Л. П. Сметанина, Ю. М. Сметанин,
О. В. Максимова

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Ижевск 2020

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Удмуртский государственный университет»
Институт математики, информационных технологий и физики
Кафедра математического анализа

Л. П. Сметанина, Ю. М. Сметанин, О. В. Максимова

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Методические рекомендации

Ижевск
2020

УДК 519.6(075.8)
ББК 22.193я7

М545

Рекомендовано к изданию Учебно-методическим советом УдГУ

Рецензент к.ф.-м.н., доцент ФГБОУ ВО «ИжГТУ им. М. Т. Калашникова» Т. С. Быкова

Авторы-составители: Л. П. Сметанина, Ю. М. Сметанин, О. В. Максимова

С502 Метод наименьших квадратов: метод. рекомендации / авторы-составители Л. П. Сметанина, Ю. М. Сметанин, О. В. Максимова. Ижевск: Изд. центр «Удмуртский университет», 2020. — 38 с.

ISBN 978-5-4312-0853-9

В методическом пособии приведены краткие теоретические сведения, изложена методика выполнения работ по аппроксимации функций различными способами и приведены примеры. Представлены алгоритмы решения задач средствами MicroSoft Excel.

Данное методическое пособие предназначено для студентов уровня бакалавриата всех направлений и всех форм обучения Института нефти и газа им. М.С.Гуцериева и Института экономики и управления в Удмуртском государственном университете.

ISBN 978-5-4312-0853-9

УДК 519.6(075.8)
ББК 22.193я7

Содержание

Введение

Данное методическое пособие подготовлено преподавателями кафедры математического анализа и отражает многолетний опыт при проведении занятий и организации самостоятельной работы студентов Института нефти и газа им. М.С.Гуцириева и Института экономики и управления в Удмуртском государственном университете.

Пособие предназначено для методического обеспечения раздела «Функции нескольких переменных» курса «Высшая математика».

В пособии содержатся краткие теоретические сведения, изложена методика выполнения работ по аппроксимации функций различными способами и приведены примеры и индивидуальные задания. Представлены алгоритмы решения задач средствами MicroSoft Excel. Индивидуальные задания могут быть использованы для организации работы на практических занятиях, а также как индивидуальные домашние задания.

Компетенции обучающегося, формируемые в результате освоения раздела:

— универсальные компетенции

УК-1 способность осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач;

— общепрофессиональные компетенции

ОПК-1 способность решать задачи, относящиеся к профессиональной деятельности, применяя методы моделирования, математического анализа, естественнонаучные и общеинженерные знания.

Пособие предназначено обеспечить студентов наглядным вспомогательным материалом и индивидуальными заданиями при изучении курса «Высшая математика» и может быть использовано для студентов других нематематических направлений УдГУ.

1. Математическое обоснование

На практике часто встает вопрос о сглаживании экспериментальной зависимости.

Пусть имеются числовые данные $(x_1; y_1)$, $(x_2; y_2)$, $\ldots$, $(x_n; y_n)$ и имеются основания предполагать, что y зависит от x.

Нужно подобрать функцию $y = f(x)$, график которой проходит как можно ближе к точкам $(x_1; y_1)$, $(x_2; y_2)$, $\ldots$, $(x_n; y_n)$. Такую функцию называют аппроксимирующей (аппроксимация — приближение), или теоретической функцией.

Зависимость между переменными x и y будем показывать в виде таблицы данных, полученных опытным путем.

x	x_1	$\ldots$	x_i	$\ldots$	x_n
y	y_1	$\ldots$	y_i	$\ldots$	y_n

Требуется наилучшим образом сгладить экспериментальную зависимость между x и y и представить ее в виде эмпирической формулы $y = f(x)$.

Одним из методов нахождения таких функций является *метод наименьших квадратов.*

Построение эмпирической формулы. Иногда вид формулы известен из физических соображений. Если нет — экспериментальные точки наносятся на график и примерно «угадывается» общий вид зависимости путем сравнения полученной кривой с графиками известных функций.

Если вид функции $y = f(x)$ обозначен, то переходят к определению параметров этой функции. Согласно методу наименьших квадратов в качестве параметров выбирают такие, чтобы сумма квадратов невязок δ_i (или отклонений) значений $f(x_i)$, найденных по эмпирической формуле $y = f(x)$, от опытных значений y_i была минимальна (рис. 1).

Полученную аппроксимирующую функцию также называют *уравнением регрессии.*

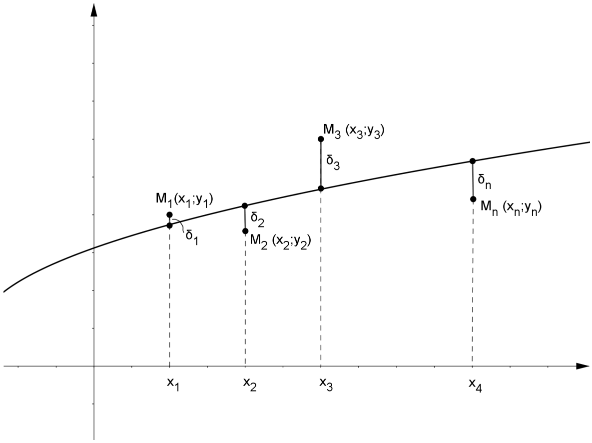


Рис.1 Эмпирическая функция

График функции называется *линией тренда* (в случае, например, линейной функции — линией линейного тренда). При этом, в общем случае, тренд — это не обязательно прямая линия.

Функции

1) $y = ax + b$,

2) $y = ax^2 + bx + c$,

3) $y = ax^b$,

4) $y = ae^{bx}$,

5) $y = \ln x + b$,

6) $y = \dfrac{1}{ax + b}$,

7) $y = \dfrac{a}{x} + b$,

8) $y = \dfrac{x}{ax + b}$,

входят в перечень наиболее часто используемых. Рассмотрим особенности применения некоторых из них.

Степенная функция: $y = ax^b$**.** Предполагаем, что в исходной таблице x и y положительны и $a > 0$. Прологарифмировав равенство, получим $\ln y = \ln a + b \ln x$. Введем новую переменную $u = \ln x$ и положим $v = \ln y$, $A = b$, $B = \ln a$.

Получим $v = Au + B$, то есть задача сводится к отысканию приближающей функции в виде линейной.

Таким образом, для нахождения искомой приближающей функции в виде степенной следует по заданной таблице данных составить новую таблицу, прологарифмировав значения x и y в исходной таблице. Найти параметры A и B по данным в преобразованной таблице.

Если выбрана **функция вида** $y = \dfrac{1}{ax+b}$, то определив $v = \dfrac{1}{y}$, получим $v = ax + b$. Составляем вместо исходной новую таблицу, в которой значения аргумента оставляем прежними, а значения функции заменяем обратными числами.

В случае **логарифмической аппроксимации** $y = a\ln x + b$ метод наименьших квадратов применяется к линейной функции $y = au + b$, где $u = \ln x$.

Линейная аппроксимация.

Пусть в качестве функции $y = f(x)$ берется линейная функция: $y = ax + b$.

Тогда $S(a;\, b) = \sum\limits_{i=1}^{n}(ax_i + b - y_i)^2 \longrightarrow min.$

Применим необходимое условие экстремума к функции двух перменных $S(a,\, b)$:

$$\begin{cases} S'_a = 0, \\ S'_b = 0. \end{cases}$$

$$\begin{cases} \sum\limits_{i=1}^{n} 2(ax_i + b - y_i)x_i = 0, \\ \sum\limits_{i=1}^{n} 2(ax_i + b - y_i) = 0. \end{cases}$$

После алгебраических преобразований получаем систему, которая носит название **нормальной**.

Запишем нормальную систему

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ a\sum\limits_{i=1}^{n} x_i + bn = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Составим определитель системы

$$\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & n \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2.$$

Определитель не равен 0, значит система имеет единственное решение.

Докажем, что рассмотренный определитель положителен при $n \geqslant 2$. Используем метод математической индукции.

Пусть $n = 2$, тогда при $x_1 \neq x_2$

$$\begin{vmatrix} x_1^2 + x_2^2 & x_1 + x_2 \\ x_1 + x_2 & 2 \end{vmatrix} = 2x_1^2 + 2x_2^2 - x_1^2 - 2x_1x_2 - x_2^2 =$$
$$= x_1^2 - 2x_1x_2 + x_2^2 = (x_1 - x_2)^2 > 0.$$

Предположим, что $k\sum\limits_{i=1}^{k} x_i^2 - \left(\sum\limits_{i=1}^{k} x_i\right)^2 > 0.$

Покажем, что неравенство справедливо для следующего k, то есть выполнено:

$$(k+1)\sum_{i=1}^{k+1} x_i^2 - \left(\sum_{i=1}^{k+1} x_i\right)^2 > 0.$$

Действительно,

$$(k+1)\sum_{i=1}^{k+1} x_i^2 - \left(\sum_{i=1}^{k+1} x_i\right)^2 =$$

$$k\sum_{i=1}^{k} x_i^2 + kx_{k+1}^2 + \sum_{i=1}^{k} x_i^2 + x_{k+1}^2 - \left(\sum_{i=1}^{k} x_i + x_{k+1}\right)^2 =$$

$$k\sum_{i=1}^{k} x_i^2 + kx_{k+1}^2 + \sum_{i=1}^{k} x_i^2 + \underline{x_{k+1}^2} - \left(\sum_{i=1}^{k} x_i\right)^2 - 2x_{k+1}\sum_{i=1}^{k} x_i - \underline{x_{k+1}^2} =$$

$$\underbrace{k\sum_{i=1}^{k} x_i - \left(\sum_{i=1}^{k} x_i\right)^2}_{>0} + kx_{k+1}^2 + \sum_{i=1}^{k}(x_i^2 - 2x_{k+1} + x_{k+1}^2 - x_{k+1}^2) =$$

$$k\sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i\right)^2 + \sum_{i=1}^{k}(x_i - x_{k+1})^2 + \underline{kx_{k+1}^2} - \underline{kx_{k+1}^2} > 0.$$

Таким образом,

$$S''_{aa} = 2\sum_{i=1}^{n} x_i^2, \qquad S''_{bb} = 2n, \qquad S''_{ab} = 2\sum_{i=1}^{n} x_i.$$

$$AC - B^2 = \begin{vmatrix} 2\sum\limits_{i=1}^{n} x_i^2 & 2\sum\limits_{i=1}^{n} x_i \\ 2\sum\limits_{i=1}^{n} x_i & 2n \end{vmatrix} = 4n\sum_{i=1}^{n} x_i^2 - 4\left(\sum_{i=1}^{n} x_i\right)^2 =$$

$$= 4\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) > 0, \qquad A > 0.$$

значит найденные a, b доставляют минимум функции $S(a;\, b)$.

Пример 1. Количество вещества, оставшегося через x минут от начала химической реакции, y задается таблицей

x_i	2	4	6	8	10	12	14
y_i	65,4	44,7	38,0	35,3	32,8	31,2	30,4

Считая зависимость между x и y линейной ($y = ax + b$), определить параметры a и b.

Вспомогательная таблица

	x_i	y_i	$x_i y_i$	x_i^2
	2	65,4	130,8	4
	4	44,7	178,8	16
	6	38,0	228,0	36
	8	35,3	282,4	64
	10	32,8	328,0	100
	12	31,2	374,4	144
	14	30,4	425,6	196
$\sum$	56	277,8	1948	560

Составим нормальную систему

$$\begin{cases} 560a + 56b = 1948, \\ 56a + 7b = 277,8. \end{cases}$$

Решая эту систему, получаем

$$a = -2,45, \qquad b = 59,2857, \qquad y = -2,45x + 59,2857.$$

Для решения систем линейных уравнений полезно вспомнить формулы Крамера.

Найдем сумму квадратов невязок в данном примере.

	x_i	$y(x_i)$	y_i	$\delta_i^2 = (y(x_i) - y_i)^2$
	2	54,3857	65,4	121,3148
	4	49,4857	44,7	22,9029
	6	44,5857	38,0	43,3714
	8	39,6857	35,3	19,2344
	10	34,7857	32,8	3,943
	12	29,8857	31,2	1,7274
	14	24,9857	30,4	29,3146
$\sum$				241,8085

Данные, представленные в таблице, показаны в виде графика на рис. 2.

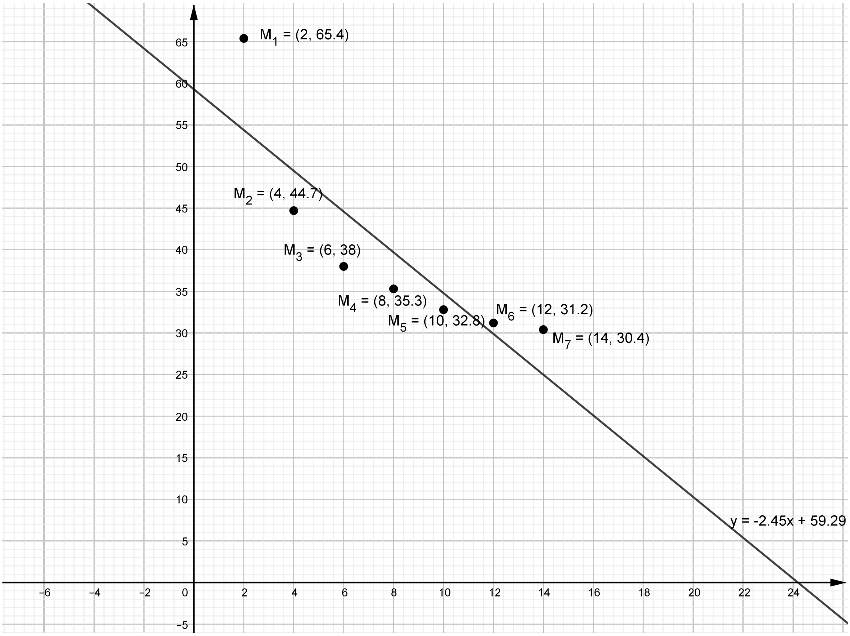


Рис.2 Линейная зависимость

Квадратичная аппроксимация.

Если рассмотреть квадратичную зависимость $y = ax^2 + bx + c$, то

$$S(a,\, b,\, c) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (ax_i^2 + bx_i + c - y_i)^2 \to \min\,.$$

Нормальная система в этом случае имеет вид

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^4 + b\sum\limits_{i=1}^{n} x_i^3 + c\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i^2 y_i, \\ a\sum\limits_{i=1}^{n} x_i^3 + b\sum\limits_{i=1}^{n} x_i^2 + c\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i + cn = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Данную систему трех линейных алгебраических уравнений можно решить используя, например, формулы Крамера.

$$a = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 y_i & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} y_i & \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}}, \quad b = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^2 y_i & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} y_i & n \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}},$$

$$c = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 y_i \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} y_i \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}}.$$

Пример 2. Имеются следующие данные о расходах на рекламу x (усл. ед.) и сбыте продукции y (усл. ед.)

x_i	1	2	3	4	5
y_i	1,6	4,0	7,4	12,0	18,0

Найти значения параметров a, b, c, считая, что между переменными существует квадратичная зависимость.

Вспомогательная таблица

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1	1,6	1	1	1	1,6	1,6
2	4,0	4	8	16	8,0	16, 0
3	7,4	9	27	81	22,2	66,6
4	12,0	16	64	256	48,0	192
5	18,0	25	125	625	90,0	450
$\sum$ 15	43,0	55	225	979	169,8	726,2

Нормальная система принимает вид

$$\begin{cases} 979a + 225b + 55c = 726,2, \\ 225a + 55b + 15c = 169,8, \\ 55a + 15b + 5c = 43. \end{cases}$$

Решение системы $a = 0,6,\ b = 0,48,\ c = 0,56.$

Искомая зависимость $y = 0,6x^2 + 0,48x + 0,56.$

Найдем сумму квадратов невязок в данном примере.

x_i	$y(x_i)$	y_i	$\delta_i^2 = (y(x_i) - y_i)^2$
1	1,64	1,6	0,0016
2	3,92	4,0	0,0064
3	7,4	7,4	0,0
4	12,08	12,0	0,0064
5	18,0	18,0	0,0
$\sum$			0,0144

Данные, представленные в таблице, показаны в виде графика на рис. 3.

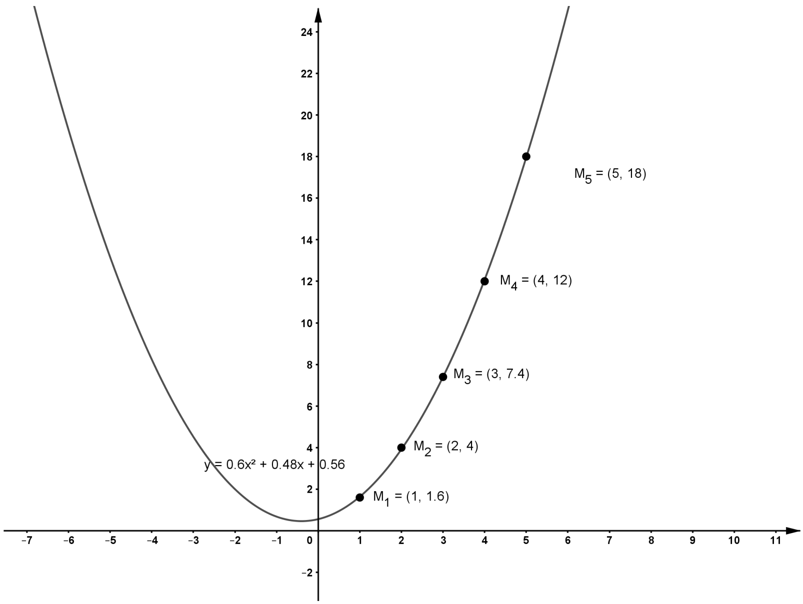


Рис.3 Квадратичная зависимость

Пример 3. Имеется четыре измерения пары переменных $(x; y)$, результаты которых приведены в таблице

x_i	1	2	3	4
y_i	0,2	0,3	1,0	1,2

Методом наименьших квадратов построить линейную зависимость $y = ax + b$. Сравнить полученную зависимость с зависимостью $y = \dfrac{1}{8}x^2$.

Аналогично **примеру 1** найдем уравнение линейной зависимости: $y = 0,37x - 0,25$.

Сравним величины $S = \sum\limits_{i=1}^{n}(f(x_i) - y_i)^2$ для найденной линейной зависимости и зависимости $y = \dfrac{1}{8}x^2$. Промежуточные вычисления представим в таблице:

x_i	y_i	$\frac{1}{8}x_i^2$	$0,37x_i - 0,25$	$S_{кв} = \left(\frac{1}{8}x_i^2 - y_i\right)^2$	$S_{лин} = (0,37x_i - 0,25 - y_i)^2$
1	0,2	0,125	0,12	0,005625	0,0064
2	0,3	0,5	0,49	0,040000	0,0361
3	1,0	1,125	0,76	0,015625	0,0576
4	1,2	2	1,23	0,640000	0,0009
$\sum$				0,701250	0,1010

Как видно $S_{лин} < S_{кв}$, следовательно, линейная зависимость предпочтительнее.

Пример 4.

Методом наименьших квадратов определить параметры экспоненциальной эмпирической функции $y = ae^{bx}$ по данным, представленным в таблице

x	1	2	3	4	5
y	2,1	6,2	13,3	24,6	117,5

Для применения метода наименьших квадратов экспоненциальная функция линеаризуется:

$$v = Ax + B, \qquad \text{где } v = \ln y, \qquad A = b, \qquad B = \ln a.$$

Вспомогательная таблица

x_i	y_i	$v_i = \ln y_i$	x_i^2	$x_i v_i$
1	2,1	0,7	1	0,7
2	6,2	1,8	4	3,6
3	13,3	2,6	9	7,8
4	24,6	3,2	16	12,8
5	117,5	4,8	25	23,8
$\sum$ 15	163,7	13,1	55	48,8

Полученные значения сумм подставляем в систему линейных уравнений для получения параметров A и B

$$\begin{cases} 55A + 15B = 48,8, \\ 15A + 5B = 13,1. \end{cases}$$

Решая систему, находим $A = 0,94$, $B = -0,2$.

Линеаризованная эмпирическая функция $v = 0,94x - 0,2$, $a = e^{-0,2} = 0,82$, $b = 0,94$.

Искомая экспоненциальная эмпирическая функция $y = 0,82e^{0,94x}$.

Если для заданной эмпирической таблицы использовались несколько эмпирических функций, то сначала для каждого класса методом наименьших квадратов определяется наилучшая функция, а затем из полученных функций выбирается та, которая имеет наименьшее среднеквадратическое отклонение — сумму квадратов невязок.

2. Аппроксимация функции средствами Excel

В Excel для построения аппроксимирующих функций или регрессий имеются две возможности:

1. Добавление выбранных регрессий (линий тренда — trendlines) на диаграмму, построенную на основе таблицы экспериментальных данных исследуемого процесса.

2. Использование встроенных статистических функций Excel, позволяющих получать регрессии (линии тренда) непосредственно на основе таблицы исходных данных.

Добавление линий тренда на диаграмму

Для таблицы данных, описывающих некоторый процесс и представленных диаграммой, в Excel имеется эффективный инструмент регрессионного анализа, позволяющий:

а) строить на основе метода наименьших квадратов и добавлять в диаграмму пять типов регрессий, которые с той или иной степенью точности моделируют исследуемый процесс — линейный, полиномиальный, логарифмический, степенной, экспоненциальный;

б) добавлять к диаграмме уравнение построенной регрессии;

в) определять степень соответствия выбранной регрессии отображаемым на диаграмме данным.

Линейная регрессия используется при моделировании характеристик, значения которых возрастают или убывают с постоянной скоростью. Это наиболее простая в построении модель исследуемого процесса. Она строится в соответствии с уравнением

$$y = ax + b,$$

где a — тангенс угла наклона линейной регрессии к оси абсцисс; b — координата точки пересечения линейной регрессии с осью ординат.

Полиномиальная регрессия полезна для описания характеристик, имеющих несколько ярко выраженных экстремумов (максимумов и минимумов). Выбор степени полинома определяется количеством экстремумов исследуемой характеристики.

Так, полином второй степени может хорошо описать процесс, имеющий только один максимум или минимум;

полином третьей степени — не более двух экстремумов;

полином четвертой степени — не более трех экстремумов и т.д.

В этом случае линия тренда строится в соответствии с уравнением

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6,$$

где коэффициенты $a_0, a_1, a_2, \ldots, a_6$ — константы, значения которых определяются в ходе построения.

Логарифмическая линия тренда применяется при моделировании характеристик, значения которых вначале быстро меняются, а затем постепенно стабилизируются. Строится в соответствии с уравнением

$$y = a \ln x + b,$$

где коэффициенты a, b — константы.

Степенная линия тренда дает хорошие результаты, если значения исследуемой зависимости характеризуются постоянным изменением скорости роста.

Строится в соответствии с уравнением:

$$y = ax^b,$$

где коэффициент a и показатель степени b — константы.

При этом, если среди данных встречаются нулевые или отрицательные значения, использовать степенную линию тренда нельзя.

Примером такой зависимости может служить график равноускоренного движения автомобиля.

Экспоненциальная регрессия используется в том случае, если скорость изменения данных непрерывно возрастает.

Строится в соответствииии с уравнением:

$$y = ae^{bx},$$

где коэффициенты a, b — константы.

Для данных, содержащих нулевые или отрицательные значения, этот вид приближения также неприменим.

При подборе линии тренда Excel автоматически расчитывает значения величины R^2, которая называется *коэффициентом детерминации* или *аппроксимации*. Этот коэффициент показывает долю тех изменений, которые в данном явлении зависят от изучаемого фактора и может принимать значения от 0 до 1. Таким образом, данный коэффициент характеризует достоверность аппроксимации.

Эта величина определяется по формуле:

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n} (y_i - y(x_i))^2}{\sum\limits_{i=1}^{n} (y(x_i))^2 - \dfrac{1}{n}\left(\sum\limits_{i=1}^{n} y(x_i)\right)^2},$$

где $y(x_i)$ — значения аппроксимирующей функции.

Чем ближе R^2 к единице, тем надежнее линия тренда аппроксимирует исследуемый процесс.

Величина R^2 показывает, насколько значение зависимой переменной определяется значениями независимых переменных. Речь идет о статистической значимости модели. Модель является очень хорошей, если R^2 превышает 0,8, и при этом сама модель имеет практическое обоснование. В случае, если все не настолько идеально, но все же выше 0,5, модель можно использовать.

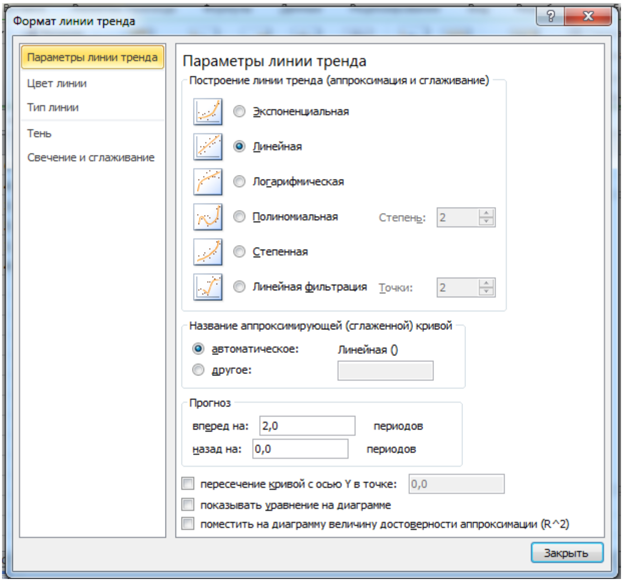
При необходимости значение R^2 всегда можно отобразить на диаграмме.

Добавить линию тренда (график аппроксимирующей функции) к ряду данных можно двумя способами:

1) активизировать построенную на основе ряда данных диаграмму, в меню «Работа с диаграммами» выбрать вкладку «Макет», далее пункт меню «Анализ» и команда «Линия тренда» команду «Добавить линию тренда» (или «Макет» $\Longrightarrow$ «Линия тренда» для Excel 2007);

2) в контекстном меню для ряда данных (правой кнопкой мыши щелкнуть по ряду данных) выбрать команду «Добавить линию тренда».

После данных действий на экране появится диалоговое окно «Формат линии тренда».



В данном диалоговом окне можно выбрать:

а) необходимый тип линии тренда (по умолчанию выбирается тип «Линейная»). Для типа «Полиномиальная» в поле «Степень» следует задать степень выбранного полинома;

б) изменить название линии тренда;

в) вывести в область диаграммы уравнение линии тренда (должен быть активизирован соответствующий пункт в диалоговом окне);

г) вывести в область диаграммы около линии тренда значение достоверности аппроксимации R^2 (должен быть активизирован соответствующий пункт в диалоговом окне).

Плюсы данного инструмента регрессионного анализа:

а) простота построения на диаграмме линии тренда;

б) наличие в перечне наиболее часто используемых типов регрессии;

в) возможность прогнозирования поведения исследуемого процесса на некоторое число шагов вперед и назад;

г) получение уравнения линии тренда (аппроксимирующей функции) в аналитическом виде без дополнительных громоздких вычислений;

д) при необходимости, возможность получения оценки достоверности проведенной аппроксимации.

Линиями тренда можно дополнить ряд данных, представленный на следующих типах диаграмм: график, гистограмма, плоские ненормированные диаграммы с областями, линейчатые, точечные, пузырьковые и биржевые. Нельзя линию тренда построить для объемной, нормированной, лепестковой, круговой и кольцевой диаграмм.

Рассмотрим работу инструментов Excel на конкретном примере из первой части.

Пример 1. Проверить результаты проведенных в первой части вычислений для следующих условий: y — количество вещества, оставшегося через x минут от начала химической реакции, задается таблицей

x_i	2	4	6	8	10	12	14
y_i	65,4	44,7	38,0	35,3	32,8	31,2	30,4

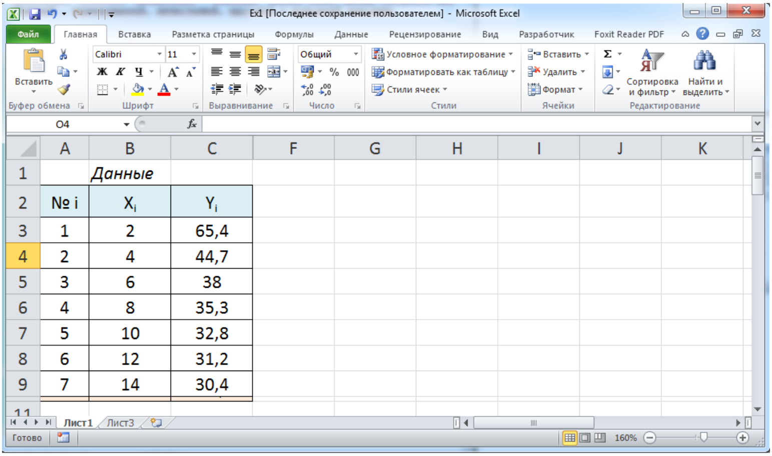
а) Считая зависимость между x и y линейной $y = ax + b$, определить параметры a и b.

б) Считая зависимость между x и y квадратичной $y = ax^2 + bc + c$, определить параметры a, b и c.

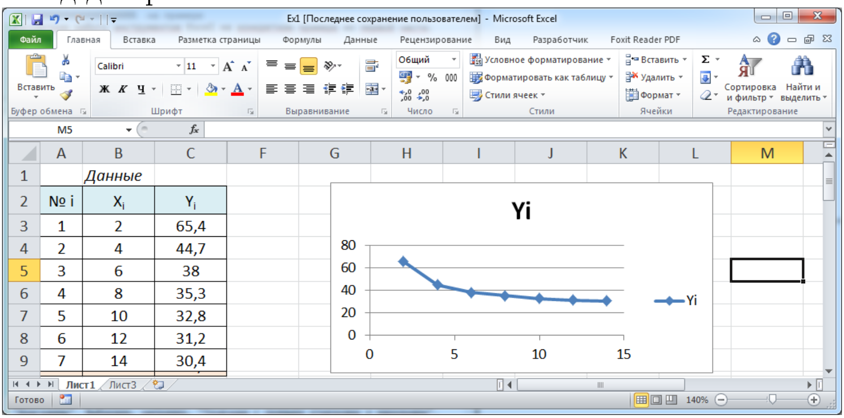
Решение. а)

Получение аппроксимирующей функции

Шаг 1. Готовим таблицу данных в документе на рабочем листе Excel.



Шаг 2. Строим диаграмму: выделив ячеки B2:C9, используем меню «Вставка» пункт «Диаграммы». Выбираем, например, «Точечная с прямыми отрезками и маркерами». При необходимости, корректируем вид диаграммы.



Шаг 3. Добавляем линию тренда (график аппроксимирующей функции) и уравнение регрессии (аппроксимирующую функцию). Также можно вывести величину достоверности аппроксимации R^2 (рис. 4).

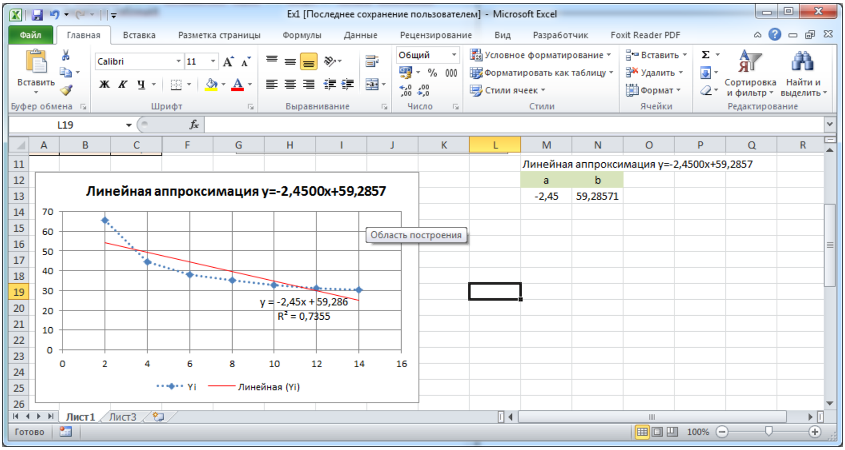


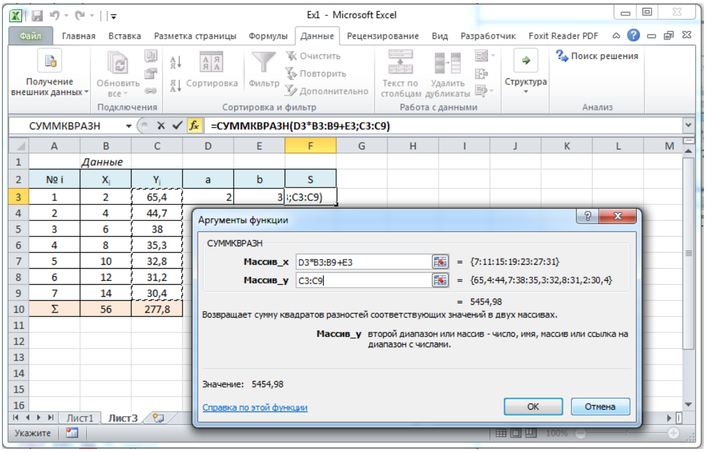
Рис. 4 Линейная аппроксимация

Получение коэффициентов аппроксимирующей функции по методу наименьших квадратов. А именно, минимизирование меры отклонений (суммы квадратов невязок) средствами Excel (Поиск решений)

Шаг 1. Готовим таблицу данных в документе на рабочем листе Excel.

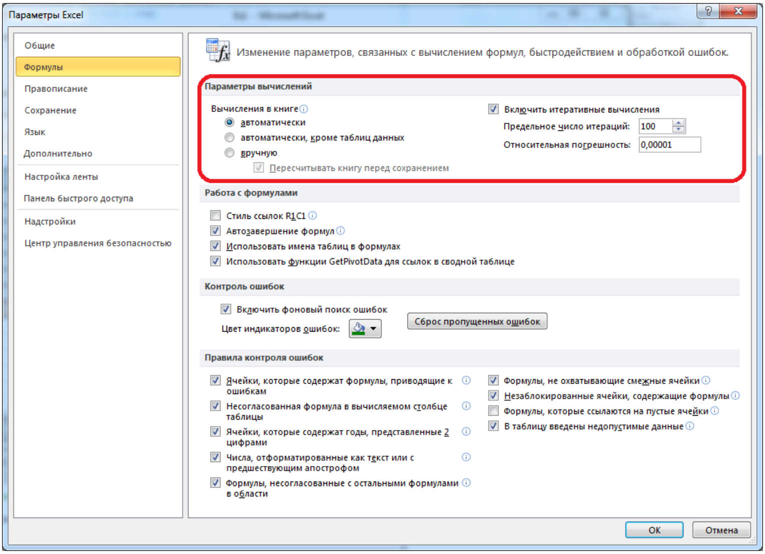
Шаг 2. Для определения коэффициентов линейной зависимости $y = ax + b$ по методу наименьших квадратов $\sum_{i=1}^{n}(ax_i + b - y_i)^2$ введем в ячейки D3 и E3 произвольные приближенные значения a и b, например 2 и 3.

В ячейку F3 введем формулу для вычисления меры отклонений $\sum_{i=1}^{n}(y(x_i) - y_i)^2$. Для этого воспользуемся функцией СУММКВРАЗН.

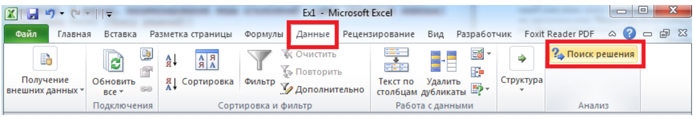


В нашем примере ячейки B2:B9 — это массив для вычисления значений аппроксимирующей функции, а C2:C9 — массив табличных (экспериментальных) значений y.

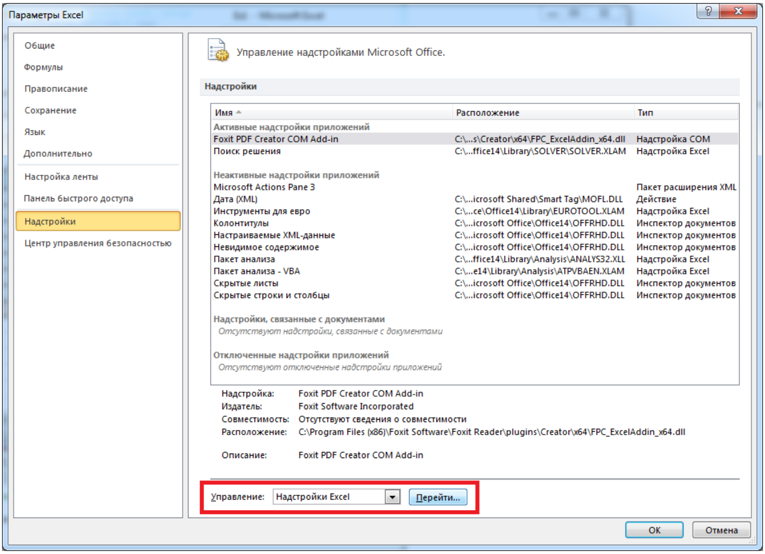
Шаг 3. Теперь минимизируем полученное значение S в ячейке F3 с помощью некоторого количества итераций. Для этого необходимо определить погрешность вычислений и предельное число итераций. Например, определим эти значения, равными 0,00001 и 100 соответственно. Воспользуемся в меню «Файл» пунктом «Параметры». В диалоговом окне «Параметры Excel» выберем «Формулы» и пункт «Параметры вычислений» (в случае Excel 2007 или Excel 2010).



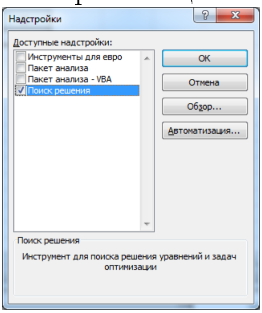
Кроме того, понадобится инструмент «Поиск решений», позволяющий выполнить поиск минимального требуемого значения. Проверить наличие этого инструмента среди активных можно на вкладке «Данные» основного меню.



В случае отсутсвия, в диалоговом окне «Параметры Excel» выбираем пункт «Надстройки». Внизу появившегося окна будет вкладка «Управление», где стоит параметр «Надстройки Excel», кнопка — «Перейти».



Выбираем опцию «Поиск решения».



Далее уточняем коэффициенты эмпирической формулы с помощью «Поиск решений». В диалоговом окне «Параметры поиска решений» для

1) «Оптимизировать целевую функцию» устанавливаем ячейку F3;

2) для пункта «До» выбираем «Минимум»;

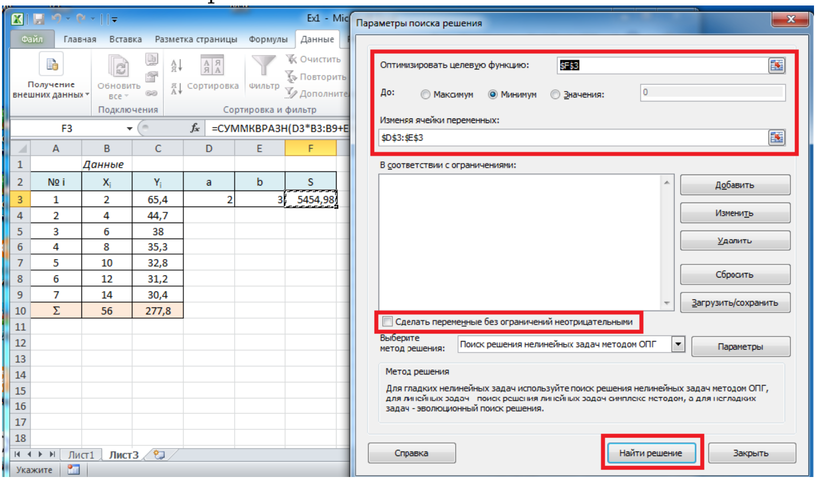
3) «Изменяя ячейки переменных», вводим ссылки на ячейки с будущими ответами. В нашем случае это D3:E3;

4) в пункте «Сделать переменные без ограничений неотрицатель-

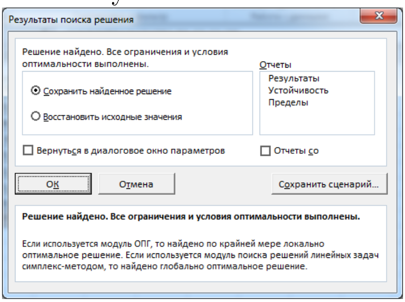
ными» необходимо снять флажок, если он будет поставлен.

Замечание: вводить ссылки на ячейки удобнее не с клавиатуры, а щелчком по соответствующей ячейке. При этом ссылки автоматически будут превращаться в абсолютные. В случае ввода ссылок на ячейки с клавиатуры — следим, чтобы в имени ячеек буквы были латинского алфавита.

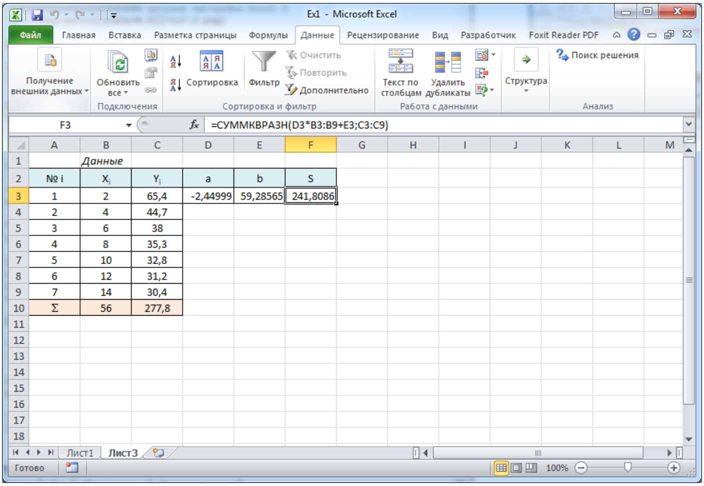
Жмем «Найти решение».



Появляется еще одно диалоговое окно «Результаты поиска решений», в котором выбираем «Сохранить найденное решение» и нажимаем кнопку «Ок».



Результаты минимизации представлены на следующем рисунке

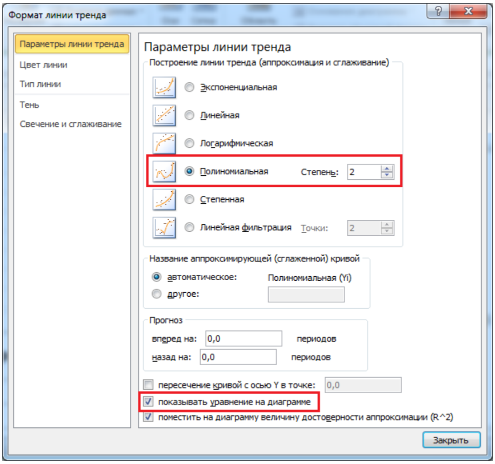


Их можно сравнить с коэффициентами уравнения линии тренда (см. рис. 4 Линейная аппроксимация).

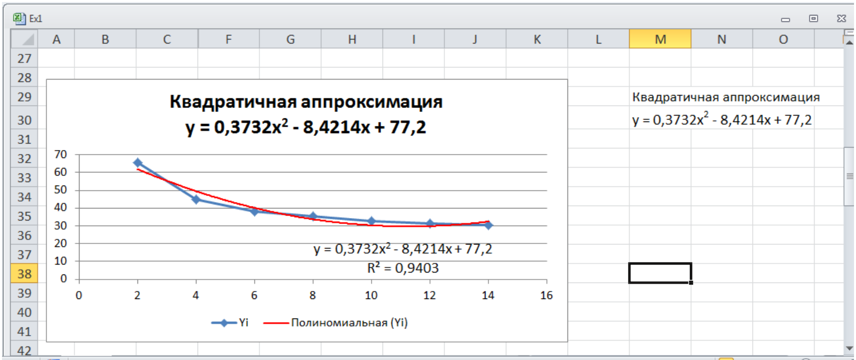
Решение: б)

Получение аппроксимирующей полиномиальной степени 2 функции (квадратичной функции)

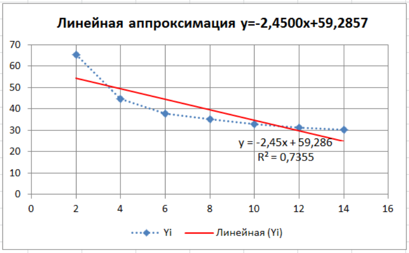
По таблице данных примера 1 добавляем квадратичную линию тренда (график аппроксимирующей функции) и уравнение регрессии (квадратичную аппроксимирующую функцию).



Корректируем параметры диаграммы: название, легенду, расположение и размер объектов.



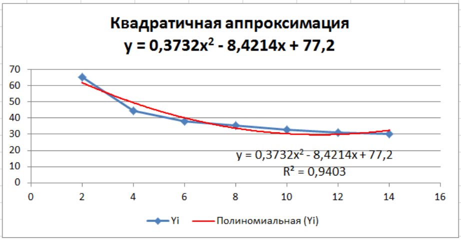
Сравним значение суммы квадратов отклонений (невязок) в случае линейной и квадратичной аппроксимации



Линейная аппроксимация y=-2,4500x+59,2857

a	b
-2,45	59,2857

X_i	Y_i	$Y(X_i)$	$(Y(Xi)-Yi)^2$
2	65,4	54,3857	121,3145
4	44,7	49,4857	22,90306
6	38	44,5857	43,37163
8	35,3	39,6857	19,23449
10	32,8	34,7857	3,943061
12	31,2	29,8857	1,727347
14	30,4	24,9857	29,31449
сумма квадратов отклонений (невязок			241,8086



Квадратичная аппроксимация

$y = 0{,}3732x^2 - 8{,}4214x + 77{,}2$

X_i	Y_i	$Y(X_i)$	$(Y(Xi)-Yi)^2$
2	65,4	61,85	12,6025
4	44,7	49,4856	22,902
6	38	40,1068	4,43861
8	35,3	33,7136	2,51666
10	32,8	30,306	6,22004
12	31,2	29,884	1,73186
14	30,4	32,4476	4,19267
сумма квадратов отклонений (невязок)			54,6043

В случае квадратичной функции эта величина меньше, значит такое приближение является более точным.

Задания для самостоятельного решения.

1) Выполнить задание, проведя необходимые вычисления.

2) Проверить вычисления и полученную зависимость средставми Excel.

3) Рассмотреть остальные варианты зависимостей в Excel и указать наиболее точно аппроксимирующую зависимость. Объяснить свой выбор.

1. В таблице приведены данные численности занятого населеня x (млн) и валового выпуска продукции y (у.е.)

x_i	80	82	83	85	86	88	89	90	91
y_i	32	34	35	36	37	38	40	39	40

Определить параметры a и b, считая зависимость линейной. Спрогнозировать выпуск продукции, если занятое население увеличиться на 10% по сравнению с последними данными.

2. Показатели стоимости основных производственных фондов x (млн. руб.) и среднесуточной производительности y приведены в таблице

x_i	2,6	2,8	2,9	3,4	4,6	5,2	6,1	7,7	10,6	14,0
y_i	19	18	20	23	26	31	37	45	53	68

Определить параметры a и b в случае линейной зависимости. Получить прогноз среднесуточной производительности при стоимости основных фондов 2 млн. руб.

3. Получены следующие данные о стоимости подключения потенциального абонента y (у.е.) в зависимости от плотности населения x(чел./ км^2) при возможном коэффициенте пропускания услуги (радиуса обслуживания базовой станции) $R = 3\,\text{км}$

x_i	10	20	30	40	50	60	70	80	90	95
y_i	1000	600	480	430	415	412	410	405	400	395

Определить параметры a, b и c, предполагая, что зависимость квадратиная $y = ax^2 + bx + c$.

4. Приведены данные о количестве пропусков занятий студентом в течение семестра x и результатах написания экзаменационной работы y (%)

x_i	1	3	5	6	8	10
y_i	85	75	70	60	50	40

Предполагая наличие линейной зависимости $y = ax + b$, найти параметры a и b. Получить прогноз результатов теста при отсутствии пропусков.

5. Приведены данные о высоте подброшенного над землей тела в зависимости от времени x (сек.), прошедшего с момента броска

x_i	0,1	0,2	0,4	0,5	0,6	0,8	0,9	1
y_i	10,2	10,37	10,6	10,76	10,8	11	11,1	11,2

Найти a, b и c, предполагая, что между x и y имеется квадратичная зависимость. Спрогнозировать высоту тела на второй секунде с момента броска.

6. В таблице приведены данные о потреблении электроэнергии y (кВт) предприятием в зависимости от времени x (час)

x_i	0,5	1	2	3	4	5	6	6,5	7	7,5
y_i	1000	1001	1004	1010	1020	1030	1050	1060	1070	1080

Найти параметры a, b и c, предположив квадратичную зависимость. Спрогнозировать потребеление электроэнергии при $x = 8$.

7. В таблице приведены данные расходов на рекламу x (тыс.у.е.) и сбыта продукции y (тыс.у.е.)

x_i	1	2	4	6	8
y_i	1,5	2,3	4,1	5,3	9,5

В предположении, что между x и y существует квадратиная зависимость, определить a, b и c. Спрогнозировать сбыт продукции при отсутствиии рекламы.

8. Имеются следующие данные о переменных x — цена товара (у.е) и y — уровень продаж (тыс.ед.). Предполагая, что между x и y существует линейная зависимость, найти эмпирическую формулу $y = ax + b$ методом наименьших квадратов.

x_i	3,0	4,0	5,0	6,0	7,0
y_i	200	160	120	90	80

9. Имеются следующие данные о переменных x — уровень потребления электроэнергии на предприятии (млн кВт $\cdot$ ч) и y — себестоимость единицы продукции. Предполагая, что между x и y существует линейная зависимость, найти эмпирическую формулу $y = ax + b$ методом наименьших квадратов.

x_i	1,2	1,3	1,4	1,5	1,6
y_i	20,0	18,8	18,2	18,1	18,0

10. Имеются следующие данные о переменных x — мощность двигателя (л.с.) и y — средний срок его эксплуатации (мес.). Предполагая, что между x и y существует линейная зависимость, найти эмпирическую формулу $y = ax + b$ методом наименьших квадратов.

x_i	30	40	50	60	70
y_i	18	20	21	24	25

11. Имеются данные о выплате пособий (тыс. руб.) за полгода.

месяцы x_i	Янв 1	Фев 2	Мар 3	Апр 4	Май 5	Июнь 6
размер пособий (тыс. руб.) y_i	1,8	1,3	1,5	1,6	1,7	1,7

Методом наименьших квадратов найти линейную эмпирическую формулу функциональной зависимости между x и y.

12. Температура x ($^\circ C$) смазочного масла в двигателе связана с температурой y ($^\circ C$) смазочного масла в коробке передач

x_i	33,7	34,5	37,0	38,0	39,0	39,3	35,5
y_i	22,1	24,0	24,4	26,0	25,8	22,4	25,8

Подобрать формулу эмпирической зависимости между x и y. Найти параметры этой формулы.

13. Прогиб балки y (м) связан с нагрузкой x (кН)

x_i	100	105	110	120	125	130
y_i	18	24	21	27	27	30

Подобрать эмпирическую формулу функциональной зависимости между x и y. Найти параметры этой формулы.

14. Электрическое сопротивление R (Ом) проволоки зависит от температуры t ($^\circ C$)

t	20,0	24,8	30,2	35,0	40,1	44,9	50,0
R	86,7	88,03	90,32	91,15	93,26	94,9	96,33

Подобрать эмпирическую формулу. Найти параметры этой формулы.

15. Известны x — процентное отношение предела текучести к прочности стали, y — процентное содержание углерода в стали

x_i	0,5	0,6	0,7	0,8	0,9
y_i	0,6	0,8	0,7	0,5	0,5

Подобрать эмпирическую формулу зависимости между x и y. Найти параметры этой формулы.

16. По экспериментальным данным построить методом наименьших квадратов линейную эмпирическую зависимость $y = ax + b$. Сравнить полученную зависимость с альтернативной и определить, какая из них лучше соответствует экспериментальным данным.

x_i	2	2,5	3	3,5	4
y_i	4,2	5,5	6,9	8	9,5

Альтернативная зависиомсть $y = 2x + 0,1x^2$.

17. По экспериментальным данным построить методом наименьших квадратов линейную эмпирическую зависимость $y = ax + b$. Сравнить полученную зависимость с альтернативной и определить, какая из них лучше соответствует экспериментальным данным.

x_i	1	2	3	4	5
y_i	1,0	1,4	1,7	2,0	2,2

Альтернативная зависиомсть $y = \sqrt{x}$.

18. По экспериментальным данным построить методом наименьших квадратов линейную эмпирическую зависимость $y = ax + b$. Сравнить полученную зависимость с альтернативной и определить, какая из них лучше соответствует экспериментальным данным.

x_i	1,0	1,5	2,0	2,5	3,0
y_i	0,50	0,30	0,25	0,18	0,12

Альтернативная зависиомсть $y = 2^{-x}$.

19. Имеются следующие экспериментальные данные о количестве единиц произведенной продукции x и издержек y (тыс.усл.ед.)

x_i	10	20	30	40	50
y_i	2,0	5,9	12,0	20,0	30,0

Функция издержек ищется в виде $y = ax + bx^2$. Определить параметры a и b методом наименьших квадратов.

20. Имеются следующие экспериментальные данные о количестве произведенного товара x (тыс.ед.) и количестве реализованного товара y (тыс.ед.)

x_i	100	120	140	160	180	200
y_i	100	114	130	146	163	180

Зависимость ищется в виде $y = 100 + a(x - 100) - b\sqrt{x - 100}$. Найти ее параметры a и b методом наименьших квадратов.

21. Имеются следующие экспериментальные данные о цене за единицу товара p (усл.ед.) и доле реализованного товара $w = \dfrac{K(x)}{x}$, где $K(x)$ — количестве реализованного товара, x — количество произведенного товара

p_i	10	12	15	16	20
w_i	1,95	0,93	0,92	0,90	0,89

Функция $w(p)$ ищется в виде $w = 1 - ap - bp^2$. Найти ее параметры a и b методом наименьших квадратов.

22. Используя метод наименьших квадратов, аппроксимировать функцию, заданную таблично

x_i	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
y_i	23,88	9,12	6,52	5,34	5,02	4,5	5,86	5,67	5,74	4,28

с помощью функции $f(x) = ae^{\frac{b}{x}}$. Найти сумму квадратов отклонений.

23. Используя метод наименьших квадратов аппроксимировать функцию, заданную таблично

x_i	0,1	0,2	0,5
y_i	10,22	5,14	2,76

с помощью функции $f(x) = \dfrac{ab}{x} + b$. Найти сумму квадратов отклонений.

24. Урожайность y (ц/га) зависит от количества внесенных удобрений x (т/га). Составлена таблица наблюдений над контрольными участками

x_i	6	7	8	9	10	11
y_i	28	30	31	32	32	33,5

Найти эмпирическую зависимость, наиболее точно отображающую результаты эксперимента.

25. Экспериментально получены значения $y = f(x)$, которые представлены в таблице

x_i	0	1	2	3	4	5
y_i	5,2	5,7	5,3	4,9	3,6	1,8

Методом наименьших квадратов определить параметры a, b, c функций $y = ax^2 + bx + c$ и $y = \dfrac{a}{x^2} + \dfrac{b}{x} + c$. Изобразить графики аппроксимирующих функций.

26. Экспериментально получены значения $y = f(x)$, которые представлены в таблице

x_i	1	2	3	4	5
y_i	0,8	-0,1	-1,2	-1,3	-1,5

Методом наименьших квадратов определить параметры a, b, c функций $y = ax^2 + bx + c$ и $y = \dfrac{a}{x^2} + \dfrac{b}{x} + c$. Изобразить графики аппроксимирующих функций.

27. Урожайность y (ц/га) зависит от количества внесенных удобрений x (т/га). Составлена таблица наблюдений над контрольными участками

x_i	6	7	8	9	10	11
y_i	28	30	31	32	32	33,5

Найти эмпирическую зависимость, наиболее точно отображающую результаты эксперимента.

Список литературы

1. Малышева Т.А. Численные методы и компьютерное моделирование. Лабораторный практикум по аппроксимации функций: учебно-методическое пособие/ Малышева Т.А.— Санкт-Петербург: Университет ИТМО, 2016.— 33 с. — ISSN 2227-8397

2. Мастяева И.Н. Численные методы: учебное пособие/ Мастяева И.Н., Семенихина О.Н.— Москва: Евразийский открытый институт, Московский государственный университет экономики, статистики и информатики, 2003.— 241 с. — ISSN 2227-8397

3. Методы математической статистики: учебное пособие/ М.Ю. Васильчик [и др.].— Новосибирск: Новосибирский государственный технический университет, 2016.— 88 с. — ISBN 978-5-7782-2811-5

4. https://microexcel.ru/metod-naimenshih-kvadratov/

5. https://excel2.ru/gruppy-statey/statisticheskiy-analiz

6. https://excel2.ru/articles/prostaya-lineynaya-regressiya-v-ms-excel

Учебное издание

Авторы-составители:
Максимова Ольга Васильевна
Сметанин Юрий Михайлович
Сметанина Людмила Петровна

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Методические рекомендации

Авторская редакция

Подписано в печать ???. Формат $60 \times 84\frac{1}{16}$.
Печать офсетная. Усл. печ. л. ??.
Тираж ?? экз. Заказ №

Издательский центр «Удмуртский университет»
426034, г. Ижевск, ул. Университетская, д. 1, корп. 4, каб. 207.
Тел./факс: +7 (3412) 500-295 E-mail: editorial@udsu.ru