

Многофакторный регрессионный анализ молочных признаков коров голштинской породы

Г. Г. Карликова¹✉, А. Ф. Контэ¹

¹Федеральный исследовательский центр животноводства – ВИЖ имени академика

Л. К. Эрнста

✉E-mail: galinakarlikova@yandex.ru

Аннотация. Цель исследований – возможность применения уравнения множественной двухфакторной регрессии для раскрытия достоверной сопряженности признаков, влияющих на уровень молочной продуктивности коров голштинской породы. **Методы исследований.** Исследования проводились в ЗАО СП «Аксиньино» Ступинского района Московской области. На основе ИАС «СЕЛЭКС» была создана база данных, включающая выборку по 11 017 головам. Данные о молочной продуктивности были взяты за 305 дней 1-й, 2-й, 3-й и максимальной лактаций. **Результаты.** За 305 дней 1-й лактации удой составил в среднем 7909,5 кг молока, за 2-ю – 8289,1 кг ($p \leq 0,001$), за 3-ю – 8446,2 кг ($p \leq 0,001$). Удой за максимальную лактацию составил 8964,3 кг молока ($p \leq 0,001$). Содержание жира и белка в молоке коров в промежутке между 1-й и 3-й лактацией составляет 4,03–4,08 и 3,22–3,23 % соответственно. Коэффициент множественной двухфакторной регрессии представляет смещение отклика от 7787,81–8239,00 (1-я, 2-я, 3-я лактации) до 8841,63 (максимальная лактация). Диаграммы рассеяния модели множественной регрессии показывают, что значение переменной величины «удой» статистически зависимо от показателей массовой доли жира и массовой доли белка в молоке. Коэффициент детерминации на уровне 0,997–0,998 свидетельствует о том, что уравнением регрессии объясняется 99,7–99,8 % дисперсии результативного признака. Значимость F -критерия Фишера говорит о высокой достоверности результатов и отсутствии случайности и наличия оправданной в нашем исследовании закономерности. Содержание жира и белка имело отрицательную связь с удоем ($p \leq 0,001$) по лактациям. Получены высокие значения взаимосвязи между уровнем удоя и содержанием белка (–0,518...–0,766), содержанием жира и белка (0,626–0,784). **Научная новизна.** Впервые проведены исследования на соответствие модели уравнения полученным экспериментальным данным и наличие количества независимых переменных (массовая доля жира и белка), включенных в уравнение для описания зависимой переменной (удой).

Ключевые слова: множественная регрессия, коровы, лактация, удой, массовая доля жира, массовая доля белка.

Для цитирования: Карликова Г. Г., Контэ А. Ф. Многофакторный регрессионный анализ молочных признаков коров голштинской породы // Аграрный вестник Урала. 2021. № 11 (214). С. 27–35. DOI: 10.32417/1997-4868-2021-214-11-27-35.

Дата поступления статьи: 08.09.2021, **дата рецензирования:** 17.09.2021, **дата принятия:** 24.09.2021.

Постановка проблемы (Introduction)

Увеличение спроса на продукцию животноводства на фоне снижения количества животных сельскохозяйственной отрасли можно нивелировать в основном за счет роста их генетического потенциала продуктивных показателей на основе новых методов разведения крупного рогатого скота. И мировая, и отечественная история скотоводства указывают на то, что только достаточно высокий уровень селекционной работы дает возможность скорого повышения как молочной, так и непосредственно мясной продуктивности крупного рогатого скота. На историческом отрезке развития племенной работы с молочным скотом применялись разные подходы и методы со своими особенностями в достижении результата – получение коров с высокой продуктивностью с целью интенсивного использования [3, с. 36]. В России молочная продуктивность

коров как хозяйственно полезный признак является приоритетом в племенной работе отрасли молочного животноводства [1]. Большое количество племенных хозяйств имеет достаточно высокие результаты в увеличении продуктивных показателей животных, однако возросшие темпы производства указывают на потребность в увеличении объемов племенной работы.

Племенная работа – это комплекс зоотехнических и организационных мероприятий, нацеленных на увеличение продуктивности, а также на улучшение племенных качеств молочного скота и использование животных в соответствии с региональными особенностями.

Генетика занимает главное положение в основах теории селекции. Именно это и характеризует этап перехода зоотехники как науки от генетики определенных особей к популяционной генетике. Поэтому на

данный момент значение популяционной генетики все возрастает и ее достижения служат основой крупномасштабной селекции. В популяционной генетике расчеты проводятся с большими группами животных [2, с. 227], [4].

Основным методом исследований служит математическая статистика, основанная на закономерностях случайного распределения большого числа вариантов, установления параметров изменения популяций, разработки эффективных программ селекции. Последние годы характеризуются большей концентрацией внимания на селекционных программах по крупному рогатому скоту, опирающихся на молекулярно-генетические методы исследования [4].

Успехи популяционной генетики в последнее время играют важную роль в практике и теории племенной работы, повышается значение биометрии, дающей возможность изучать морфологическую и физиологическую изменчивости признаков, выявлять влияние детерминации отдельных факторов на успех селекции, и непосредственно биотехнологии (генная инженерия, получение химерных животных, трансплантация эмбрионов и др.) [3, с. 37].

При селекции по какому-либо из признаков необходимо установить степень и тенденции взаимосвязи с другими признаками, направленными на выявление и математическое выражение изменений и зависимостей. Путем регрессионного анализа выявляются и математически выражаются результаты научного эксперимента и зависимости между ними [5].

Конструирование модели регрессии характеризуется итерационный процесс, нацеленный на нахождение эффективных переменных, не зависящих друг от друга, для объяснения зависимых переменных, которые необходимо понять и смоделировать, опираясь на метод регрессии, с целью определения эффективных предсказателей из исследуемых величин.

Регрессионное уравнение не только задает форму исследуемой связи, но также указывает на степень изменчивости одного из показателей, сопровождаемого изменчивостью другого показателя. При построении уравнения многофакторной регрессии решается вопрос о специфике данной модели, в соответствии с чем выбираются факторы влияния и определяется вид уравнения регрессии. Добавление в модель уравнения того или иного количества факторов обуславливается природой взаимосвязи данного моделируемого показателя с другими признаками [5], [6, с. 5].

После определения линейного регрессионного уравнения в целом оценивается значимость его как формулы, а также его отдельных параметров. Оценить значимость регрессионного уравнения – следовательно, найти соответствие модели анализа, экспликацию взаимозависимости переменных, опытным данным, а также установить, достаточно ли число объясняющих переменных уравнения (одна, две или несколько) для характеристики зависимой переменной [7, с. 126]. Уточнение значимости регрессионного уравнения в основном основывается на F -критерии Фишера [8, с. 9], [9].

Расчет корреляционного значения как один из статистических методов анализа совместно с биологическими методами предоставляет возможность селекционерам установить взаимосвязь молочных признаков, учитываемых при разведении животных [10], [11]. Как известно, отдельные признаки обладают различной степенью корреляционного отношения. Положительная генетическая корреляция обеспечивает селекцию выдающихся животных по отдельным признакам, что обуславливает непосредственное улучшение не только одного признака, но и группы взаимосвязанных с ним. Отрицательная же корреляция указывает на то, что селекция по отдельному признаку может привести к ухудшению коррелируемого с ним другого признака [12]. Корреляционные связи существуют среди как количественных, так и качественных признаков [13], [14].

Для определения достоверного влияния совокупности факторов на параметры молочности коров мы посчитали необходимым рассмотреть множественную регрессию, в связи с чем целью проводимой работы явилось определение факторов, влияющих на удои коров голштинской породы при помощи уравнения множественной двухфакторной регрессии для раскрытия достоверной сопряженности признаков.

Задачи исследования:

- определить характер прямой регрессии между признаками молочной продуктивности;
- произвести расчет коэффициентов множественной регрессии параметров молочности;
- дать характеристику генетическим корреляциям между удоем и массовой долей жира и белка;

Методология и методы исследования (Methods)

Объектом наших исследований были коровы голштинской породы ЗАО СП «Аксиньино» Ступинского района Московской области. В целях проведения исследования молочной продуктивности коров была создана выборка данных на основе базы ИАС «СЕ-ЛЭКС», включающая данные по 11 017 головам. Нами использовались данные первичного племенного и зоотехнического учета. Данные о молочной продуктивности были взяты за 305 дней первой, второй, третьей и наивысшей лактации. Одними из главных селекционируемых признаков были удои, жирномолочность и белкомолочность, соотнесенные к удою по каждой из лактаций.

Принимая во внимание, что в нашем исследовании используются две независимые переменные (МДЖ и МДБ) и между ожиданием (зависимой переменной – удои) и каждой из переменных (k) существует линейная зависимость, модель множественной линейной регрессии будет иметь следующий вид (1):

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (1)$$

где $a_0, a_1, a_2, \dots, a_n$ – значения коэффициентов регрессии,

x_1, x_2, \dots, x_n – независимые переменные.

В данной формуле:

a_0 – смещение,

a_1 – отклонение прямой y , зависящей от значения x_1 , если x_2, \dots, x_n – это константы;

a_2 – отклонение прямой y , зависящей от значения x_2 , если x_1, \dots, x_n – константы;

a_n – отклонение прямой y , зависящей от значения x_n , если x_1, x_2, \dots, x_{n-1} , служащих константами.

Расчет коэффициентов множественной регрессии производился методом наименьших квадратов системы нормальных уравнений (2):

$$\begin{cases} na_0 + a_1 \sum x_1 + \dots + a_p \sum x_p = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + \dots + a_p \sum x_1 x_p = \sum y x_1 \\ \dots \\ a_0 \sum x_p + a_1 \sum x_1 x_p + \dots + a_p \sum x_p^2 = \sum y x_p \end{cases} \quad (2)$$

В нашем случае используется двухфакторная модель, и система нормальных уравнений имеет следующий вид:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum y x_1 \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum y x_2 \end{cases} \quad (3)$$

Расчеты проведены с использованием программы STATISTICA 10.

Результаты (Results)

В исследовании мы использовали данные молочной продуктивности коров по первой, второй, третьей и максимальной лактациям. Удой служил зависимой

переменной, а жирномолочность и белковомолочность – независимые переменные.

Анализируя данные продуктивного использования животных, необходимо отметить, что с возрастанием номера лактации удои увеличиваются. Исследуемые коровы за 305 дней первой лактации в среднем дали 7909,5 кг молока, то есть без 90 кг молока 8000. За вторую лактацию удои составил 8289,1 кг, что на 379,6 кг ($P \leq 0,001$) больше по сравнению с первой. По третьей лактации продукция молока возросла до 8446,2 кг молока, что на 157,1 кг больше, чем за вторую лактацию, и на 536,7 кг ($P \leq 0,001$) больше, чем за первую. За максимальную лактацию от коров получено на 1054,8 кг больше, чем за первую лактацию, на 675,2 кг больше, чем за вторую, и на 518,1 кг ($P \leq 0,001$) больше по сравнению с третьей лактацией.

Белковомолочность и жирномолочность служат основными показателями, характеризующими продуктивность молочных животных. Жирномолочность и белковомолочность в промежутке между первой и третьей лактациями особо не меняется.

Так, процент массовой доли жира за первую лактацию составил 4,04, за вторую – 4,05, за третью лактацию снижается – 4,03. За период максимально продуктивной лактации процент жира составил 4,08.

Таблица 1
Молочная продуктивность голштинских коров

Показатели	Лактация			
	1-я	2-я	3-я	Максимальная
Удой, кг	7909,5 ± 32	8289,1 ± 37	8446,2 ± 47	8964,3 ± 33
Массовая доля жира, %	4,04 ± 0,005	4,05 ± 0,006	4,03 ± 0,007	4,08 ± 0,005
Массовая доля белка, %	3,22 ± 0,003	3,23 ± 0,003	3,22 ± 0,004	3,22 ± 0,003

Table 1
Dairy productivity of Holstein cows

Indicators	Lactation			
	1 st	2 nd	3 rd	Maximum
Milk yield, kg	7909.5 ± 32	8289.1 ± 37	8446.2 ± 47	8964.3 ± 33
Mass fraction of fat, %	4.04 ± 0.005	4.05 ± 0.006	4.03 ± 0.007	4.08 ± 0.005
Mass fraction of protein, %	3.22 ± 0.003	3.23 ± 0.003	3.22 ± 0.004	3.22 ± 0.003

Таблица 2
Уравнения множественной регрессии по молочным показателям ($n = 11\ 017$)

Лактация	Уравнения
1-я	$y = 7787,81 - 519,19x_1 - 1747,48x_2$
2-я	$y = 8123,84 - 591,94x_1 - 1756,33x_2$
3-я	$y = 8239,00 - 802,77x_1 - 1542,16x_2$
Максимальная	$y = 8841,63 - 661,60x_1 - 1893,83x_2$

Примечание. X_1 – содержание жира в молоке; x_2 – содержание белка в молоке.

Table 2
Multiple regression equations for milk indicators ($n = 11\ 017$)

Lactation	Equations
1 st	$y = 7787,81 - 519,19x_1 - 1747,48x_2$
2 nd	$y = 8123,84 - 591,94x_1 - 1756,33x_2$
3 rd	$y = 8239,00 - 802,77x_1 - 1542,16x_2$
Maximum	$y = 8841,63 - 661,60x_1 - 1893,83x_2$

Note. X_1 – fat content in milk; x_2 – protein content in milk.

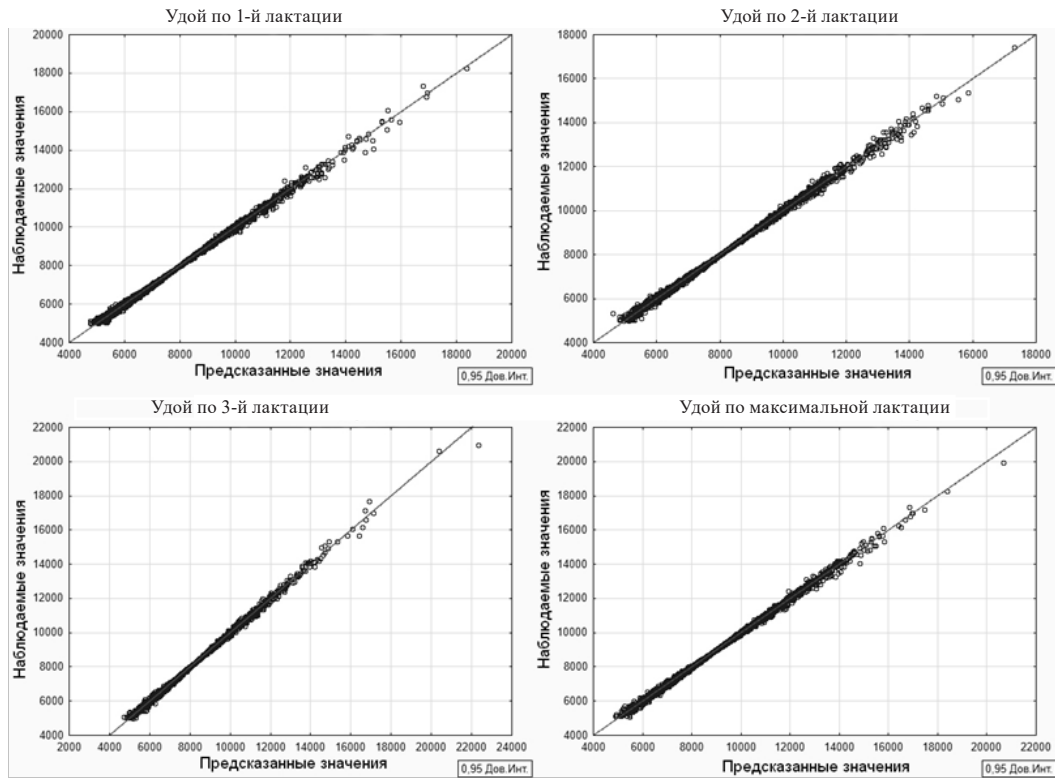


Рис. 1. Наблюдаемые и предсказанные значения

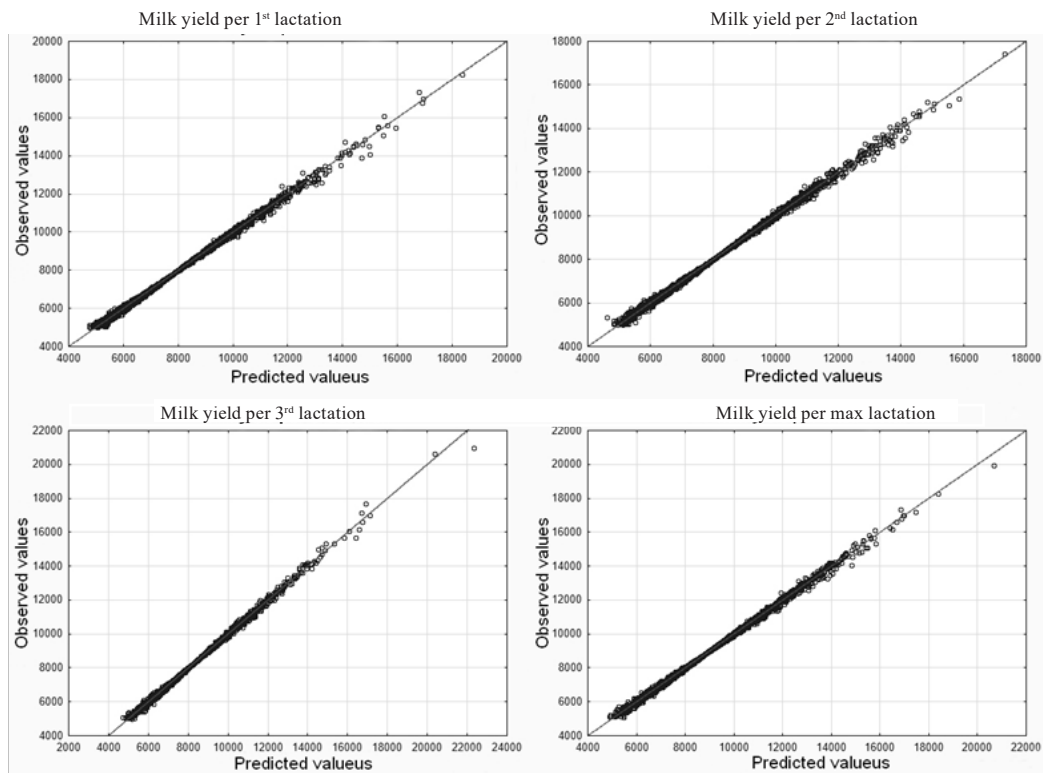


Fig. 1. Observed and predicted values

Касательно белкомолочности необходимо отметить, что за первую лактацию содержание составило 3,22 %, за вторую – 3,23 % и за третью – 3,22 %. За период максимальной лактации процент массовой доли белка составил 3,22, что сравнимо с результатами этого показателя за первую и третью лактации. В ходе проведенных расчетов получили значения коэффициентов уравнения двухфакторной регрессии $a_{0...n}$, при

подстановке которых в уравнение (1) находим данные уравнения регрессии (таблица 2).

Значение коэффициента множественной двухфакторной регрессии a_0 является предсказанным значением показателей, перечисленных в первой части таблицы 2 при значении факторов x_1, x_2 , равном нулю, и представляет смещение отклика, равного 7787,81–8841,63.

Значения коэффициентов регрессии $a_1 \dots a_2$ зависят от значения знака перед ними. Положительное значение коэффициента регрессии указывает на положительное взаимодействие фактора с исследуемой переменной, а отрицательное значение указывает на обратное влияние фактора. В нашем случае полученные уравнения регрессии показывают, что показатель «удой» по всем лактациям имеет сильное, но отрицательное взаимодействие с содержанием жира и белка.

Для визуализации связи между прогнозируемым значением Y и переменной X нами были построены диаграммы рассеяния (разброса, поля корреляции) по каждой из лактаций. Данные изображали ординатами в двумерной плоскости. Координаты (X и Y), определяющие нахождение каждой ординаты, представляют значения двух переменных. Наши исследования показывают, что значения переменных обладают сильной взаимосвязью, что указывает на то, что множество точек данных обладает определенной формой (расположены на прямой линии или кривой, характеризуемой уравнением), как представлено на рис. 1. Точки очень тесно укладываются вдоль прямой линии, что является общей тенденцией для построенных нами графиков по всем лактациям, взятых нами для расчетов.

Построенные диаграммы рассеяния модели множественной регрессии довольно наглядно отражают изменения молочных признаков по всем взятым лактациям.

Диаграммы рассеяния показывают, что значение переменной величины «удой» в наших исследованиях статистически зависит от показателей жирномолочности и белкомолочности в течение лактации.

Визуально оценив результаты диаграмм, мы можем в последующем применить регрессионный анализ для нахождения зависимости между признаками.

В наших исследованиях характер прямой регрессии имеет отрицательное значение, то есть с увеличением определенного x_n на какую-то величину в уравнении регрессии при условии, что остальные значения $x_{1 \dots n}$ будут равны 0, R понизится на величину b_n (при условии отрицательного знака «-» перед ними).

На характер адекватности полученных уравнений могут указывать показатели регрессионной статистики (таблица 3).

Анализируемый показатель «удой» по лактациям имеет высокий уровень линейной зависимости (множественный коэффициент корреляции $R > 0,70$).

Как известно, коэффициент детерминации (R^2) оценивает «силу влияния» переменной x на y . При оценке моделей регрессии он характеризует соответствие модели данным. По мнению исследователей, приемлемый коэффициент детерминации моделей должен быть на уровне 50 % (тогда в данном случае множественный коэффициент корреляции превышает 70 % по модулю). Модели, обладающие коэффициентом детерминации свыше 80 %, достаточно эффективные (значение коэффициента корреляции более 90 %). Коэффициент детерминации, равный единице, указывает на функциональную созависимость переменных.

Интерпретация результатов наших исследований: $R^2 > 0,95$ – высокая точность аппроксимации. Коэффициент детерминации, полученный в наших расчетах на уровне 0,997–0,998, свидетельствует о том, что регрессионное уравнение характеризуется 99,7–99,8 % дисперсии результативного признака, а доля прочих факторов составляет лишь 1,2–1,3 %.

Рассчитанная высокая регрессионная зависимость говорит о том, что при селекции можно учитывать 1–2 показателя из полученных уравнений. Однако созависимости признаков могут обладать самыми разными формами. Также при наиболее кропотливом анализе связей следует учитывать, что на результативный признак обычно влияет не один фактор, а несколько.

При сопоставлении моделей с разным набором факторов используется скорректированный детерминационный коэффициент, обладающий несмещенными оценками дисперсий, для того чтобы число регрессоров не оказывало влияния на статистику. Этот показатель всегда меньше единицы, теоретически он может быть и меньше нуля. Но применение показателя в сравнении вполне обоснованно.

Таблица 3
Регрессионная статистика молочных показателей

Показатели	Удой по лактациям, кг			
	1-я	2-я	3-я	Максимальная
Множественный R	0,99896	0,9989	0,99891	0,99911
Множественный R^2	0,99792	0,9979	0,99782	0,99822
Скорректированный R^2	0,99792	0,9978	0,99781	0,99822
$F(4,3181)$	382 778,11	298 139,87	210 004,44	481 123,92
Стандартная ошибка оценки	82,3944	86,4204	94,6586	81,0248

Table 3
Regression statistics of dairy indicators

Indicators	Milk yield by lactation, kg			
	1 st	2 nd	3 rd	Maximum
Multiple R	0.99896	0.9989	0.99891	0.99911
Multiple R^2	0.99792	0.9979	0.99782	0.99822
Corrected R^2	0.99792	0.9978	0.99781	0.99822
$F(4.3181)$	382 778.11	298 139.87	210 004.44	481 123.92
Standard evaluation error	82.3944	86.4204	94.6586	81.0248

Выявление взаимосвязей между показателями молочной продуктивности коров

Признаки	Удой, кг			
	1-я лактация	2-я лактация	3-я лактация	Максимальная лактация
Содержание жира, %	-0,581 ± 0,006	-0,220 ± 0,003	-0,379 ± 0,007	-0,469 ± 0,004
Содержание белка, %	-0,606 ± 0,006	-0,766 ± 0,010	-0,518 ± 0,009	-0,672 ± 0,006

Table 4

Identification of interrelations between indicators of dairy productivity of cows

Indicators	Milk yield, kg			
	1 st lactation	2 nd lactation	3 rd lactation	Maximum lactation
Fat content, %	-0,581 ± 0,006	-0,220 ± 0,003	-0,379 ± 0,007	-0,469 ± 0,004
Protein content, %	-0,606 ± 0,006	-0,766 ± 0,010	-0,518 ± 0,009	-0,672 ± 0,006

Значимость уравнения регрессии оценивается в основном на основе F -критерия Фишера. Оценка значимости F указывает на достаточно высокую достоверность значений результатов и отсутствие случайности в нашем исследовании.

Регрессионное уравнение всегда сопряжено со значением тесноты связи. Применению линейной регрессии сопутствует линейный коэффициент корреляции. Необходимо обратить непосредственное внимание на корреляционные связи между удоем, содержанием жира и белка (таблица 4).

Анализируя данные таблицы 4, необходимо отметить, что, как и в случае расчета коэффициента регрессии, содержания жира и белка имеют достоверно отрицательную связь ($p \leq 0,001$) с удоем по всем наблюдаемым лактациям. При этом отмечена умеренная связь удоя и содержания жира по первой лактации (-0,581) удоя и содержания белка (-0,606). По второй лактации связь удоя и содержания жира довольно слабая (-0,220), при этом сильная связь удоя и содержания белка (-0,766). Связь удоя и содержания жира по третьей лактации слабая (-0,379), но отмечена умеренная связь удоя и содержания белка (-0,518). По максимальной лактации к умеренным относится как связь удоя и содержания жира (-0,469), так и связь удоя и содержания белка (-0,672).

Что касается связи между содержанием жира и белка, то она характеризуется значениями от умеренных до сильных положительных (0,626–0,784).

Обсуждение и выводы (Discussion and Conclusion)

Современные технологии теоретического анализа данных служат основой управления новейшими процессами популяционной генетики на современном этапе, от этого зависит и принятие конструктивных решений по дальнейшему преобразованию. Статистика дает грамотный подход к сбору данных, характеристике их структуры, осмыслению и определению закономерности массы явлений, выявлению сложного процесса, скрываемого количеством информации. Она дает возможность исследовать ее, а также прийти к доказательно правильному решению. Статистика в науке обладает комплексом методов, требуемых для

осуществления последовательной исследовательской работы. Применение методов статистического анализа дает понимание характера изменчивости показателей и способствует получению эффективных и правильных результатов и решений.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель [15], [16].

Точный контроль факторов, вовлеченных в регрессию, служит главным из этапов применения методологии регрессии в практике. Принципы подхода к установлению факторов в зависимости от значений корреляций могут быть различными. Непосредственно они могут приводить к разным методикам конструирования уравнения множественной регрессии. В зависимости от того, какая методика составления регрессионного уравнения принята, изменяется последовательность ее решения [17].

Для графического представления взаимосвязей между показателями были построены диаграммы, наглядно показывающие данные и дающие возможность проверки предположений, до установления значений корреляции [18], [19, р. 1552]. Обращая внимание на созависимость между признаками, вполне можно предсказать наличие у них корреляционной связи. В то время как одна из переменных растет, другая тоже увеличивается [20, р. 4486].

Как известно, знак коэффициента корреляции очень важен для интерпретации полученной связи. При отрицательном коэффициенте корреляции наибольшему значению одного из признаков будет сопоставлено меньшее значение другого. Это говорит о том, что при увеличении одной из переменных будет способствовать уменьшению другой переменной [21, р. 11599].

Построив модель множественной линейной регрессии, найдя параметры выбранного уравнения, оценив надежность уравнения регрессии и его параметров, проведя интерпретацию полученных результатов, мы пришли к следующим выводам:

– при селекции по признаку увеличения удоя необходимо установить степень и направленность взаимосвязей с такими признаками, как жирномолочность и белкомолочность, поскольку в противном случае при росте удоя может наблюдаться их снижение;

– генетический тренд признаков молочной продуктивности животных исследуемой популяции указывает на большую их сопряженность.

Благодарности (Acknowledgements)

Выражаем благодарность за предоставление данных специалистам ЗАО ПЗ «Аксиньино» Ступинского района Московской области. Исследования выполнены в рамках Государственного задания Минобрнауки России 0445-2021-0016.

Библиографический список

1. Лепехина Т. В. Корреляционная связь и наследуемость основных хозяйственно-полезных признаков у коров разных генераций: автореф. дис. ... канд. биол. наук [Электронный ресурс]. URL: <https://www.disscat.com/content/korrelyatsionnaya-svyaz-i-nasleduemost-khozyaistvenno-poleznykh-priznakov-u-korov-raznykh-g> (дата обращения: 23.08.2021).
2. Контэ А. Ф., Игнатъева Л. П. Многофакторный регрессионный анализ показателей типа телосложения коров-перволеток черно-пестрой голштинизированной породы Подмосковья // Вестник Ульяновской ГСХА. 2020. № 4 (52). С. 226–231.
3. Дунин И. М. Наши усилия должны быть направлены на снижение зависимости от импортного племенного материала и генетических технологий // Аграрная наука. 2021. № 5. С. 36–37.
4. Гомидова В. С. Множественная регрессия и корреляция [Электронный ресурс]. URL: <http://scienceforum.ru/2015/article/2015008364> / <https://scienceforum.ru/2015/article/2015008364> (дата обращения: 19.08.2021).
5. Тарасова Е. И. Гены-маркеры продуктивных характеристик молочного скота [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/geny-markery-produktivnyh-harakteristik-molochnogo-skota-obzor> (дата обращения: 19.08.2021).
6. Максимова Т. Г., Попова И. Н. Эконометрика: учебно-методическое пособие. Санкт-Петербург: Университет ИТМО, 2018. 70 с.
7. Мартынова Е. Н., Исупова Ю. В. Экстерьерные особенности и молочная продуктивность голштинизированных коров холмогорской породы разных генераций // Пермский аграрный вестник. 2018. № 1 (21). С. 125–131.
8. Шилова З. В., Шилов О. И. Теория вероятностей и математическая статистика: учебное пособие. Киров: Изд-во ВГГУ, 2015. 158 с.
9. Norouziyan M. A., Bayatani H., Alavijeh M. V. Comparison of artificial neural networks and multiple linear regression for prediction of dairy cow locomotion score [Электронный ресурс]. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8094148> (дата обращения: 23.08.2021).
10. Popescu A. Regression modelling in predicting milk production depending on dairy bovine livestock [Электронный ресурс]. URL: <https://www.researchgate.net/publication/292361817> (дата обращения: 23.08.2021).
11. Yamazaki T., Takeda H., Hagiya K., Yamaguchi S., Sasaki O. Prediction of random-regression coefficient for daily milk yield after 305 days in milk by using the regression-coefficient estimates from the first 305 days [Электронный ресурс]. URL: <https://pdfs.semanticscholar.org/b5a4/b2bfd68ef8a61bfa7e8ae817e9a60f9cf4f6.pdf> (дата обращения: 23.08.2021).
12. Кольцов С.Н. Регрессионный анализ [Электронный ресурс]. URL: <https://docplayer.com/29963508-Regressionnyy-analiz-kolcov-s-n.html> (дата обращения: 23.08.2021).
13. Токарев Д. В. Линейная регрессия [Электронный ресурс]. URL: <https://masters.donntu.org/2005/kita/tokarev/library/linreg.htm> (дата обращения: 23.08.2021).
14. Вильвер Д. С. Взаимосвязь хозяйственно-полезных признаков коров различных генотипов [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/vzaimosvyaz-hozyaystvenno-poleznykh-priznakov-korov-razlichnyh-genotipov> (дата обращения: 23.08.2021).
15. Галочкин В. Т., Латыш А. Р. Исследование уравнения множественной линейной регрессии [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/issledovanie-uravneniya-mnozhestvennoy-lineynoy-regressii/viewer> (дата обращения: 05.09.2021).
16. Корреляция и регрессия. Информационная грамотность [Электронный ресурс]. URL: <https://biconsult.ru/services/korrelyatsiya-i-regressiya> (дата обращения: 05.09.2021).
17. Линейная регрессия и множественная регрессия: в чем разница [Электронный ресурс]. URL: <https://nesrakonk.ru/what-difference-between-linear-regression-and-multiple-regression/> (дата обращения: 05.09.2021).
18. Гедранович А. Б. Модели парной и множественной регрессии [Электронный ресурс]. URL: http://www.miu.by/rus/kaf_ep/kaf_download/24169_90115326.pdf (дата обращения: 05.09.2021).
19. Ayalew W., Aliy M., Negussie E. Estimation of genetic parameters of the productive and reproductive traits in Ethiopian Holstein using multi-trait models // Asian Australasian Journal of Animal Sciences. 2017. № 30 (11). Pp. 1550–1556. DOI: 10.5713/ajas.17.0198.
20. Lu H., Wang Y., Bovenhuis H. Phenotypic and genetic effects of season on milk production traits in dairy cattle in the Netherlands // Journal of Dairy Science. 2020. Vol. 104. Iss. 4. Pp. 4486–4497.

21. Lu H., Bovenhuis H. Phenotypic and genetic effects of pregnancy on milk production traits in Holstein-Friesian cattle // Journal of Dairy Science. 2020. Vol. 103. Iss. 12. Pp. 11597–11604.

Об авторах:

Галина Геннадьевна Карликова¹, доктор сельскохозяйственных наук, старший научный сотрудник отдела популяционной генетики и генетических основ разведения животных, ORCID 0000-0002-9021-1404, AuthorID 667690; galinakarlikova@yandex.ru

Александр Федорович Конгэ¹, кандидат сельскохозяйственных наук, научный сотрудник отдела популяционной генетики и генетических основ разведения животных, ORCID 0000-0003-4877-0883, AuthorID 849809; alexandrconte@yandex.ru

¹ Федеральное исследовательское учреждение животноводства – ВИЖ имени академика Л. К. Эрнста, Дубровицы, Россия

Multivariate regression analysis of dairy characteristics of Holstein cows

G. G. Karlikova¹✉, A. F. Konte¹

¹ L. K. Ernst Federal Research Center for Animal Husbandry, Dubrovitsy, Russia

✉ E-mail: galinakarlikova@yandex.ru

Abstract. The purpose of the research is the possibility of applying the equation of multiple two-factor regression to reveal the reliable conjugacy of signs affecting the level of milk productivity of Holstein cows. **Research methods.** The research was carried out in CJSC BP “Aksin’ino” of Stupinskiy district of the Moscow region. Based on the IAS “SELEX”, a database was created, including a sample of 11 017 heads. Data on milk productivity were taken for 305 days of 1st, 2nd, 3rd and maximum lactation. **Results.** For 305 days of the 1st lactation, milk yield averaged 7909.5 kg of milk, for the 2nd – 8289.1 kg ($p \leq 0,001$) and the 3rd lactation – 8446.2 kg ($p \leq 0,001$). Milk yield for maximum lactation was 8964.3 kg of milk ($p \leq 0,001$). The fat and protein content in cow milk between the 1st and 3rd lactation is 4.03–4.08 and 3.22–3.23 %. The multiple two-factor regression coefficient represents the response bias from 7787.81–8239.00 (1st, 2nd, 3rd lactation) to 8841.63 (maximum lactation). The scattering diagrams of the multiple regression model show that the value of the variable "milk yield" is statistically dependent on the indicators of the mass fraction of fat and the mass fraction of protein in milk. The coefficient of determination at the level of 0.997–0.998 indicates that the regression equation explains 99.7–99.8 % of the variance of the effective feature. The significance of Fischer's *F*-test indicates the high reliability of the results and the absence of randomness and the presence of a pattern justified in our study. The fat and protein content had a negative relationship with milk yield ($p \leq 0,001$) by lactation. High values of the relationship between milk yield and protein content (–0.518...–0.766), fat and protein content (0.626–0.784) were obtained. **Scientific novelty.** For the first time, studies were conducted on the correspondence of the equation model to the experimental data obtained, and the presence of the number of independent variables (mass fraction of fat and protein) included in the equation to describe the dependent variable (milk yield).

Keywords: multiple regression, cows, lactation, milk yield, fat mass fraction, protein mass fraction.

For citation: Karlikova G. G., Konte A. F. Mnogofaktornyy regressiionnyy analiz molochnykh priznakov korov golshhtinskoy porody [Multivariate regression analysis of dairy characteristics of Holstein cows] // Agrarian Bulletin of the Urals. 2021. No. 11 (214). Pp. 27–35. DOI: 10.32417/1997-4868-2021-214-11-27-35. (In Russian.)

Date of paper submission: 08.09.2021, **date of review:** 17.09.2021, **date of acceptance:** 24.09.2021.

References

1. Lepekhina T. V. Korrelyatsionnaya svyaz' i nasleduyemost' osnovnykh khozyaystvenno-poleznykh priznakov u korov raznykh generatsiy: avtoref. dis. ... kand. biol. nauk [Korrelyatsionnaya svyaz' i nasleduyemost' osnovnykh khozyaystvenno-poleznykh priznakov u korov raznykh generatsiy: abstract of dissertation ... candidate of biological sciences] [e-resource]. URL: <https://www.dissercat.com/content/korrelyatsionnaya-svyaz-i-nasleduyemost-khozyaystvenno-poleznykh-priznakov-u-korov-raznykh-ge> (date of reference: 23.08.2021). (In Russian.)
2. Konte A. F., Ignat'yeva L. P. Mnogofaktornyy regressiionnyy analiz pokazateley tipa teloslozheniya korov-perivotelok cherno-pestroy golshhtinizirovannoy porody Podmoskov'ya [Multifactorial regression analysis of body type indicators of first-calf cows of the black-and-white Holstein breed of the Moscow region] // Vestnik of Ulyanovsk State Agricultural Academy. 2020. No. 4 (52). Pp. 226–231. (In Russian.)

3. Dunin I. M. Nashi usiliya dolzhny byt' napravlen na snizheniye zavisimosti ot importnogo plemennogo materiala i geneticheskikh tekhnologiy [Our efforts should be aimed at reducing dependence on imported breeding material and genetic technologies] // *Agrarian science*. 2021. No. 5. Pp. 36–37. (In Russian.)
4. Gomidova V. S. Mnozhestvennaya regressiya i korrelyatsiya [Multiple regression and correlation] [e-resource] // *Studencheskiy nauchnyy forum: materialy VII Mezhdunarodnoy studencheskoy nauchnoy konferentsii*. URL: <http://scienceforum.ru/2015/article/2015008364/ahttps://scienceforum.ru/2015/article/2015008364> (date of reference: 20.08.2020). (In Russian.)
5. Tarasova E. I. Geny-markery produktivnykh kharakteristik molochnogo skota [Genes-markers of productive characteristics of dairy cattle]. URL: <https://cyberleninka.ru/article/n/geny-markery-produktivnyh-harakteristik-molochnogo-skota-obzor> (date of reference: 19.08.2021). (In Russian.)
6. Maksimova T. G., Popova I. N. Ekonometrika: uchebno-metodicheskoye posobiye [Econometrics: educational and methodical manual. Saint Petersburg: ITMO University, 2018. 70 p. (In Russian.)
7. Martynova E. N., Isupova Yu. V. Ekster'yernyye osobennosti i molochnaya produktivnost' golshtinizirovannykh korov kholmogorskoy porody raznykh generatsiy [Exterior features and milk productivity of Holstein cows of the Kholmogorsky breed of different generations] // *Perm Agrarian Journal*. 2018. No. 1 (21). Pp. 125–131. (In Russian.)
8. Shilova Z. V., Shilov O. I. Teoriya veroyatnostey i matematicheskaya statistika: uchebnoye posobiye [Probability theory and mathematical statistics: textbook]. Kirov: Izd-vo VGGU, 2015. 158 p. (In Russian.)
9. Norouzian M. A., Bayatani H., Alavijeh M. V. Comparison of artificial neural networks and multiple linear regression for prediction of dairy cow locomotion score [e-resource]. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8094148> (date of reference: 23.08.2021).
10. Popescu A. Regression modelling in predicting milk production depending on dairy bovine livestock [e-resource]. URL: <https://www.researchgate.net/publication/292361817> (date of reference: 23.08.2021).
11. Yamazaki T., Takeda H., Hagiya K., Yamaguchi S., Sasaki O. Prediction of random-regression coefficient for daily milk yield after 305 days in milk by using the regression-coefficient estimates from the first 305 days [e-resource]. URL: <https://pdfs.semanticscholar.org/b5a4/b2bfd68ef8a61bfa7e8ae817e9a60f9cf4f6.pdf> (date of reference: 23.08.2021).
12. Kol'tsov S. N. Regressionnyy analiz [Regression analysis] [e-resource]. URL: <https://docplayer.com/29963508-Regressionnyy-analiz-kolcov-s-n.html> (date of reference: 23.08.2021). (In Russian.)
13. Tokarev D. V., Koval'chenko I. D., Borodkin L. I., et al. Lineynaya regressiya [Linear regression] [e-resource]. URL: <https://masters.donntu.org/2005/kita/tokarev/library/linreg.htm> (date of reference: 23.08.2021). (In Russian.)
14. Vil'ver D. S. Vzaimosvyaz' khozyaystvenno-poleznykh priznakov korov razlichnykh genotipov [The relationship of economically useful traits of cows of different genotypes] [e-resource]. URL: <https://cyberleninka.ru/article/n/vzaimosvyaz-hozyaystvenno-poleznyh-priznakov-korov-razlichnyh-genotipov> (date of reference: 23.08.2021). (In Russian.)
15. Galochkin V. T. Issledovaniye uravneniya mnozhestvennoy lineynoy regressii [e-resource]. [Investigation of the multiple linear regression equation] URL: <https://cyberleninka.ru/article/n/issledovanie-uravneniya-mnozhestvennoy-lineynoy-regressii/viewer> (date of reference: 23.08.2021). (In Russian.)
16. Korrelyatsiya i regressiya. Informatsionnaya gramotnost' [Correlation and regression. Information literacy]. [e-resource]. URL: <https://biconsult.ru/services/korrelyatsiya-i-regressiya> (date of reference: 23.08.2021). (In Russian.)
17. Lineynaya regressiya i mnozhestvennaya regressiya: v chem raznitsa [Linear regression and multiple regression: what is the difference] [e-resource]. URL: <https://nesrakonk.ru/what-difference-between-linear-regression-and-multiple-regression> (date of reference: 05.09.2021). (In Russian.)
18. Gedranovich A. B. Modeli parnoy i mnozhestvennoy regressii [Paired and multiple regression models] [e-resource]. URL: http://www.miu.by/rus/kaf_ep/kaf_download/24169_90115326.pdf (date of reference: 05.09.2021). (In Russian.)
19. Ayalew W., Aliy M., Negussie E. Estimation of genetic parameters of the productive and reproductive traits in Ethiopian Holstein using multi-trait models // *Asian Australasian Journal of Animal Sciences*. 2017. № 30 (11). Pp. 1550–1556. DOI: 10.5713/ajas.17.0198.
20. Lu H., Wang Y., Bovenhuis H. Phenotypic and genetic effects of season on milk production traits in dairy cattle in the Netherlands // *Journal of Dairy Science*. 2020. Vol. 104. Iss. 4. Pp. 4486–4497.
21. Lu H., Bovenhuis H. Phenotypic and genetic effects of pregnancy on milk production traits in Holstein-Friesian cattle // *Journal of Dairy Science*. 2020. Vol. 103. Iss. 12. Pp. 11597–11604.

Authors' information:

Galina G. Karlikova¹, doctor of agricultural sciences, senior staff scientist of population genetics and animal breeding department, ORCID 0000-0002-9021-1404, AuthorID 667690; galinakarlikova@yandex.ru

Aleksandr F. Konte¹, candidate of agricultural sciences, scientific researcher of population genetics and animal breeding department, ORCID 0000-0003-4877-0883, AuthorID 849809; alexandrconte@yandex.ru

¹ L. K. Ernst Federal Research Center for Animal Husbandry, Dubrovitsy, Russia