Классификация животных с учетом селекционных индексов

Катков Константин Александрович, кандидат технических наук, доцент
Северо-Кавказский федеральный научный аграрный центр (г. Михайловск)

В работе рассмотрены вопросы классификации овечьего поголовья с использованием процедуры формирования селекционных индексов и последующего дискриминантного анализа. С помощью индексной оценки группы баранов производителей формируется обучающая выборка для последующей классификации более многочисленной группы потомков. Использование интегрированного математического пакета Matlab с встроенной функцией дискриминантного анализа упрощает и ускоряет процесс классификации.

Ключевые слова: *селекция, индексная оценка, дискриминантный анализ, классификация.*

DOI: 10.5281/zenodo.3473935

Успешное ведение селекционной работы в животноводстве предполагает использование в селекционном процессе животных с наилучшими показателями хозяйственно полезных признаков. Если селекция ведется по какому-либо одному признаку, то выбор животных с наилучшими показателями не вызывает затруднений. В то же время, часто отбор животных необходимо вести сразу по нескольким ХПП. В этом случае задача усложняется [1, с.18; 2, с.26]. Решением проблемы является использование селекционных индексов [3, с.52]. Выбор набора признаков для составления индекса является одной из задач селекционера. Итогом индексной оценки животных будет безразмерная числовая величина, которая является комплексной оценкой каждого животного, учитывающей сразу несколько ХПП, участвующих в формировании селекционного индекса. Путем сравнения с каким-либо пороговым значением индекс позволит ранжировать животных и классифицировать всю совокупность оцениваемых животных на особей, которые потенциально представляют интерес для селекционной работы, и особей, подлежащих выбраковке [5, с.34]. Первую группу обозначим, как 1-й класс, вторую – 2-й класс.

Также классификацию животных можно провести, не прибегая к построению селекционных индексов. Для этой цели можно использовать дискриминантный анализ. При таком анализе проводится исследование различий животных по выбранным признакам. Дискриминантный анализ позволяет определить признаки, которые вносят наиболее существенный вклад в выявленные достоверные различия между животными. Итогом дискриминантного анализа является прогноз принадлежности оцениваемых животных к 1-му или 2-му классу.

Использование математических методов в овцеводстве ставит перед исследователями ряд проблем. Большинство ХПП, используемых в животноводстве, имеют тенденцию: «чем больше значение признака, тем лучше животное». В тонкорунном овцеводстве для признака «тонина шерсти» использование такого подхода неприемлемо [2, с.26]. Здесь наблюдается обратная тенденция, и наилучшим считается животное с минимальным значением тонины. Это требует отдельного подхода к обработке информации, связанной с этим признаком [6, с.10].

Еще одна проблема связана с большим поголовьем овечьих стад. Необходимость анализа показателей сотен голов животных требует большого объема вычислительных работ, и обойтись в этом вопросе без использования современных информационных технологий весьма затруднительно. В настоящее время существует большое количество специализированных математических пакетов, которые позволяют решить эту проблему.

Также следует учесть, что при использовании дискриминантного анализа актуальной задачей является построение обучающей выборки, которая позволит классифицировать совокупность оцениваемых животных.

В настоящем исследовании предлагается объединенное использование дискриминантного анализа и селекционных индексов для классификации овечьего поголовья с целью повышения эффективности селекционного процесса.

Постановка задачи. Имеется группа баранов-производителей. Также имеется более многочисленная группа потомков этих производителей мужского пола. Требуется оценить всю совокупность животных и разбить их на два класса в соответствии с перспективой дальнейшей селекционной работы. В качестве хозяйственно полезных признаков, на основании которых будет проводиться классификация, используются: живая масса (ЖМ) животных, длина шерсти (ДШ), настриг чистой шерсти (НШ) и тонина шерсти (Т).

Для решения поставленной задачи предлагается путем построения селекционных индексов для баранов производителей определить обучающую выборку для дискриминантного анализа. Затем, используя полученную обучающую выборку, классифицировать всю группу потомков. Тем самым определяются наилучшие и наихудшие из потомков. Оставшуюся группу «середняков» можно классифицировать с помощью селекционных индексов. Такой подход позволит: более точно определиться с обучающей выборкой для классификации животных дискриминантным анализом; провести прогнозирование классификации потомков с целью выявления наилучших животных; сократить выборку потомков для проведения дальнейшей индексной оценки.

Решение. Исходными данными для решения поставленной задачи являются показатели продуктивности баранов цигайской породы. Бараны-производители обозначаются S1, S2, ..., S10. Потомки обозначаются: C1, C2 и т.д. Показатели желательного типа (целевые показатели по породе) и значения

коэффициентов наследования (h^2) приведены в таблице 1.

Для проведения индексной оценки предлагается сформировать селекционные индексы на основе двух подходов: индексы на основе селекционного дифференциала (I^D) и на основе селекционного отношения (I^R). Так как в составе индексов присутствует признак «тонина шерсти», то для построения индекса на основе селекционного дифференциала предлагается значения тонины для каждого животного предварительно умножить на -1. Для построения индекса на основе селекционного отношения предлагается селекционное отношение для тонины предварительно возвести в степень -1. Алгоритм формирования индексов подробно описан в [5, с.32; 6, с. 12].

Таблица 1 – Целевые показатели, значения $h2$ баранов-производителей

	Хозяйственно полезные признаки			
	ЖМ	ДШ	НШ	Т
Коэффициент h^2	0.37	0.34	0.39	0.39
Целевые показатели	110 кг	14 см	5 кг	29 мкм

Результаты. В результате индексной оценки были получены индексы I^R и I^D, которые позволили разбить группу баранов производителей на 2 класса и сформировать обучающую выборку для последующего дискриминантного анализа. Вид обучающей выборки с показателями ХПП баранов-производителей, классом животного и значением селекционных индексов представлен в таблице 2.

Таблица 2 – Обучающая выборка и значения селекционных индексов

№ барана	Обучающая выборка					Индексы	
	ЖМ, кг	ДШ, см	НШ, кг	Т, мкм	Класс	I^R	I^D
S3	100	12	5,9	32	1	0,9606	0,0895
S5	110	14	5,3	33	1	0,9866	0,1958
S6	110	13	4,2	30	1	0,9346	0,4451
S7	110	15	5,3	32	1	1,0117	0,5429
S9	96	19	6,07	33	1	1,0839	0,4075
S1	94	15	3,35	33	0	0,8722	-0,5511
S2	88	16	3,6	32	0	0,8949	-0,2767
S4	78	14	3,9	30	0	0,8619	-0,2404
S8	94	18	3,4	33	0	0,9346	-0,1376
S10	80	13	4,7	33	0	0,8666	-0,7750

В таблице 2 принадлежность животного к 1-му классу обозначена цифрой «1», принадлежность ко 2-му классу – цифрой «0».

Таким образом, данные таблицы 2 позволяют сформировать обучающую выборку для дискриминантного анализа, которая будет состоять из матрицы исходных данных X и вектора принадлежности к классу *class*:

$$X = \begin{bmatrix} 100 & 12 & 5,9 & 32 \\ 110 & 14 & 5,3 & 33 \\ 110 & 13 & 4,2 & 30 \\ 110 & 15 & 5,3 & 32 \\ 96 & 19 & 6,07 & 33 \\ 94 & 15 & 3,35 & 33 \\ 88 & 16 & 3,6 & 32 \\ 78 & 14 & 3,9 & 30 \\ 94 & 18 & 3,4 & 33 \\ 80 & 13 & 4,7 & 33 \end{bmatrix}, \qquad class = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

Теперь можно применить процедуру дискриминантного анализа ко всей совокупности потомков классифицированных выше производителей. Алгоритм дискриминантного анализа подробно описан в [9].

В рамках данного исследования использовался интегрированный математический пакет Matlab, в составе инструментов которого есть встроенная функция дискриминантного анализа [10, с.125]. Синтаксис вызова этой функции имеет вид: CL = classify (P, X, class).

Здесь P – матрица исходных данных 45-ти потомков размерностью 45x4, матрица X и вектор class представлены в (1).

В результате вызова функции classify получаем вектор CL размерностью 45x1, где каждый элемент соответствует отдельному потомку. Если потомок принадлежит к 1-му классу, то соответствующий элемент вектора CL равен 1, если второму – то 2. Разбивка на классы производится на основании сравнения значения дискриминантной функции (F) с константой дискриминации (C). В рассматриваемом примере константа дискриминации составила величину $C = 239,01$.

Значения дискриминантной функции для потомков C1, ..., C45 представлены на рисунке 1. На этом рисунке значения функции F обозначены круглыми маркерами, для удобства восприятия соединенными пунктирной линией. Значение константы дискриминации (C) обозначено прямой сплошной линией на графике.

Анализ графика на рисунке 1 позволяет сразу выделить животных, принадлежащих к 1-му классу. Таковыми окажутся животные с номерами: C1, C6, C7, C10, C11, C15, C32, C37, C40, C42, C45. В таблице 2 показатели этих животных выделены жирным цветом, а их номера – вертикальной штриховкой. Также можно выделить животных с наихудшими показателями. Таковыми будут особи с номерами: C4, C9, C13, C21 (номера выделены заливкой в табл. 2). Дискриминантный анализ позволил из группы,

состоящей из 45 голов, выделить 11 наилучших животных и 4-х наихудших, сократив тем самым исходную выборку на 33%. Теперь, при необходимости, оставшиеся животные могут быть оценены с помощью индексной оценки. Вероятно, что по результатам такой оценки к первому классу могут быть отнесены особи с номерами С14, С22, С26, С29, С41 (рисунок 1, горизонтальная штриховка номеров в табл.2).

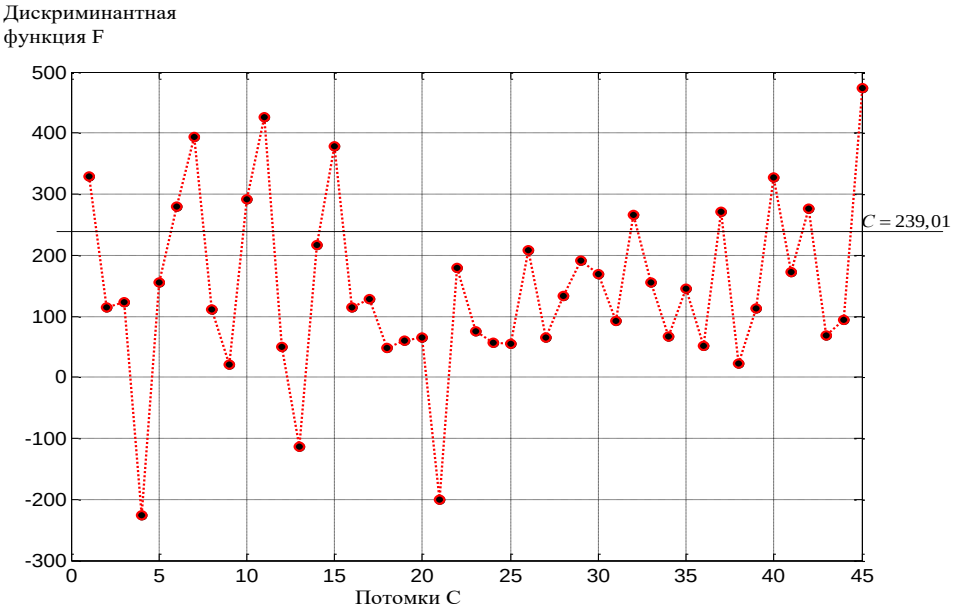

Рис. 1 – Значения дискриминантной функции для группы потомков

Таблица 2 – Данные ХПП и классификация потомков

№ потомка	ЖМ,кг	ДШ, см	НШ, кг	Т, мкм	Класс
С1	**97**	**15**	**5,4**	**32**	**1**
С2	77	14	4,5	31	2
С3	82	13	3,9	30	2
С4	63	14	2,4	32	2
С5	86	13	3,9	30	2
С6	**100**	**11**	**4,5**	**31**	**1**
С7	**95**	**13**	**6**	**31**	**1**
С8	83	10	4,5	32	2
С9	96	14	3,3	35	2
С10	**88**	**12**	**6,2**	**33**	**1**
С11	**102**	**15**	**6**	**32**	**1**
С12	80	16	4,8	35	2
С13	76	12	3,6	35	2
С14	83	13	4,8	30	2
С15	100	15	6	33	1
С16	90	12	4,2	33	2
С17	90	15	4,2	33	2
С18	73	12	3,9	30	2
С19	77	15	4,2	32	2
С20	83	13	4,2	33	2
С21	79	13	2,4	35	2
С22	77	15	4,8	30	2
С23	92	13	3,9	34	2
С24	85	16	4,8	36	2
С25	89	13	3,6	33	2
С26	93	14	4,5	32	2
С27	80	10	3,6	30	2
С28	93	12	4,8	35	2
С29	90	14	3,9	30	2
С30	97	12	4,2	33	2
С31	88	11	4,5	34	2
С32	**96**	**14**	**4,2**	**30**	**1**
С33	97	15	3,9	33	2

С34	86	14	3,6	32	2
С35	89	16	5,4	36	2
С36	96	13	3	33	2
С37	90	14	5,4	32	1
С38	80	15	3,9	33	2
С39	96	14	3,6	33	2
С40	99	16	4,5	30	1
С41	93	13	4,5	33	2
С42	90	14	4,8	30	1
С43	78	14	3,6	30	2
С44	94	13	3,6	33	2
С45	100	15	6	30	1

Дискриминантный анализ позволяет определить степень влияния на дискриминантную функцию каждой из переменных. Для этого необходимо определить стандартизованные коэффициенты детерминации (b_j) согласно [8, с.102]:

$$b(j) = a(j)\sqrt{\frac{S(j,j)}{p-n}}, \qquad (2)$$

где $a(j)$ – коэффициенты детерминации; $S(j,j)$ – диагональные элементы матрицы ковариации; p – число признаков; n – количество классов.

В рассматриваемом примере стандартизованные коэффициенты равны

$$b = \begin{bmatrix} 44 & 8{,}53 & 49{,}08 & -31{,}58 \end{bmatrix}.$$

Таким образом, можно утверждать, что наибольшее влияние на значения дискриминантной функции оказывает настриг чистой шерсти (третий элемент вектора b).

Заключение. Проведенное исследование показывает, что комбинация индексной оценки и дискриминантного анализа позволит достаточно быстро провести классификацию животных по данным обучающей выборки. При этом в оцениваемой группе животных выявляются явные «лидеры» и «аутсайдеры». Исключение таких животных из анализируемой выборки позволит сократить ее и уменьшить дальнейшие вычислительные затраты. Стоит отметить, что вполне возможно, что дискриминантный анализ полностью удовлетворит селекционера и последующая индексная оценка оставшейся выборки может не понадобиться. Наличие в Matlab встроенной функции дискриминантного анализа существенно упрощает работу и сокращает время ее выполнения.

Литература:

1.Шерстная продуктивность потомков от производителей импортной селекции / Ефимова Н.И., Скорых Л.Н., Копылов И.А. // Сборник научных трудов Всероссийского научно-исследовательского института овцеводства и козоводства. 2015. Т. 2. № 8. С. 17-21.

2.Копылов И.А., Скорых Л.Н., Ефимова Н.И. Мясность молодняка овец породы советский меринос и их помесей с австралийскими баранами // Овцы, козы, шерстяное дело. 2017. № 2. С. 26-27.

3.Интенсификация племенного отбора в свиноводстве / Михайлов Н.В., Каратунов Г.А., Третьякова О.Л., Костылев Э.В. пос. Персияновский: ДонГАУ, 1999. 100 с.

4.Оценка племенной ценности баранов-производителей методом BLUP / К.А. Катков, С.С. Бобрышов, Л.Н. Скорых, В.Б. Копылов, М.А. Афанасьев // Главный зоотехник. 2018. №5. С. 25-32.

5.Использование селекционных индексов при подборе родительских пар в овцеводстве / Катков К.А., Скорых Л.Н., Остапчук П.С., Емельянов Е.А., Куевда Т.А. // Главный зоотехник. 2019. №7(192). С. 30-37.

6.Два подхода к формированию селекционных индексов в овцеводстве / Катков К.А., Скорых Л.Н., Остапчук П.С., Емельянов Е.А., Паштецкая А.В. // Вестник АПК Ставрополья. 2019. №2(34) С.8-14.

7.Using genetic markers in breeding sheep / Degtyarev D.Y., Skorykh L.N., Kovalenko D.V., Emelyanov S.A., Konik N.V. // Research Journal of Pharmaceutical, Biological and Chemical Sciences. 2016. Vol. 7. № 4. P. 2137-2139.

8.Румишинский Л.З. Математическая обработка результатов эксперимента. М.: Изд-во «Наука», 1974. 192 с.

9.StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft. WEB: http://www.statsoft.ru/home/textbook/default.htm.

10. Основы компьютерного моделирования. Учебное пособие / К.А. Катков, И.П. Хвостова, В.И. Лебедев, Е.Н. Косова, А.А. Плетухина, О.Л. Серветник, О.В. Вельц, М.Г. Крамаренко. Ставрополь: изд-во СКФУ, 2013. 220 с.