ПРИБЛИЖЕНИЕ ТАБЛИЧНЫХ ФУНКЦИЙ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

Постановка задачи аппроксимации

По результатам экспериментов получена таблица с произвольным расположением аргументов: (x_i, y_i), $i = \overline{1,n}$. Аналитическое выражение табличной функции может быть неизвестным. На основе этой таблицы требуется найти формулу $F = F(x)$, ***приближённо*** описывающую зависимость между экспериментальными данными таблицы. При этом отклонение значений в точках x_i, $i = \overline{1,n}$, вычисленные по формуле $F = F(x)$, от экспериментальных данных y_i должны быть минимальными.

Дано: точки наблюдения (x_i, y_i) $i = \overline{1,n}$ (**их количество *n***).

x	x_1	x_2	…	x_n
y	y_1	y_2	…	y_n

Найти функцию $F(x)$ и значение $F(x^*)$, $x^* \notin [x_1; x_n]$:

1. $F_i = F(x_i) \approx y_i, \forall i = \overline{1,n}$
2. $\rho(\mathbf{y}, \mathbf{F}) \to \min$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, $\mathbf{F} = (F_1, F_2, \ldots, F_n)$

Определение. Полученное функциональное соотношение $F(x)$ для приближённого описания зависимости между экспериментальными данными таблицы (x_i, y_i), $i = \overline{1,n}$ называется эмпирической формулой, а функция $F(x)$ – эмпирической функцией.

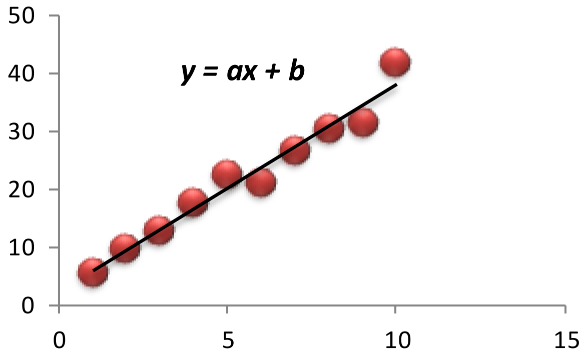
Возможным вариантом решения этой задачи является интерполирование. Однако этот способ, требующий обязательного совпадения табличных значений табличной и приближающей функций во всех табличных аргументах мало пригоден. При большом количестве узлов он является неудобным и сложным, ибо потребует отыскания многочлена большой степени.

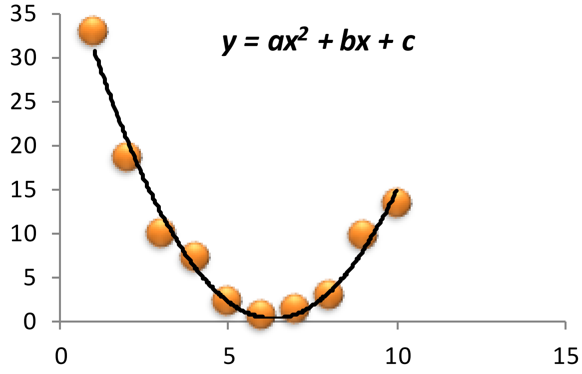
Кроме того, экспериментальные данные в силу ряда причин могут иметь трудно учитываемые случайные или систематические ошибки. В этих условиях интерполирование становится сомнительным.

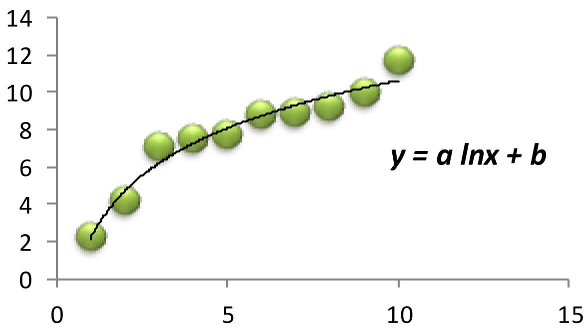
Часто с помощью какой-либо простой функции с проходящим около табличных точек графиком, удаётся добиться эффекта сглаживания ошибок и получить более точное приближение.

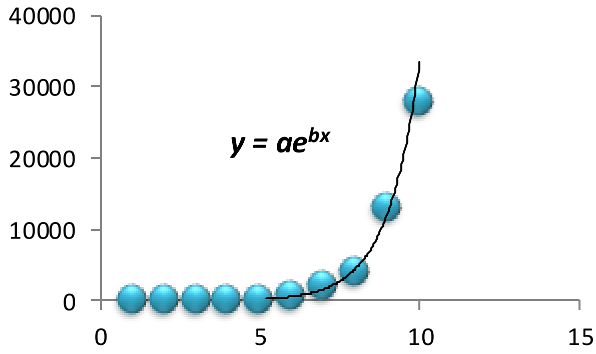
Поиск эмпирической формулы $F(x)$ начинается с определения класса функций, которые лучше всего отражают связь между табличными данными. Эффективным методом для этого являются графические изображения. На координатной плоскости отмечаются определяемые данной функцией точки, а затем по характеру их расположения подбирается вид приближения из числа известных элементарных функций.

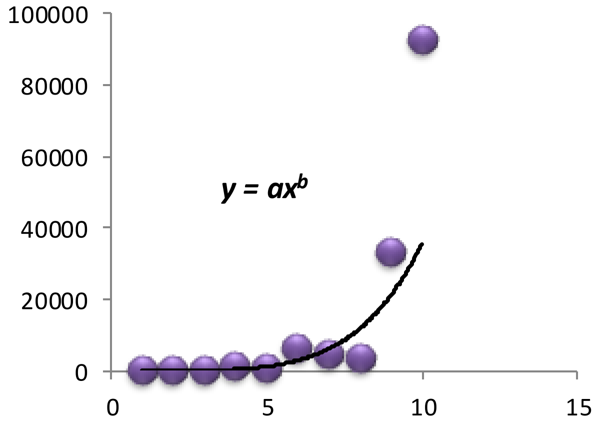
В перечень наиболее часто используемых классов функций входят, например, линейные $y=a_1x+a_0$, полиномиальные $y=a_2x^2+a_1x+a_0$ логарифмические $y=a\ln x+b$, экспоненциальные $y=ae^{bx}$, степенные $y=ax^b$, дробно-рациональные $y=\dfrac{1}{ax+b}$, гиперболические $y=a+\dfrac{b}{x}$ и другие функциональные зависимости.

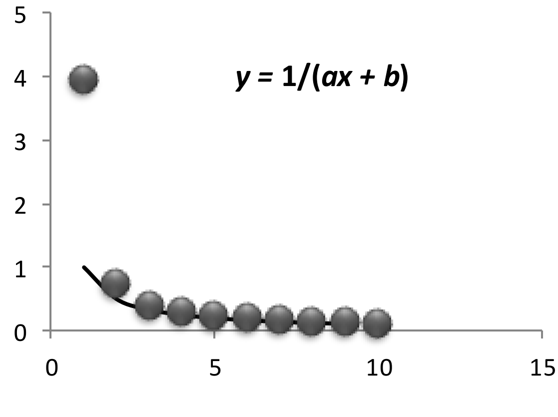












Метод наименьших квадратов для поиска приближённой зависимости

Поиск числовых параметров выбранной эмпирической зависимости $F(x)$ сводится к решению задачи поиска минимального значения метрики $\rho(\mathbf{y},\mathbf{F})=\sqrt{\sum_{i=1}^{n}\left(y_i-F(x_i)\right)^2}$.

Определение. Квадрат метрики $\rho^2(\mathbf{y},\mathbf{F})=\sum_{i=1}^{n}\left(y_i-F(x_i)\right)^2$ называется среднеквадратическим отклонением.

Задача приближения по методу наименьших квадратов. Требуется определить такие значения параметров $a_0,a_1,\ldots a_n$ выбранной эмпирической зависимости $F(x,a_0,a_1,\ldots a_n)$, при которых квадрат метрики $\rho^2(\mathbf{y},\mathbf{F})$ (среднеквадратическое отклонение) будет наименьшим:

$$\rho^2(\mathbf{y},\mathbf{F})=\sum_{i=1}^{n}\left(y_i-F(x_i,a_0,a_1,\ldots a_n)\right)^2\to\min$$

Эта задача является задачей поиска минимума функции многих переменных $a_0,a_1,\ldots a_n$:

$$S(x,a_0,a_1,\ldots a_n)=\sum_{i=1}^{n}\left(y_i-F(x_i,a_0,a_1,\ldots a_n)\right)^2\to\min$$

Необходимое условие экстремума функции многих переменных – равенство нулю всех частных производных функции:

$$\forall i=\overline{1,n}\ :\ \frac{\partial S}{\partial a_i}=0\,.$$

Поскольку

$$\frac{\partial S}{\partial a_i}=\sum_{i=1}^{n}\frac{\partial\left(y_i-F(x_i,a_0,a_1,\ldots a_n)\right)^2}{\partial a_i}=2\sum_{i=1}^{n}\left(y_i-F(x_i,a_0,a_1,\ldots a_n)\right)\frac{\partial F(x_i,a_0,a_1,\ldots a_n)}{\partial a_i}=0\,,$$

параметры $a_0,a_1,\ldots a_n$ определяются из системы уравнений:

$$\sum_{i=1}^{n}\left(y_i-F(x_i,a_0,a_1,\ldots a_n)\right)\frac{\partial F(x_i,a_0,a_1,\ldots a_n)}{\partial a_i}=0\,,\ i=\overline{1,n} \qquad (1)$$

Линейная аппроксимация

Пусть на основании графического изображения в качестве эмпирической функции выбирается линейная функция $F(x)=a_1x+a_0$.

Система (1) примет вид:

$$\left.\begin{aligned}\sum_{i=1}^{n}\left(y_i-(a_1x_i+a_0)\right)\frac{\partial F(a_1x_i+a_0)}{\partial a_0}=0\\ \sum_{i=1}^{n}\left(y_i-(a_1x_i+a_0)\right)\frac{\partial F(a_1x_i+a_0)}{\partial a_1}=0\end{aligned}\right\},\quad \left.\begin{aligned}\sum_{i=1}^{n}\left(y_i-(a_1x_i+a_0)\right)\cdot 1=0\\ \sum_{i=1}^{n}\left(y_i-(a_1x_i+a_0)\right)\cdot x_i=0\end{aligned}\right\}.$$

Отсюда получается система линейных уравнений для определения параметров a_0, a_1 линейной аппроксимации:

$$\left.\begin{array}{lcccc} a_1\sum\limits_{i=1}^{n} x_i^2 & + & a_0\sum\limits_{i=1}^{n} x_i & = & \sum\limits_{i=1}^{n} x_i y_i \\ a_1\sum\limits_{i=1}^{n} x_i & + & na_0 & = & \sum\limits_{i=1}^{n} y_i \end{array}\right\}. \qquad (2)$$

Задача 1

Методом наименьших квадратов определить параметры линейной эмпирической функции $F(x) = a_1 x + a_0$ по экспериментальным данным, представленным в таблице. Найти $F(6)$.

x	**y**
1	1,6
2	1
3	3
4	3,3
5	5,4

Решение

Количество точек наблюдения *n* = 5.

Для определения параметров линейной зависимости $F(x) = a_1 x + a_0$ составляется вспомогательная таблица:

	x	**y**	**x²**	**xy**
	1	1,6	1	1,6
	2	1	4	2
	3	3	9	9
	4	3,3	16	13
	5	5,4	25	27
Σ	**15**	**14**	**55**	**53**

Полученные значения сумм подставляются в систему линейных уравнений для определения параметров a_0, a_1 линейной аппроксимации:

$$\left.\begin{array}{lcccc} 55a_1 & + & 15a_0 & = & 52{,}8 \\ 15a_1 & + & 5a_0 & = & 14{,}3 \end{array}\right\}$$

Комментарий к решению системы. Точные методы решения этой системы линейных уравнений известны: решение системы с помощью обратной матрицы, с помощью формул Крамера, метод исключения неизвестных – метод Гаусса. В данном случае проще применить метод Гаусса. Действительно, если умножить второе уравнение системы на (-3) и сложить его с первым, то легко определяется параметр a_1:

$$\left.\begin{array}{rcrcr} 55a_1 & + & 15a_0 & = & 52{,}8 \\ -45a_1 & - & 15a_0 & = & -42{,}9 \end{array}\right\},\ \left.\begin{array}{rcrcr} 55a_1 & + & 15a_0 & = & 52{,}8 \\ 10a_1 & & & = & 9{,}9 \end{array}\right\},\ \left.\begin{array}{rcrcr} 55a_1 & + & 15a_0 & = & 52{,}8 \\ a_1 & & & = & 0{,}99 \end{array}\right\}.$$

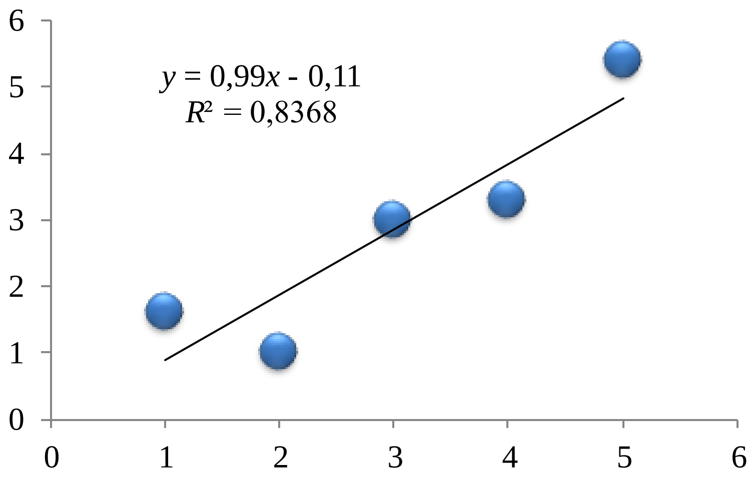
При подстановке $a_1 = 0{,}99$ в любое уравнение исходной системы получается $a_0 = -0{,}11$.

Искомая эмпирическая функция имеет вид: $F(x) = 0{,}99x - 0{,}11$.

$F(6) = 0{,}99 \cdot 6 - 0{,}11 = 5{,}83$

Реализация решения примера в MS Excel

Услуга Мастера диаграмм Построение линии тренда реализует метод наименьших квадратов для поиска коэффициентов эмпирической функции и построения её графика. Параметры линейной зависимости определяются с помощью Мастера функций: с помощью функции ЛИНЕЙН из категории Статистические функции.



Квадратическая (параболическая, полиномиальная 2-й степени) аппроксимация

Пусть на основании графического изображения в качестве эмпирической функции выбирается линейная функция $F(x) = a_2x^2 + a_1x + a_0$.

Система (1) примет вид:

$$\left.\begin{array}{l} \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot\dfrac{\partial F\left(a_2x_i^2 + a_1x_i + a_0\right)}{\partial a_0} = 0 \\ \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot\dfrac{\partial F\left(a_2x_i^2 + a_1x_i + a_0\right)}{\partial a_1} = 0 \\ \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot\dfrac{\partial F\left(a_2x_i^2 + a_1x_i + a_0\right)}{\partial a_2} = 0 \end{array}\right\},\ \left.\begin{array}{l} \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot 1 = 0 \\ \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot x_i = 0 \\ \sum\limits_{i=1}^{n}\left(y_i - \left(a_2x_i^2 + a_1x_i + a_0\right)\right)\cdot x_i^2 = 0 \end{array}\right\}.$$

Отсюда получается система линейных уравнений для определения параметров a_0, a_1, a_2 параболической аппроксимации:

$$\left.\begin{array}{l} a_2\sum_{i=1}^{n} x_i^4 + a_1\sum_{i=1}^{n} x_i^3 + a_0\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i \\ a_2\sum_{i=1}^{n} x_i^3 + a_1\sum_{i=1}^{n} x_i^2 + a_0\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a_2\sum_{i=1}^{n} x_i^2 + a_1\sum_{i=1}^{n} x_i + na_0 = \; = \sum_{i=1}^{n} y_i \end{array}\right\}.$$

Экспоненциальная аппроксимация

Пусть на основании графического изображения в качестве эмпирической функции выбирается экспоненциальная функция $F(x) = ae^{bx}$.

Для применения метода наименьших квадратов экспоненциальная функция линеаризуется:

$\ln F = \ln ae^{bx}$, $\ln F = \ln a + bx$.

Метод наименьших квадратов применяется к линейной функции $\hat{F} = a_1 X + a_0$, где $a_0 = \ln a$, $a_1 = b$, $\hat{F} = \ln F$, $X = x$.

Задача 2

Методом наименьших квадратов определить параметры экспоненциальной эмпирической функции $F(x) = ae^{bx}$ по экспериментальным данным, представленным в таблице. Найти минимальную величину $\rho^2(\mathbf{y}, \mathbf{F})$. Найти $F(6)$.

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

Решение

Количество точек наблюдения $n = 5$.

Для применения метода наименьших квадратов экспоненциальная функция линеаризуется:

$$\ln F = \ln ae^{bx}, \; \ln F = \ln a + bx.$$

Метод наименьших квадратов применяется для определения коэффициентов линейной функции $\hat{F} = a_1 X + a_0$, аппроксимирующей экспериментальные данные (X_i, Y_i), где $X_i = x_i$, $Y_i = \ln y_i$, $a_0 = \ln a$, $a_1 = b$, $\hat{F} = \ln F$, $i = \overline{1,n}$.

Для определения параметров линейной зависимости $\hat{F} = a_1 X + a_0$ составляется вспомогательная таблица:

	x	**y**	**X = x**	**Y= lny**	X^2	**XY**	**F**	$(y-F)^2$
	1	2,1	1	0,7	1	0,7	2,13	0,00
	2	6,2	2	1,8	4	3,6	4,82	1,90
	3	13,3	3	2,6	9	7,8	10,90	5,76
	4	24,6	4	3,2	16	12,8	24,65	0,00
	5	117,5	5	4,8	25	23,8	55,73	3815,82
Σ	**15**	**163,7**	**15**	**13,1**	**55**	**48,8**		**3823,493**

Полученные значения сумм подставляются в систему линейных уравнений для определения параметров a_0, a_1 линейной аппроксимации:

$$\left.\begin{array}{rcrcl} 55a_1 & + & 15a_0 & = & 48,8 \\ 15a_1 & + & 5a_0 & = & 13,1 \end{array}\right\}.$$

Решением системы являются значения $a_1 = 0,94$ и $a_0 = -0,20$. Линеаризованная эмпирическая функция имеет вид: $\hat{F} = 0,94x - 0,20$.

Определение параметров экспоненциальной ***эмпирической*** зависимости производится из ранее принятых соотношений:

$$a_0 = \ln a,\ a_1 = b.$$

Откуда $a = e^{a_0} = e^{-0,20} = 0,82$, $b = a_1 = 0,94$. Искомая экспоненциальная эмпирическая функция будет иметь вид: $F(x) = ae^{bx} = 0,94 \cdot e^{0,82x}$.

Логарифмическая аппроксимация

Пусть на основании графического изображения в качестве эмпирической функции выбирается логарифмическая функция $F = a\ln x + b$.

Метод наименьших квадратов применяется к линейной функции $\hat{F} = a_1 X + a_0$, где $a_0 = b$, $a_1 = a$, $\hat{F} = F$, $X = \ln x$.

Степенная аппроксимация

Пусть на основании графического изображения в качестве эмпирической функции выбирается степенная функция $F(x) = ax^b$.

Для применения метода наименьших квадратов степенная функция линеаризуется:

$$\ln F(x) = \ln ax^b,\ \ln F(x) = \ln a + b\ln x.$$

Метод наименьших квадратов применяется для определения коэффициентов линейной функции $\hat{F} = a_1 X + a_0$, аппроксимирующей экспериментальные данные (X_i, Y_i), где $X_i = \ln x_i$, $Y_i = \ln y_i$, $a_0 = \ln a$, $a_1 = b$, $\hat{F} = \ln F$, $i = \overline{1,n}$.

Для определения параметров линейной зависимости $\hat{F} = a_1 X + a_0$ составляется вспомогательная таблица:

	x	y	X = lnx	Y= lny	X^2	XY	F	$(y-F)^2$
	1	2,1	0,0	-0,5	0,0	0	2,13	0,00
	2	6,2	0,7	0,6	0,5	0,407423	4,82	1,90
	3	13,3	1,1	1,7	1,2	1,852699	10,90	5,76
	4	24,6	1,4	2,3	1,9	3,164054	24,65	0,00
	5	117,5	1,6	2,9	2,6	4,704647	55,73	3815,82
Σ	**15**	**163,7**	**4,8**	**7,0**	**6,2**	**10,1**		**3823,493**

Полученные значения сумм подставляются в систему линейных уравнений для определения параметров a_0, a_1 линейной аппроксимации:

$$\left.\begin{array}{cccc} 6,2a_1 & + & 4,8a_0 & = & 10,1 \\ 4,8a_1 & + & 5a_0 & = & 7,0 \end{array}\right\}.$$

Решением системы являются значения $a_1 = 2,14$ и $a_0 = -0,65$. Линеаризованная эмпирическая функция имеет вид: $\hat{F} = 2,14x - 0,65$.

Определение параметров степенной ***эмпирической*** зависимости производится из ранее принятых соотношений:

$$a_0 = \ln a,\ a_1 = b.$$

Откуда $a = e^{a_0} = e^{-0,65} = 0,52$, $b = a_1 = 2,14$. Искомая степенная эмпирическая функция будет иметь вид: $F(x) = ax^b = 0,52x^{2,14}$.

Вопрос выбора наилучшей эмпирической зависимости

Если для таблицы (x_i, y_i) можно указать несколько классов эмпирических функций, то сначала из каждого класса методом наименьших квадратов ищется наилучшая функция, а затем из них выбирается та, которая даёт наименьшее среднеквадратическое отклонение $\rho(\mathbf{y}, \mathbf{F})$.

Решение задачи 5 из Типового расчёта №2

Задание 5. Приближение табличных функций по методу наименьших квадратов

Найти функцию, являющуюся наилучшим приближением к данной табличной функции по методу наименьших квадратов.

x	y
1	2,6
2	6,4
3	19
4	38,9
5	85,8
6	322,3
7	796,9

Для исследования в MS Excel использовать

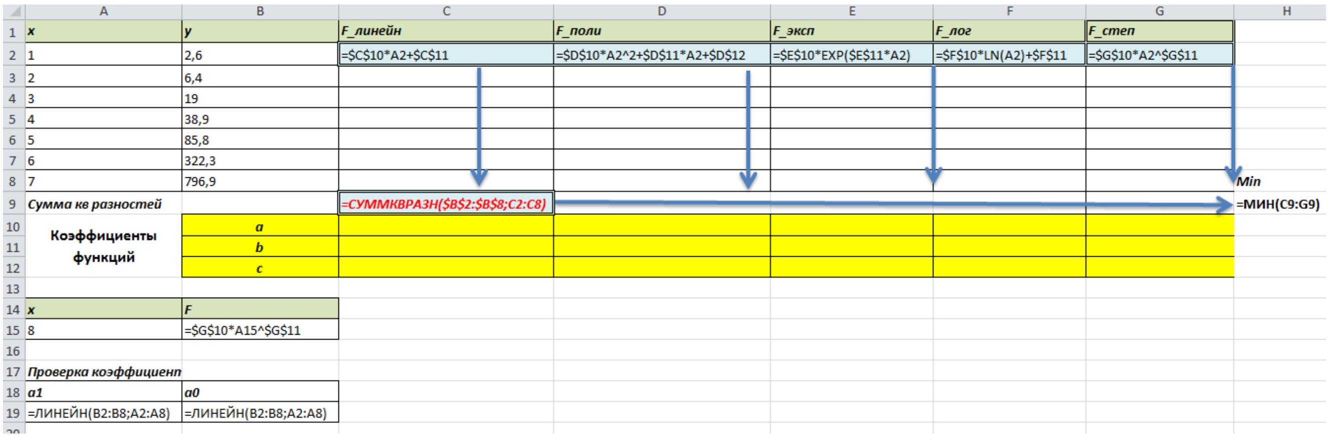
1) линейную функцию;
2) полиномиальную функцию 2-й степени;
3) экспоненциальную функцию;
4) логарифмическую функцию;
5) степенную функцию.

Выбрав наиболее подходящую зависимость, найти прогнозное значение $F(8)$.

Решение

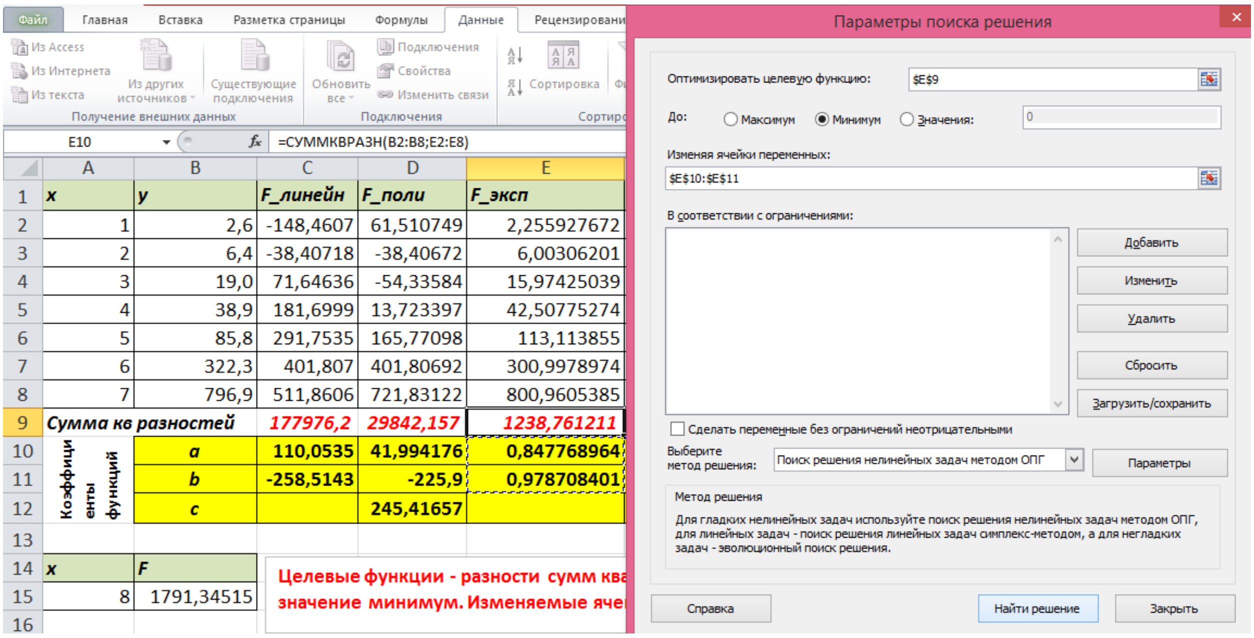
Порядок работы:

1. Ввести двумерный массив данных (x_i, y_i), $i=\overline{1,7}$ в таблицу MS Excel.
2. Сформировать массивы предполагаемых эмпирических зависимостей, используя абсолютные ссылки на параметры функций (пусть ячейки с этими параметрами пока не заполнены).
3. Сформировать суммы квадратов разностей значений эмпирической функции и соответствующих эмпирических данных.

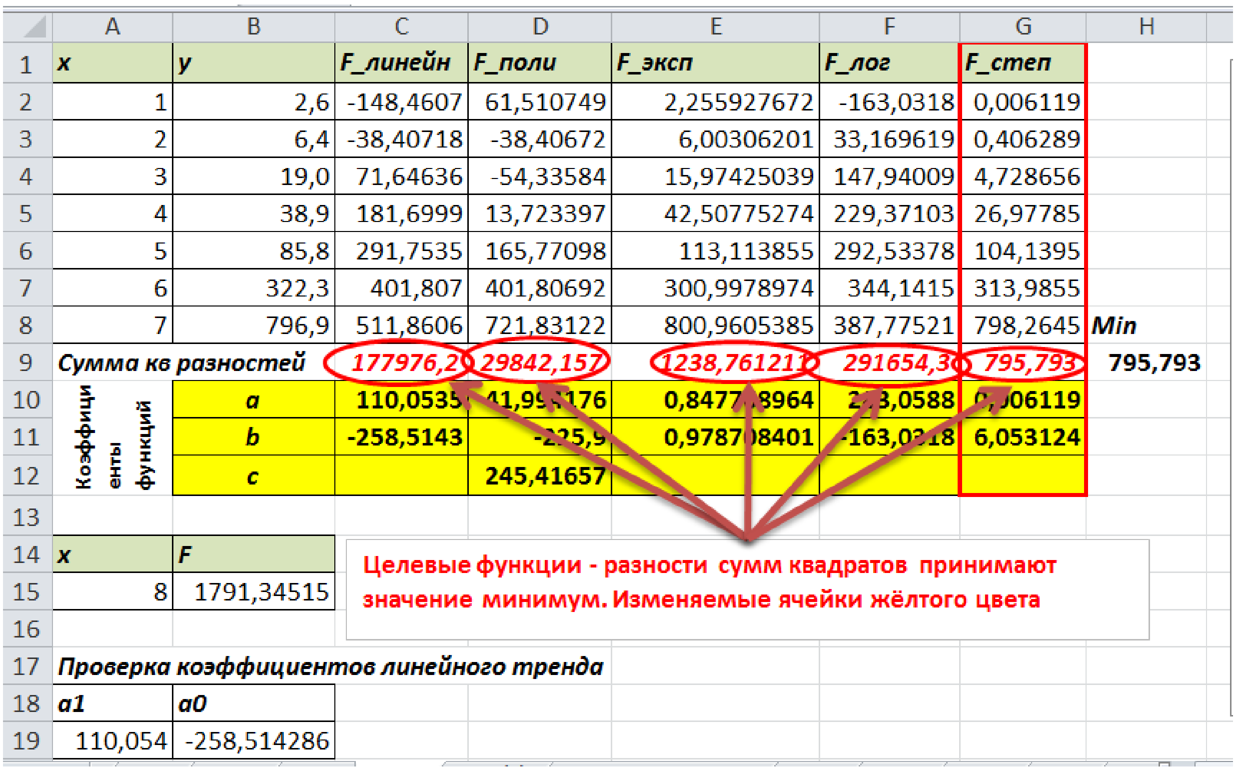


	A	B	C	D	E	F	G	H
1	*x*	*y*	*F_линейн*	*F_поли*	*F_эксп*	*F_лог*	*F_степ*	
2	1	2,6	=$C$10*A2+$C$11	=$D$10*A2^2+$D$11*A2+$D$12	=$E$10*EXP($E$11*A2)	=$F$10*LN(A2)+$F$11	=$G$10*A2^$G$11	
3	2	6,4						
4	3	19						
5	4	38,9						
6	5	85,8						
7	6	322,3						
8	7	796,9						*Min*
9	*Сумма кв разностей*		=СУММКВРАЗН($B$2:$B$8;C2:C8)					=МИН(C9:G9)
10	**Коэффициенты функций**	*a*						
11		*b*						
12		*c*						
13								
14	*x*	*F*						
15	8	=$G$10*A15^$G$11						
16								
17	*Проверка коэффициенп*							
18	*a1*	*a0*						
19	=ЛИНЕЙН(B2:B8;A2:A8)	=ЛИНЕЙН(B2:B8;A2:A8)						

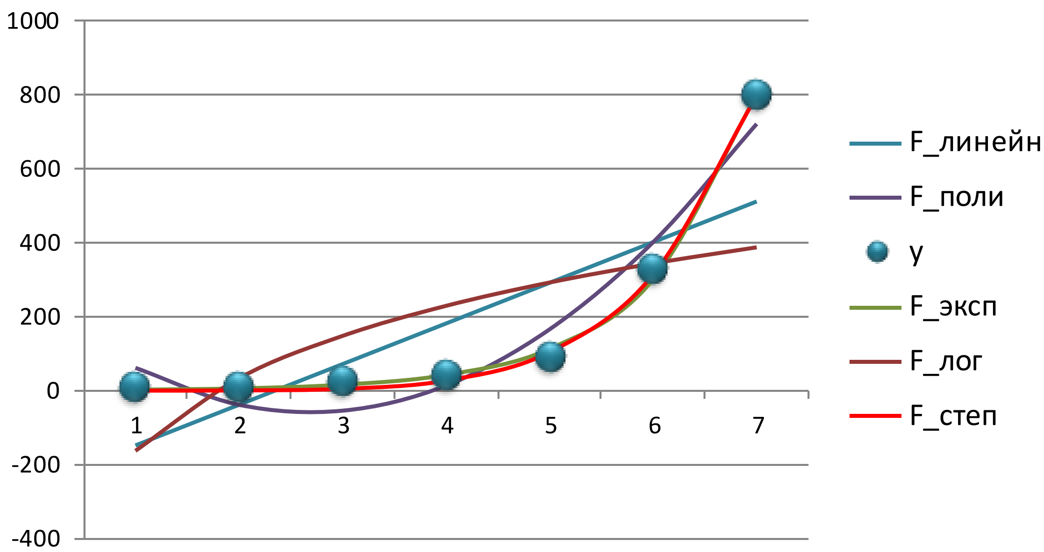
4. С помощью Надстройки MS Excel Поиск решения провести минимизацию целевых функций - сумм квадратов разностей, используя в решении метод ОПГ.



5. Из полученных минимальных значений сумм квадратов выбрать наименьшее значение. Соответствующую функцию можно принимать за искомую эмпирическую зависимость, по которой вычисляется прогнозное значение.



6. Построить графики полученных эмпирических функций.


1000
800
600
400
200
0
-200
-400
1
2
3
4
5
6
7
F_линейн
F_поли
y
F_эксп
F_лог
F_степ