают 48 %; биологическая фильтрация снижала асимметрию до диапазона 0,1–1,0. Продуктивные признаки показали устойчивое распределение, близкое к нормальному, и взаимную избыточность ($r = 0,94...0,96$): в индекс целесообразно включать один из этих признаков. Линейная зависимость межотельного интервала от сервис-периода ($r = 0,997$) исключает их одновременное использование. Содержание соматических клеток и индекс осеменения показали слабую связь с остальными признаками, что повышает их ценность как самостоятельных компонентов индекса.

Ключевые слова: селекционный индекс, красные породы, красная степная, молочный скот, качество данных, информативность признаков, воспроизводительные качества, соматические клетки, биологическая фильтрация, корреляционный анализ, племенной учет

Благодарности. Работа выполнена в рамках государственного задания Министерства сельского хозяйства Российской Федерации (тема № 082-00066-26-00 на 2026 год от 26 января 2026 г.).

Для цитирования: Шевчук А. П., Князева Т. А. Оценка качества и информативности признаков при формировании селекционного индекса красных молочных пород КРС // Аграрный вестник Урала. 2026. Т. 26, № 09. С. 1769–1782. <https://doi.org/10.32417/1997-4868-2026-26-09-1769-1782>.

Дата поступления статьи: 19.06.2026, **дата рецензирования:** 06.07.2026, **дата принятия:** 20.07.2026.

Assessment of trait quality and informativeness for constructing a selection index in red dairy cattle breeds

A. P. Shevchuk✉, T. A. Knyazeva

All-Russian Research Institute of Animal Breeding, Lesnye Polyany settlement, Moscow region, Russia

✉E-mail: vniiplek.kras@outlook.com

Abstract. The reliability of a selection index depends directly on the quality of the underlying herd-recording data and on the soundness of the traits chosen for its structure. **The purpose** of the study was to assess the quality and informativeness of economically important traits in red dairy cattle breeds in order to select indicators suitable for inclusion in a multi-trait selection index. **Methodology and methods.** The dataset comprised 43 987 records from 18 061 cows across 16 breeding farms (Selex Dairy Cattle herd-management information system; birth years 2011–2024; lactations one through five). Seventeen traits of milk production, reproduction, and udder health were subjected to descriptive statistical analysis, outlier detection by biological and statistical (Tukey fences) criteria, log-transformation for variance stabilisation, and Pearson correlation analysis; all processing was carried out in the R environment. **Scientific novelty.** For the first time in a red-breed population, an end-to-end audit of the structure and quality of primary herd-recording data was performed as a distinct stage preceding index construction, with quantification of the proportion of biologically implausible records for each trait and justification of trait selection by the criteria of informativeness and non-redundancy. **Results.** Reproductive traits and somatic cell count were found to exhibit pronounced positive (right) skewness (coefficient up to 2.48) and a high proportion of outliers, reaching 48 % for age at first conception; biological filtering reduced skewness to the 0.1–1.0 range. Production traits showed stable, near-normal distributions and mutual redundancy ($r = 0.94 \dots 0.96$), making it appropriate to include only one of them in the index. The near-perfect linear dependence between days open and calving interval ($r = 0.997$) precludes their simultaneous use. Somatic cell count and number of services showed weak associations with the other traits, which enhances their value as independent components of the index.

Keywords: selection index, red breeds, Red Steppe, dairy cattle, data quality, trait informativeness, reproductive performance, somatic cell count, biological filtering, correlation analysis, herd recording

Acknowledgements. This study was carried out within the framework of the state assignment of the Ministry of Agriculture of the Russian Federation (topic No. 082-00066-26-00 on 2026 from 26th January 2026).

For citation: Shevchuk A. P., Knyazeva T. A. Assessment of trait quality and informativeness for constructing a selection index in red dairy cattle breeds. *Agrarian Bulletin of the Urals*. 2026; 26 (09): C. 1769–1782. <https://doi.org/10.32417/1997-4868-2026-26-09-1769-1782>. (In Russ.)

Date of paper submission: 19.06.2026, **date of review:** 06.07.2026, **date of acceptance:** 20.07.2026.

Постановка проблемы (Introduction)

Эффективность селекционно-племенной работы в молочном скотоводстве определяется не только выбором статистического метода оценки генетического потенциала, но и качеством первичной информации. Переход отечественной системы оценки к методологии BLUP и комплексным селекционным индексам, закрепленный решениями Коллегии Евразийской экономической комиссии [1], существенно расширил перечень учитываемых признаков: к традиционным показателям молочной продуктивности (удой, выход молочного жира и белка) добавлены воспроизводительные качества и характеристика здоровья вымени, прежде всего содержание соматических клеток [1; 2]. Подобное расширение целевой функции отбора соответствует мировой практике, где комплексные индексы (немецкий RZG, североамериканский NTM, амери-

канские TPI и NM\$, канадский LPI) интегрируют продуктивные, функциональные, экстерьерные и репродуктивные субиндексы [2; 3].

Воспроизводительные показатели, возрастные характеристики и интервалы между технологическими событиями регистрируют в хозяйствах вручную, нередко с разной степенью полноты и достоверности [4]. Международные рекомендации по генетической оценке (ICAR, INTERBULL) прямо указывают на необходимость предварительного контроля логической непротиворечивости данных, выявления биологически невозможных значений и количественной оценки качества записей до начала расчетов [5; 6]. Игнорирование этого этапа ведет к смещению дисперсионных компонентов, искажению генетических корреляций и, как следствие, к некорректному ранжированию животных.

Дополнительную сложность создает сама структура взаимосвязей между признаками. Часть показателей дублирует друг друга по информационному содержанию (например, абсолютный выход компонентов молока тесно связан с удоем), часть связана определением (сервис-период и межотельный интервал), а между продуктивностью и воспроизводством существует известный физиологический антагонизм, многократно отмеченный в исследованиях [7–9]. Включение в индекс избыточных или функционально зависимых признаков не повышает точность оценки, но усложняет интерпретацию весовых коэффициентов и может приводить к двойному учету одной и той же биологической информации.

Для красных молочных пород, разведение которых в России до настоящего времени велось преимущественно по отдельным признакам без применения принципов индексной селекции, задача предварительной оценки качества и информативности исходных данных приобретает особую актуальность [10; 11]. Прежде чем конструировать комплексный индекс, необходимо установить, какие признаки обладают достаточной достоверностью регистрации, какова доля биологически недостоверных записей по каждому из них, насколько устойчивы распределения показателей и в какой мере они дублируют друг друга по информационному содержанию.

Целью настоящего исследования является оценка качества и информативности хозяйственно полезных признаков красных молочных пород крупного рогатого скота для отбора показателей, пригодных к включению в структуру комплексного селекционного индекса.

Методология и методы исследования (Methods)

В исследовании использованы данные 16 племенных хозяйств, полученные из информационно-аналитической системы «Селэкс. Молочный скот». Общий объем выборки составил 43 987 записей, сформированных по 18 061 корове (включая как живых, так и выбывших животных) с датой рождения в период с 2011 по 2024 годы. В анализ включены записи по лактациям с первой по пятую.

Структура выборки охватывала показатели молочной продуктивности, воспроизводства и технологические характеристики. Объем доступных наблюдений варьировал по признакам и составлял: для сервис-периода 43 954 наблюдения, сухостойного периода – 36 825, межотельного интервала – 25 820, продолжительности стельности – 25 820, возраста при отеле – 43 987, возраста первого осеменения – 43 987, возраста результативного осеменения – 25 820, интервала от отеля до первого осеменения – 43 987, количества осеменений – 43 987, содержания соматических клеток – 651, легкости отела – 43 987, номера лактации – 43 987, удоя –

43 987, массовой доли жира – 43 987, выхода молочного жира – 43 987, массовой доли белка – 43 280, выхода молочного белка – 43 280.

Средние значения основных признаков и их стандартное отклонение составили: сервис-период – 129,1 дня (SD = 86,81 дня), сухостойный период – 58,9 дня (SD = 12,14 дня), межотельный интервал – 402,6 дня (SD = 81,74 дня), продолжительность стельности – 276,6 (SD = 17,31 дня). Возраст при первом отеле составил 43,4 месяца (SD = 19,06 месяца), возраст первого осеменения – 34,4 месяца (SD = 19 месяцев), возраст плодотворного осеменения – 45,3 месяца (SD = 17,52 месяца). Интервал от отеля до первого осеменения составил в среднем 115,5 дня (SD = 95,12 дня). Среднее количество осеменений – 1,8 (SD = 1,25).

Показатели удоя составили в среднем 6762,1 кг (SD = 1591,72 кг), массовая доля жира – 4,18 % (SD = 0,33 %), выход молочного жира – 282,9 кг (SD = 72,33 кг), массовая доля белка – 3,30 % (SD = 0,23 %), выход молочного белка – 225,6 кг (SD = 61,43 кг). Среднее содержание соматических клеток составило 378 тыс/мл (SD = 229 тыс/мл). Легкость отела характеризовалась средним значением 1,6 балла (SD = 0,99 балла).

Распределение большинства признаков характеризовалось выраженной положительной асимметрией, особенно для показателей воспроизводства и содержания соматических клеток в молоке (коэффициент асимметрии до 2,47 и выше), что отражало наличие экстремальных значений в исходных данных. Для показателей продолжительности стельности отмечалась отрицательная асимметрия.

Методология и методы исследования (Methods)

Исходные данные импортировали в среду R в текстовом формате с учетом кодировки и десятичного разделителя. Все переменные первоначально загружали как символьные с последующим приведением к соответствующим типам данных. Даты технологических событий (отел, осеменение, запуск, первое осеменение в лактации) преобразовывали в формат даты. Числовые признаки очищали от некорректных значений (пустые строки, служебные обозначения пропусков, нечисловые символы) и приводили к числовому типу.

Для количественных признаков задавали биологически допустимые диапазоны: сервис-период – 40–300 дней, сухостойный период – 30–90 дней, межотельный интервал – 300–600 дней, продолжительность стельности – 260–295 дней, возраст при отеле – 20–60 месяцев, возраст первого осеменения – 13–30 месяцев, возраст плодотворного осеменения – 14–60 месяцев, интервал до первого осеменения – 30–300 дней, индекс осеменения – 1–6, содержание соматических клеток – 10 000–3 000 000 кл/мл, легкость отела – 1–5 баллов, номер лактации – 1–5, удой – 2500–13 000 кг,

массовая доля жира – 3,2–5,1 %, выход молочного жира – 100–600 кг, массовая доля белка – 2,8–3,8 %, выход молочного белка – 100–400 кг. Значения за пределами указанных диапазонов заменяли пропущенными.

Для оценки качества данных дополнительно формировали индикаторы выбросов и рассчитывали долю исключенных наблюдений по каждому признаку. Пропущенные значения не заполняли, все расчеты выполняли на основе доступных наблюдений.

Для показателя содержания соматических клеток применяли корректировку: значения менее 100 кл/мл исключали, а значения 100–2000 интерпретировали как представленные в тысячах и умножали на 1000.

Распределения признаков анализировали с использованием гистограмм и диаграмм размаха. Рассчитывали основные описательные статистики (среднее, стандартное отклонение, медиана, квар-

тили, минимальные и максимальные значения). Вариабельность оценивали по стандартному отклонению и межквартильному размаху, асимметрию – по коэффициенту асимметрии и расхождению между средним и медианой.

Выбросы определяли по двум критериям: биологическому (выход за установленные диапазоны) и статистическому (границы $Q1 - 1,5 \cdot IQR$ и $Q3 + 1,5 \cdot IQR$). Для исходных данных сохраняли флаги выбросов и рассчитывали их долю.

Для снижения асимметрии и стабилизации дисперсии применяли логарифмическое преобразование $\log(x)$. Исходные данные при этом сохраняли и использовали для интерпретации результатов.

Линейные взаимосвязи между признаками оценивали с использованием коэффициента корреляции Пирсона на основе попарно доступных наблюдений. Результаты представляли в виде корреляционной матрицы с визуализацией интенсивности связей.

Таблица 1
Характеристика признаков до биологической очистки

Признак	Количество наблюдений	Среднее значение	Стандартное отклонение	Медиана	Минимум	Максимум	Асимметрия
Продолжительность сервис-периода, дни	43 954	129,2	86,8	105	2	1 405	2,47
Межотельный интервал, дни	25 820	402,6	81,7	381	164	1 664	2,48
Продолжительность стельности, дни	25 820	276,6	17,3	280	2	785	–6,22
Возраст плодотворного осеменения, мес.	43 987	34,4	19,0	30	3	165	1,34
Количество осеменений	43 987	1,80	1,25	1	0	15	2,32
Количество соматических клеток, тыс.	651	385	230	337	2	1 670	2,25
Удой, кг	43 987	6 762	1592	6618	814	15 552	0,35
Доля жира, %	43 985	4,18	0,33	4,14	3,0	6,8	0,6
Доля белка, %	43 280	3,30	0,23	3,24	0	5,68	2,28

Table 1
Characteristics of traits before biological data editing

Trait	Number of observations	Mean value	Standard deviation	Median	Min	Max	Skew
Duration of service period, days	43 954	129.2	86.8	105	2	1 405	2.47
Calving interval, days	25 820	402.6	81.7	381	164	1 664	2.48
Duration of pregnancy, days	25 820	276.6	17.3	280	2	785	–6.22
Age at first conception, months	43 987	34.4	19.0	30	3	165	1.34
Number of inseminations	43 987	1.80	1.25	1	0	15	2.32
Somatic cell count, thousand	651	385	230	337	2	1 670	2.25
Milk yield, kg	43 987	6 762	1592	6618	814	15 552	0.35
Fat, %	43 985	4.18	0.33	4.14	3.0	6.8	0.6
Protein, %	43 280	3.30	0.23	3.24	0	5.68	2.28

Таблица 2

Характеристика признаков после биологической очистки

Признак	Количество наблюдений	Среднее значение	Стандартное отклонение	Медиана	Минимум	Максимум	Асимметрия
Продолжительность сервис-периода, дни	41 804	116,5	60,8	101	20	300	0,91
Межотельный интервал, дни	24 906	394,8	62,0	379	300	600	1,04
Продолжительность стельности, дни	22 711	280,3	5,7	281	260	295	–0,57
Возраст плодотворного осеменения, мес.	22 859	20,5	5,5	19	12	30	0,27
Количество осеменений	43 535	1,74	1,09	1	1	6	1,67
Количество соматических клеток, тыс.	640	392	226	340	102	1 670	2,4
Удой, кг	43 873	6 768	1 571	6 621	2 509	12 991	0,36
Доля жира, %	43 715	4,17	0,32	4,14	3,21	5,10	0,32
Доля белка, %	41 572	3,28	0,17	3,23	2,80	3,80	1,17

Table 2

Characteristics of traits after biological data editing

Trait	Number of observations	Mean value	Standard deviation	Median	Min	Max	Skew
Duration of service period, days	41 804	116.5	60.8	101	20	300	0.91
Calving interval, days	24 906	394.8	62.0	379	300	600	1.04
Duration of pregnancy, days	22 711	280.3	5.7	281	260	295	–0.57
Age at first conception, months	22 859	20.5	5.5	19	12	30	0.27
Number of inseminations	43 535	1.74	1.09	1	1	6	1.67
Somatic cell count, thousand	640	392	226	340	102	1 670	2.4
Milk yield, kg	43 873	6 768	1 571	6 621	2 509	12 991	0.36
Fat, %	43 715	4.17	0.32	4.14	3.21	5.10	0.32
Protein, %	41 572	3.28	0.17	3.23	2.80	3.80	1.17

Результаты (Results)

Большинство изученных признаков характеризуется выраженной положительной асимметрией и высоким уровнем вариабельности (таблица 1). Наиболее заметно асимметрия выражена у показателей воспроизводства: сервис-периода (2,47), межотельного интервала (2,48), интервала от отела до первого осеменения (1,90), индекса осеменения (2,32). Высокая асимметрия отмечена и для содержания соматических клеток в молоке (2,25), среди показателей которых присутствует значительное число экстремально высоких значений.

Распределение продолжительности стельности отличалось выраженной отрицательной асимметрией (–6,22), что указывает на наличие некорректных или экстремальных значений в исходных данных. Аналогичные искажения наблюдались для признаков возраста и других показателей, таких как легкость отела и номер лактации.

Доля значений, выходящих за биологически допустимые пределы, существенно различалась между признаками (таблица 3). Наибольший уро-

вень выбросов выявлен для возраста первого осеменения (48,03 %), интервала от отела до первого осеменения (19,65 %), возраста при первом отеле (17,21 %) и возраста плодотворного осеменения (17,13 %), что указывает на систематические отклонения или неоднородность исходных данных по этим признакам. У признака «продолжительность стельности» доля исключенных значений составила 12,04 %, преимущественно за счет нижних границ. У большинства продуктивных признаков доля выбросов была незначительной и не превышала 4 %.

Особого внимания требуют показатели, характеризующие возраст и интервал от отела до осеменения: для них выявлена высокая доля выбросов. Наибольшее отклонение у возраста первого осеменения (48,03 %): практически половина наблюдений выходила за пределы биологически допустимого диапазона. Значительная доля выбросов зафиксирована также для возраста при первом отеле (17,21 %), возраста результативного осеменения (17,13 %) и интервала от отела до первого осеменения (19,65 %). Структура выбросов асимметрична:

для возраста первого осеменения и возраста при первом отеле основная часть исключенных значений приходилась на верхнюю границу – это возможно, когда большинство животных с существенной задержкой развития или имеются ошибки регистрации данных. Для интервала до первого осеменения значительная доля выбросов пришлась на нижнюю границу (15,04 %): здесь возможны как некорректная фиксация дат, так и случаи раннего осеменения, не соответствующие физиологическим нормам.

Высокая доля исключенных значений по этим признакам означает: исходные данные неоднородны, и для дальнейшего анализа требуется биологическая фильтрация.

После фильтрации распределение большинства признаков существенно нормализовалось. Коэффициент асимметрии снизился до диапазона 0,1–1,0 для большинства признаков: экстремальные наблюдения стали влиять заметно меньше. Для сервис-периода асимметрия снизилась с 2,47 до 0,91, для межотельного интервала – с 2,48 до 1,04, для интервала от отела до первого осеменения – с 1,90 до 0,84.

После устранения экстремальных значений показателя «продолжительность стельности» распределение нормализовалось (коэффициент асимметрии – 0,57), стандартное отклонение в среднем снизилось с 17,31 до 5,75 дня.

Таблица 3
Доля значений, исключенных при биологической фильтрации

Признак	Обрезка слева, %	Обрезка справа, %	Всего, %
Сервис-период, дни	0,228	4,664	4,891
Сухостойный период, дни	1,13	2,376	3,506
Межотельный интервал, дни	0,662	2,878	3,54
Продолжительность стельности, дни	11,143	0,899	12,041
Возраст первого отела, мес.	0,073	17,137	17,21
Возраст плодотворного осеменения, дни	0,155	47,878	48,032
Возраст первого осеменения, мес.	0	17,134	17,134
Дней до первого осеменения	15,039	4,615	19,654
Индекс осеменения	0,025	1,003	1,028
Количество соматических клеток, тыс.	1,69	0	1,69
Легкость отелов	0	1,485	1,485
Номер лактации	0	2,899	2,899
Удой, кг	0,202	0,057	0,259
Массовая доля жира, %	0,018	0,596	0,614
Выход молочного жира, кг	0,166	0,061	0,227
Массовая доля белка, %	0,014	3,933	3,946
Выход молочного белка, кг	0,628	0,691	1,319

Table 3
Proportion of records excluded during biological filtering

Trait	Lower trim, %	Upper trim, %	Total, %
Service period, days	0.228	4.664	4.891
Dry period, days	1.13	2.376	3.506
Calving interval, days	0.662	2.878	3.54
Duration of pregnancy, days	11.143	0.899	12.041
Age at first calving, months	0.073	17.137	17.21
Age at fruitful insemination, days	0.155	47.878	48.032
Age at first insemination, days	0	17.134	17.134
Days to first insemination	15.039	4.615	19.654
Insemination index	0.025	1.003	1.028
Somatic cell count, thousands	1.69	0	1.69
Calving ease	0	1.485	1.485
Lactation number	0	2.899	2.899
Milk yield, kg	0.202	0.057	0.259
Fat content, %	0.018	0.596	0.614
Fat yield, kg	0.166	0.061	0.227
Protein content, %	0.014	3.933	3.946
Protein yield, kg	0.628	0.691	1.319

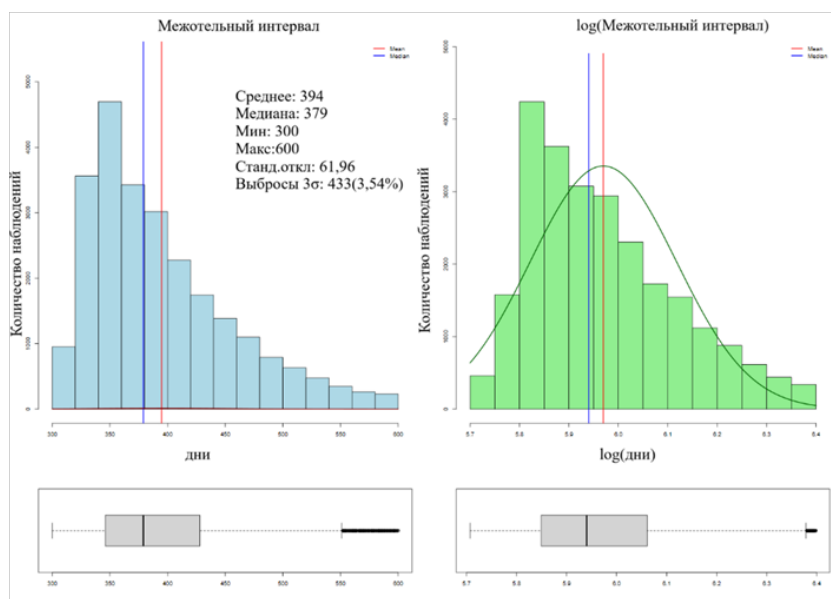


Рис. 1. Сравнение распределений межотельного интервала до и после логарифмирования

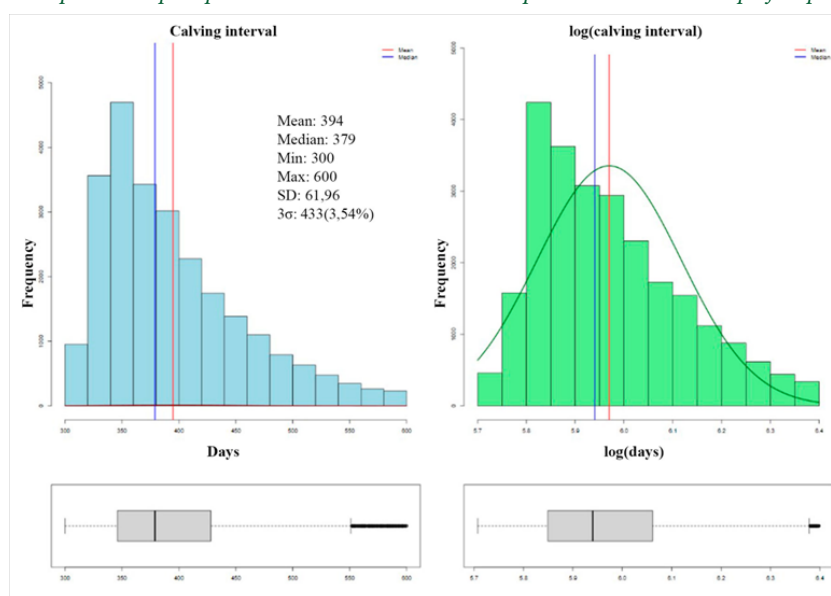


Fig. 1. Distribution of calving interval before and after log-transformation

Продуктивные признаки (удой, выход жира и белка) изначально были близки к нормальному распределению (асимметрия меньше 0,6), и очистка данных эту структуру практически не изменила.

Несмотря на фильтрацию, для отдельных признаков (содержание соматических клеток, число осеменений) сохраняется умеренная положительная асимметрия, обусловленная биологической природой показателей [12].

После исключения значений, выходящих за биологически допустимые границы, выполнен анализ распределений признаков на основе гистограмм и диаграмм размаха для исходных и логарифмированных данных.

Для признаков воспроизводства исходные распределения сохраняют правостороннюю асим-

метрию: наблюдаются концентрация значений в нижней части диапазона и вытянутая граница распределения вправо. После логарифмирования распределения становятся более симметричными, уменьшается длина правого распределения, значения уплотняются в центральной части. На диаграммах размаха это проявляется сокращением количества крайних точек и уменьшением длины линии диапазона значений.

Наиболее заметный эффект логарифмирования наблюдается для признака «дни от отела до первого осеменения»: исходное распределение сильно растянуто, с выраженной правой границей распределения, тогда как после преобразования распределение становится компактнее, а медиана смещается ближе к центру распределения (рис. 2).

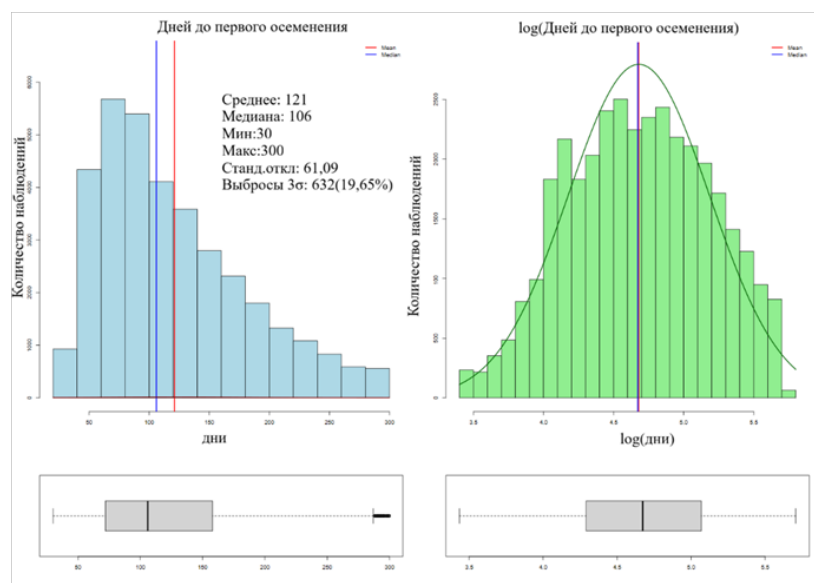


Рис. 2. Сравнение распределений дней до первого осеменения до и после логарифмирования

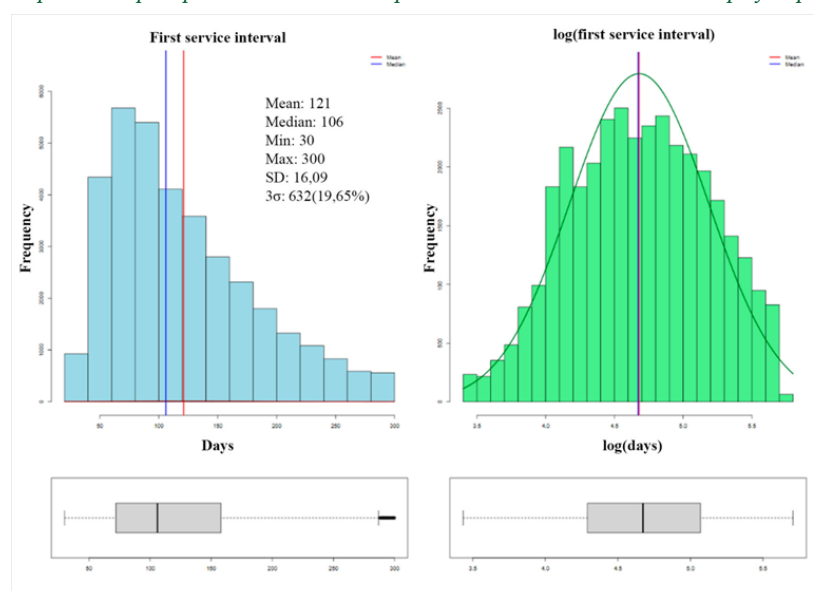


Fig. 2. Distribution of interval to first service before and after log-transformation

Для содержания соматических клеток исходное распределение имеет резко выраженную асимметрию с большим разбросом значений (рис. 3). Логарифмирование существенно изменяет форму распределения: уменьшается разброс, снижается влияние отдельных экстремальных значений, структура становится более равномерной.

Для признаков возраста при первом отеле и возраста результативного осеменения исходные распределения были близки к симметричным (рис. 4). Логарифмирование изменило форму распределения незначительно и не привело к заметному улучшению структуры. Это демонстрируют диаграммы размаха: положение медианы и длина линии диапазона значений практически не меняются.

Продуктивные признаки (удой, выход молочного жира и молочного белка) изначально характеризуются слабой асимметрией и относительно равно-

мерным распределением. После логарифмирования форма распределения мало меняется и остается близкой к исходной. Массовая доля жира и белка распределены компактно, с небольшим разбросом и ограниченным диапазоном значений: логарифмирование не дает заметных изменений структуры.

В матрицу парных корреляций включены восемь признаков, рассматриваемых как непосредственные кандидаты на включение в селекционный индекс. Признаки, применявшиеся только для контроля качества данных, и признаки, дублирующие включенные по информационному содержанию, в основную матрицу не вошли (таблица 4).

Взаимосвязи между продуктивными показателями устойчивы независимо от преобразования данных. Между показателями молочной продуктивности выявлены высокие положительные связи.

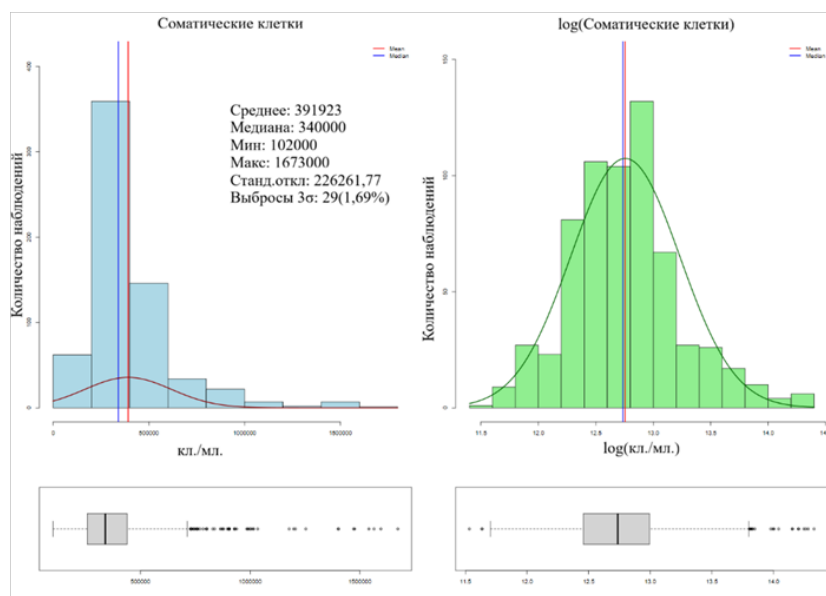


Рис. 3. Распределение признака содержание соматических клеток до и после логарифмирования

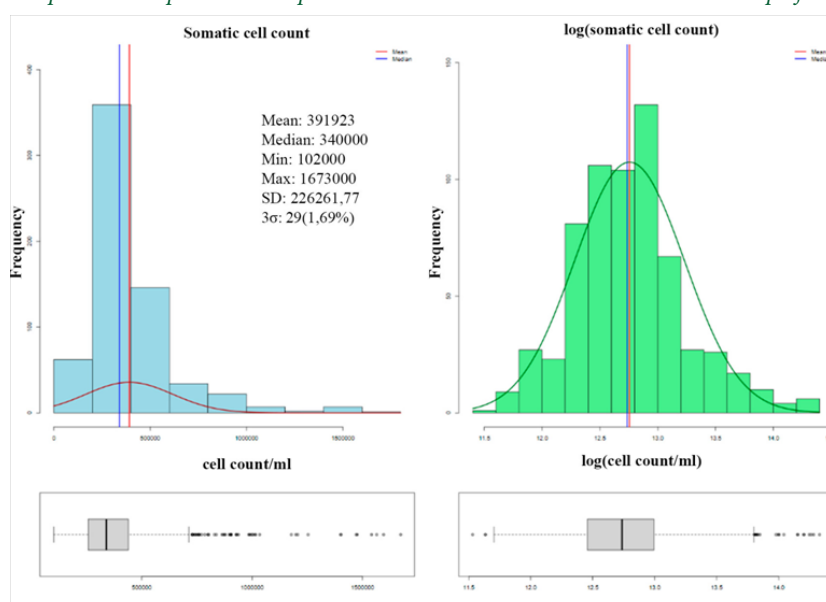


Fig. 3. Distribution of somatic cell count before and after log-transformation

Удой тесно связан с выходом молочного жира ($r = 0,956$) и выходом молочного белка ($r = 0,956$). Показатели выхода молочного жира и молочного белка сильно коррелируют между собой: $r = 0,936$.

Связь между продуктивными и воспроизводственными признаками выражена слабее. Удой имеет умеренную положительную связь с сервис-периодом ($r = 0,271$) и межотельным интервалом ($r = 0,261$). Сравнимые значения получены для показателей выхода молочного жира и молочного белка с продолжительностью сервис-периода и межотельного интервала ($r \approx 0,23...0,25$), что означает, что высокая продуктивность приводит к удлинению межотельного периода.

Между сервис-периодом и межотельным интервалом установлена линейная связь ($r = 0,997$), отражающая прямую зависимость этих признаков.

Возраст первого плодотворного осеменения слабо связан с продуктивными признаками ($r \approx 0,18...0,23$) и умеренно – с индексом осеменения ($r = 0,414$), что приводит к увеличению возраста наступления стельности при увеличении числа осеменений.

Содержание соматических клеток в молоке слабо связано с продуктивностью ($r \approx 0,13...0,21$); с воспроизводственными признаками связь слабая отрицательная (с сервис-периодом $r = -0,104$; межотельным интервалом $r = -0,109$). Число осеменений на оплодотворение не связано с продуктивными признаками ($r < 0,08$). Наблюдается слабая положительная связь этого показателя с воспроизводительными качествами ($r \approx 0,12$) и умеренная – с возрастом первого плодотворного осеменения ($r = 0,414$).

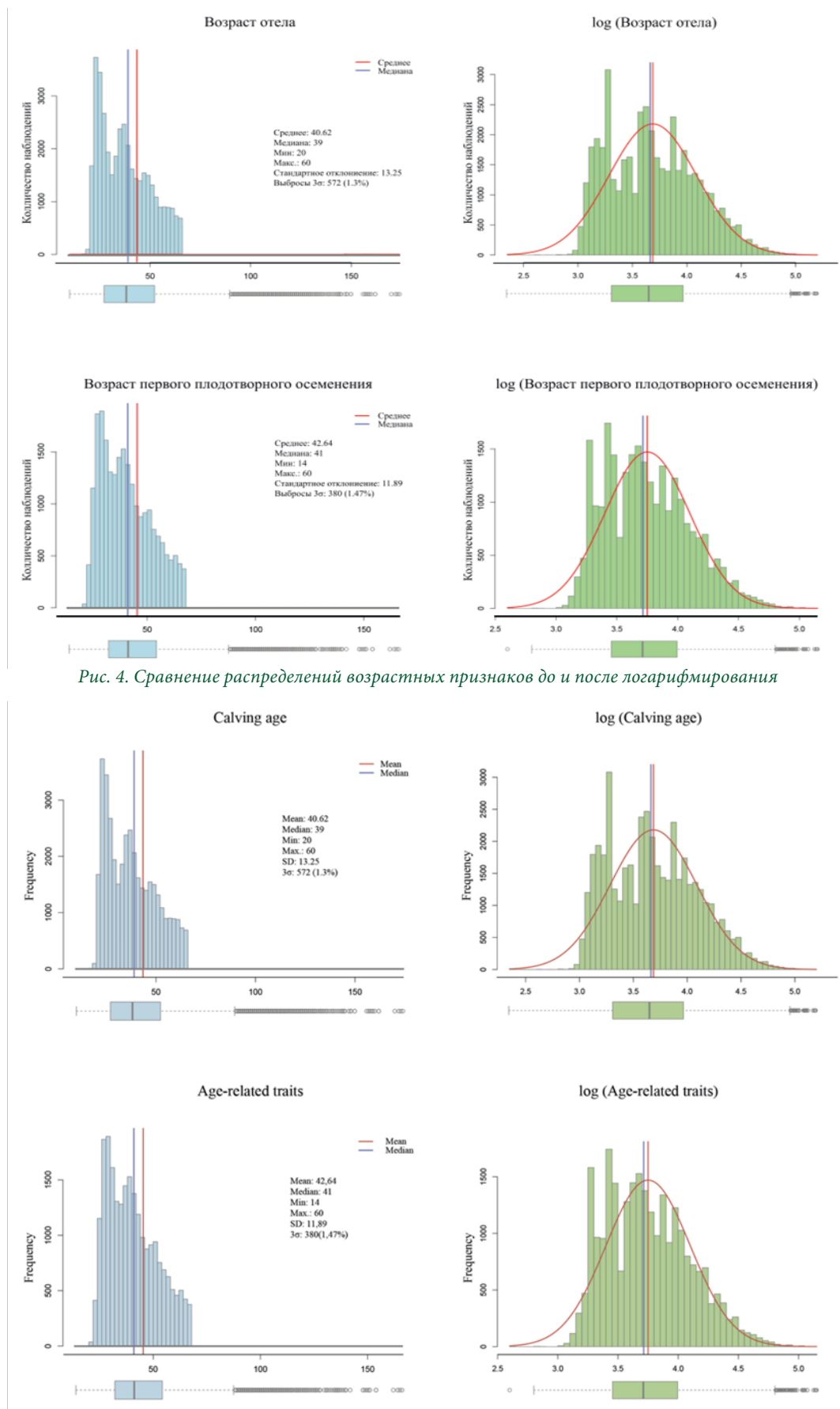


Рис. 4. Сравнение распределений возрастных признаков до и после логарифмирования

Fig. 4. Distribution of age-related traits before and after log-transformation

Таблица 4

Матрица парных корреляций признаков после биологической фильтрации

	1. Удой, кг	2. Выход жира, кг	3. Выход белка, кг	4. Сервис-период, дни	5. Межотельный интервал, дни	6. Возраст первого плодотворного осеменения, мес.	7. Соматические клетки, тыс. кл/мл	8. Кратность осеменения
1. Удой, кг	1	0,956	0,956	0,271	0,261	0,213	0,129	0,055
2. Выход жира, кг	0,956	1	0,936	0,253	0,247	0,176	0,128	0,018
3. Выход белка, кг	0,956	0,936	1	0,240	0,230	0,229	0,208	0,074
4. Сервис-период, дни	0,271	0,253	0,240	1	0,997	0,171	-0,104	0,117
5. Межотельный интервал, дни	0,261	0,247	0,230	0,997	1	0,173	-0,109	0,125
6. Возраст первого плодотворного осеменения, мес.	0,213	0,176	0,229	0,171	0,173	1	0,148	0,414
7. Соматические клетки, тыс. кл/мл	0,129	0,128	0,208	-0,104	-0,109	0,148	1	0,052
8. Кратность осеменения	0,055	0,018	0,074	0,117	0,125	0,414	0,052	1

Table 4

Matrix of pairwise trait correlations after biological filtering

	1. Milk yield, kg	2. Fat yield, kg	3. Protein yield, kg	4. Service period, days	5. Calving interval, days	6. Age at first conception, months	7. Somatic cell count, thousand cells/ml	8. Conception rate (services per conception)
1. Milk yield, kg	1	0.956	0.956	0.271	0.261	0.213	0.129	0.055
2. Fat yield, kg	0.956	1	0.936	0.253	0.247	0.176	0.128	0.018
3. Protein yield, kg	0.956	0.936	1	0.240	0.230	0.229	0.208	0.074
4. Service period, days	0.271	0.253	0.240	1	0.997	0.171	-0.104	0.117
5. Calving interval, days	0.261	0.247	0.230	0.997	1	0.173	-0.109	0.125
6. Age at first conception, months	0.213	0.176	0.229	0.171	0.173	1	0.148	0.414
7. Somatic cell count, thousand cells/ml	0.129	0.128	0.208	-0.104	-0.109	0.148	1	0.052
8. Conception rate (services per conception)	0.055	0.018	0.074	0.117	0.125	0.414	0.052	1

Результаты анализа распределений, доли выбросов и структуры взаимосвязи между признаками легли в основу оценки пригодности их показателей для построения селекционного индекса.

Распределение продуктивных признаков – удоя, выхода молочного жира и молочного белка – характеризуется как стабильное, с низкой долей выбросов и высокой согласованностью показателей между собой ($r = 0,94 \dots 0,96$). Отмечается как высокая информативность, так и высокая избыточность, а потому включение всех трех показателей в индекс дублирует информацию. Значит, при формировании индекса достаточно учитывать или удой, или один из компонентных показателей.

Между сервис-периодом и межотельным интервалом связь практически линейная ($r = 0,997$) – увеличение сервис-периода удлиняет межотельный интервал. Считаем, что одновременно их использо-

вать в индексе нецелесообразно, один из признаков можно взять как производный показатель.

Показатели воспроизводства (сервис-период, межотельный интервал, число дней от отела до первого осеменения) отличаются высокой вариабельностью и выраженной асимметрией исходных распределений. После фильтрации эти признаки сохраняют вариативность и могут быть использованы как характеристика воспроизводительной функции (при условии строгого контроля качества данных) [4; 6].

Содержание соматических клеток в молоке отличается устойчивой положительной асимметрией и сравнительно слабой связью с продуктивными и воспроизводственными признаками. Это указывает на независимый характер изменчивости показателей рассматриваемого признака, его избыточность по отношению к продуктивным и воспроизводительным признакам. Это делает содержание

соматических клеток потенциально ценным дополнительным компонентом индекса, вносящим самостоятельную информацию, не дублируемую другими показателями.

Признак «число осеменений» проявляет умеренную связь с возрастом плодотворного осеменения ($r = 0,41$) и слабую связь с остальными признаками, что позволяет рассматривать этот показатель как отдельный индикатор эффективности воспроизводства.

Обсуждение и выводы (Discussion and Conclusion)

В результате проведенных исследований установлено, что качество первичных записей в базе данных красных молочных пород оказалось крайне неоднородным по группам признаков, и эту неоднородность нельзя игнорировать при построении индекса. Если по удою и компонентам молока доля недостоверных значений не достигает 1 %, то по возрасту первого осеменения исключается почти половина выборки. Два признака, одинаково важные для селекционной цели, имеют несопоставимую надежность исходных данных, что является проблемой, которую прежняя оценка только по удою позволяла не замечать.

Особенно показательны признаки воспроизводства. Их низкая наследуемость хорошо известна. Так, J. Seo и J. Lee для корейской популяции голштинской породы получили оценки наследуемости репродуктивных признаков в пределах 0,007–0,035, то есть почти вся фенотипическая изменчивость этих признаков формируется под влиянием среды [4].

В. J. Hayes и J. E. Pryce подчеркивают, что при низкой наследуемости селекционируемых признаков точность племенной оценки в первую очередь зависит не от модели, а от объема и корректности фенотипических записей [8]. Отсюда следует важный вывод: для признаков с низкой наследуемостью ошибки регистрации более влиятельны, чем для высоконаследуемых признаков.

Показатель наследуемости удоя ($h^2 \approx 0,3$) означает, что большая доля фенотипической изменчивости признака имеет аддитивно-генетическую природу и ошибки учета нивелируются. Каждая некорректная дата осеменения напрямую искажает оценку, так как сервис-период и индекс осеменения имеют низкую генетическую компоненту в общей дисперсии.

Отмеченные нами 48 % выбросов по возрасту первого осеменения и 19 % по интервалу от отела до

первого осеменения – это не дефект записей базы, а фактор, способный обнулить и без того слабый генетический сигнал репродуктивных признаков.

Преимущественно односторонний характер выбросов (основная их часть для возраста первого осеменения и возраста при первом отеле – на верхней границе, для интервала от отела до первого осеменения – на нижней) указывает не на случайные технические ошибки, а на систематические искажения: запоздалую фиксацию событий в хозяйствах либо чрезмерно раннее осеменение. Это относится к практике ручного ввода репродуктивных данных операторами на местах, на уязвимость которой прямо указывают рекомендации ICAR и Interbull [5; 6]. Систематический сдвиг показателей не усредняется при увеличении выборки, а при попадании в модель направленно смещает оценку.

Что касается содержания количества соматических клеток в молоке, то, несмотря на ценные свойства этого признака (информационную независимость от продуктивности и показателей воспроизводства), выводы по нему основываются всего на 651 записи против почти 44 000 по остальным показателям. Это менее 1,5 % выборки. Переносить заключение о качестве данных по количеству соматических клеток на всю популяцию красных пород пока преждевременно. Скудость данных по количеству соматических клеток в молоке – отдельная проблема племенного учета в стадах молочного скота, где систематический контроль здоровья вымени еще не осуществляется. Прежде чем включать показатель количества соматических клеток в индекс, потребуется накопить репрезентативный массив измерений.

Таким образом, полученные результаты задают конкретный порядок действий при подготовке фенотипических данных красных пород к индексной оценке. Рекомендуем разнести признаки по уровню доверия: показатели продуктивности пригодны к расчетам почти без потерь; показатели воспроизводительных признаков требуют обязательной биологической фильтрации относительно с физиологическими границами; содержание соматических клеток в молоке требует накопления данных. Такой дифференцированный подход к качеству записей предпочтительнее, чем единообразная обработка показателей всех признаков, участвующих при формировании селекционного индекса.

Библиографический список

1. Решение Коллегии Евразийской экономической комиссии от 22 августа 2023 г. № 125 «О внесении изменений в Методику оценки племенной ценности крупного рогатого скота молочного направления продуктивности» [Электронный ресурс]. URL: <https://docs.eaeunion.org/documents/414/7601> (дата обращения: 16.06.2026).
2. Cole J. B. Evolution of selection objectives in dairy cattle breeding // Journal of Dairy Science. 2023. Vol. 106, No. 4. Pp. 2345–2356. DOI: 10.3168/jds.2022-22381.

3. Miglior F., Muir B. L., Van Doormaal B. J. Modern total merit indices in dairy breeding // *Journal of Dairy Science*. 2021. Vol. 104, No. 2. Pp. 1476–1489. DOI: 10.3168/jds.2020-18769.
4. Seo J., Lee J. Estimation of genetic parameters for reproductive traits in Korean dairy cattle // *Animal Bioscience*. 2024. Vol. 38, No. 1. Pp. 12–22. DOI: 10.5713/ab.24.0432.
5. Guidelines for dairy cattle genetic evaluation: ICAR Technical Series. Rome: International Committee for Animal Recording, 2023 [Электронный ресурс]. URL: <https://www.icar.org/Guidelines/09-Dairy-Cattle-Genetic-Evaluation.pdf> (дата обращения: 05.06.2026).
6. Egger-Danner C., Simianer H., Cole J. B. International gaps in genomic evaluation of dairy cattle // *Interbull Bulletin*. 2022. No. 61. Pp. 15–28.
7. Weller J. I., Gershoni M., Ezra E. Breeding dairy cattle for female fertility and production in the age of genomics // *Veterinary Sciences*. 2022. Vol. 9, No. 8. Pp. 434. DOI: 10.3390/vetsci9080434.
8. Hayes B. J., Pryce J. E. Prediction accuracy under selection in dairy cattle // *Animal Genetics*. 2022. Vol. 53, No. 2. Pp. 165–176. DOI: 10.1111/age.13168.
9. Захарова Л. Н. Оценка коров-первотелок красной степной породы по показателям молочной продуктивности и живой массы в условиях Якутии // *Главный зоотехник*. 2021. № 6. С. 35–43. DOI: 10.33920/sel-03-2106-05.
10. Министерство сельского хозяйства Российской Федерации. Разработка унифицированной программы расчета племенной ценности крупного рогатого скота молочного и мясного направлений продуктивности: отчет о НИР [Электронный ресурс]. Москва, 2023. URL: <https://eec.eaeunion.org/upload/iblock/64f/OTCHET.pdf> (дата обращения: 05.06.2026).
11. Сермягин А. А., Белоус А. А., Контэ А. Ф. [и др.]. Валидация геномного прогноза племенной ценности быков-производителей по признакам молочной продуктивности дочерей на примере популяции черно-пестрого и голштинского скота // *Сельскохозяйственная биология*. 2020. Т. 55, № 2. С. 257–268. DOI: 10.15389/agrobiology.2020.2.257rus.
12. Pralle R. S., White H. M. Symposium review: Big data, machine learning, and digital phenotyping in dairy cattle research and management // *Journal of Dairy Science*. 2020. Vol. 103, No. 4. Pp. 3702–3713. DOI: 10.3168/jds.2019-17228.
13. Тусупов О. К., Алтухова Н. С., Харитонов С. Н. Племенная ценность [Электронный ресурс] // *Большая российская энциклопедия: научно-образовательный портал*. 2023. URL: <https://bigenc.ru/c/plemennaia-tsennost-9baee2> (дата обращения: 05.06.2026).
14. Berry D. P. Invited review: Incorporating resilience traits in dairy cattle selection // *Animal*. 2022. Vol. 16, Suppl. 2. Article number 100395. DOI: 10.1016/j.animal.2021.100395.
15. Сермягин А. А. Племенная оценка скота как основа животноводства [Электронный ресурс] // *Аграрные новости*. 2024. № 3. URL: <https://agri-news.ru/zhurnal/2024/3-2024/voprosy-plemennoy-otsenki/> (дата обращения: 05.06.2026).

Об авторах:

Арсений Павлович Шевчук, научный сотрудник лаборатории разведения отечественных и зарубежных пород скота, Всероссийский научно-исследовательский институт племенного дела, п. Лесные Поляны, Московская область, Россия; ORCID 0009-0002-7257-3097, AuthorID 1197315.

E-mail: vniiplem.kras@outlook.com

Татьяна Александровна Князева, ведущий научный сотрудник лаборатории разведения отечественных и зарубежных пород скота, Всероссийский научно-исследовательский институт племенного дела, п. Лесные Поляны, Московская область, Россия; ORCID 0000-0003-3022-9538, AuthorID 626830.

E-mail: red-step@mail.ru

References

1. *Decision of the Board of the Eurasian Economic Commission of 22 August 2023 No. 125 “On amendments to the Methodology for assessing the breeding value of dairy cattle”* [Internet]. Available from: <https://docs.eaeunion.org/documents/414/7601> (cited 2026 Jun 16). (In Russ.)
2. Cole J. B. Evolution of selection objectives in dairy cattle breeding. *Journal of Dairy Science*. 2023; 106 (4): 2345–2356. DOI: 10.3168/jds.2022-22381.
3. Miglior F., Muir B. L., Van Doormaal B. J. Modern total merit indices in dairy breeding. *Journal of Dairy Science*. 2021; 104 (2): 1476–1489. DOI: 10.3168/jds.2020-18769.
4. Seo J., Lee J. Estimation of genetic parameters for reproductive traits in Korean dairy cattle. *Animal Bioscience*. 2024; 38 (1): 12–22. DOI: 10.5713/ab.24.0432.

5. *Guidelines for dairy cattle genetic evaluation: ICAR Technical Series* [Internet]. Rome: International Committee for Animal Recording, 2023 [cited 2026 Jun 05]. Available from: <https://www.icar.org/Guidelines/09-Dairy-Cattle-Genetic-Evaluation.pdf>.
6. Egger-Danner C., Simianer H., Cole J. B. International gaps in genomic evaluation of dairy cattle. *Interbull Bulletin*. 2022; 61: 15–28.
7. Weller J. I., Gershoni M., Ezra E. Breeding dairy cattle for female fertility and production in the age of genomics. *Veterinary Sciences*. 2022; 9 (8): 434. DOI: 10.3390/vetsci9080434.
8. Hayes B. J., Pryce J. E. Prediction accuracy under selection in dairy cattle. *Animal Genetics*. 2022; 53 (2): 165–176. DOI: 10.1111/age.13168.
9. Zakharova L. N. Evaluation of first-calf heifers of Red Steppe breed by indicators of milk productivity and live weight under the environments of Yakutia. *Head of Animal Breeding*. 2021; 6: 35–43. DOI: 10.33920/sel-03-2106-05. (In Russ.)
10. Ministry of Agriculture of the Russian Federation. *Development of a unified program for calculating the breeding value of dairy and beef cattle: research report* [Internet]. Moscow, 2023 [cited 2026 Jun 05]. Available from: <https://ecc.eaeunion.org/upload/iblock/64f/OTCHET.pdf>. (In Russ.)
11. Sermyagin A. A., Belous A. A., Konte A. F., et al. Validation of genomic prediction of breeding value of sires for daughters' milk production traits in the population of Black-and-White and Holstein cattle. *Sel'skokhozyaystvennaya biologiya*. 2020; 55 (2): 257–268. DOI: 10.15389/agrobiology.2020.2.257eng. (In Russ.)
12. Pralle R. S., White H. M. Symposium review: Big data, machine learning, and digital phenotyping in dairy cattle research and management. *Journal of Dairy Science*. 2020; 103 (4): 3702–3713. DOI: 10.3168/jds.2019-17228.
13. Tusupov O. K., Altukhova N. S., Kharitonov S. N. Breeding value. *Great Russian Encyclopedia: scientific and educational portal* [Internet]. 2023 [cited 2026 Jun 05]. Available from: <https://bigenc.ru/c/plemennaia-tsennost-9baee2>. (In Russ.)
14. Berry D. P. Invited review: Incorporating resilience traits in dairy cattle selection. *Animal*. 2022; 16 (2): 100395. DOI: 10.1016/j.animal.2021.100395.
15. Sermyagin A. A. Breeding evaluation of cattle as the basis of animal husbandry. *Agrarnye Novosti* [Internet]. 2024 [cited 2026 Jun 05]; 3. Available from: <https://agri-news.ru/zhurnal/2024/3-2024/voprosy-plemennoy-otsenki>. (In Russ.)

Authors' information:

Arseniy P. Shevchuk, researcher of the laboratory for breeding red breeds of cattle, All-Russian Scientific Research Institute of Breeding, Lesnye Polyany settlement, Moscow region, Russia;
ORCID 0009-0002-7257-3097, AuthorID 1197315. *E-mail*: vniiplem.kras@outlook.com

Tatyana A. Knyazeva, leading researcher of the laboratory for breeding red breeds of cattle, All-Russian Scientific Research Institute of Breeding, Lesnye Polyany settlement, Moscow region, Russia;
ORCID 0000-0003-3022-9538, AuthorID 626830. *E-mail*: red-step@mail.ru