Волкова Светлана Николаевна
д-р с.-х. наук, профессор
Сивак Елена Евгеньевна
д-р с.-х. наук, профессор
ФГБОУ ВО «Курская государственная сельскохозяйственная академия им. И.И. Иванова»
г. Курск, Курская область

ПРИМЕНЕНИЕ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ К ПОДБОРУ ПРОИЗВОДСТВЕННЫХ ФУНКЦИЙ

Аннотация*: в статье представлены методы обработки экспериментальных данных, применяемые при заполнении лабораторных, расчетно-графических, контрольных работ, в научно-исследовательской работе, при курсовом и дипломном проектировании на различных факультетах, поскольку методы позволяют исследовать закономерности моделей независимо от их конкретной природы.*

Ключевые слова*: производственные функции, метод наименьших квадратов, формулы, экспериментальные данные, параметры, уравнения.*

За всю многовековую историю математика не была унитарной наукой. Математика – это язык, это ключ к познанию диалектического понимания природы. «Это большой город, чьи предместья не перестают разрастаться несколько хаотическим образом на окружающем его пространстве, в то время как центр периодически перестраивается, следуя каждый раз все более и более ясному плану и стремясь к все более и более величественному расположению, в то время как старые кварталы с их лабиринтами переулков сносятся для того, чтобы проложить к окраинам улицы все более прямые, все более широкие, все более удобные». В подтверждение тому теория вероятностей – математическая наука о количественных закономерностях моделей случайных явлений независимо от их конкретной природы и математическая статистика, занимающаяся разработкой методов сбора, регистрации и обработки результатов наблюдений с целью познания закономерностей случайных массовых явлений.

Методы обработки экспериментальных данных, изложенные в статье, следует применять при заполнении лабораторных, расчетно-графических, контрольных работах; в научно-исследовательской работе, при курсовом и дипломном проектировании на различных факультетах, поскольку методы позволяют исследовать закономерности моделей независимо от их конкретной природы [1, с. 35].

Подбор производственных функций и определение их параметров методом наименьших квадратов: Требуется составить уравнение динамики роста урожайности по ниже приведенным данным, характеризующим динамику роста урожайности зерновых культур (таблица 1), овощей (таблица 2), сеянных трав (таблица 3), кукурузы, плодов (таблица 4.) в учхозе СХА по годам в зависимости от удобрения (таблица 5).

Цели работы:

1) освоить метод наименьших квадратов;

2) ознакомиться с методикой решения подбора эмпирических формул на компьютере;

3) составить уравнение по указанным данным [2, с. 78].

При выполнении работы за X принимают одну из колонок таблицы 5, соответствующую порядковому номеру студента в журнале n и номер группы N. За Y принимают столбец, соответствующий порядковому номеру студента в журнале из таблиц 1–4.

Например, пусть студент из первой группы (N = 1) и его порядковый номер в журнале 9 (n = 9). Тогда он выбирает 1-ый столбец из таблицы 5. соответствующей n = 9 с наименованием с наименованием «Год» и столбец под номером 9 из таблицы 1 «Озимые» и получает таблицу 6.

Таблица 1

Урожайность зерновых (ц/га)

Озимая рожь	Озимая пшеница	Ячмень	Овес	Вика	Яровые зерновые	Зерновая	Горох	Озимые
1	2	3	4	5	6	7	8	9
17,9	21,2	11,9	31,2	13,9	13,3	17,6	11,4	21
30,1	27,4	13,5	30,6	25,4	19,4	22,3	8,2	27,7
34,5	27,2	30,9	37,5	25,5	30,0	28,8	22,2	27,9
33,9	27,7	32	26,3	24,4	29	29	18,7	28,7
28,3	25,7	32,9	25,2	18,5	28,1	27	12,6	26
28,6	31,6	38,8	30,2	21,5	33,2	32,2	14,6	31
39,1	43,3	41,1	26,6	18,6	38	40,4	–	42,6

Таблица 2

Урожайность овощей

Картофель	Овощи	Огурцы	Помидоры	Капуста	Кормовая свекла
10	11	12	13	14	15
90,6	90	106	57	–	340
108	495	185	15	719	428
155,4	245,8	190	130	422,7	326,1
157,2	456,7	114	9	756,7	636,7
117,5	387,1	50	27,5	660,1	570
223,8	160,6	121	141	252,2	605,6
179	120,6	45	239	130,7	399

Таблица 3

Урожайность сеянных трав

Однолет. травы	з/корма	Многолетн. травы	з/корма	Естествен. сенокосы	Однол. травы	з/корма
16	17	18	19	20	21	22
41,1	68,1	30,1	203,7	31,5	40	180
49,7	145	53,1	317	34,6	41	185
43,7	83,4	50,9	640,8	31	41,1	190
44,2	145,5	46	331,8	17	42	195
46,4	104,7	54,3	306	17,2	43	191,7
35,8	149,3	51,8	205,8	32	45	181,7

59,4	157,6	58,2	440,2	31,3	47	182

Таблица 4

Урожайность плодов и кукурузы

Плоды	Кукуруза 1	Кукуруза 2
23	24	25
82,1	167,7	290
58,9	115	300
36,3	438,7	310
67,4	392,9	301
60,2	242	305
58,2	430	295
24,4	473	315

Таблица 5

Урожайность по годам (кг/га)

Год	Удобрения n =1,2,3,4,5,6,7,8,9				Год	Удобрения n =23			
	N	P	K	Всего		N	P	K	Всего
1	2	3	4	5	1	2	3	4	5
2010	98	41	53	173	1981	120	7	38	166
2011	74	53	114	242	1982	148	36	48	234
2012	75	48	63	186	1983	98	45	112	255
2013	128	85	157	371	1984	165	52	168	355
2014	155	81	114	350	1985	107	37	50	195
2015	150	113	115	378	1986	118	90	45	253
2016	231	101	143	475	1987	166	111	127	404
n =10,11,12,13,14					n =24,25				
1	225	175	178	578	1	100	54	63	218
2	280	106	213	600	2	165	56	92	313
3	233	120	190	543	3	168	65	187	421
4	213	159	234	606	4	304	35	177	517
5	139	107	150	396	5	195	70	105	370
6	135	203	203	541	6	410	318	296	1024
7	120	124	168	412	7	248	162	92	502
n = 15									
1	153	123	133	410					
2	280	106	213	600					
3	233	120	190	543					
4	230	85	223	562					
5	353	276	380	1010					
6	204	70	70	344					
7	370	206	170	746					
n =16, 17, 18, 19, 20, 21, 22									

1	92	23	6	122
2	105	29	32	166
3	104	25	9	138
4	143	80	128	351
5	144	42	57	244
6	228	28	36	292
7	155	100	71	326

Таблица 6

Экспериментальные данные

X	2010	2011	2012	2013	2014	2015	2016
Y	21	27,7	27,9	28,7	26	31	42,6

Можно упростить и вместо 2010 ввести – 1, 2011 – 2 и т. д. Таким образом, заменив год на цифры 1, 2, 3, 4, 5, 6, 7, получим таблицу 7.

Таблица 7

Экспериментальные данные

X	1	2	3	4	5	6	7
Y	21	27,7	27,9	28,7	26	31	42,6

Зависимость У от Х находят в виде $y = ax + в$ (линейная), или $y = ax^2 + вx + c$ (квадратичная), или другой, используя метод наименьших квадратов при определении параметров а, в, с.

Рассмотрим теперь более сложный случай [3, с.10]. Пусть предполагается, что зависимость признака Y от признака X имеет вид $y = f(x, \alpha_1, \alpha_2, \ldots \alpha_m)$, где x – значения признака X, y – значения признака Y; $\alpha_1, \alpha_2, \ldots, \alpha_m$ подлежащие определению параметры, и что в результате эксперимента были получены следующие эмпирические данные таблицы 8

Таблица 8

Эмпирические данные

Значение признака X	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
Значение признака Y	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n

Метод наименьших квадратов (МНК) утверждает, что наивероятнейшие значения параметров $\alpha_1, \alpha_2, \ldots, \alpha_m$ дает минимум функции

$$S = \sum_{n=1}^{n} (y_i - f(x_i, \alpha_1, \alpha_2, \ldots \alpha_m))^2 \qquad (1)$$

Если $f(x, \alpha_1, \alpha_2, \ldots \alpha_m)$ имеет непрерывные частные производные по всем своим параметрам, то необходимым условием минимума функции представляет систему m уравнений с m – неизвестными:

$$\frac{\partial S}{\partial \alpha_1} = 0; \frac{\partial S}{\partial \alpha_1} = 0; \ldots \frac{\partial S}{\partial \alpha_m} = 0 \qquad (2)$$

Нахождение функциональной зависимости между признаками Y и X на основании (1) называют выравниванием эмпирических данных вдоль кривой $y = f(x, \alpha_1, \alpha_2, \ldots, \alpha_m)$.

Если $f(x, \alpha_1, \alpha_2, \ldots, \alpha_m) = \alpha_1 x + \alpha_2$, то этой кривой будет прямая линия $y = \alpha_1 x + \alpha_2$. В этом случае система (2) может быть преобразована в так называемую нормальную систему МНК при выравнивании по прямой:

$$\begin{aligned} &\alpha_1 \sum_{i=1}^{n} x_i^2 + \alpha_2 \sum_{1}^{n} x_i = \sum_{1}^{n} x_i y_i \\ &\alpha_1 \sum_{i=1}^{n} x_i + \alpha_2 n = \sum_{1}^{n} y_i \end{aligned} \qquad (3)$$

Система (2) при выравнивании по параболе $y = \alpha_1 x^2 + \alpha_2 x + \alpha_3$ может быть преобразована к виду:

$$\begin{cases} \alpha_1 \sum_{i=1}^{n} x_i^4 + \alpha_2 \sum_{1}^{n} x_i^3 + \alpha_3 \sum_{1}^{n} x_i^2 = \sum_{1}^{n} x_i^2 y_i \\ \alpha_1 \sum_{i=1}^{n} x_i^3 + \alpha_2 \sum_{1}^{n} x_i^2 + \alpha_3 \sum_{1}^{n} x_i = \sum_{1}^{n} x_i y_i \\ \alpha_1 \sum_{1}^{n} x_i^2 + \alpha_2 \sum_{1}^{n} x_i + \alpha_3 n = \sum_{1}^{n} y_i \end{cases}$$

Пусть в результате эксперимента получена таблица 8.

Таблица 8

Вспомогательная таблица

x_i	23	29	25	80	42	28	100
y_i	40	41	41,1	42	43	45	47

Здесь y_i – урожайность однолетних трав в учхозе СХА; x_i – количество внесенных удобрений.

Построим на плоскости точки с соответствующими координатами (23; 40); (29; 41) …; (100; 47) рисунок 1. Плавно соединим точки [4, с.79]. Точки с координатами (80; 42) и (28; 450 удалены от предполагаемой линии. Поэтому при подборе зависимости $y = ax + b$, эти значения исключим из рассмотрения. По оставшимся пяти значениям подберем формулу и найдем параметры a и b методом наименьших квадратов.

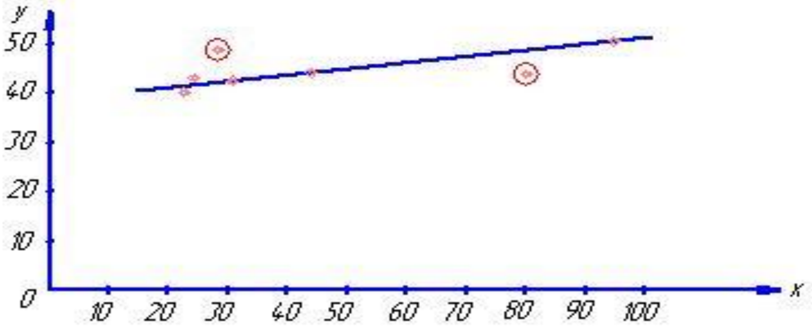

Рис. 1. Графическое изображение эмпирических данных

Так как выравниваются эмпирические данные вдоль прямой, то пользуемся системой (3) [5, с. 32]. Результаты вычислений удобно вести с помощью таблицы 9.

Таблица 9

Схема расчета параметров a и b МНК

N	x_i	y_i	$x_i y_i$	x_i^2	$y_i = ax_i + b$	$\varepsilon_i = Y_i - y_i$	ε_i^2
1	23	40	920	529	40,6598	0,6598	0,43534
2	25	41,1	1027,5	625	40,892	-0,274	0,07576
3	29	41	1189	841	41,1674	0,1679	0,02819
4	42	43	1806	1764	42,2672	-0,7328	0,536996
5	100	47	4700	10000	47,714	0,174	0,030276
$\sum$	219	212,1	9642,5	13759	–	–	1,1058743

В систему (3) для n = 5, имеющую вид:

$$\begin{cases} a\sum_{1}^{5} x_i^2 + b\sum_{1}^{5} x_i = \sum_{1}^{5} x_i y_i \\ a\sum_{1}^{5} x_i + 5b = \sum_{1}^{5} y_i \end{cases} \tag{4}$$

подставляем данные из табл. 3.6., получаем:

$$\begin{cases} 13759a + 219b = 9642{,}5 \\ 219a + 5b = 212{,}1 \end{cases} \tag{5}$$

Решаем систему (5), находим a=0,084602; b=38,714428.

Уравнение динамики роста урожайности (производственная функция) имеет вид: y = 0,084602 x + 38,714428 (6).

Оценим погрешность полученного результата и заполним соответствующие столбцы в таблице 9. Для этого по формуле (31) для каждого x_i находим y_i. Затем вычисляем $\varepsilon_i = Y_i - y_i$, и находим ε_i^2. Таким образом, параметры a и b в уравнении (6), рассчитанные МНК, получены с точностью 1,1.

Рассмотрим пример криволинейной корреляции. Пусть в результате эксперимента получена таблица 10.

Таблица 10

Урожайность (ц/га) озимой пшеницы по годам

x	2010	2011	2012	2013	2014	2015	2016
y	31,9	27,4	27,2	27,7	25,7	31,6	43,3

Построим на координатной плоскости точки с координатами (x_i, y_i). В дальнейшем года заменим на цифры 1, 2, 3, 4, 5, 6, 7, что не повлияет на результат исследования.

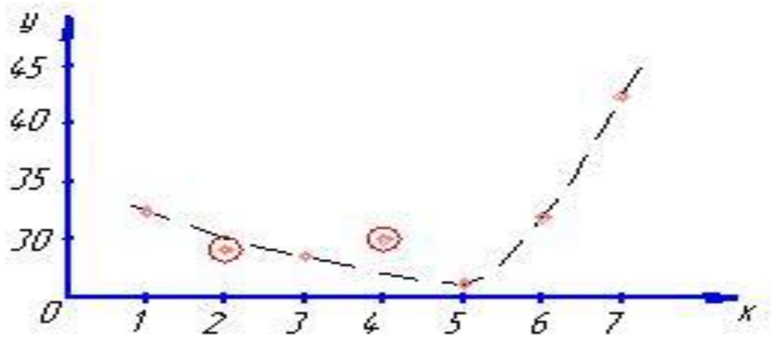

Рис. 2. Эмпирические данные и линия регрессии

Плавно соединив эмпирические данные линией, будем выравнивать по параболе y = ax + вx + c, исключив из рассмотрения точки, соответствующие 2011 и 2014 гг. Параметры a, в, и c определим МНК по формуле (3), имеющий вид для n = 5:

$$\begin{cases} a\sum_{1}^{5} x_i^4 + b\sum_{1}^{5} x_i^3 + c\sum_{1}^{5} x_i^2 = \sum_{1}^{5} x_i^2 y_i \\ a\sum_{1}^{5} x_i^3 + b\sum_{1}^{3} x_i^2 + c\sum_{1}^{5} x_i = \sum_{1}^{5} x_i y_i \\ a\sum_{1}^{5} x_i^2 + b\sum_{1}^{5} x_i + 5c = \sum_{1}^{5} y_i \end{cases} \qquad (7)$$

Расчеты проведем в таблице 11.

Таблица 11

Схема расчетов параметров а, в и с МНК

п/п	x_i	y_i	x_i^4	x_i^3	x_i^2	$x_i^2 y_i$	$x_i y_i$	$y_i = ax_i^2 + bx_i + c$	b_i	ε_i^2
1	2	3	4	5	6	7	8	9	10	11
1	1	319	1	1	1	31,9	31,9	33,2	1,3	1,69
2	2	272	16	8	4	108,3	54,4	25,97	-123	1,51
3	3	257	81	27	9	231,3	77,1	25,24	-0, 46	0,21
4	4	316	256	64	16	505,6	126,4	31,31	-0,29	0,08
5	5	433	625	125	25	1082,5	216,5	44,08	0,78	0,6
Σ	15	159,7	977	225	55	1960,1	506,3	159,8	–	4,09

Подставляя в формулу данные из таблицы, получим:

$$\begin{cases} 977a + 225b + 55c = 1960{,}1 \\ 225a + 55b + 15c = 506{,}3 \\ 55a + 15b + 5c = 159{,}7 \end{cases} \qquad (8)$$

Решая, находим а = 3,35; в = −17,38; с = 47,23.

Динамика роста урожайности озимой пшеницы по годам описывается производственной функцией

$$y = 3{,}35x^2 - 17{,}38x + 47{,}23 \qquad (9)$$

Оценим погрешность полученного результата. Для этого найдем для каждого x_i значения y_i по формуле (9), затем вычислим разность между теоретическим значением Y_i и эмпирическим y_i, т.е. $\varepsilon_i = Y_i - y_i$ и найдем ε_i^2, заполнив соответствующие столбцы 9, 10, 11 в табл. 3.3.4. В результате проведенного исследования делаем вывод, что параметры а, в и с в формуле определены МНК) [6, с. 12].

Следует отметить, что рассмотренные в примерах математические модели-упрощенные, о чем свидетельствует и полученная погрешность. Но даже на этих моделях можно сделать прогноз на ближайшую перспективу, варьируя переменные величины в рамках модели.

Выполнять данную работу следует с применением вычислительной техники разного уровня от микрокалькуляторов до ППЭВМ с привлечением к расчетам пакета прикладных программ, т.е. от частичного автоматизирования до полной автоматизации процесса.

На практике, работая со студентами дневного отделения, есть студенты, которые по заданному алгоритму сами составляют программы и автоматизируют всю лабораторную работу. Но интерпретацию результатов все равно им приходилось делать самостоятельно, получив модель компьютерного варианта. В случае разницы просчитанных работ с помощью калькулятора и с помощью ППЭВМ при защите работ ребятам надо было это объяснить результаты. Это активизировало их к более глубокому изучению самих цифр, точности их расчетов, заложенных в программный продукт.

Считаем, что такие работы полезны и для школьников в их научных исследованиях, а не только в вузе. Поскольку, нет ничего увлекательнее, выравнивать экспериментальные данные к любой линии, в том числе и кусочно-заданной на некоторых промежутках. А получив аналитическую модель, обучающийся превращается в руководителя того процесса, который изучает. Поскольку может получать не только интересующие его значения на построенном отрезке, но и знать изменения результата при изменении фактора, а следовательно строить краткосрочный прогноз при стабильной ситуации развития изучаемого процесса) [7, с. 21].

Возможен коллективный проект, когда весь класс или группа принимает участие и делает общий вывод по результатам исследования, а именно какую культуру для выращивания можно рекомендовать в рассмотренных условиях, для какой необходимо дополнительное исследование по объему выборки, а какую следует повременить с рекомендацией к производству.

Работая с цифрой и ее точностью, в случае обработки эксперимента, прививается навык исследовательской деятельности и творческого поиска в решении поставленной задачи.

Список литературы

1. Волкова С.Н. К вопросу оценки качества прогнозов моделирования экосистем / С.Н. Волкова, Т.И. Романова, М.И. Пашкова [и др.] // Вестник Курской государственной сельскохозяйственной академии. – 2017. – №3. – С. 38–44.

2. Волкова С.Н. Нелинейные взаимодействия и их моделирование в социально-экологических системах / С.Н. Волкова, Е.Е. Сивак, М.И. Пашкова [и др.] // Вестник Курской государственной сельскохозяйственной академии. – 2016. – №3. – С. 77–80.

3. Сивак Е.Е. Новые нетрадиционные культуры — перспектива развития сельского хозяйства // Аграрная наука. – 2006. – №7. – С. 9–10.

4. Волкова С.Н. Последствия антропогенного воздействия в развитии сельского хозяйства / С.Н. Волкова, Ю.И. Майоров, Е.Е. Сивак [и др.] // Вестник Курской государственной сельскохозяйственной академии. – 2012. – №2. – С. 78–80.

5. Шлеенко А.В. Прогнозирование рисков, разрушающих естественные экосистемы / А.В. Шлеенко, С.Н. Волкова, Е.Е. Сивак [и др.] // Известия Юго-Западного государственного университета. – 2014. – №1 (52). – С. 30–34.

6. Волкова С.Н. Разработка технологии биоэнергетики / С.Н. Волкова, Е.Е. Сивак, В.В. Морозова [и др.] // Известия Юго-Западного государственного университета. Серия: Техника и технологии. – 2017. – Т. 2. – С. 11–14.

7. Волкова С.Н. Улучшение структуры землепользования / / С.Н. Волкова, Е.Е. Сивак, В.В. Морозова [и др.] //Вестник Курской государственной сельскохозяйственной академии. – 2017. – №1. – С. 20–24.