

ЛАБОРАТОРНАЯ РАБОТА №11

СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Цель работы

Изучение практического применения метода наименьших квадратов для построения аппроксимирующих кривых при статистической обработке результатов механических испытаний.

1 Теоретическая часть

Метод наименьших квадратов (МНК, [англ. Ordinary Least Squares, OLS](#)) — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. Он может использоваться для «решения» переопределенных систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределенных) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции. МНК является одним из базовых методов [регрессионного анализа](#) для оценки неизвестных параметров регрессионных моделей по выборочным данным.

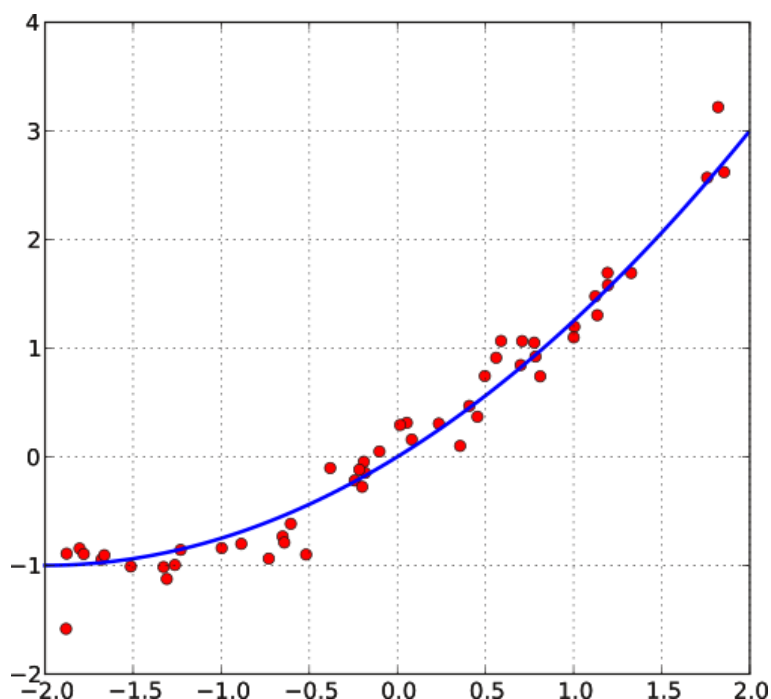


Рис. 1. Пример кривой, проведённой через точки, имеющие нормально распределённое отклонение от истинного значения

Иными словами, метод наименьших квадратов позволяет по экспериментальным данным подобрать такую аналитическую функцию, которая проходит настолько близко к экспериментальным точкам, насколько это возможно.

В общем случае **задачу** можно сформулировать следующим образом.

Пусть в результате эксперимента были получены некая экспериментальная зависимость $y(x)$, представленная в таблице 1.

Таблица 1

x	x_1	x_2	x_3	...	x_{n-1}	x_n
y	y_1	y_2	y_3	...	y_{n-1}	y_n

Необходимо построить аналитическую зависимость $f(x, a_1, a_2, \dots, a_m)$, наиболее точно описывающую результаты эксперимента. Для построения параметров функции $f(x, a_1, a_2, \dots, a_m)$ будем использовать метод наименьших квадратов. Идея метода наименьших квадратов заключается в том, что функцию $f(x, a_1, a_2, \dots, a_m)$ необходимо подобрать таким образом, чтобы сумма квадратов отклонений измеренных значений y_i от расчётных $y_i = f(x_i, a_1, a_2, \dots, a_m)$ была бы наименьшей (см. рис. 2):

$$S(a_0, a_1, a_2, \dots, a_m) = \sum_{i=1}^N (f(x_i, a_0, a_1, \dots, a_m) - y_i)^2 \rightarrow \min \quad (1)$$

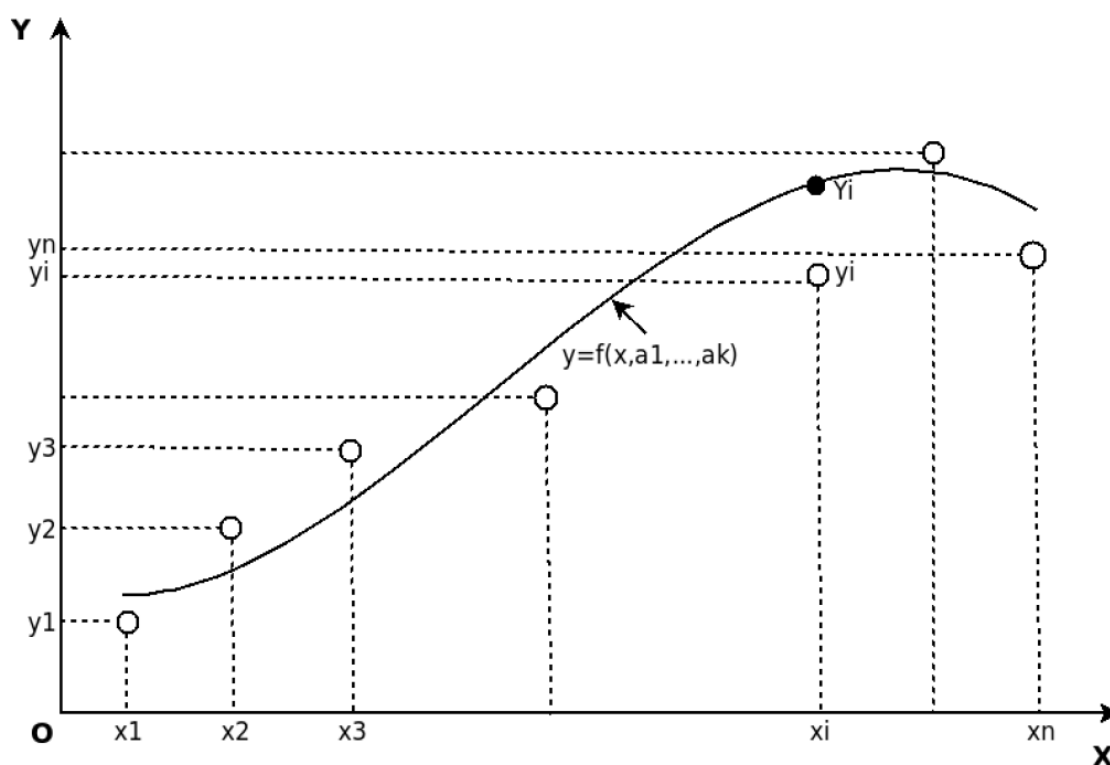


Рис. 2. Геометрическая интерпретация метода наименьших квадратов

Задача состоит из двух **этапов**:

1. По результатам эксперимента определить внешний вид подбираемой зависимости.
2. Подобрать коэффициенты зависимости $y = f(x, a_1, a_2, \dots, a_m)$.

Математически задача подбора коэффициентов зависимости сводится к определению коэффициентов a_i из условия (1).

Методика аппроксимации результатов механических испытаний методом наименьших квадратов

Рассмотрим *аппроксимацию* некоей экспериментальной зависимости $y = f(x)$.

Уравнение аппроксимирующей функции будем искать в виде

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m, \quad (2)$$

здесь $a_0, a_1, a_2, \dots, a_m$ – неизвестные константы.

Аппроксимация по *методу наименьших квадратов* сводится к отысканию минимума функции

$$S(a_0, a_1, a_2, \dots, a_m) = \sum_{i=1}^N (f(x_i, a_0, a_1, \dots, a_m) - y_i)^2 \rightarrow \min, \quad (3)$$

здесь y_i – значения эмпирически полученных данных, N – количество экспериментальных точек.

При нахождении минимума функции (3) ее частные производные по $a_0, a_1, a_2, \dots, a_m$ приравниваются к нулю. В итоге получается система линейных уравнений:

$$\begin{aligned} Na_0 + a_1 \sum_{i=1}^N x_i + a_2 \sum_{i=1}^N x_i^2 + \dots + a_m \sum_{i=1}^N x_i^m &= \sum_{i=1}^N y_i, \\ a_0 \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2 + a_2 \sum_{i=1}^N x_i^3 + \dots + a_m \sum_{i=1}^N x_i^{m+1} &= \sum_{i=1}^N x_i y_i, \\ \dots \dots \dots & \dots \dots \dots \\ a_0 \sum_{i=1}^N x_i^m + a_1 \sum_{i=1}^N x_i^{m+1} + a_2 \sum_{i=1}^N x_i^{m+2} + \dots + a_m \sum_{i=1}^N x_i^{2m} &= \sum_{i=1}^N x_i^m y_i \end{aligned} \quad (4)$$

При нахождении функции, аппроксимирующей экспериментальные данные, величины x_i, y_i при $i=1, \dots, N$ известны. Для нахождения неизвестных коэффициентов a_0, a_1, \dots, a_m , необходимо решить систему линейных уравнений относительно этих коэффициентов.

Одним из способов решения системы линейных уравнений является матричный способ. Составим на основании системы (4) матрицу системы X и матрицы столбцы неизвестных и свободных членов A и Y :

$$X = \begin{pmatrix} N & \sum_{i=1}^N x_i & \dots & \sum_{i=1}^N x_i^m \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \dots & \sum_{i=1}^N x_i^{m+1} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{i=1}^N x_i^m & \sum_{i=1}^N x_i^{m+1} & \dots & \sum_{i=1}^N x_i^{2m} \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ a_m \end{pmatrix}; \quad Y = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \\ \cdot \\ \sum_{i=1}^N x_i^m y_i \end{pmatrix}. \quad (5)$$

Тогда справедливо выражение

$$X \cdot A = Y \text{ или } \begin{pmatrix} N & \sum_{i=1}^N x_i & \dots & \sum_{i=1}^N x_i^m \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \dots & \sum_{i=1}^N x_i^{m+1} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{i=1}^N x_i^m & \sum_{i=1}^N x_i^{m+1} & \dots & \sum_{i=1}^N x_i^{2m} \end{pmatrix} \times \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ a_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \\ \cdot \\ \sum_{i=1}^N x_i^m y_i \end{pmatrix} \quad (6)$$

Выражения (6) и (4) тождественны. Решение матричного уравнения (6) можно найти в виде

$$A = X^{-1}Y. \quad (7)$$

Матричным методом можно решать только те системы, в которых *число уравнений совпадает с числом неизвестных*. Это обусловлено тем, что для нахождения обратной матрицы X^{-1} матрица X должна быть квадратной.

В качестве примера возьмем изменение относительного прогиба для пятислойной консольной композитной балки, вызванное отклонениями угла укладки внешнего слоя (табл. 2);

Таблица 2

Относительное изменение прогиба ΔU_z^{omn} в зависимости от отклонений угла укладки внешнего слоя при различном количестве слоев

Консольная балка				
Отклонение угла укладки, градусы	Относительное изменение прогиба $\Delta U_z^{omn}, \%$			
	Количество слоев			
	5	9	13	17
-10	20,0	8,8	6,9	5,0
-5	9,2	1,9	1,2	0,7
-3	5,3	0,6	0,1	0,0
0	0,0	0,0	0,0	0,0
3	4,5	0,8	1,3	1,1
5	7,1	1,8	2,7	2,3
10	11,9	5,6	7,5	6,4
Двухопорная балка				
Отклонение угла укладки, градусы	Относительное изменение прогиба ΔU_z^{omn}			
	Количество слоев			
	5	9	13	17
-10	27,8	13,3	8,4	5,6
-5	12,4	3,4	1,4	0,6
-3	7,0	1,2	0,1	0,1
0	0,0	0,0	0,0	0,0
3	5,1	0,8	1,7	1,5
5	7,3	2,2	3,5	3,0
10	9,5	7,4	9,5	8,0

матрицы X и Y , вычисленные по формулам (5) на основе табл. 2, при $m = 6$ имеют вид:

$$X = \begin{pmatrix} 7 & 0 & 268 & 0 & 21412 & 0 & 2032708 \\ 0 & 268 & 0 & 21412 & 0 & 2032708 & 0 \\ 268 & 0 & 21412 & 0 & 2032708 & 0 & 200,8 \cdot 10^6 \\ 0 & 21412 & 0 & 2032708 & 0 & 200,8 \cdot 10^6 & 0 \\ 21412 & 0 & 2032708 & 0 & 200,8 \cdot 10^6 & 0 & 20,0 \cdot 10^9 \\ 0 & 2032708 & 0 & 200,8 \cdot 10^6 & 0 & 20,0 \cdot 10^9 & 0 \\ 2032708 & 0 & 200,8 \cdot 10^6 & 0 & 20,0 \cdot 10^9 & 0 & 2 \cdot 10^{12} \end{pmatrix},$$

$$Y = \begin{pmatrix} 58,0 \\ -94,3 \\ 3685,8 \\ -8423,8 \\ 330061,8 \\ -820623 \\ 32171711 \end{pmatrix}.$$

Искомая матрица A равна

$$A = X^{-1}Y = \begin{pmatrix} -3 \cdot 10^{-11} \\ -74,7 \cdot 10^{-3} \\ 0,696 \\ -6,2 \cdot 10^{-3} \\ -18,0 \cdot 10^{-3} \\ 2,9 \cdot 10^{-5} \\ 126 \cdot 10^{-6} \end{pmatrix}.$$

Таким образом, имеем

$$a_0 = -3 \cdot 10^{-11}, a_1 = -74,7 \cdot 10^{-3}, a_2 = 0,696, a_3 = -6,2 \cdot 10^{-3},$$

$$a_4 = -18,0 \cdot 10^{-3}, a_5 = 2,9 \cdot 10^{-5}, a_6 = 126 \cdot 10^{-6}.$$

Функция (2) принимает вид

$$f(x) = -3 \cdot 10^{-11} - 74,7 \cdot 10^{-3}x + 0,696x^2 - 6,2 \cdot 10^{-3}x^3 -$$

$$-18,0 \cdot 10^{-3}x^4 + 2,9 \cdot 10^{-5}x^5 + 126 \cdot 10^{-6}x^6.$$

2 Практическая часть

1 В соответствии с таблицей 2 построить графические зависимости относительного изменения прогиба от величины отклонения угла укладки и количества слоев для двухопорных и консольных балок, определить необходимый порядок аппроксимирующего полинома и

вычислить уравнения аппроксимирующих кривых для 5-тислойных двухопорной и консольной балок.

2 Построить графики по результатам эксперимента по определению модуля межслойного сдвига, найти аппроксимирующую кривую и определить значения истинного модуля упругости E_{1n} и модуля межслойного сдвига G_{13} .

3 Некоторые тонкости работы в программе Microsoft Excel

3.1 Умножение матриц в программе MS Excel

В программе MS EXCEL можно производить умножение матрицы на матрицу. Осуществляется это так, как показано на рис. далее.

Умножение матриц

Задание. Даны матрицы A и B (рис.4). Найти их произведение $C=A*B$.

	A	B	C	D	E	F	G
1		1	4	7		9	6
2	A=	2	5	8	B=	8	5
3		3	6	9		7	4

Рисунок 3.

Ход работы:

1. Выделяем мышкой при нажатой левой кнопке соответствующий диапазон ячеек D5:E7 (строк такое же количество как в матрице A, а столбцов такое же количество как в матрице B).
- 2.

	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						

Рис. 4.

3. Вызываем мастер функций (меню «Вставка-Функция» или значок «Вставить функцию» в строке формул) и в категории «Полный алфавитный перечень» находим функцию «МУМНОЖ» и нажимаем ОК.
4. В появившемся окне вводим диапазон значений исходных матриц A и B (рис.5).

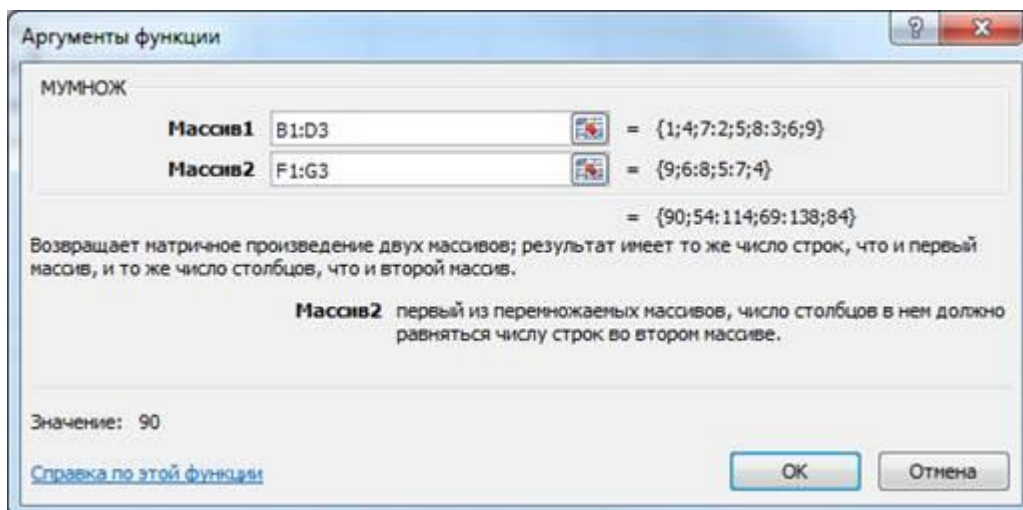


Рисунок 5.

5. Для получения результата нажимаем сочетание клавиш «Shift»+«Ctrl»+«Enter».
- 6.

	90	54
C=A*B=	114	69
	138	84

Рисунок 6

3.2 Нахождение обратной матрицы в MS Excel

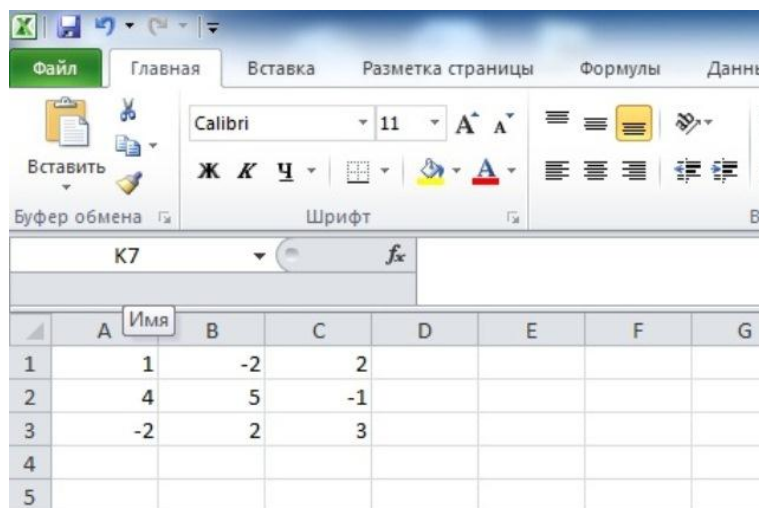
Нахождение обратной матрицы также возможно в EXCEL.

Условие существования обратной матрицы: **Матрица имеет обратную только тогда, когда ее определитель не равен нулю.**

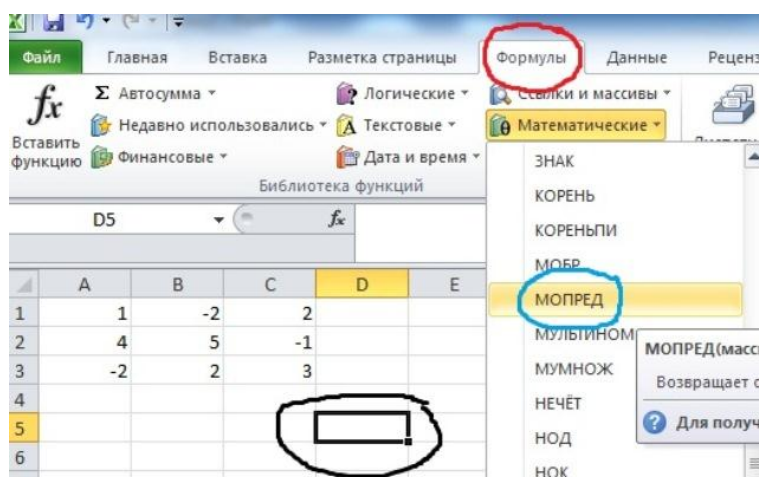
3.2.1 Нахождение определителя матрицы в MS Excel

$$\begin{vmatrix} 1 & -2 & 2 \\ 4 & 5 & -1 \\ -2 & 2 & 3 \end{vmatrix}$$

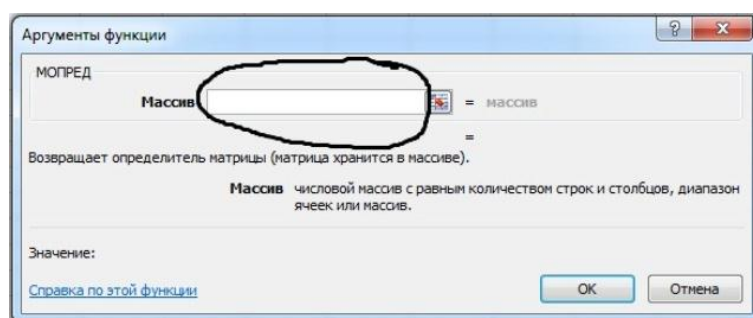
Создаем на рабочем столе документ MS EXCEL, открываем его и прямо с первой ячейки начинаем вписывать элементы определителя построчно



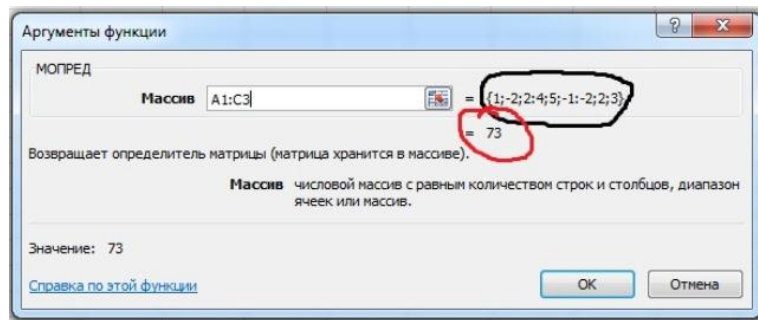
Нажмем на какую-нибудь ячейку, в которой нам запишут ответ (мой выбор обведен черным). Далее нажимаем Формулы, затем Математические и в выпадающем окне выбираем **МОПРЕД**.



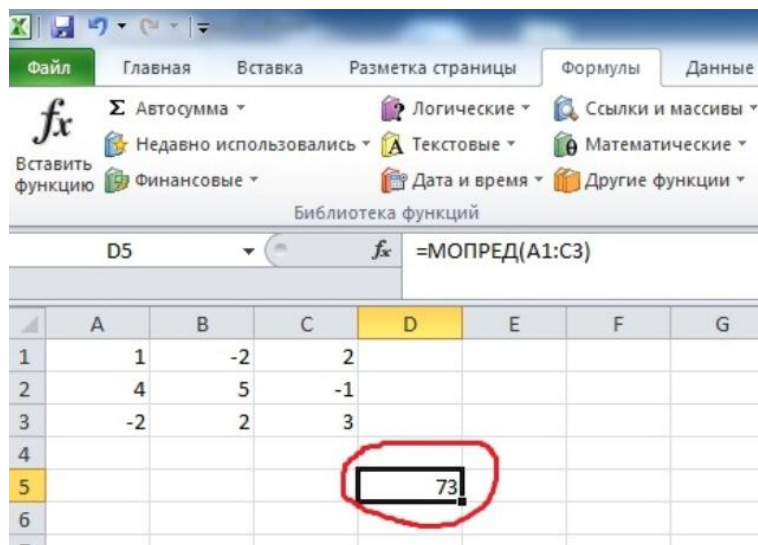
После нажатия появляется такое окно.



В этом окне, там где обведено, надо вписать адреса ячеек, на которых находятся элементы определителя. У нас элементы определителя записаны в ячейках начиная с A1 и заканчивая в C3. Впишем эти значения разделяя двоеточием, используя только латинские буквы.



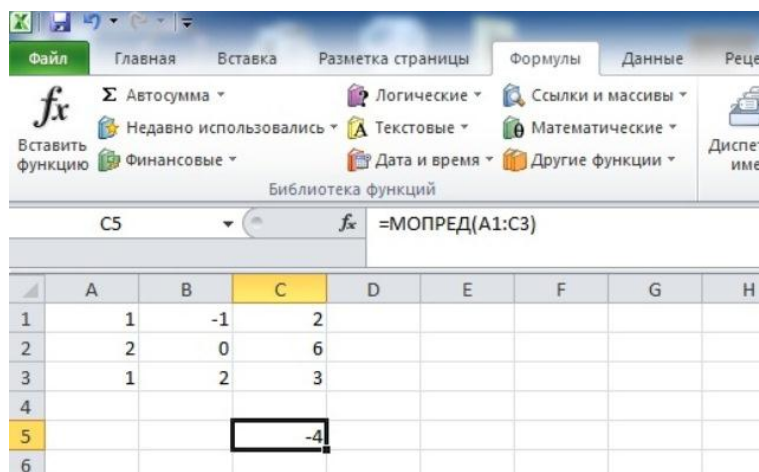
Если все записали правильно, то автоматически в обведенной черным области появятся элементы определителя, а в обведенной красным области появится ответ. Чтобы он записался в выбранную нами ячейку, нажимаем ОК.



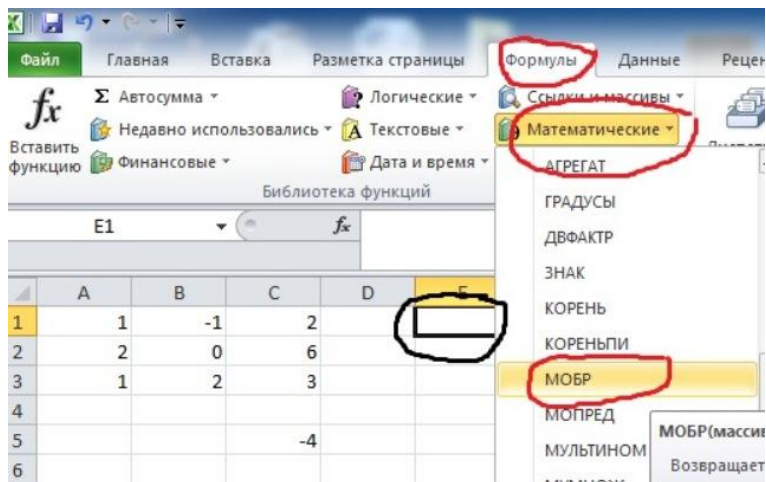
3.2.2 Нахождение обратной матрицы

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 0 & 6 \\ 1 & 2 & 3 \end{pmatrix}$$

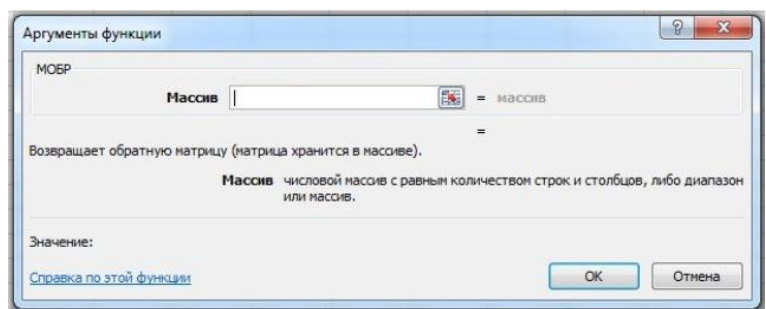
Первым делом проверяем, есть ли у матрицы определитель.



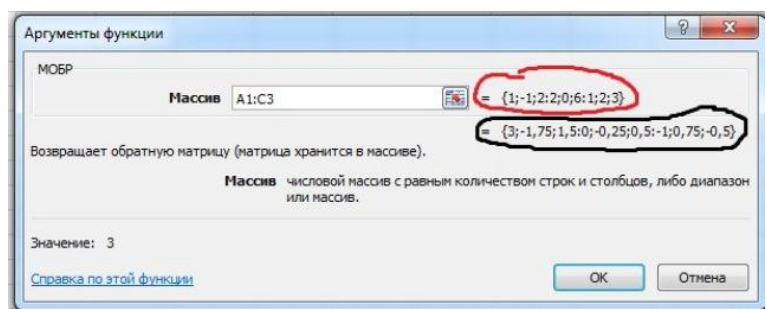
Он оказался равен **-4**, а это значит, что у нашей матрицы есть обратная (если бы определитель оказался равен нулю, то матрица не имеет обратную, следовательно нет смысла считать далее). Теперь отметим ячейку, с которой начнем записывать ответ. Например, E1. Нажимаем **Формулы**, затем **Математические** и в появившемся окне находим **МОБР**



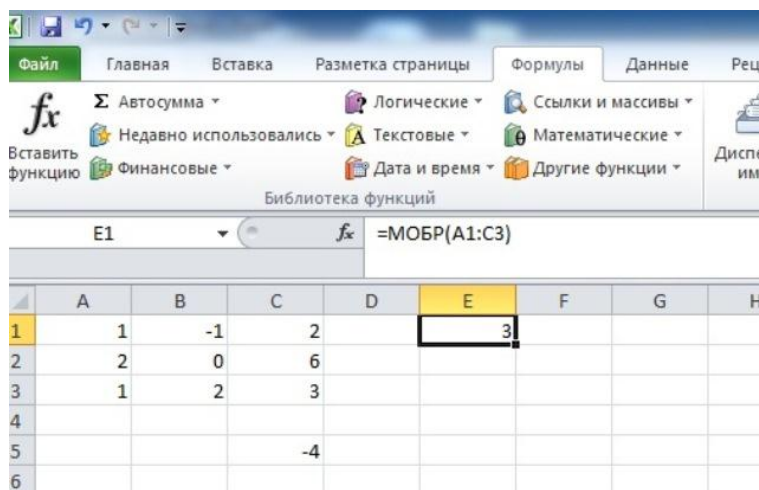
После нажатия появляется вот такое окно, в котором надо вписать адреса ячеек, в которых находятся элементы матрицы в **Массив**



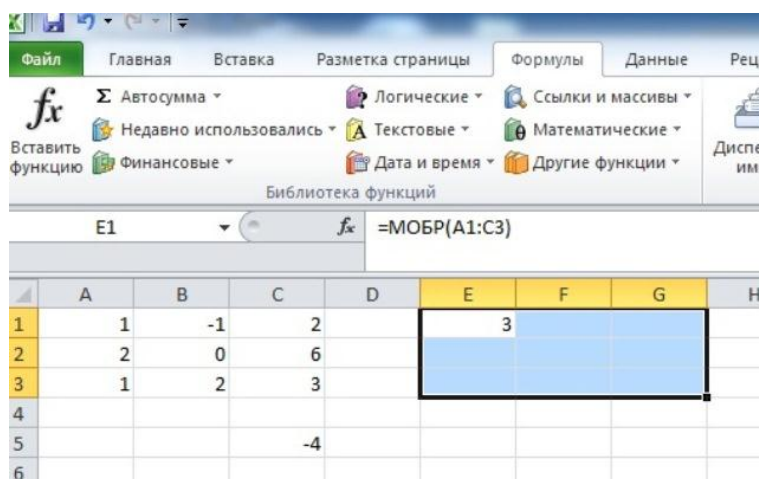
У нас элементы записаны в ячейки начиная с **A1** и заканчивая в **C3**, поэтому так и записываем (смотрите картинку)



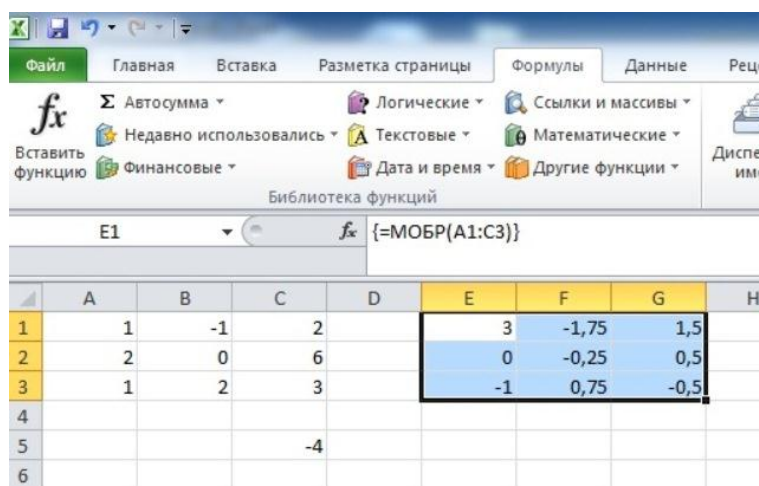
Если все сделали правильно, то автоматически заполнится место, обведенное красным и запишется ответ, который обведен черным. В таком виде ответ трудно переадресовать и поэтому нажимаем **ОК**. В ячейке, которую мы застолбили под ответ, появилось число 3, Это только первый элемент полученной обратной матрицы.



Чтобы виден был весь ответ, выполняем следующие действия: Начиная с ячейки **E1** выделяем три строки и три столбца (именно столько было у исходной матрицы и столько же будет у обратной)



нажимаем клавишу **F2**, а затем на одновременно на три клавиши **Ctrl+Shift+Enter**.



В выделенном месте появляются, теперь уже все, элементы обратной матрицы.

3.3 Метод наименьших квадратов в MS EXCEL – функция ТЕНДЕНЦИЯ

Условия задачи: молодая пара, которые, с недавних пор, живут вместе и совместно делят столик для косметических принадлежностей в ванной. Молодой человек начал замечать, что половина его столика неумолимо сокращается, сдавая свои позиции муссам для волос и соевым комплексам. За последние несколько месяцев парень внимательно следил за тем, с какой скоростью увеличивается число предметов на ее части стола. В таблице ниже представлено число предметов девушки на столике в ванной, накопившихся за последние несколько месяцев.

	А	В
1	Месяц	Число предметов
2	1	8
3	2	6
4	3	10
5	4	6
6	5	10
7	6	13
8	7	9
9	8	11
10	9	15
11	10	17
12		

Вопрос: какое количество предметов будет на столике для косметических принадлежностей в ванной через 16 месяцев.

Для решения поставленной задачи воспользуемся методом наименьших квадратов. В Excel имеется функция для расчета значения по **методу наименьших квадратов**. Это функция называется **ТЕНДЕНЦИЯ**. Синтаксис у нее следующий:

ТЕНДЕНЦИЯ (известные значения Y; известные значения X; новые значения X; конст)

где:

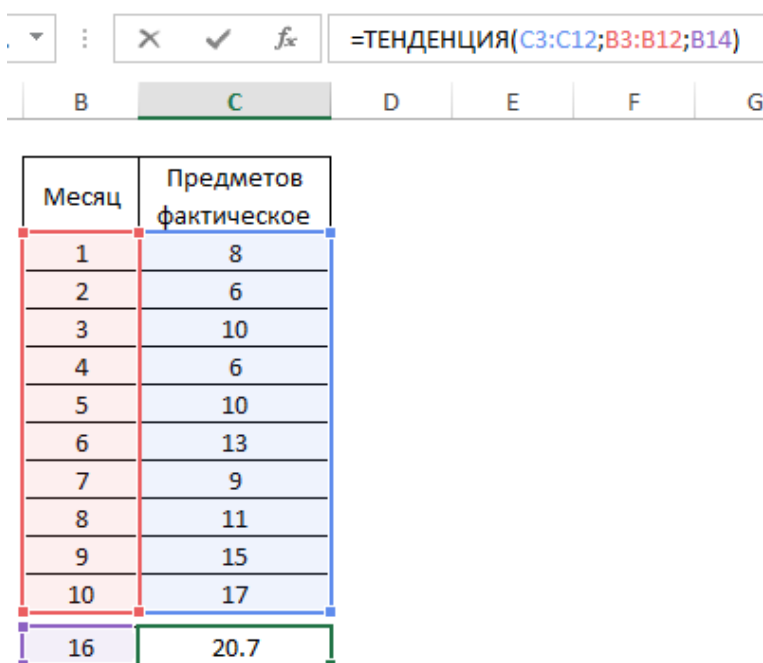
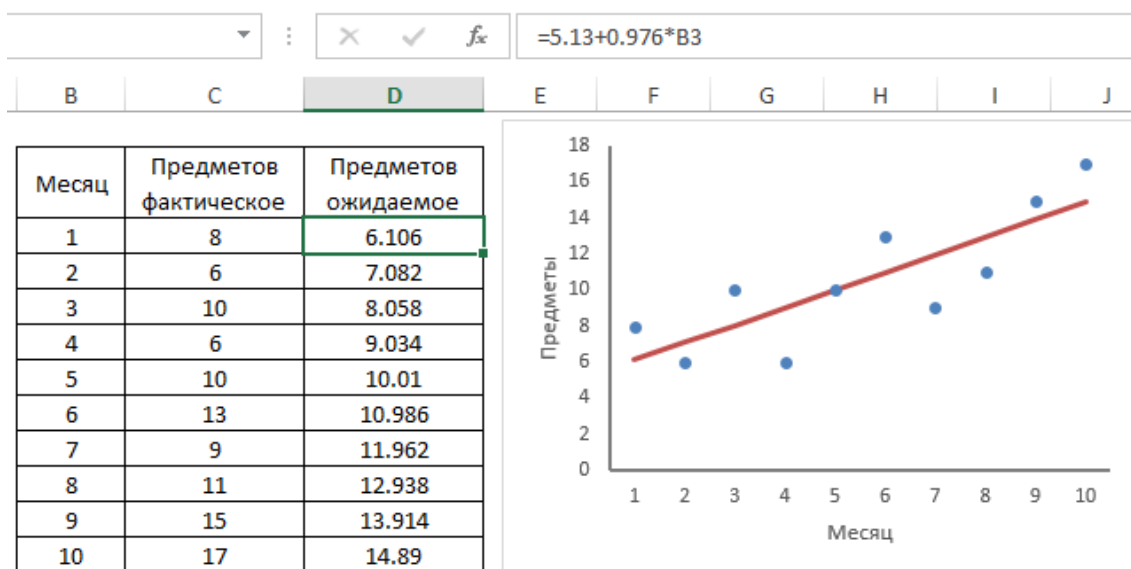
известные значения Y – массив зависимых переменных, в нашем случае, количество предметов на столике

известные значения X – массив независимых переменных, в нашем случае это месяц

новые значения X – новые значения X (месяца) для которого **функция ТЕНДЕНЦИЯ** возвращает ожидаемое значение зависимых переменных (количество предметов)

конст — необязательный. Логическое значение, которое указывает, требуется ли, чтобы константа *b* была равна 0.

Например, на рисунке показана функция **ТЕНДЕНЦИЯ**, используемая для определения ожидаемого количества предметов на столике в ванной для 16-го месяца.



3.4 Построение аппроксимирующих кривых средствами MS EXCEL

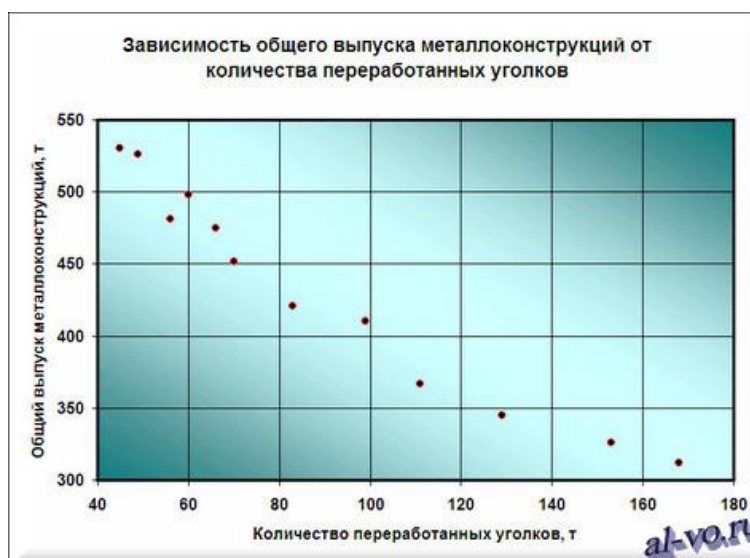
ПРИМЕР: Производственный участок изготавливает строительные металлоконструкции из листового и профильного металлопроката. Участок работает стабильно, заказы однотипные, численность рабочих колеблется незначительно. Есть данные о выпуске продукции за предыдущие 12 месяцев и о количестве переработанного в эти периоды времени металлопроката по группам: листы, двутавры, швеллеры, уголки, трубы круглые, профили прямоугольного сечения, круглый прокат. После предварительного анализа исходных данных возникло предположение, что суммарный месячный выпуск металлоконструкций существенно зависит от количества уголков в заказах. Проверим это предположение.

Прежде всего, несколько слов об аппроксимации. Мы будем искать закон – аналитическую функцию, то есть функцию, заданную уравнением, которое лучше других описывает зависимость общего выпуска металлоконструкций от количества уголкового проката в выполненных заказах. Это и есть аппроксимация, а найденное уравнение называется аппроксимирующей функцией для исходной функции, заданной в виде таблицы.

1. Включаем Excel и помещаем на лист таблицу с данными статистики.

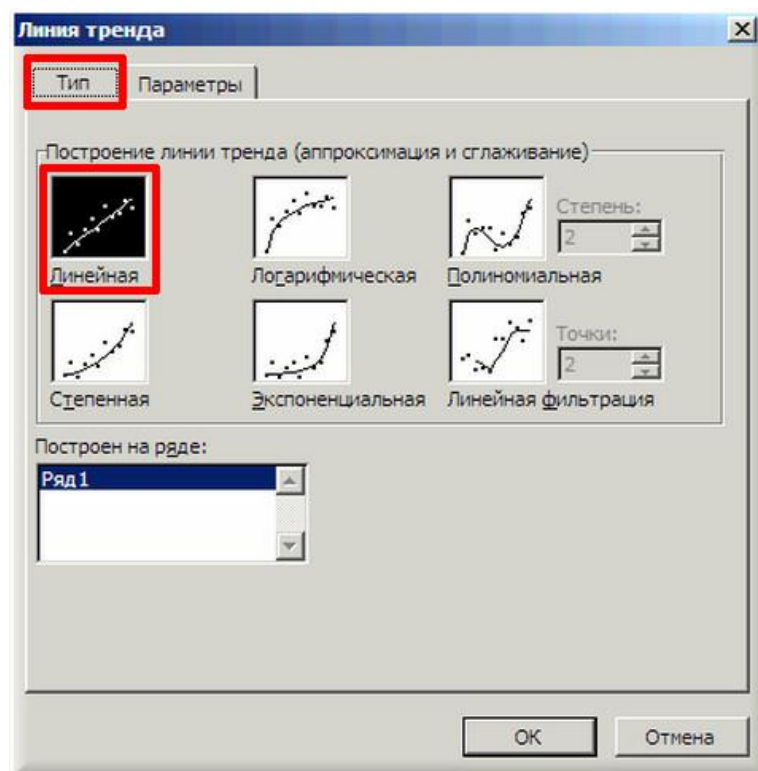
	А	В	С
1	a1-vo.ru	Статистические данные	
2	Период	Количество переработанных уголков, т	Общий выпуск металлоконструкций, т
3	Январь	60	498
4	Февраль	66	475
5	Март	153	326
6	Апрель	45	530
7	Май	99	410
8	Июнь	111	367
9	Июль	129	345
10	Август	168	312
11	Сентябрь	83	421
12	Октябрь	49	526
13	Ноябрь	56	481
14	Декабрь	70	452

2. Далее строим и форматируем точечную диаграмму, в которой по оси X задаем значения аргумента – количество переработанных уголков в тоннах. По оси Y откладываем значения исходной функции – общий выпуск металлоконструкций в месяц, заданные таблицей.

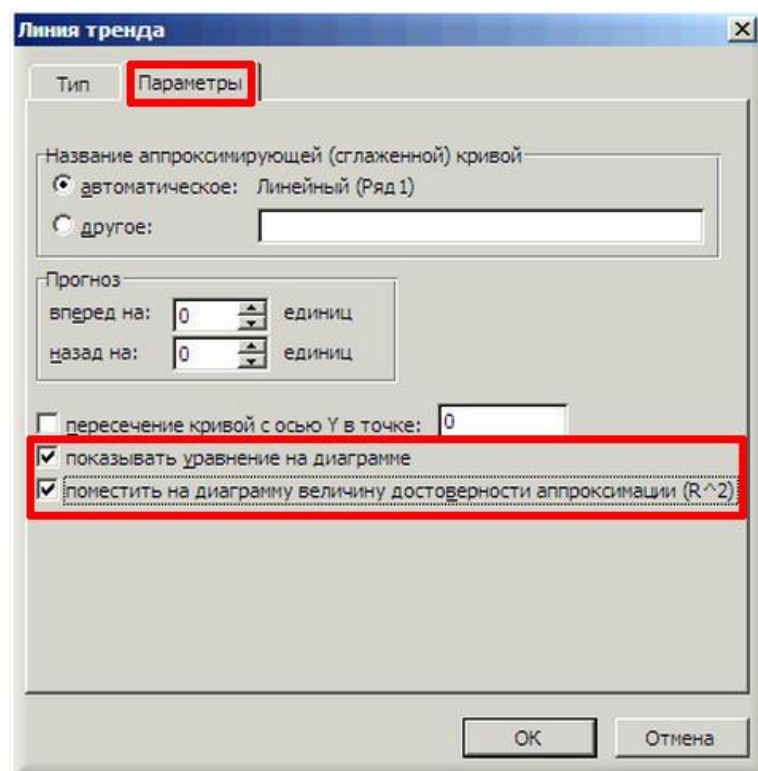


3. «Наводим» мышь на любую из точек на графике и щелчком правой кнопки вызываем контекстное меню. В выпавшем меню выбираем «Добавить линию тренда...».

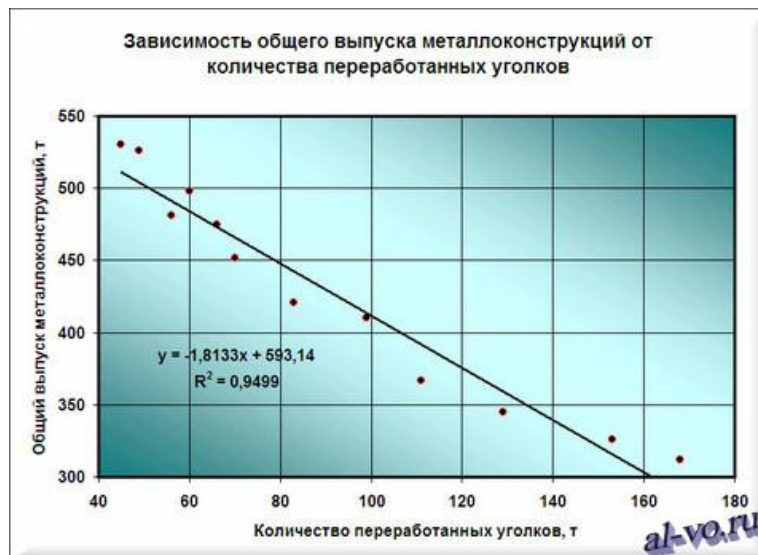
4. В появившемся окне «Линия тренда» на вкладке «Тип» выбираем «Линейная».



5. Далее на вкладке «Параметры» ставим 2 галочки и нажимаем «ОК».



6. На графике появилась прямая линия, аппроксимирующая нашу табличную зависимость.



Мы видим кроме самой линии уравнение этой линии и, главное, мы видим значение параметра R^2 – **величины достоверности аппроксимации!** Чем ближе его значение к 1, тем наиболее точно выбранная функция аппроксимирует табличные данные!

Величина достоверности аппроксимации по формуле:

$$R^2 = 1 - \frac{\sum_1}{\sum_2}, \quad (8)$$

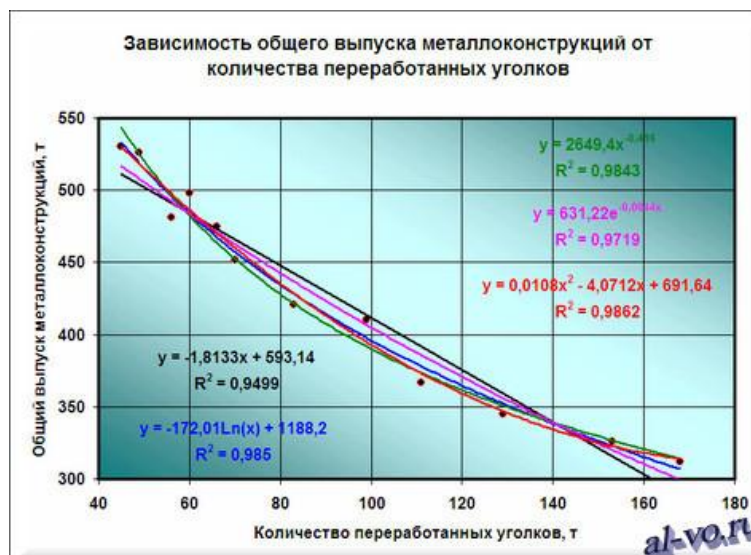
где

$$\sum_1 = \sum_{i=1}^N (f(x_i) - y_i)^2, \quad \sum_2 = \sum_{i=1}^N y_i^2 - \frac{\sum_{i=1}^N y_i^2}{N}, \quad (9)$$

здесь y_i – экспериментальные значения, полученные для аргумента x_i , а $f(x_i)$ – значения функции для x_i , полученные по уравнению аппроксимирующей кривой.

Если $R^2 \geq 0,95$, то говорят о высокой точности аппроксимации, если $0,8 \leq R^2 < 0,95$, то аппроксимация удовлетворительна, если $0,6 \leq R^2 < 0,8$, то аппроксимация слабая. Если $R^2 < 0,6$, то точность аппроксимации недостаточна.

7. Строим линии тренда, используя степенную, логарифмическую, экспоненциальную и полиномиальную аппроксимации по аналогии с тем, как мы строили линейную линию тренда.



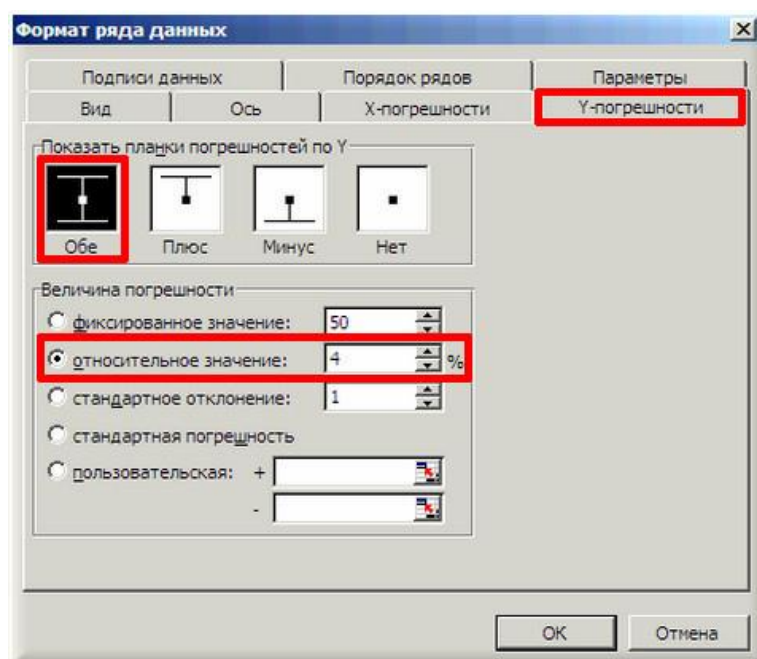
Лучше всех из выбранных функций аппроксимирует наши данные полином второй степени, у него максимальный коэффициент достоверности R^2 .

Однако хочу вас предостеречь! Если вы возьмете полиномы более высоких степеней, то, возможно, получите еще лучшие результаты, но кривые будут иметь замысловатый вид.... Здесь важно понимать, что мы ищем функцию, которая имеет физический смысл. Что это означает? Это означает, что нам нужна аппроксимирующая функция, которая будет выдавать адекватные результаты не только внутри рассматриваемого диапазона значений X , но и за его пределами, то есть ответит на вопрос: «Какой будет выпуск металлоконструкций при количестве переработанных за месяц уголков меньше 45 и больше 168 тонн!» Поэтому я не рекомендую увлекаться полиномами высоких степеней, да и параболу (полином второй степени) выбирать осторожно!

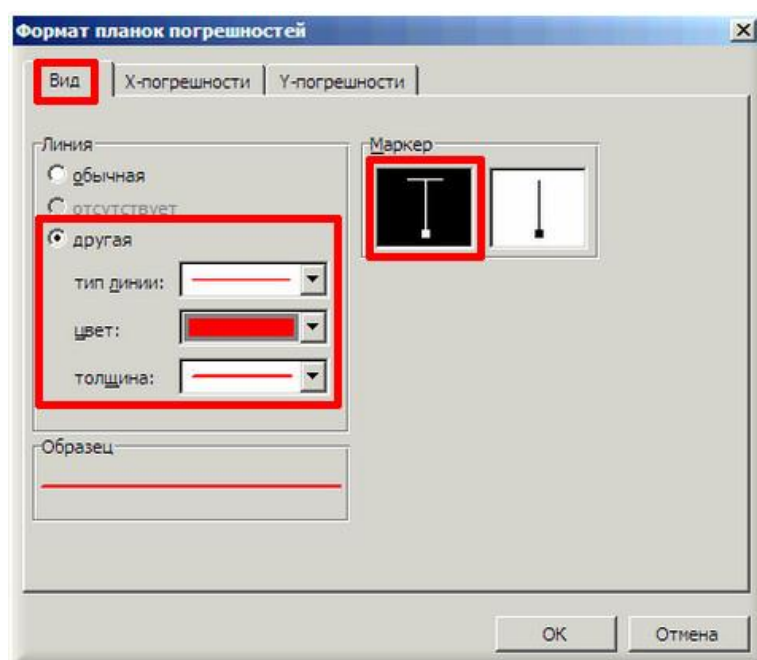
Итак, нам необходимо выбрать функцию, которая не только хорошо интерполирует табличные данные в пределах диапазона значений $X=45\dots168$, но и допускает адекватную экстраполяцию за пределами этого диапазона. Я выбираю в данном случае логарифмическую функцию, хотя можно выбрать и линейную, как наиболее простую. В рассматриваемом примере при выборе линейной аппроксимации в excel ошибки будут больше, чем при выборе логарифмической, но не на много.

8. Удаляем все линии тренда с поля диаграммы, кроме логарифмической функции. Для этого щелкаем правой кнопкой мыши по ненужным линиям и в выпавшем контекстном меню выбираем «Очистить».

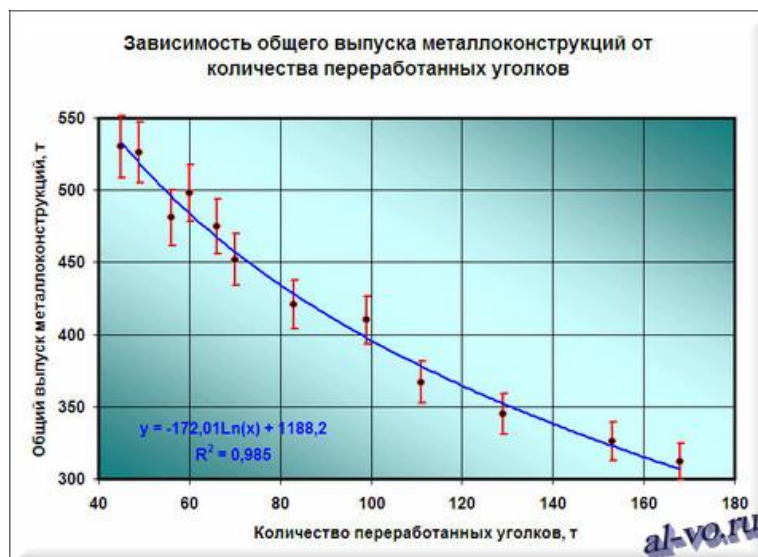
9. В завершении добавим к точкам табличных данных планки погрешностей. Для этого правой кнопкой мыши щелкаем на любой из точек на графике и в контекстном меню выбираем «Формат рядов данных...» и настраиваем данные на вкладке «**Y-погрешности**» так, как на рисунке ниже.



10. Затем щелкаем по любой из линий диапазонов погрешностей правой кнопкой мыши, выбираем в контекстном меню «**Формат полос погрешностей...**» и в окне «**Формат планок погрешностей**» на вкладке «**Вид**» настраиваем цвет и толщину линий.



Аналогичным образом форматируются любые другие объекты диаграммы в Excel!
Окончательный результат диаграммы представлен на следующем снимке экрана.



Итоги. Результатом всех предыдущих действий стала полученная формула аппроксимирующей функции $y = -172,01 \cdot \ln(x) + 1188,2$. Зная ее, и количество уголков в месячном наборе работ, можно с высокой степенью вероятности ($\pm 4\%$ — смотри планки погрешностей) спрогнозировать общий выпуск металлоконструкций за месяц! Например, если в плане на месяц 140 тонн уголков, то общий выпуск, скорее всего, при прочих равных составит 338 ± 14 тонн.

Для повышения достоверности аппроксимации статистических данных должно быть много. Двенадцать пар значений – это маловато.

Из практики скажу, что **хорошим результатом** следует считать нахождение аппроксимирующей функции с коэффициентом достоверности $R^2 > 0,87$. **Отличный результат** – при $R^2 > 0,94$.

Контрольные вопросы:

- 1) Суть метода наименьших квадратов.
- 2) В каких случаях можно найти обратную матрицу для заданной?
- 3) Как правильно выбрать необходимый порядок аппроксимирующего полинома?

Список литературы:

- 1) ru.wikipedia.org
- 2) интернет-ресурсы