УДК 356.489.123.127

ПОСТРОЕНИЕ ДВУХФАКТОРНОГО ДИСПЕРСИОННОГО КОМПЛЕКСА В ЗООТЕХНИИ

М. Н. БОРИСЕВИЧ

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

(Поступила в редакцию 30.01.2020)

В зоотехнической практике и научных исследованиях часто возникают ситуации, когда требуется подвергнуть сравнительному анализу одновременно не две, а несколько групп животных: например, при испытании пород, линий, при оценке производителей по качеству потомства, определении стандартности линии или отобранной для каких-либо целей группы животных, а также в других случаях, связанных с изучением влияния различных факторов (биологических, кормовых, гигиенических) на рост, развитие, продуктивность, здоровье животных и иные интересующие зоотехника (или биолога) признаки. В большинстве случаев важно не только установить факт воздействия на изучаемый объект того или иного фактора или даже группы факторов (т.е. установить достоверность влияния), но и выяснить степень этого воздействия, оценить его относительную силу, например, у особей всего стада, группы особей, выборочной совокупности (малой или большой). Для этого пользуются дисперсионным анализом. В этом случае, как правило, обработке подвергаются выборочные данные, оформленные в специальный статистический комплекс.

В статье рассматривается алгоритм построения двухфакторного статистического комплекса применительно к оценке доли влияния метода разведения (фактор А) и уровня кормления (фактор В) на проявление у свиней гетерозиса по показателям их плодовитости. Данные предоставлены кафедрой животноводства УО ВГАВМ. В проведенном сотрудниками опыте были использованы чистопородное разведение (А1), инбридинг (А2), межпородное скрещивание (А3). Животных содержали на двух типах рациона: со средним (В1) и высоким (В2) уровнем питательных веществ. В опыте было 45 свиноматок, от которых учтено 45 опоросов и получено 468 поросят.

Ключевые слова: *статистический комплекс, метод разведения, уровень кормления, свиньи, гетерозис, плодовитость.*

In livestock practice and scientific research, situations often arise when it is necessary to perform comparative analysis not simultaneously, but several groups of animals: for example, when testing breeds, lines, when evaluating producers according to the quality of the offspring, determining the standardness of the line or selected for any goals of a group of animals, as well as in other cases related to the study of the influence of various factors (biological, feed, hygiene) on the growth, development, productivity, animal health and other interests ehnika (or biologist) signs. In most cases, it is important not only to establish the fact of the impact of a particular factor or even a group of factors on the studied object (i.e., to establish the reliability of the effect), but also to determine the degree of this effect and evaluate its relative strength, for example, in individuals of the whole herd, group individuals, sample population *(small or large). To do this, use variance analysis. In this case, as a rule, sample data processed in a special statistical complex is processed.*

The article discusses the algorithm for constructing a two-factor statistical complex in relation to assessing the share of the influence of the breeding method (factor A) and the level of feeding (factor B) on the manifestation of heterosis in pigs according to their fertility indicators. Data provided by the department of animal husbandry UO VGAVM. In the experiment conducted by the staff, pure-breed breeding (A1), inbreeding (A2), and interbreeding (A3) were used. Animals were kept on two types of diet: with an average (B1) and high (B2) level of nutrients. In the experiment there were 45 sows, from which 45 farrowing animals were taken into account and 468 piglets were obtained.

***Key words**: statistical complex, breeding method, feeding level, pigs, heterosis, fecundity.*

Введение. Изменчивость свойственна всем живым существам и является одним из факторов эволюции, поскольку обеспечивает материал для естественного отбора, создавая новые варианты признаков, а также множество их прежних комбинаций. Она обусловлена наследственностью и служит основой для выведения новых пород животных. Выражается изменчивость в огромном разнообразии признаков. Поэтому следует отличать изменчивость признаков (разнообразие) от понятия изменчивости как способности живых существ приобретать новые признаки и свойства в процессе онтогенеза. Изменчивость признака представляет собой степень его варьирования у группы объектов, которая зависит от выраженности данного признака у отдельных особей этой группы [1].

Все хозяйственно полезные признаки у сельскохозяйственных животных подразделяют на качественные (описываются словами, например масть, пол, тип телосложения, тип движения лошади и др.) и количественные (измеряются, подсчитываются и выражаются цифрами, например, удой, живая масса, настриг шерсти, плодовитость, резвость и др.). И поскольку существуют количественные и качественные признаки, то, соответственно, различают количественную и качественную изменчивость.

Многие качественные признаки имеют только два возможных альтернативных состояния, например, пол – мужской или женский, животные – здоровые или больные, рогатые или комолые. Некоторые качественные признаки могут иметь несколько состояний, например, по типу телосложения животные могут быть крепкой, грубой, рыхлой, нежной и плотной конституции. Отдельные качественные признаки могут иметь количественные показатели, например, количество лейкоцитов у здоровых и больных животных. Количественная изменчивость может быть двух типов: непрерывная и прерывная (дискретная). При непрерывной изменчивости между значениями признака (вариантами)

нет резких границ и переходов, все определяется точностью измерений (удой, живая масса, содержание жира и белка в молоке и др.). То есть переход от одного количественного уровня признака к другому составляет непрерывный ряд величин.
Если различия между вариантами определяются целыми числами, то это прерывная (дискретная) изменчивость. Например, число поросят у одной свиноматки (плодовитость) выражается целым числом (9, 10, 11 и т.д.), или яйценоскость у кур, выраженная в количестве штук снесённых яиц. В зоотехнической и ветеринарной практике изучение корреляционной (соотносительной) изменчивости имеет большое значение. Сельскохозяйственные животные обладают большим разнообразием морфологических, физиологических, хозяйственно полезных признаков, из которых многие имеют важное значение для практики животноводства и на их улучшение и совершенствование направлена селекционно-племенная работа. В то же время большое число признаков не играет практической роли и не является объектом селекционного воздействия. Поэтому в практической работе селекционеру важно знать не только признаки, связанные между собой, но и направление этой связи, чтобы получить желаемый результат. Например, молочная продуктивность и мясные качества (способность к откорму) у крупного рогатого скота находятся в отрицательной взаимосвязи, поэтому пока не выведены породы, сочетающие высокие значения этих признаков [2].

Изменчивость признаков живых организмов изучают разными методами. Одним из них является дисперсионный подход [3], основу которого составляют приемы вариационно-статистического анализа. Данный метод основывается на анализе массовых данных. Известно, что цифровые данные, собранные в процессе массовых исследований, хотя и имеют определенное значение, ещё недостаточны для того, чтобы сделать из них необходимые выводы. Фактически такие данные представляют собой сырой материал, который нуждается в математической обработке. В массовых исследованиях невозможно извлечь нужную информацию, оценить надёжность отдельных суммарных показателей, убедиться в достоверности или недостоверности наблюдаемых между ними различий, если данные не упорядочены, не систематизированы и математически не обработаны. Эту задачу в зоотехнии решает дисперсионный анализ [3] – обработке подвергаются выборочные данные, оформленные в специальный статистический комплекс.

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

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.

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. По-сути, обработке подвергаются выборочные данные, оформленные в специальный статистический комплекс.

Основная часть. В данной статье рассматривается алгоритм построения двухфакторного статистического комплекса применительно к оценке доли влияния метода разведения (фактор А) и уровня кормления (фактор В) на проявление у свиней гетерозиса по показателям их плодовитости. Данные предоставлены кафедрой животноводства УО ВГАВМ. В проведенном сотрудниками опыте были использованы чистопородное разведение (А1), инбридинг (А2), межпородное скрещивание (А3). Животных содержали на двух типах рациона: со средним (В1) и высоким (В2) уровнем питательных веществ. В опыте было 45 свиноматок, от которых учтено 45 опоросов (n) и получено 468 поросят. Исходные экспериментальные данные приведены в табл.1.

Исходя из данной таблицы ставилась задача получить следующие заключения, используя математический аппарат дисперсионного анализа: о доле влияния общефакториальной и случайной дисперсий с подтверждением их достоверности (заключение 1); о доле влияния метода разведения, уровня кормления и совместного использования метода и уровня с оценкой соответствующих им достоверностей (заключение 2).

С точки зрения привлекаемого метода при анализе любого признака исследователь встречается с влиянием не только учтённых факторов, но и случайных [4].

Учтённые (или организованные) факторы (X) – это факторы, которые изучаются и контролируются в опыте.

Неучтённые (или случайные) факторы (Z) – те факторы, которые не учитываются и не контролируются схемой опыта, но всеже оказывают влияние на вариабельность признака.

Двухфакторные статистические комплексы применительно к данной задаче позволяют выявить не только долю влияния каждого фактора (A и B), но и долю их совместного влияния, вызывающего дополнительную дисперсию (C_{AB}). Она не является результатом простой суммы действия обоих факторов, а возникает благодаря совместному и интегрированному воздействию обоих факторов, т.е. AB [5].

Таблица 1. **Исходные экспериментальные данные для построения двухфакторного статистического комплекса дисперсионного анализа**

Метод разведения (A)	A_1		A_2		A_3	
Уровень кормления (B)	B_1	B_2	B_1	B_2	B_1	B_2
Плодовитость маток (X_i)	9	10	8	9	9	9
	11	9	9	10	10	12
	10	11	9	11	9	11
	9	10	11	12	11	10
		11	9	10	12	13
		10		10	11	11
		10		10		12
		11		10		12
				10		15
				11		10
						10
						11

Поэтому в двухфакторных комплексах факториальная дисперсия C_x разлагается на три составляющие дисперсии:

$$C_x = C_A + C_B + C_{AB}.$$

В равномерных и пропорциональных комплексах (к которым относится и наш комплекс) дисперсия C_{AB} определяется простой разностью:

$$C_{AB} = C_x - C_A - C_B.$$

Алгоритм двухфакторного статистического комплекса сводится к построению табл. 2, в которой в качестве исходной присутствует табл. 1. Приведенная в табл.2 информация позволяет получить нужное нам суждение о доле влияния общефакториальной и случайной дисперсий и установить их достоверность, что имеет прямое отношение к первому заключению, названному выше.
Этим мы сейчас и займемся.

При проведении дисперсионного анализа пользуются следующими величинами [6]:
С – дисперсия признака, или изменчивость, которая может быть:

C_y – общей, C_x – частной или факториальной, C_z – случайной или остаточной;

σ^2 – варианса (или средний квадрат), отражающая меру дисперсии вариант в вариационном ряду; различают вариансы:

σ_y^2 – общую, σ_x^2 – факториальную, σ_z^2 – остаточную;

η_x^2 – доля изменчивости признака, обусловленная влиянием изучаемого фактора; η_z^2 – доля изменчивости признака, обусловленная влиянием случайных факторов, F_x – критерий достоверности Фишера.

Дисперсионный анализ осуществляется в несколько этапов путем обработки числового массива таблицы статистического комплекса (табл. 2) и последующим составлением сводной таблицы (табл. 3) [6].

Первый этап заключается в получении общей, факториальной и остаточной дисперсий C_y ; C_x ; C_z по формулам:

$$C_y = \sum X_i^2 - H = 5182 - 4867{,}2 = 314{,}8;$$
$$C_x = \sum h_x - H = 4886{,}8 - 4867{,}2 = 19{,}6;$$
$$C_z = \sum X_i^2 - \sum h_x = 5182 - 4886{,}8 = 295{,}2.$$

Проверяем, выполняется ли равенство:

$$C_y = C_x + C_z$$

т.е. 314,8 = 19,6 + 295,2, получаем тождество, подтверждающее правильность расчетов для величин С.

Второй этап состоит в вычислении долей каждой частной дисперсии в общей дисперсии, для чего рассчитываются величины η_x^2 ; η_z^2 .

Таблица 2. **Двухфакторный статистический комплекс дисперсионного анализа**

Факторы, показатели, порядок вычислений	Градации по факторам						
Метод разведения (А)	A_1		A_2		A_3		$k_A=3$
Уровень кормления (В)	B_1	B_2	B_1	B_2	B_1	B_2	$k_B=2$
Плодовитость	9	10	8	9	9	9	
маток (X_i)	11	9	9	10	10	12	$\sum X = 468$
	10	11	9	11	9	11	
	9	10	11	12	11	10	$H = \frac{(\sum X_i)^2}{n} = 4867,2$
		11	9	10	12	13	
		10		10	11	11	
		10		10		12	
		11		10		12	
				10		15	
				11		10	
n_i	4	8	5	10	6	12	$\sum n_i = n = 45$
пропорциональность отношения (ni)	1	2	1	2	1	2	-
$\sum X_i$	39	82	46	103	62	136	$\sum\sum X_i = \sum X = 468$
$(\sum X_i)^2$	1521	6724	2116	10609	3844	18496	-
$h_x = \frac{(\sum X_i)^2}{n_i}$	380,2	840,5	423,2	1060,9	640,7	1541,3	$\sum h_z = 4886,8$
$\sum X_i^2$	383	965	428	1067	648	1691	$\sum X_i^2 = 5182$
$\overline{X}_i = \frac{\sum X_i}{n_i}$	9,75	10,25	9,2	10,3	10,33	11,33	$X_o = \frac{468}{45} = 10,4$

Первая определяет общефакториальную долю влияния:

$$\eta_x^2 = \frac{C_x}{C_y} = \frac{19,6}{314,8} = 0,0622.$$

Она свидетельствует о том, что 6,22 % всей изменчивости обусловлено действием только учтенных в опыте факторов.

Вторая величина характеризует долю остаточной изменчивости:

$$\eta_z^2 = \frac{C_z}{C_y} = \frac{295,2}{314,8} = 0,9378.$$

Величина этой доли составляет 93,78 %, что в свою очередь означает, 93,78 %, изменчивости обусловлено неучтенными (случайными) факторами.

Подтвердим полученные расчеты проверкой равенства:

$$\eta_y^2 = 1 = \eta_x^2 + \eta_z^2 .$$

Получаем очевидное тождество 0,0622 + 0,9378 = 1, следовательно, расчеты выполнены правильно.

Третий этап сводится к корректированию полученных дисперсий на число степеней свободы (ν), вычисляемых для каждой дисперсии по формулам:

$$\nu_x = \text{k-1}$$
$$\nu_z = \text{n-k}$$
$$\nu_y = \text{n-1},$$

где n – количество всех вариант статистического комплекса; k – количество групп (градаций).

Правильность вычисления ν_y сверяем по формуле:

$$\nu_y = \nu_x + \nu_z .$$

Тождество 44 = 5 + 39 = 44 подтверждает корректность расчетов.

Корректированную дисперсию (средний квадрат, или вариансу) рассчитываем по формулам:

$$\sigma_x^2 = \frac{C_x}{\nu_x} = \frac{19,6}{5} = 3,92;$$

$$\sigma_z^2 = \frac{C_z}{\nu_z} = \frac{295,2}{39} = 7,5.$$

Четвертый этап дисперсионного анализа направлен на определение достоверности полученного значения факториальной дисперсии. Делается это для того, чтобы количественно подтвердить, достоверно ли влияние данного фактора на варьирующий признак.

Для этого используем коэффициент Фишера F_x (общефакториальный критерий достоверности), который получается в результате деления факториальной дисперсии на остаточную:

$$F_x = \frac{\sigma_x^2}{\sigma_z^2} = \frac{3,92}{7,50} = 0,52.$$

Находим табличные значения F_{st} при $\nu_1 = \nu_x = 5$ и $\nu_2 = \nu_z = 39$:

$$F_{st} = \begin{cases} 5,3 \text{ при } P = 0,999 \\ 3,5 \text{ при } P = 0,99 \\ 2,5 \text{ при } P = 0,95 \end{cases} .$$

Поскольку вычисленное $F_x < F_{st}$ для всех уровней вероятности (Р), а именно 0,52<2,5, то следует принять гипотезу, что общефакториальная дисперсия недостоверна. На этом формулировку заключения 1 можно считать законченной.

Теперь продолжим осуществлять ту часть дисперсионного анализа, которая поможет нам получить заключение 2, упоминавшееся выше. Она предполагает разложение общефакториальной дисперсии (C_x) на три составляющие C_A, C_B и C_{AB}. Сопоставляя полученные частные средние $(\overline{\overline{X}}_i)$, можно заметить некоторую закономерность в уровне плодовитости свиноматок по градациям факторов А и В, а именно, более высокая плодовитость наблюдается по группам промышленного скрещивания A_3B_1 и A_3B_2; наиболее низкая – по группам инбридинга A_2B_1 и A_2B_2.

Высокий уровень кормления повышает плодовитость во всех группах (A_1B_2; A_2B_2; A_3B_2).

Для доказательства доли влияния факторов А, В и АВ во всех комбинациях на показатель плодовитости и определения достоверности действия каждого из факторов продолжаем дисперсионный анализ по факторам. Составляем вспомогательную таблицу статистического комплекса (табл. 3), в которой обобщаем для каждой градации одного фактора суммарные показатели по градациям другого фактора. Обработку данных таблицы делаем последовательно для каждой градации факторов А и В.

В результате получаем величины $\sum h_A$ и $\sum h_B$ и частные средние по фактору без учета уровня другого фактора.

Величину дисперсии от совместного влияния факторов А и В выражаем значением C_{AB}, которое в равномерных и пропорциональных комплексах, а следовательно, и для нашего примера определяется простой разностью: $C_{AB} = C_x – C_A – C_B$. Частные дисперсии C_A, C_B и C_{AB} вычисляем по формулам:

$$C_A = \sum h_A - H = 4870{,}1 – 4867{,}2 = 2{,}9;$$

$$C_B = \sum h_B - H = 4875{,}3 – 4867{,}2 = 8{,}1;$$

$$C_{AB} = C_x - C_A - C_B = 19{,}6 – 2{,}9 – 8{,}1 = 8{,}6.$$

Таблица 3. **Вспомогательная таблица для дисперсионного анализа**

Градации по факторам	Объединяемые градации	n_i	$\sum X_i$	$(\sum X_i)^2$	$h_i = \frac{(\sum X_i)^2}{n_i}$	$X_i = \frac{\sum X_i}{n_i}$
A_1	B_1+B_2	12	121	14641	1220	10,83
A_2	B_1+B_2	15	149	22201	1466,6	9,93
A_3	B_1+B_2	18	198	39304	2183,5	11
по фактору А	-	45	468	-	4870,1	-
B_1	$A_1+A_2+A_3$	15	147	21609	1440,6	9,8
B_2	$A_1+A_2+A_3$	30	321	103041	3434,7	10,7
по фактору В	-	45	468	-	4875,3	-

Находим доли влияния А, В и АВ на результативный признак:

$$\eta_A^2 = \frac{C_A}{C_y} = \frac{2,9}{314,8} = 0,009, \text{ или } 0,9\%;$$

$$\eta_B^2 = \frac{C_B}{C_y} = \frac{8,1}{314,8} = 0,025, \text{ или } 2,57\%;$$

$$\eta_{AB}^2 = \frac{C_{AB}}{C_y} = \frac{8,6}{314,8} = 0,027, \text{ или } 2,73\%.$$

Проверяем правильность выполненных расчетов:

$$\eta_x^2 = \eta_A^2 + \eta_B^2 + \eta_{AB}^2$$

Тождество 6,22 = 0,9 + 2,57 + 2,73 подтверждает их корректность.
Определяем число степеней свободы по факторам:

$$\nu_A = k_A - 1 = 3 - 1 = 2;$$
$$\nu_B = k_B - 1 = 2 - 1 = 1;$$
$$\nu_{AB} = \nu_A \times \nu_B = 2 \times 1 = 2.$$

Рассчитываем частные варианcы:

$$\sigma_A^2 = \frac{C_A}{\nu_A} = \frac{2,9}{2} = 1,45;$$

$$\sigma_B^2 = \frac{C_B}{\nu_B} = \frac{8,1}{1} = 8,1;$$

$$\sigma_{AB}^2 = \frac{C_{AB}}{\nu_{AB}} = \frac{8,6}{1} = 8,6.$$

Подтверждаем достоверность влияния частных факторов, рассчитав $F_{эмп}$ и табличное F_{st}:

$$F_A = \frac{\sigma_A^2}{\sigma_z^2} = \frac{1,45}{7,50} = 0,193, \qquad \nu_A = 2; \nu_z = 39,$$

$$F_B = \frac{\sigma_B^2}{\sigma_z^2} = \frac{8,10}{7,50} = 1,08, \qquad \nu_B = 1; \nu_z = 39,$$

$$F_{AB} = \frac{\sigma_{AB}^2}{\sigma_z^2} = \frac{8,60}{7,50} = 1,14, \qquad \nu_{AB} = 1; \nu_z = 39,$$

при соответствующих:

$$F_{st} = \begin{cases} 8,5 \\ 5,2; \\ 3,2 \end{cases}$$

$$F_{st} = \begin{cases} 12,9 \\ 7,3; \\ 4,1 \end{cases}$$

$$F_{st} = \begin{cases} 12,9 \\ 7,3. \\ 4,1 \end{cases}$$

Сопоставляя вычисленные F_A, F_B и F_{AB} с табличными F_{st}, приходим к заключению, что все они меньше, чем F_{st}, для всех уровней вероятности. Следовательно, влияние раздельно действующих факторов А и В и их совместное действие АВ недостоверны и составляют небольшую долю в общей изменчивости признака.

Заключение. Двухфакторный анализ показал, что на плодовитость свиноматок не оказывают достоверного влияния ни метод разведения (его доля влияния составила лишь 0,9 %), ни уровень кормления (доля влияния 2,57 %), ни совместное использование первого и второго (доля совместного влияния также ничтожно мала и составляет лишь 2,73 %).

Сформулированный вывод заслуживает внимания специалистов. Он может быть количественно подтвержден в рамках двухфакторного 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. Гордиенко, Н. В. Новейшая энциклопедия животноводства для профессионалов и любителей / Н. В. Гордиенко. – М.: Удача, 2008. – 448 с.

2. Солдатов, А. П. Основы животноводства. Издание 3-е, переработанное и дополненное. / А. П. Солдатов. – Москва: Мир, 2012. – 146 с.

3. Любищев, А. А. Дисперсионный анализ в биологии / А. А. Любищев. – М.: Книга по Требованию, 2012. – 101 с.

4. Ибрагимов, Н. Х. Азбука группового анализа / Н. Х. Ибрагимов. – М.: [не указано], 2010. – 287 с.

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

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