УДК 356.489.123.127

ПОСТРОЕНИЕ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО КОМПЛЕКСА В ЗООТЕХНИИ

М. Н. БОРИСЕВИЧ

УО «Витебская ордена «Знак Почета» государственная академия ветеринарной медицины»,
г. Витебск, Республика Беларусь, 210026

(Поступила в редакцию 28.01.2020)

При решении многих зоотехнических задач, имеющих отношение к животноводству, часто прибегают к построению так называемого статистического комплекса, основу которого составляет дисперсионный анализ. Так, например, для познания закономерностей изменчивости объектов какой-либо совокупности важно вычленить долю влияния отдельных факторов. Для одного фактора (в нескольких градациях) необходимую информацию о его силе и достоверности можно получить, например, путём привлечения однофакторного дисперсионного комплекса.

В данной статье этот подход предложен для оценки влияния породы свиноматок на их плодовитость. С его помощью можно оценить межгрупповые различия на предмет их достоверности – обусловлены ли они изучаемым фактором (породой свиноматок) или являются случайными. Если различия достоверны и обусловлены влиянием наследственности матерей, то представляется возможным оценить также и коэффициент наследуемости.

В задаче учтённым фактором является порода свиноматок, поэтому градаций (групп) в комплексе три (лакомб, ландрас, донская). Поскольку во всех группах одинаковое число наблюдений – по 10 свиноматок, то следует строить равномерный статистический комплекс при малом числе наблюдений (n = 10). Поэтому решение задачи сводится к построению статистической таблицы дисперсионного анализа для указанных выше градаций. Анализ табличных данных позволяет сформулировать следующее заключение, имеющее важное практическое значение – достоверное влияние породы свиноматок на их плодовитость лишь на 25 % обусловлено изменчивостью генотипов и в значительно большей степени (75 %) зависит от случайных факторов.

***Ключевые слова:** статистический комплекс, однофакторный, дисперсионный, порода свиноматок, плодовитость.*

In solving many zootechnical problems related to animal husbandry, they often resort to the construction of the so-called statistical complex, the basis of which is analysis of variance. So, for example, in order to understand the laws of the variability of objects of an aggregate, it is important to isolate the proportion of influence of individual factors. For one factor (in several gradations), the necessary information about its strength and reliability can be obtained, for example, by involving a one-factor dispersion complex.

In this article, this approach is proposed to assess the influence of the pig-uterus breed on their fertility. With its help, it is possible to evaluate intergroup differences in terms of their reliability – whether they are due to the studied factor (sows breed) or are random. If the differences are significant and due to the influence of the heredity of the mothers, then it seems possible to evaluate the coefficient of heritability as well.

In the task, the sow breed is the factor taken into account, therefore there are three gradations (groups) in the complex (lacombe, landras, Don). Since in all groups the same number of observations is 10 sows each, a uniform statistical complex should be built with a small number of observations (n = 10). Therefore, the solution of the problem is reduced to the construction of a statistical table of variance analysis for the above gradations. The analysis of tabular data allows us to formulate the following conclusion, which is of great practical importance - the reliable influence of sows on their fertility is only 25 % due to variability of genotypes and to a much greater extent (75 %) depends on random factors.
***Key words**: statistical complex, one-factor, dispersion, sows breed, fecundity.*

Введение. В практической деятельности зоотехников при проведении экспериментальных исследований возникает необходимость в установлении влияния некоторых факторов на полученные результаты. Существует ряд статистических методов, позволяющих определить силу, направление, закономерности влияния факторов на результат как в генеральной, так и в выборочной совокупностях (например, расчет критерия Фишера, корреляционный анализ, регрессия, критерий согласия Пирсона и др.) [1]. Дисперсионный анализ был разработан и предложен английским ученым, математиком и генетиком Рональдом Фишером в 20-х годах XX века [2].

В зоотехнии его используют для изучения влияния одного или нескольких факторов на результативный признак. Он основан на принципе «отражения разнообразий значений факторных на разнообразие значений результативного признака» и устанавливает силу влияния факторов в выборочных совокупностях [3].

Сущность метода дисперсионного анализа заключается в измерении отдельных дисперсий (общей, факториальной, остаточной) и определении силы (доли) влияния изучаемых факторов (оценки роли каждого из факторов, либо их совместного влияния) на результативные признаки [4].

В общем случае это статистический подход, связанный с оценкой связи между факторными и результативным признаками в различных группах, отобранных случайным образом. В его основе лежит анализ отклонений всех единиц исследуемой совокупности от среднего арифметического. В качестве меры отклонений выбирается дисперсия – средний квадрат отклонений. При этом отклонения, вызываемые воздействием факторного признака (фактора), сравниваются с величиной отклонений, обусловленных случайными обстоятельствами. Если отклонения, вызываемые факторным признаком, более существенны, чем случайные отклонения, то считается, что факторный признак оказывает существенное влияние на признак результативный [5].

Для вычисления дисперсии значения отклонений каждой варианты (каждого зарегистрированного числового значения признака) от среднего арифметического возводят в квадрат. Тем самым избавляются от отрицательных знаков. Затем эти отклонения (разности) суммируют и делят на число наблюдений, т.е. усредняют отклонения. Таким образом получают значения дисперсий [6].

Важным методическим значением для применения дисперсионного анализа в зоотехнии является правильное формирование выборки. В зависимости от поставленной цели и задач выборочные группы (контрольная и экспериментальная) могут формироваться случайным образом независимо друг от друга Такие выборки называются независимыми.

Нередко результаты воздействия факторов исследуются у одной и той же выборочной группы (например, у одних и тех же животных) до и после воздействия (кормления, лечения, профилактики, реабилитационных мероприятий), такие выборки называются зависимыми.

Дисперсионный анализ, в котором проверяется влияние одного фактора, называется однофакторным (одномерный анализ). При изучении влияния более чем одного фактора используют многофакторный дисперсионный анализ (многомерный анализ). При этом следует четко различать факторные и результативные признаки: факторные признаки влияют на изучаемое явление, результативные – изменяются под влиянием факторных признаков. При проведении дисперсионного анализа могут использоваться как качественные, так и количественные признаки [7].

Среди методов дисперсионного анализа можно назвать два наиболее распространенных: метод Фишера, применяется в однофакторном анализе, когда совокупная дисперсия всех наблюдаемых значений раскладывается на дисперсию внутри отдельных групп и дисперсию между группами; и метод общей линейной модели – в его основе лежит корреляционный или регрессионный анализ, используемый в многофакторном построении [8].

Обычно в зоотехнических исследованиях используются только однофакторные, максимум двухфакторные дисперсионные комплексы. Что касается многофакторных комплексов, то их можно использовать с оговоркой – анализируя последовательно комплексы одно- или двухфакторные, выделяемые из всей наблюдаемой совокупности. Условия применения дисперсионного анализа укладываются в следующие положения: 1) задачей исследования является определение силы влияния одного (до трех) факторов на результат или определение силы совместного влияния различных факторов; 2) изучаемые факторы должны быть независимыми (несвязанными) между собой; 3) подбор групп для исследования проводится рандомизированно (путем случайного отбора, от английского random, т.е. выбранных наугад); 4) можно применять как количественные, так и качественные признаки [9].

При проведении однофакторного дисперсионного анализа рекомендуется (как необходимое условие применения): 1) нормальность распределения анализируемых групп или соответствие выборочных групп генеральным совокупностям с нормальным распределением; 2) независимость (несвязанность) распределения наблюдений в группах; 3) наличие частоты (повторность) наблюдений [10].

Нормальность распределения определяется кривой Гаусса (Де Мавура), которую можно описать функцией y = f(x), так как она относится к числу законов распределения, используемых для приближенного описания явлений, носящих случайный, вероятностный характер. Предмет зоотехнических исследований – явления вероятностного характера, нормальное распределение в таких исследованиях встречается весьма часто. Важными в этом смысле являются принципы применения метода дисперсионного анализа. Сводятся они к следующему. Сначала формулируется нулевая гипотеза, то есть предполагается, что исследуемые факторы не оказывают никакого влияния на значения результативного признака, а выявленные различия являются случайными. Затем определяется, какова вероятность получения наблюдаемых (или более сильных) различий при условии справедливости нулевой гипотезы. Если эта вероятность мала (максимальную приемлемую вероятность отвергнуть верную нулевую гипотезу определяют уровнем значимости $\alpha = 0,05$), то нулевая гипотеза отвергается и заключается, что результаты исследования статистически значимы. Однако это еще не значит, что доказано действие именно изучаемых факторов (вопрос, прежде всего, планирования исследования), но все же маловероятно, что результат обусловлен случайными факторами [11]. При выполнении всех условий применения дисперсионного анализа, разложение общей дисперсии математически выглядит следующим образом:

$$C_{общ.} = C_{факт} + C_{ост.},$$

где $C_{общ.}$ – общая дисперсия наблюдаемых значений (вариант), характеризуется разбросом вариант от общего среднего, описывает вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию; $C_{факт}$ – факторная (межгрупповая) дисперсия, характеризуется различием средних в каждой группе и зависит от влияния исследуемого фактора, по которому дифференцируется каждая группа; $C_{ост.}$ – остаточная (внутригрупповая) дисперсия, характеризует рассеяние вариант внутри групп, отражает случайную вариацию, т.е. ту часть вариации, происходящую под влиянием неуточненных факторов и не зависящую от признака, положенного в основу группировки. Вариация изучаемого признака зависит от силы влияния каких-то неучтенных случайных факторов, как от организованных (заданных исследователем), так и от случайных (неизвестных) факторов [12].

В данной статье представленный выше дисперсионный подход предложен для оценки влияния породы свиноматок на их плодовитость. В основу статьи положены данные табл.1.

Таблица 1. **Породы свиноматок и их плодовитость**

Порода свиноматок	Плодовитость свиноматок, гол									
Лакомб	10	10	9	9	9	10	9	11	10	10
Ландрас	11	11	11	9	11	9	12	11	12	11
Донская	10	11	10	10	9	11	11	10	10	11

Основная часть. С помощью дисперсионного метода можно оценить межгрупповые различия на предмет их достоверности – и в частности, обусловлены ли эти различия изучаемым фактором (породой свиноматок) или являются случайными. Если же различия достоверны и обусловлены влиянием наследственности матерей, то представляется возможным оценить также и коэффициент наследуемости.

В задаче учтённым фактором является порода свиноматок, поэтому градаций (групп) в комплексе три (s_1 – лакомб, s_2 – ландрас, s_3 – донская). Поскольку во всех группах одинаковое число наблюдений – по 10 свиноматок, то следует строить равномерный статистический комплекс при малом числе наблюдений ($n = 10$). Поэтому решение задачи сводится к построению таблицы статистического комплекса дисперсионного анализа для названных выше градаций. Построение данной таблицы сопряжено с рядом расчетных шагов, так или иначе связанных с вычислением соответствующих статистических величин. Мы не станем следовать этому правилу, а приведем окончательный вид таблицы (таблица 2) со всем числовым массивом, а затем поясним механизм получения всех расчетных значений.

В табл. 2 отражены варианты признака (X_i) для трёх градаций s_1 – лакомб, s_2 – ландрас, s_3 – донская.

В каждой группе (градации) определены суммы вариант $\sum X_i$ и рассчитаны частные средние величины $\overline{X}_i$ по формуле для малой выборки. Сумма вариант в первой градации равна 97, во второй – 108, в третьей – 103. Соответственно, частная средняя первой группы 9,7 гол.; второй – 10,8; третьей – 10,3 гол.

Для всех вариант комплекса рассчитано среднее арифметическое значение $\overline{X}_o$ по формуле:

$$\overline{X}_o = \frac{\sum\sum X_i}{\sum n_i},$$

где $\sum\sum X_i$ – сумма суммированных вариант во всех группах; $\sum n_i$ – количество особей во всех группах.

В нашем примере сумма вариант в первой группе равна 97, во второй – 108, в третьей – 103. В каждой группе по 10 голов, поэтому количество особей во всех группах равно 30. Полученные значения подставляем в формулу для расчёта среднего арифметического значения для всех вариант статистического комплекса $\overline{X}_o$ (общей средней):

$$\overline{X}_o = \frac{\sum\sum X_i}{\sum n_i} = \frac{97+108+103}{10+10+10} = 10{,}27 \text{ гол.}$$

Для каждой градации определены отклонения частных средних каждой группы от общей средней величины $(\overline{X}_i - \overline{X}_o)$. Найденные значения возведены в квадрат и вычислена их сумма $\sum(\overline{X}_i - \overline{X}_o)^2$ – в дальнейшем она будет использована для нахождения факториальной дисперсии C_x.

Сопоставление между собой частных средних даёт основание предположить, что степень их разнообразия может служить показателем силы влияния фактора. Как уже известно, частные средние (т.е. средние отдельных групп – выборок) могут отличаться друг от друга за счёт случайности выборок. Поэтому чем больше степень разнообразия частных средних по сравнению с уровнем их разнообразия, обусловленным случайностями, тем сильнее влияние фактора. Можно рассмотреть вопрос о влиянии фактора и в другом контексте: чем меньше разнообразие частных средних по сравнению с уровнем их случайного разнообразия, тем сильнее влияние фактора при его стабилизирующем уравнивающем действии.

Во всех группах определены отклонения средней арифметической от каждой варианты $(X_i - \overline{\overline{X}}_i)$ и полученные значения возведены в квадрат $(X_i - \overline{\overline{X}}_i)^2$. Данные занесены в соответствующие столбцы табл. 2.

В каждой группе определена сумма квадратов отклонений $(X_i - \overline{\overline{X}}_i)^2$.

Полученные результаты по группам суммированы – $\sum\sum (X_i - \overline{\overline{X}}_i)^2$ – эта сумма отражает меру случайной дисперсии C_z. На этом заканчивается заполнение табл.2.

Таблица 2. **Таблица статистического комплекса дисперсионного анализа для трёх градаций породы свиноматок**

№ строк	№ пп.	лакомб - s_1			ландрас - s_2			донская - s_3		
		X_i	$X_i-\overline{X_i}$	$(X_i-\overline{X}_i)^2$	X_i	$X_i-\overline{X_i}$	$(X_i-\overline{X}_i)^2$	X_i	$X_i-\overline{X_i}$	$(X_i-\overline{X}_i)^2$
1	**1**	**10**	**0,3**	**0,09**	**11**	**0,3**	**0,09**	**10**	**0,3**	**0,09**
2	**2**	**10**	**0,3**	**0,09**	**11**	**0,3**	**0,09**	**11**	**0,3**	**0,09**
3	**3**	**9**	**-0,7**	**0,49**	**11**	**-0,7**	**0,49**	**10**	**-0,7**	**0,49**
4	**4**	**9**	**-0,7**	**0,49**	**9**	**-0,7**	**0,49**	**10**	**-0,7**	**0,49**
5	**5**	**9**	**-0,7**	**0,49**	**11**	**-0,7**	**0,49**	**9**	**-0,7**	**0,49**
6	**6**	**10**	**0,3**	**0,09**	**9**	**0,3**	**0,09**	**11**	**0,3**	**0,09**
7	**7**	**9**	**-0,7**	**0,49**	**12**	**-0,7**	**0,49**	**11**	**-0,7**	**0,49**
8	**8**	**11**	**0,3**	**1,69**	**11**	**0,3**	**1,69**	**10**	**0,3**	**1,69**
9	**9**	**10**	**0,3**	**0,09**	**12**	**0,3**	**0,09**	**10**	**0,3**	**0,09**
10	**10**	**10**	**0,3**	**0,09**	**11**	**0,3**	**0,09**	**11**	**0,3**	**0,09**
11	n_i	**10**	-	$\sum = 4{,}10$	**10**	-	$\sum = 9{,}60$	**10**	-	$\sum = 4{,}10$
12	$\sum X_i$	**97**			**108**			**103**		
13	$\overline{X}_i$	**9,7**			**10,8**			**10,3**		
14	$\overline{X_O}$	**10,27**								
15	$\overline{X_i}-\overline{X_o}$	**-0,57**			**0,53**			**0,03**		
16	$(\overline{X_i}-\overline{X_o})^2$	**0,3249**			**0,28**			**0,009**		
17	$\sum(\overline{X}_i-\overline{X}_o)^2$	**0,6067 (для C_x)**								
18	$\sum(X_i-\overline{X_i})^2 = C_z$	**17,8**								

Перейдем теперь к составлению сводной таблицы дисперсионного анализа – табл.3.

Таблица 3. **Сводная таблица дисперсионного анализа**

Источник варьирования	Число степеней свободы	Сумма квадратов (дисперсия)	Средний квадрат (варианса)	Доля влияния
Группы (x) межгрупповое варьирование	2	6,067	3,034	0,25
Объекты (z) внутригруппвовое варьирование	27	17,8	0,659	0,75
Общее варьирование (y)	29	23,867	-	1

Изначально рассчитываем число степеней свободы:

$$\nu_x = k - 1, \quad \nu_z = n - k, \quad \nu_y = n - 1$$

где k – число групп, n = общее число особей.

Затем определяем межгрупповую дисперсию:

$$C_x = \sum (\overline{X}_i - \overline{X}_o)^2 \times n_o,$$

где расчет n_o осуществляется, если в группах задано неодинаковое количество вариант, поэтому n_o уместно назвать поправкой для межгрупповой дисперсии. В общем случае n_o определяется по формуле:

$$n_o = \frac{1}{a-1} \times \left(n - \frac{\sum n_i^2}{n} \right),$$

где a – число групп, n = общее число вариант, n_i – число вариант в каждой группе. В нашей задаче принято одинаковое число вариант, поэтому $n_o = 10$.

Дальше рассчитываем дисперсии:

$$C_z = \sum (X_i - \overline{X}_i)^2$$

$$C_y = C_x + C_z$$

и заносим в столбец «сумма квадратов (дисперсия)».

На следующем этапе вычисляем вариансы – факториальную и случайную:

$$\sigma_x^2 = \frac{C_x}{\nu_x};$$

$$\sigma_z^2 = \frac{C_z}{\nu_z};$$

и также заносим в столбец «средний квадрат (варианса)».

В завершение оценим долю факторов: учтенного, неучтенного (или случайного) и изменчивого. Для этого воспользуемся формулами:

$$\eta_x^2 = \frac{C_x}{C_y};$$

$$\eta_z^2 = \frac{C_z}{C_y}$$

$$\eta^2_y = \eta^2_x + \eta^2_z$$

Мы подошли к завершающей стадии исследований. Однако на данном этапе считать их полностью законченными нельзя. Во-первых, потому, что неизвестно, насколько они достоверны, а во-вторых, можно ли ими руководствоваться для оценки силы влияния породы свиноматок на их плодовитость и если можно, то согласно какого алгоритма.

Решением этих двух задач мы сейчас и займемся.

Первая задача связана с определением достоверности воздействующего фактора (породы свиноматок) и его влияния на изменчивость признака (плодовитость свиноматок), а именно, докажем математически строго, достоверно или недостоверно влияние породы свиноматок на их плодовитость. Для этого рассчитаем эмпирический критерий достоверности Фишера ($F_{эмп.}$) и сравним его со стандартным значением (F_{st}) согласно принятым в животноводстве методикам [13].

Критерий Фишера ($F_{эмп.}$) найдем как отношение варианcы факториальной σ_x^2 (межгрупповой) к вариансе случайной σ_z^2 (внутригрупповой):

$$F_{эмп.} = \frac{\sigma_x^2}{\sigma_z^2}.$$

Вычисленную величину $F_{эмп.} = 4{,}604$ сравним с табличными значениями (F_{st}), представленными в [13] для заданных степеней свободы (равных 2 по горизонтали и 27 по вертикали):

[redacted]

Приходим к заключению – рассчитанный критерий Фишера больше, чем минимальное требование для достоверности результата – $F_{эмп.} = 4{,}604 > F_{st.} = 3{,}3$ при 0,95, но меньше второго порога вероятности – $F_{эмп.} = 4{,}604 < F_{st.} = 5{,}5$ при $P=0{,}99$. Следовательно, нулевая гипотеза о недостоверности воздействующего фактора (породы свиноматок) может быть отвергнута, а значит, может иметь место противоположная ей гипотеза – о достоверном влиянии породы свиноматок на их плодовитость.

Поскольку влияние изучаемого фактора на изменчивость признака достоверно, то совершенно уместно перейти к расчёту коэффициента наследуемости с тем, чтобы установить долю или силу влияния этого фактора на изменчивость признака, а точнее на плодовитость свиноматок. Это и есть вторая наша задача.

Доля влияния наследственности в общей фенотипической изменчивости признака (по-другому, коэффициент наследуемости изучаемого признака) в дисперсионном анализе выражается в долях от единицы или в процентах и вычисляется по простой математической формуле:

$$h^2 = \eta^2_x = \frac{C_x}{C_y}.$$

Она приведена в сводной табл. 3:

$$h^2 = \eta^2_x = 0{,}25 = 25\%.$$

Заключение. Следовательно, можно заключить, что достоверное влияние породы свиноматок на их плодовитость лишь на 25 % обусловлено изменчивостью генотипов и в значительно большей степени (75 %) зависит от случайных факторов. Сформулированный вывод заслуживает внимания специалистов. Он может быть количественно подтвержден в рамках однофакторного дисперсионного анализа. При этом возможности этого подхода не ограничиваются примером, рассмотренным в данной статье. Они гораздо шире и значительно более действеннее для самых разных приложений в животноводстве.

ЛИТЕРАТУРА

1. Любищев, А. А. Дисперсионный анализ в биологии / А. А. Любищев. – М., 2012. – 101 с.

2. Шеффе, Г. Дисперсионный анализ / Г. Шеффе. – М.: Главная редакция физико-математической литературы издательства «Наука», 2011. – 512 с.

3. Елинов, Н. П. Основы биотехнологии. / Н. П. Елинов. – Спб: Изд. Фирма «Наука», 1995. – 600 с.

4. Ашмарин, И. В. Статистические методы в микробиологическом исследовании / И. В. Ашмарин, А. А. Воробьев – Л.: Медгиз, 1962. – 160 с.

5. Гмурман, В. Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2003. – 523с.

6. Рассел, Джесси Т-критерий Стьюдента: моногр. / Джесси Рассел. - М., 2013. – 741 с.

7. Адлер, Ю. П. Введение в планирование эксперимента. – М., 1969. – 159 с.

8. Адлер, Ю. П. Предпланирование эксперимента. – М.: Знание, 1980. – 72 с.

9. Айвазян, С. А. Статистическое исследование зависимостей. – М.: Металлургия, 1968. – 227 с.

10. Айвазян, С. А., Бежаева Е. И., Староверов О. В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240 с.

11. Айвазян, С. А., Еников И. С., Мешалкин Л. Д. Прикладная статистика: Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. – 472 с.

12. Азгальдов, Д. Д. Теория и практика оценки качества товаров. – М.: Экономика, 1982. – 256 с.

13. Гордиенко, Н. В. Новейшая энциклопедия животноводства для профессионалов и любителей / Н. В. Гордиенко. – М.: Удача, 2008. – 448 с.