DOI: 10.31279/2222-9345-2018-7-30-86-90
УДК 636.32/.38.082.2

К. А. Катков, Г. Т. Бобрышова, Л. Н. Скорых, О. С. Копылова, М. А. Афанасьев

Katkov K. A., Bobryshova G. T., Skorykh L. N., Kopylova O. S., Afanasyev M. A.

АЛГОРИТМ ПРОВЕРКИ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ РАЗЛИЧИЙ ХОЗЯЙСТВЕННО ПОЛЕЗНЫХ ПРИЗНАКОВ МЕЖДУ РАЗЛИЧНЫМИ ГЕНЕАЛОГИЧЕСКИМИ ГРУППАМИ ЖИВОТНЫХ

ALGORITHM TO TEST THE STATISTICAL SIGNIFICANCE OF DIFFERENCES IN ECONOMICALLY USEFUL TRAITS BETWEEN VARIOUS GENEALOGICAL GROUPS OF ANIMALS

Приведена последовательность действий, которую необходимо выполнить селекционеру при проведении оценки племенной ценности животных. Оцениваемой группой животных являются бараны-производители. Оценка проводится по значению хозяйственно полезных признаков потомков этих производителей. В качестве таких признаков выступают живая масса потомков и их настриг шерсти. Собранные опытным путем исходные данные должны подвергнуться статистическому анализу, в ходе которого необходимо исключить аномальные измерения. Затем следует определить достоверность различий между значениями хозяйственно полезных признаков в различных генеалогических группах животных. Показано, что для определения достоверности различий необходимо провести последовательно три статистических теста. Если тест на сравнение разности двух средних с доверительными границами дает результат о случайности различий, то необходимо перейти ко второму тесту – сравнению по методу Стьюдента. Если и этот тест покажет случайный характер различий, то необходимо перейти к третьему тесту – проверке по критерию Фишера. Если хотя бы один статистический тест указывает на достоверность различий между средними показателями, то принимается гипотеза о значимом характере различий между средними значениями хозяйственно полезных признаков в исследуемых генеалогических группах. Проведен сравнительный анализ этих тестов. Показано, что наибольшей достоверностью обладает проверка статистической значимости различий по критерию Фишера. Обладая данными о неслучайности различий, можно приступить к определению гарантированно-минимума превосходства одной генеалогической группы животных над другой.

Ключевые слова: статистические тесты, доверительный интервал, нулевая гипотеза, альтернативная гипотеза.

The article presents the sequence of actions that must be performed by the breeder when assessing the breeding value of animals. The estimated groups of animals are the tupping rams. The evaluation made on the value of the economically useful traits of the descendants from these stud rams. As economically useful traits are the live weight of offspring and of their wool clip. The baseline data collected by the experiment should undergo statistical analysis, during which anomalous measurements should be excluded. Then it is necessary to define the reliability of the differences between the values of economically useful traits in various genealogical groups of animals. The article shows that in order to determine the accuracy of the differences, it is necessary to conduct three statistical tests sequentially. If the test for comparing the difference between two averages with confidence limits gives the result of about the randomness of the differences, then it is necessary to go to the second test – comparison according to the Student's t-test. If this test also shows the random nature of the differences, then you need to go to the third test – check by Fisher criterion. If at least one statistical test indicates the reliability of the differences between the average indices then the hypothesis about the significant character of the differences between the average values of economically useful traits in the studied genealogical groups is accepted. The comparative analysis of these tests was carried out. It is shown that the verification of statistical significance of differences by the Fisher criterion is the most reliable. Possessing data on the non-randomness of differences, it is possible to start the definition of the guaranteed superiority minimum of one genealogical group of animals over another.

Key words: statistical tests, confidence interval, null hypothesis, alternative hypothesis.

Катков Константин Александрович – кандидат технических наук, доцент, ведущий научный сотрудник лаборатории информационных технологий Всероссийский научно-исследовательский институт овцеводства и козоводства – филиал ФГБНУ «Северо-Кавказский федеральный научный аграрный центр» г. Ставрополь
Тел.: 8(8652)71-81-55
E-mail: kkatkoff@mail.ru

Бобрышова Галина Тимофеевна – кандидат сельскохозяйственных наук, заместитель директора Всероссийский научно-исследовательский институт овцеводства и козоводства – филиал ФГБНУ «Северо-Кавказский федеральный научный аграрный центр» г. Ставрополь
Тел.: 8(8652)71-81-55
E-mail: science@vniiook.ru

Katkov Konstantin Aleksandrovich – Ph.D of Technical Sciences, Associate professor, Leading Researcher at the Laboratory of Information Technologies All-Russian Research Institute of Sheep Breeding and Goat Breeding – branch FSBSI «North Caucasian Federal Agricultural Research Centre» Stavropol
Tel.: 8(8652)71-81-55
E-mail: kkatkoff@mail.ru

Bobryshova Galina Timofeevna – Ph.D of Agricultural Sciences, Assistant Director All-Russian Research Institute of Sheep Breeding and Goat Breeding – branch FSBSI «North Caucasian Federal Agricultural Research Centre» Stavropol
Tel.: 8(8652)71-81-55
E-mail: science@vniiook.ru

Скорых Лариса Николаевна –

доктор биологических наук,
ведущий научный сотрудник отдела овцеводства
Всероссийский научно-исследовательский институт
овцеводства и козоводства – филиал
ФГБНУ «Северо-Кавказский федеральный научный
аграрный центр»
г. Ставрополь
Тел.: 8(8652)71-81-55
E-mail: smu.sniizhk@yandex.ru

Копылова Оксана Сергеевна –

кандидат физико-математических наук, доцент
кафедры физики
ФГБОУ ВО «Ставропольский государственный
аграрный университет»
г. Ставрополь
Тел.: 8-918-867-45-64
E-mail: zolterxp@list.ru

Афанасьев Михаил Анатольевич –

ассистент кафедры физики
ФГБОУ ВО «Ставропольский государственный
аграрный университет»
г. Ставрополь
Тел.: 8(8652)71-81-55
E-mail: ssau_physics@mail.ru

Skorykh Larisa Nikolayevna –

Doctor of Biological Sciences, Leading Researcher
of the Department of Sheep Breeding
All-Russian Research Institute of Sheep Breeding
and Goat Breeding – branch
FSBSI «North Caucasian Federal
Agricultural Research Centre»
Stavropol
Tel.: 8(8652)71-81-55
E-mail: smu.sniizhk@yandex.ru

Kopylova Oksana Sergeevna –

Ph.D of Physical and Mathematical Sciences,
Associate Professor of the Department of Physics
FSBEI HE «Stavropol State Agrarian University»
Stavropol
Tel.: 8-918-867-45-64
E-mail: zolterxp@list.ru

Afanasyev Mikhail Anatolyevich –

Assistant of the Department of Physics
FSBEI HE «Stavropol State Agrarian University»
Stavropol
Tel.: 8(8652)71-81-55
E-mail: ssau_physics@mail.ru

Целью любой программы селекции является максимально возможное повышение генетического потенциала животных какой-либо популяции по хозяйственно полезным признакам. Оно может быть достигнуто путем использования животных с высокой генетической ценностью. Одним из путей выявления племенной ценности животных является их оценка по значению хозяйственно полезных признаков их потомков [1, 2]. Важным этапом в подобной работе является проведение статистического анализа измеренных опытным путем значений хозяйственно полезных признаков потомков исследуемой группы производителей. Другими словами, необходимо определить достоверность различий между значениями хозяйственно полезных признаков в различных генеалогических группах этих потомков. Если различия между группами потомков носят неслучайный характер, то можно приступать к оценке племенной ценности производителей. В рамках данной работы было проведено исследование группы из 264 ярок, являющихся потомками семи

баранов-производителей. Производителей обозначим: $S_1, S_2, S_3, S_4, S_5, S_6$ и S_7 . Для поголовья овец в качестве хозяйственно полезных признаков (ХПП) выступают живая масса и настриг шерсти.

После сбора исходных данных необходимо провести их статистический анализ. Исходные данные были ранжированы, и на основании такого ранжирования построены двойные вариационные ряды. После этого была проведена проверка закона распределения хозяйственно полезных признаков потомков. Значения статистических параметров оцениваемой совокупности потомков представлены в таблице 1. Если коэффициент асимметрии (K_A) и коэффициент эксцесса (K_E) меньше, чем их ошибки репрезентативности (m_A и m_E соответственно), более чем в 3 раза [3, 4, 5], то можно принять гипотезу о нормальности распределения признаков, где ошибки репрезентативности определяются выражениями:

$$m_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}; m_E = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}, (1)$$

где n – величина выборки ($n=264$).

Таблица 1 – Описательная статистика исходных данных

Показатель	Живая масса (кг)	Настриг шерсти (кг)
Среднее арифметическое	52,568±0,476	5,237±0,063
Дисперсия выборки	59,874	1,186
Медиана	51,5	5
Мода	50	5
Эксцесс	0,398	-1,087
Асимметричность	0,599	0,136
Минимальное значение	36	3,2
Максимальное значение	76	9,2

Используя данные из таблицы 1, получим следующие значения отношений:

для живой массы:

$$\begin{cases} |K_A|/m_A = 1,2796 < 3 \\ |K_E|/m_E = 0,4848 < 3 \end{cases};$$

для настрига шерсти:

$$\begin{cases} |K_A|/m_A = 0,2853 < 3 \\ |K_E|/m_E = 1,308 < 3 \end{cases}. \quad (2)$$

Таким образом, результаты (2) позволяют принять гипотезу о нормальном распределении значений ХПП в исследуемой совокупности потомков.

На следующем этапе необходимо приступить к проверке статистической значимости различий. Для этого необходимо рассчитать средние значения хозяйственно полезных признаков в каждой из генеалогических групп, ошибки этих средних значений, СКО, коэффициенты вариации и доверительные интервалы. Затем необходимо сравнить средние показатели в группах по принципу «каждый с каждым». При сравнении средних показателей необходимо определить следующие величины: разницу средних значений и ошибку этой разности; доверительную границу по каждому из хозяйственно полезных признаков; критерии достоверности различий Стьюдента и Фишера для требуемого уровня достоверности.

Численность групп в исследуемой совокупности животных примерно одинакова: S_1 – 39 голов, S_2 – 42, S_3 – 31, S_4 – 38, S_5 – 38, S_6 – 39, S_7 – 37 голов, – поэтому можно ожидать, что на основании анализа рассчитанных данных будет сделан верный вывод о значимости различий значений хозяйственно полезных признаков между группами потомков.

Теперь необходимо применить ряд статистических тестов. Для выявления генеалогических групп, показатели которых отличаются от средних показателей остальных не случайно, проведем парное сравнение всех групп по принципу «каждый с каждым». В качестве нулевой гипотезы (H_0) примем гипотезу об отсутствии различий между средними показателями хозяйственно полезных признаков в разных генеалогических группах. Альтернативная гипотеза – гипотеза о значимости различий (утверждает наличие различий или существование связи). Альтернативная гипотеза (H_A) является «рабочей» гипотезой исследования. В соответствии с этой гипотезой различия достаточно значимы (не случайны).

Для принятия решений о том, какую из гипотез (H_0 или H_A) следует принять, используют статистические критерии, которые включают в себя методы расчета определенного показателя (эмпирического значения критерия), на основании которого принимается решение об отклонении или принятии гипотезы, а также правила принятия решения. Рассчитанное эмпирическое значение критерия сравнивается с известным (например, заданным таблично) эталонным

числом, называемым критическим значением критерия.

Критические значения приводятся, как правило, для нескольких уровней значимости α : 5 % (0,05), 1 % (0,01) или еще более высоких. Если полученное эмпирическое значение критерия оказывается меньше или равно критическому, то гипотеза H_0 не может быть отклонена – считается, что на заданном уровне значимости характеристики распределений совпадают. Если эмпирическое значение критерия оказывается строго больше критического, то гипотеза H_0 отвергается и принимается альтернативная гипотеза H_A – характеристики распределений считаются различными с достоверностью различий $1 - \alpha$.

Если эмпирическое значение критерия для данного числа степеней свободы ($n - 1$) оказывается ниже критического уровня, соответствующего выбранному значению α (порогу вероятности), то гипотеза H_0 не опровергается, и это означает, что выявленная разница недостоверна (случайна).

В качестве статистических тестов будем использовать следующие: 1) сравнение разности средних с доверительной границей случайных колебаний; 2) метод Стьюдента – проверка соответствия критерия достоверности различий критическому значению нормированного отклонения; 3) проверка по критерию достоверности Фишера.

Если хотя бы один статистический тест указывает на достоверность различий между средними показателями, то принимается альтернативная гипотеза H_A о значимом характере различий между средними значениями хозяйственно полезных признаков в исследуемых генеалогических группах.

1-й тест. Обозначим разность между средними значениями величин $\bar{x}$ и $\bar{y}$ как ΔD . При сравнительной оценке средних показателей большое значение имеет средняя ошибка разности, наблюдаемой между ними. По этой ошибке можно установить степень достоверности наблюдаемых различий. Ошибка разности двух средних, согласно [6], определяется выражением

$$m_D = \pm \sqrt{m_{\bar{x}}^2 + m_{\bar{y}}^2}. \quad (3)$$

Тогда разность между средними (ΔM) определяется выражением

$$\Delta M = \Delta D \pm m_D. \quad (4)$$

Если величина ΔM превосходит доверительную границу случайных колебаний, определяемую как

$$g = t_S m_D, \quad (5)$$

где $t_S = 1,96$ – коэффициент Стьюдента для вероятности 95 %, то можно сделать вывод о достоверности этой разности.

2-й тест. Здесь сначала определяется разность между средними значениями хозяйственно полезных признаков в разных генеалогических группах:

$$\Delta D = \bar{x} - \bar{y}. \quad (6)$$

Далее определяется критерий достоверности разности средних

$$t_D = \Delta D / m_D. \quad (7)$$

После этого значение t_D подставляют в неравенство

$$t_D \geq t_{St}, \quad (8)$$

где t_{St} – стандартные значения критерия достоверности, которые находятся по таблице Стьюдента по числу степеней свободы ($\nu = n_1 + n_2 - 2$), где n_1 и n_2 – объемы сравниваемых выборок [6, 7].

Если условие (8) выполняется, то разность между средними значениями достоверна, то есть нулевая гипотеза отвергается и принимается альтернативная гипотеза. В противном случае разность между средними случайна, не достоверна.

3-й тест. Практика показывает, что описанные выше два способа не всегда бывают точны и могут выдавать ошибку. Это связано с тем, что неточно оценивается вариабельность величин. Более точно разница между показателями вариации двух величин может быть оценена с помощью критерия Фишера, который представляет собой отношение двух дисперсий

$$F = \sigma_x^2 / \sigma_y^2. \quad (9)$$

Числителем всегда берется большая вариация. Затем значение F сравнивается с табличным критическим значением для заданного уровня достоверности. Если выполняется неравенство

$$F \geq F_k, \quad (10)$$

то разность между средними значениями достоверна, то есть нулевая гипотеза отвергается и принимается альтернативная гипотеза.

В противном случае разность между средними значениями хозяйственно полезных признаков случайна, не достоверна. Таблицы с критическими значениями критерия Фишера доступны в [6, 7].

Если хотя бы один статистический тест указывает на достоверность различий между средними показателями, то принимается альтернативная гипотеза H_A о значимом характере различий между средними значениями хозяйственно полезных признаков в исследуемых генеалогических группах. На основании приведенных выше статистических тестов были проведены сравнения средних показателей хозяйственно полезных признаков всех генеалогических групп между собой по принципу «каждый с каждым». Анализ полученных данных показывает, что проверку достоверности различий между средними значениями необходимо проводить по нескольким статистическим тестам. Так, если дисперсии величин сильно отличаются друг от друга, а средние значения в то же время близки, то только проверка по критерию Фишера может дать правильный результат и поможет избежать ошибки.

Следует заметить, что генеалогические группы следует сравнивать по принципу «каждый с каждым». Ниже, в таблице 2, приведены примеры такого сравнения для нескольких групп в исследуемом стаде. В тех парах, где различия между значениями хозяйственно полезных признаков достоверны (не случайны), цветом выделена ячейка с названием анализируемого признака. Там, где различия случайны, выделения признака нет.

В случае, когда результаты оценки достоверности подтверждаются сравнением разности средних с доверительной границей случайных колебаний, можно определить гарантированный минимум превосходства одной группы над другой в среднем на одну особь.

Таблица 2 – Проверка достоверности различий между средними значениями хозяйственно полезных признаков

Признак	1 тест		2 тест		3 тест	
	Разница средних, кг	Доверительная граница, кг	Критерий достоверности различий Стьюдента		Критерий достоверности Фишера	
			t_s	t_D	F_k	F
<i>Сравнение потомков отцов S_1 и S_2</i>						
Живая масса	2,394±1,467	2,876	1,99	1,612	1,53	1,032
Настриг шерсти	0,102±0,178	0,348	1,99	0,568	1,53	1,046
<i>Сравнение потомков отцов S_1 и S_7</i>						
Живая масса	2,396±1,439	2,820	1,99	1,638	1,53	1,275
Настриг шерсти	0,034±0,167	0,327	1,99	0,198	1,53	1,588
<i>Сравнение потомков отцов S_2 и S_3</i>						
Живая масса	7,399±1,539	3,016	1,99	4,733	1,62	1,025
Настриг шерсти	0,149±0,238	0,466	1,99	0,651	1,74	2,071
<i>Сравнение потомков отцов S_3 и S_5</i>						
Живая масса	8,534±1,827	3,581	1,99	4,471	1,76	1,809
Настриг шерсти	0,020±0,283	0,555	1,99	0,068	1,76	1,134

Для этого используется метод доверительных границ. Гарантированный минимум превосходства, согласно [6], определяется выражением

$$\Delta = \Delta D - g. \quad (11)$$

Таким образом, последовательность действий при проверке статистической значимости различий хозяйственно полезных признаков должна быть следующей:

1. Статистическая проверка исходных данных, включающая в себя исключение аномаль-

ных измерений и проверку нормальности распределения измеренных величин.

2. Проведение статистических тестов для каждой пары генеалогических групп и принятие одной из двух гипотез (H_0 или H_A) о значимости различий между хозяйственно полезными признаками в этих группах.

3. Определение гарантируемого минимума превосходства одной генеалогической группы над другой.

Имея данные о неслучайности различий, можно приступать к оценке племенной ценности баранов-производителей.

Литература

1. Чижова Л. Н., Селионова М. И., Силкина С. Ф. Способ оценки, прогноза продуктивности сельскохозяйственных животных в раннем возрасте на основе биохимических тест-систем. Ставрополь, 2010. 41 с.
2. Селионова М. И., Бобрышова Г. Т. Овцеводство Ставропольского края, настоящее и будущее // Овцы, козы, шерстяное дело. 2016. № 1. С. 4–7.
3. Данко П. Е., Попов А. Г., Кожевникова Т. Я. Высшая математика в упражнениях и задачах : учеб. пособие для вузов. В 2-х ч. Ч. II. М. : Высшая школа, 1999. 416 с.
4. Румшинский Л. З. Математическая обработка результатов эксперимента : Справочное руководство. М. : Главная редакция физико-математической литературы изд-ва «Наука», 1971. 192 с.
5. Вентцель Е. С. Теория вероятностей : учебник для вузов. 6-е изд. М. : Высшая школа, 1999. 576 с.
6. Лакин Г. Ф. Биометрия. Изд. 4-е, перераб. и доп. М. : Высшая школа, 1990. 350 с.
7. Таблица критических значений [Электронный ресурс]. URL: http://gym42.ru/stat/Book/Data/page_7.htm (дата обращения: 22.01.2018).

References

1. Chizhova L. N., Selionova M. I., Silkina S. F. Method of estimation, prediction of productivity of farm animals in early age on the basis of biochemical test systems. Stavropol, 2010. 41 p.
2. Selionova M. I., Bobryshova G. T. Sheep breeding of the Stavropol territory, the present and the future // Sheeps, Goats, Wool Business. 2016. № 1. P. 4–7.
3. Danko P. E., Popov A. G., Kozhevnikova T. Ya. Higher mathematics in exercises and tasks: a textbook for universities. In 2 v. V II. M. : Higher school, 1999. 416 p.
4. Rumshinskiy L. Z. Mathematical processing of experimental results: reference guide. M. : Main edition of physical and mathematical literature publishing house «Nauka», 1971. 192 p.
5. Venttsel' E. S. Probability theory. Textbook for high schools. 6 edition. M. : Higher school, 1999. 576 p.
6. Lakin G. F. Biometrics. Edition 4 – supplemented and revised. M. : Higher school, 1990. 350 p.
7. Table of critical values [Electronic resource]. URL: http://gym42.ru/stat/Book/Data/page_7.htm (date of access: 22.01.2018).