

**ОСОБЕННОСТИ ПРИМЕНЕНИЯ МЕТОДА
ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ
СТАЦИОНАРНОГО ОПЫТА
ПРИ ЧАСТИЧНОЙ ИХ УТРАТЕ**

**С.В. Фролова, Л.И. Никонорова,
Л.В. Бобрович, Н.Е. Макова**

*ФГОУ ВПО «Мичуринский государственный аграрный
университет». г. Тамбов*

Рецензент А.И. Завражнов

Ключевые слова и фразы: группировка экспериментальных данных; коэффициент вариации; матрица дисперсионного анализа; обработка данных.

Аннотация: Рассмотрен подход к обработке экспериментальных данных с биологическими объектами методом приведения неортогональных схем к схемам ортогональным, за счет удаления вариантов, содержащих выпавшие данные.

Интересным для всех исследователей является вопрос обработки и использования – для сравнений и сопоставлений опытных вариантов – оставшихся данных при утрате части экспериментального материала по каким-либо непредвиденным причинам. В настоящее время, в связи с широким воздействием различных стресс-факторов окружающей среды на объекты исследований в сельском хозяйстве, в частности, в плодоводстве, и, прежде всего, на сами плодовые деревья (морозные зимы, чрезмерное выпадение осадков, заболачивание местности, нанесение вреда грызунами и т.д.), исследователи сталкиваются с проблемой утраты некоторых данных в поставленных экспериментах. Насколько неразрешим в данном случае вопрос о возможности работы с оставшимися данными с целью получения достоверных обоснованных выводов? Существуют разные методы решения данной задачи.

В тех случаях, когда исследователям приходится иметь дело со случаями выпадения одной-двух повторностей в опыте, например, при учете урожая, могут оказаться очень полезными простейшие статистические приемы, позволяющие теоретически приблизительно вычислить необходимые значения [1]. Совершенно очевидно, что в результате выпадения

Бобрович Л.В. – доктор сельскохозяйственных наук, доцент, заведующая кафедрой агроэкологии и защиты растений МичГАУ; Фролова С.В., Никонорова Л.И., Макова Н.Е. – сотрудники МичГАУ, г. Мичуринск.

данных опыт становится неполным и проводить сравнения вариантов при разном количестве повторных делянок по традиционной – ортогональной – схеме дисперсионного анализа в них нельзя. Поэтому либо необходимо использовать неортогональную схему, либо схему такого опыта необходимо привести к схеме опыта с полным набором делянок.

Таким образом, существует несколько возможных вариантов решения этой проблемы, что частично было рассмотрено в работах [3, 4, 7]. В настоящей работе рассматривается еще один, заслуживающий, на наш взгляд, внимания, метод – так называемый метод вычеркивания столбцов при обработке данных с использованием дисперсионного анализа [2]. Смысл методики его применения состоит в том, что при потере какого-либо значения (по конкретному варианту и повторности), в матрице дисперсионного анализа производится вычеркивание всего столбца, содержащего утраченный элемент.

Рассмотрим применение данного метода на конкретном примере. В исследованиях с различными привойно-подвойными комбинациями саженцев яблони в питомнике МичГАУ были получены данные по диаметру штамбиков, количеству листьев, площади листовой пластины саженцев яблони трех сортов на полукарликовом клоновом подвое 54-118. В качестве примера в табл. 1 представлены данные по диаметру штамбиков саженцев изучаемых сортов, послужившие в качестве исходной матрицы для проведения дисперсионного анализа.

Допустим, в эксперименте могли быть утрачены два значения (например, пятый элемент в первом варианте и второй элемент во втором варианте – мы имитировали процесс выпадения этих данных совершенно случайным образом, используя при этом таблицу случайных чисел).

Тогда, следуя требованиям применения рассматриваемого метода, мы должны исключить из исходной матрицы полностью два столбца – второй и пятый (табл. 2).

Таблица 1

Диаметр штамбиков саженцев яблони, мм
(исходная матрица с двумя выпавшими значениями)

Сорт (вариант)	Повторности									
	1	2	3	4	5	6	7	8	9	10
Северный синап	12,5	11,0	13,2	10,5	13,3	10,0	14,5	15,0	14,8	14,9
Спартан	10,1	10,9	10,5	11,5	9,0	10,7	9,0	9,0	10,5	10,3
Уэлси	10,2	9,5	10,2	10,2	10,0	9,8	11,5	10,0	12,3	11,0

Таблица 2

Диаметр штамбиков саженцев яблони, мм
(исходная матрица с выделением выпадающих столбцов)

Сорт (вариант)	Повторности									
	1	2	3	4	5	6	7	8	9	10
Северный синап	12,5	11,0	13,2	10,5	13,3	10,0	14,5	15,0	14,8	14,9
Спартан	10,1	10,9	10,5	11,5	9,0	10,7	9,0	9,0	10,5	10,3
Уэлси	10,2	9,5	10,2	10,2	10,0	9,8	11,5	10,0	12,3	11,0

Таблица 3

Диаметр штабиков саженцев яблони, мм
(матрица с вычеркнутыми столбцами)

Сорт (вариант)	Повторности							
	1	2	3	4	5	6	7	8
Северный синап	12,5	13,2	10,5	10,0	14,5	15,0	14,8	14,9
Спартан	10,1	10,5	11,5	10,7	9,0	9,0	10,5	10,3
Уэлси	10,2	10,2	10,2	9,8	11,5	10,0	12,3	11,0

Учитывая особенности метода вычеркивания столбцов, содержащих утраченные элементы, матрица для проведения дисперсионного анализа в этом случае принимает вид, представленный в табл. 3.

Мы проводили расчеты с помощью программы ASTRA (разработанной доцентом кафедры информатики МичГАУ, канд. с.-х. наук С.В. Фроловой), созданной специально для обработки экспериментальных данных различными математическими методами, в том числе и методом дисперсионного анализа, и получили следующие значения наименьшей существенной разницы для сравнения опытных вариантов: $HCP_{05} = 1,5$ мм, а для исходной матрицы без выпадения 1,26.

В публикациях разных авторов, посвященных рассматриваемой проблеме [1, 2, 5, 6], рассматривались стационарные опыты с одним или двумя выпадениями, и рекомендации, естественно, давались только для соответствующих случаев.

Хочется обратить внимание исследователей на то, что два выпадения для матрицы размером 3×5 при применении данного метода, то есть при удалении двух столбцов из исходной матрицы для проведения дисперсионного анализа, составит около 13 % выпавших данных от числа исходных, в матрице размером 3×10 – 6,6 %, а размером 3×100 – всего лишь 0,7 %.

Очень важным, на наш взгляд, представляется определить процент ошибки вычисляемого значения HCP_{05} при возможной утрате экспериментальных данных, если известен процент их выпадения, чтобы исследователи могли иметь представление о достоверности проводимых сравнений. В рассматриваемом примере ошибка определения HCP_{05} составила 16 %.

При группировке экспериментальных данных в матрицы для проведения соответствующих исследований мы использовали новую компьютерную программу GLORIA (разработанную доцентом кафедры информатики МичГАУ, канд. с.-х. наук С.В. Фроловой), которая из генеральных совокупностей различных изучаемых показателей формирует случайным образом повторности по каждому опытному варианту, которые далее обрабатываются методом дисперсионного анализа.

Таким образом, с применением названной программы, HCP_{05} для сравнения вариантов опыта в рассматриваемом примере было рассчитано как без выпадения данных (то есть по полной матрице размером 3×100 исходных значений – далее будем называть ее исходной), так и с учетом их выпадения (случайным образом – также при помощи программы) на уровне до 10, 20 и 30 % исходных данных.

В литературе рекомендуют применение рассматриваемого метода, как уже было отмечено, лишь в случаях с 1-2 выпавшими датами. Тем не менее, часто процент утраченных в исследованиях данных бывает значительно большим. В этой связи представляется интересным рассмотреть возможность применения данного метода именно в таких случаях.

Например, если предположить, что выпасть могло 3 значения (выпавшими в этом случае окажется 10 % данных – табл. 4), а после удаления соответствующих столбцов – уже 30 % и матрица опыта примет вид, представленный в табл. 5.

Проведение дисперсионного анализа в этом случае выглядит следующим образом (табл. 6–9).

$$C = \frac{(\sum x_1)^2}{an} = 0,03;$$

$$\sum \varrho_0 = (\sum x_1)^2 - C = 63,51;$$

Таблица 4

Диаметр штабиков саженцев яблони, мм
(исходная матрица с тремя выпавшими значениями)

Сорт (вариант)	Повторности									
	1	2	3	4	5	6	7	8	9	10
Северный синап	12,5	11,0	13,2	10,5	13,3	10,0	14,5	15,0	14,8	14,9
Спартан	10,1	10,9	10,5	11,5	9,0	10,7	9,0	9,0	10,5	10,3
Уэлси	10,2	9,5	10,2	10,2	10,0	9,8	11,5	10,0	12,3	11,0

Таблица 5

Диаметр штабиков саженцев яблони, мм
(матрица с вычеркнутыми столбцами)

Сорт (вариант)	Повторности						
	1	2	3	4	5	6	7
Северный синап	12,5	11,0	13,2	10,5	15,0	14,8	14,9
Спартан	10,1	10,9	10,5	11,5	9,0	10,5	10,3
Уэлси	10,2	9,5	10,2	10,2	10,0	12,3	11,0

Таблица 6

Сводная таблица для проведения дисперсионного анализа
(исходная матрица)

Вариант	Повторности							$\sum A$	\bar{X}
Северный синап	12,5	11,0	13,2	10,5	15,0	14,8	14,9	91,9	13,1
Спартан	10,1	10,9	10,5	11,5	9,0	10,5	10,3	72,8	10,4
Уэлси	10,2	9,5	10,2	10,2	10,0	12,3	11,0	73,4	10,5
$\sum \Pi$	32,8	31,4	33,9	32,2	34,0	37,6	36,2	238,1	$\bar{X}_0 = 11,3$

Таблица 7

Таблица отклонений

Вариант	Повторности							$\sum A$
Северный синап	1,2	-0,3	1,9	-0,8	3,7	3,5	3,6	12,8
Спартан	-1,2	-0,4	-0,8	0,2	-2,3	-0,8	-1,0	-6,3
Уэлси	-1,1	-1,8	-1,1	-1,1	-1,3	1,0	-0,3	-5,7
$\sum \Pi$	-1,1	-2,5	0,0	-1,7	0,1	3,7	2,3	$\sum x_1 = -0,8$

Таблица 8

Таблица квадратов отклонений

Вариант	Повторности							$(\sum A)^2$
Северный синап	1,44	0,09	3,61	0,64	13,69	12,25	12,96	163,84
Спартан	1,44	0,16	0,64	0,04	5,29	0,64	1,00	39,69
Уэлси	1,21	3,24	1,21	1,21	1,69	1,00	0,09	32,49
$\sum \Pi$	1,21	6,25	0	2,89	0,01	13,69	5,29	$\sum (\sum A)^2 = 236,02$ $(\sum x_1)^2 = 0,64$

Таблица 9

Таблица дисперсионного анализа

Варьирование	Сумма квадратов $\sum Q$	Число степеней свободы df	Дисперсия S^2	F_{Φ}	F_{05}
Общее	63,5	20	–	–	–
Повторностей	4,2	6			
Вариантов	33,7	2	16,9	7,9	3,9
Случайных погрешностей	25,6	12	2,1		

$$\sum Q_{\Pi} = \sum (\sum \Pi)^2 / n - C = 4,16;$$

$$\sum Q_A = \sum (\sum A)^2 / a - C = 33,7;$$

$$\sum Q = \sum Q_0 - \sum Q_A - \sum Q_{\Pi} = 25,62.$$

$$HCP_{05} = t_{05} S_d; \quad S_d = \sqrt{\frac{2S^2}{n}} = \sqrt{\frac{2 \cdot 2,14}{7}} = 0,78; \quad HCP_{05} = 2,18 \cdot 0,78 = 1,7.$$

Таким образом, значение HCP_{05} в данном случае составляет 1,7 мм. Процент ошибки определения HCP_{05} при этом оказывается равным 35 %.

Аналогичные приведенным по показателю диаметра штампика саженцев яблони трех сортов в питомнике нами были сделаны расчеты для

таких показателей роста и развития саженцев, как площадь листовой пластинки и количество листьев.

Данные по полученным значениям $НСР_{05}$ для соответствующих по размеру матриц по каждому из этих показателей сведены в табл. 10, где представлены также и ошибки $НСР_{05}$ при выпадении определенного процента исходных экспериментальных данных. Эти сведения помогут исследователям ориентироваться, насколько возможна сравнительная оценка опытных вариантов в случаях частичной утраты экспериментальных (учетных) данных в размере 10–30 %. Следует отметить, что уже при 20 % выпавших дат, фактически происходит потеря (за счет вычеркивания столбцов при применении рассматриваемого метода) более 50 % данных в целом по опыту и достоверная сравнительная оценка его результатов становится невозможной.

Для того чтобы подход к количеству анализируемых выборок, формируемых из исходных матриц после частичной утраты опытных данных был объективным, что вызвано необходимостью достоверной оценки применения рассматриваемого метода вычеркивания столбцов в исследованиях, мы рассчитывали его с использованием формулы [1]:

$$n_{\text{lim}} = \frac{t^2 \left[(1 \pm k) \sigma^2 \right]}{\Delta^2}, \quad (1)$$

где n_{lim} – необходимое количество повторностей в доверительном интервале; t – критерий Стьюдента (2,00 для $P = 0,05$); σ^2 , σ – дисперсия и среднее квадратическое отклонение (по пробным выборкам конкретных показателей); k – коэффициент (по В.С. Пугачеву, 1979); Δ – допустимая погрешность (определяется исследователем в соответствии со смысловым значением изучаемого показателя, обычно до 10 % от средних арифметических величин).

Как известно из работ разных авторов [5, 6], обработка выборок различных объемов (малых – до 30 единиц, и больших – более 30 единиц) требует различных методических подходов, и полученные в результате обработки выборок малых и больших объемов из единой исходной генеральной совокупности статистические показатели могут в итоге значительно различаться. В этой связи значительный интерес для исследователей представляет возможный процент ошибки при уменьшении объемов изучаемых экспериментальных выборок из-за потери части опытных данных.

В этой связи в наших исследованиях были рассмотрены выборки малого объема – на примере растений малины, а также малого и большого объема – на примере растений малины и саженцев яблони в плодовом питомнике по показателям роста, указанным в табл. 10, предоставленные Л.И. Никоноровой и Н.Е. Маковой по результатам своих научных исследований.

Обработка этих данных показала, что процент ошибки вычисления $НСР_{05}$ при утрате даже 10 % опытных данных значительно колебался в зависимости от размера исследуемых выборок, возрастая с уменьшением их объема (см. табл. 10).

Таблица 10

**Значения НСР₀₅ и процент ошибки их вычисления
при 10%-ной потере данных в опыте**

Показатели	НСР ₀₅			НСР ₀₅		
	Общая по опыту	При 10 % выпавших данных	Ошибка вычисления НСР ₀₅ , %	Общая по опыту	При 10 % выпавших данных	Ошибка вычисления НСР ₀₅ , %
<i>Саженьцы яблони</i>						
НСР ₀₅	Для матриц (130×3)			Для матриц (100×3)		
Диаметр штамбика	0,67	0,63	7,3	0,7	0,7	10,1
НСР ₀₅	Для матриц (10×3)			Для матриц (30×3)		
Количество листьев	22,8	22,1	23,6	17,5	12,7	26,5
НСР ₀₅	Для матриц (10×3)			Для матриц (30×3)		
Площадь листьев	8,2	9,8	18,6	6,03	5,3	13,9
<i>Двухлетние побеги малины</i>						
НСР ₀₅	Для матриц (10×3)			Для матриц (30×3)		
Количество листьев	5,5	5,7	23,4	3,9	3,6	8,2
НСР ₀₅	Для матриц (10×3)			Для матриц (30×3)		
Высота побегов	22,5	24,5	17,1	16,1	14,4	15,4
НСР ₀₅	Для матриц (10×3)			Для матриц (30×3)		
Диаметр штамбика	0,23	0,21	6,3	0,13	0,11	1,17

Таким образом, при использовании рассматриваемого в настоящей работе метода применения дисперсионного анализа, из-за его особенностей, исследователям не следует безоговорочно отказываться от работы с выборками малого объема (в случае потери по каким-либо причинам части опытных данных), так как это зависит от самих изучаемых показателей и их вариабельности, и они дают важную научную информацию, но, тем не менее, следует помнить о возможных ошибках при сравнении опытных вариантов и, при возможности, отдавать предпочтение все-таки работе с выборками возможно больших объемов, ограничивая их разумными пределами, исходя, например, из формулы (1), дающей интервальные значения необходимых повторностей в исследованиях для достоверной оценки опытных данных.

Изменчивость самого изучаемого показателя можно оценить количественно по величине коэффициента вариации, что представляет интерес для дальнейшего изучения. В наших исследованиях коэффициент вариации составил в среднем 35 % по показателю диаметра штамбика саженцев яблони, 26 % по количеству листьев на саженце, 20 % по площади листовой пластинки.

Настоящая работа является продолжением исследований в направлении особенностей применения дисперсионного анализа в исследованиях с

биологическими объектами, в том числе с плодовыми растениями, проводимыми коллективом авторов МичГАУ в течение последнего десятилетия.

Работа выполнена при финансовой поддержке РФФИ, грант № 06-07-96317 р_центр_а.

Список литературы

1. Доспехов, Б.А. Методика полевого опыта / Б.А. Доспехов. – М. : Колос, 1965. – 422 с.
2. Молостов, А.С. Методика полевого опыта / А.С. Молостов. – М. : Колос, 1966. – 180 с.
3. Никонорова, Л.И. К вопросу обработки экспериментальных данных в плодоводстве при их частичной утрате / Л.И. Никонорова, С.В. Фролова, Л.В. Бобрович // Актуальные проблемы развития АПК : сб. тр. всерос. конф. – Саратов, 2006. – С. 27–30.
4. Никонорова, Л.И. Использование неортогональных схем дисперсионного анализа в стационарном опыте в исследованиях по плодоводству в зависимости от необходимого числа повторностей / Л.И. Никонорова, Л.В. Бобрович, С.В. Фролова // Вопросы современной науки и практики. Университет им. В.И. Вернадского. – 2006. – № 3(5). – С. 27–32.
5. Потапов, В.А. Методы обработки экспериментальных данных в плодоводстве / В.А. Потапов, В.И. Кашин, А.Г. Курсаков. – М. : Колос, 1997. – 144 с.
6. Снедекор, Дж.У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии / Дж.У. Снедекор // пер. в англ. В.Н. Перегудова. – М. : Изд-во с.-х. литературы, журналов и плакатов, 1961. – 561 с.
7. Фролова, С.В. К вопросу обработки опытных данных при их частичной утрате в стационарном опыте / С.В. Фролова, Л.И. Никонорова, Л.В. Бобрович // Экологические проблемы отраслей народного хозяйства : сб. тр. междунар. конф. – Пенза, 2006. – С. 197–200.

Peculiarities of Dispersion Analysis Method Application for Processing Stationary Experimental Partly Lost Data

S.V. Frolova, L.I. Nikonorova, L.V. Bobrovich, N.E. Makova

Michurinsk State Agrarian University, Michurinsk

Key words and phrases: experimental data grouping; variation coefficient; disperse analysis matrix; data processing.

Abstract: The paper deals with the approach to processing experimental data on biological entities by the method of transferring non-orthogonal schemes into orthogonal ones through deleting the variants containing missing data.

© С.В. Фролова, Л.И. Никонорова,
Л.В. Бобрович, Н.Е. Макова, 2008