7 АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Первоначально данные исследований представляют в виде таблиц. Однако табличные данные не имеют наглядности и не могут быть использованы в математических моделях, описывающих тот или иной процесс. Указанных недостатков лишены эмпирические формулы, отражающие приближенно с определенным уровнем достоверности зависимость между изучаемыми величинами, параметрами и т. д. Этот процесс называется аппроксимацией.

Прежде всего, необходимо нанести на координатную сетку данные опыта и провести через полученные точки кривую таким образом, чтобы она по возможности близко проходила от всех экспериментальных точек. Количество точек выбирается с учетом следующих правил:

1. Каждый перегиб кривой должен быть описан по меньшей мере тремя опытами, а каждый участок близкий к прямолинейному - двумя опытами. Близко к назначенным пределам следует поставить два «концевых» опыта.

2. Если требуется установить не только общие закономерности, но и возможно более точно численные значения функций, каждый перегиб кривой должен быть обоснован минимум пятью опытами.

При этом необходимо не только учитывать физическую сущность рассматриваемого процесса, но и использовать соображения о том, как должна вести себя кривая в некоторых характерных точках, а именно: при значениях аргумента близких к нулю может ли кривая проходить через начало координат, пересекает ли она координатные оси, имеет ли асимптоты и т. д. Таким образом, первый этап математической обработки данных состоит в выборе эмпирической формулы, графическое изображение которой согласуется в общих чертах с размещением экспериментальных точек на координатной сетке.

Задачей дальнейшей математической обработки является определение числовых значений входящих в формулу параметров. В большинстве случаев зависимость между переменными можно задать множеством эмпирических формул и только глубокое знание физической сущности изучаемого процесса позволяет остановиться на одной из них.

Метод подбора числовых значений, входящих в формулу параметров основан на принципе наименьших квадратов. Сущность метода состоит в том, что из множества возможных эмпирических зависимостей, тип которых уже определен $\mathbf{y=f(x)}$ выбирается та, для которой сумма квадратов отклонений, замеряемых по оси $\mathbf{y}$ является наименьшей.

Впервые этот метод предложил Гаусс. Рассмотрим применение метода наименьших квадратов для получения эмпирических формул, описывающих различные формы зависимостей.

7.1 Линейная функция

Пусть дано $\mathbf{n}$ точек с координатами $(\mathbf{x_i,y_i})$, представляющие собой данные эксперимента – таблица 7.1.

Таблица 7.1 Опытные данные

i	x_i	y_i	x_i^2	$x_i \cdot y_i$
1	x_1	y_1		
2	x_2	y_2		
…	…	…	…	…
n	x_n	y_n		
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$		

Очевидно, эмпирическую формулу нужно искать в виде линейной функции **y=ax+b**, причем коэффициенты **a** и **b** нужно подобрать так, чтобы суммарное отклонение $\sum_{i=1}^{n} d_i^2$ принимало минимальное значение. Задача нахождения минимального значения функции есть классическая задача дифференциального исчисления. Ординате 1-ой точки (x_i,y_i) на прямой **y=ax+b** соответствует точка (x_i,ax_i+b). Поэтому имеем отклонение $d_i=y_i-(ax_i+b)$ - рисунок 7.1.

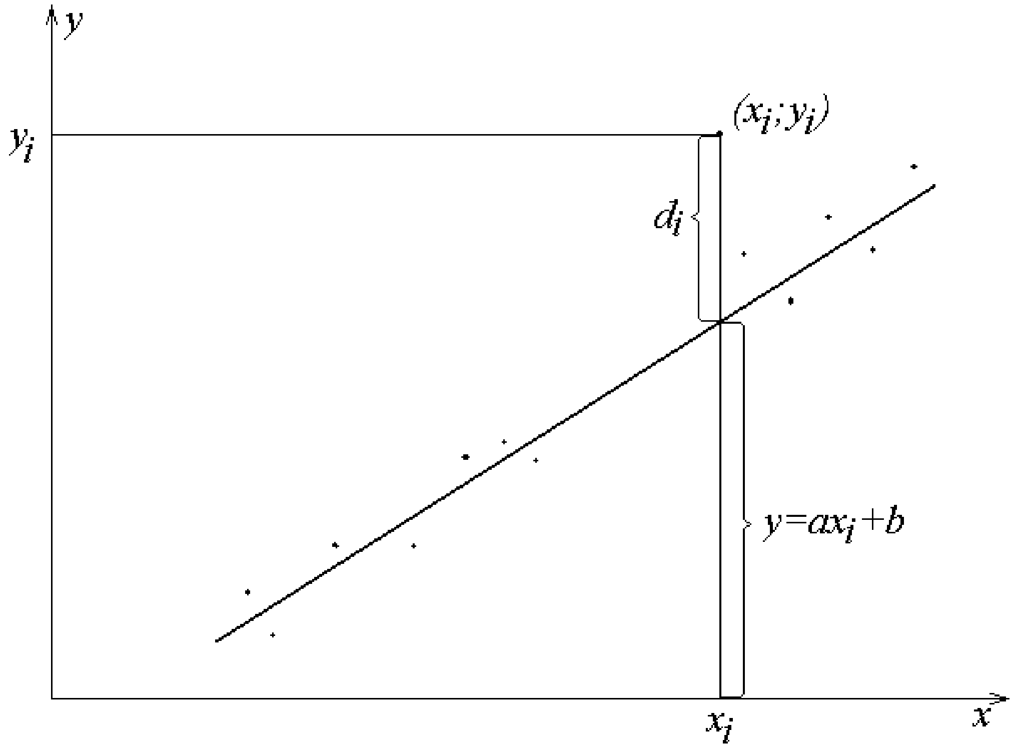

Рисунок 7.1 График линейной функции **y=ax+b**

Сумма квадратов отклонений составит

$$\mathbf{S} = \sum_{i=1}^{n} \mathbf{d_i} = \sum \left[\mathbf{y_i} - (\mathbf{a} \cdot \mathbf{x_i} + \mathbf{b})\right]^2, \qquad (7.1)$$

Для нахождения экстремума функции двух переменных необходимо осуществить следующую последовательность операций.

1. Найти частные производные функции **S** по переменным **a** и **b**, для чего продифференцируем **S** по **a** и **b**.

$$\left.\begin{aligned}\frac{dS}{da}&=\sum_{i=1}^{n}2\left[y_i-(ax_i+b)(-x_i)\right]=-2\sum_{i=1}^{n}x_i\left[y_i-(ax_i+b)\right]\\ \frac{dS}{db}&=\sum_{i=1}^{n}2\left[y_i-(ax_i+b)(-1)\right]=-2\sum_{i=1}^{n}\left[y_i-(ax_i+b)\right]\end{aligned}\right\}\quad(7.2)$$

2. Приравнять частные производные нулю и решая полученную систему уравнений, найти критические точки.

$$\left.\begin{aligned}\frac{dS}{da}&=0,\\ \frac{dS}{db}&=0.\end{aligned}\right\}\quad(7.3)$$

$$\sum_{i=1}^{n}x_i\left[y_i-(ax_i+b)\right]=0,$$

или

$$a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x=\sum x_i\cdot y_i\,.\quad(7.4)$$

$$\sum_{i=1}^{n}\left[y_i-(ax_i+b)\right]=0,$$

или

$$a\sum_{i=1}^{n}x_i+b\cdot n=\sum y_i^2\,.\quad(7.5)$$

Выражения (7.4) и (7.5) представляют собой систему двух линейных уравнений с двумя неизвестными - **a** и **b**.

Для составления нормальной системы удобно пользоваться таблицей 7.1.

Прямая **y=ax+**b при найденных из выражений (7.4) и (7.5) числовых значений **a** и **b** есть прямая наилучшего приближения к точкам (x_i,y_i) по критерию метода наименьших квадратов. Эту прямую называют также линией регрессии **y** на **x**.

Пример. Определить зависимость износа двигателя от наработки – таблица 7.2.

Таблица 7.2 Расчет коэффициентов линейного уравнения регрессии

i	Наработка, тыс. ч x_i	Износ, мм y_i	x_i^2	$x_i \cdot y_i$
1	0,2	0,05	0,04	0,010
2	0,8	0,08	0,64	0,064
3	1,4	0,17	1,96	0,238
4	2,0	0,21	4,0	0,420
5	2,6	0,27	6,76	0,702
6	3,2	0,37	10,24	1,184
7	3,8	0,40	14,44	1,520
8	4,4	0,42	19,36	1,848
Σ	18,4	1,97	57,44	3,988

Нормальная система имеет вид

$$\left.\begin{aligned} 57{,}44\mathbf{a} + 18{,}4\mathbf{b} &= 5{,}986 \\ 18{,}4\mathbf{a} + 8\mathbf{b} &= 1{,}97 \end{aligned}\right\}$$

В результате решения имеем

$\mathbf{a} = 0{,}09623$; $\mathbf{b} = 0{,}02492$.

Эмпирическая формула имеет вид

$\mathbf{y}=0{,}09623\mathbf{x} + 0{,}02492$.

7.2 Квадратичная функция

Эмпирическую формулу нужно искать в виде квадратичной функции

$$\mathbf{y=ax^2+bx+c}. \tag{7.6}$$

Сумма квадратов отклонений имеет вид

$$\mathbf{S}=\sum_{i=1}^{n}\left[y_i-\left(ax_i^2+bx_i+c\right)\right]^2, \tag{7.7}$$

Частные производные по **a**, **b** и **c**.

$$\frac{ds}{da}=\sum_{i=1}^{n}2\left[y_i-\left(ax_i^2+bx_i+c\right)\left(-x^2\right)\right]=$$

$$=-2\sum_{i=1}^{n}x_i^2\left[y_i-\left(ax_i^2+bx_i+c\right)\right], \tag{7.8}$$

$$\frac{ds}{db}=\sum_{i=1}^{n}2\left[y_i-\left(ax_i^2+bx_i+c\right)\left(-x_i\right)\right]=$$

$$=-2\sum_{i=1}^{n}x_i\left[y_i-\left(ax_i^2+bx_i+c\right)\right], \tag{7.9}$$

$$\frac{ds}{dc}=\sum_{i=1}^{n}2\left[y_i-\left(ax_i^2+bx_i+c\right)\left(-1\right)\right]=$$

$$=-2\sum_{i=1}^{n}\left[y_i-\left(ax_i^2+bx_i+c\right)\right], \tag{7.10}$$

Приравнивая частные производные нулю, получаем:

$$a\sum_{i=1}^{n}x_i^4+b\sum_{i=1}^{n}x_i^3+c\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_i^2\cdot y_i, \tag{7.11}$$

$$a\sum_{i=1}^{n}x_i^3+b\sum_{i=1}^{n}x_i^2+c\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_i\cdot y_i, \tag{7.12}$$

$$a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i+cn=\sum_{i=1}^{n}y_i. \tag{7.13}$$

При составлении нормальной системы данные опыта удобно свести в таблицу 7.3.

Таблица 7.3 Расчет коэффициентов квадратичного уравнения регрессии

i	x_i	x_i^2	x_i^3	x_i^4	y_i	$x_i \cdot y_i$	$x_i^2 \cdot y_i$
1	x_1	x_1^2	x_1^3	x_1^4	y_1	$x_1 \cdot y_1$	$x_1^2 \cdot y_1$
2	x_2	x_2^2	x_2^3	x_2^4	y_2	$x_2 \cdot y_2$	$x_2^2 \cdot y_2$
…	…	…	…	…	…	…	…
n	x_n	x_n^2	x_n^3	x_n^4	y_n	$x_n \cdot y_n$	$x_n^2 \cdot y_n$
$\sum_{i=1}^{n}$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i^3$	$\sum_{i=1}^{n} x_i^4$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i \cdot y_i$	$\sum_{i=1}^{n} x_i^2 \cdot y_i$

Решая нормальную систему, найдем значения **a**, **b**, **c** для эмпирической формулы.

Рассмотренный прием можно использовать для аппроксимации опытных данных многочленами более высоких степеней. Коэффициенты многочлена в этом случае также определяют из системы уравнений, получаемых приравниванием нулю частных производных от суммы квадратов отклонений по соответствующим параметрам. Этот метод можно применять в случае, когда аппроксимация неполиноминальная.

7.3 Гиперболическая функция

Уравнение гиперболической функции

$$\mathbf{y} = \mathbf{a} + \frac{\mathbf{b}}{\mathbf{x}}. \tag{7.14}$$

Сумма квадратов отклонений имеет вид

$$\mathbf{S} = \sum_{i=1}^{n}\left(\mathbf{a} + \frac{\mathbf{b}}{\mathbf{x}_i} - \mathbf{y}_i\right)^2. \tag{7.15}$$

Частные производные по **a** и **b** равны

$$\frac{dS}{da}=\sum_{i=1}^{n}2(a+\frac{b}{x_i}-y_i)(1)=2\sum_{i=1}^{n}(a+\frac{b}{x_i}-y_i), \qquad (7.16)$$

$$\frac{dS}{db}=\sum_{i=1}^{n}2(a+\frac{b}{x_i}-y_i)(\frac{1}{x_i})=2\sum_{i=1}^{n}(a+\frac{b}{x_i}-y_i)\frac{1}{x_i}. \qquad (7.17)$$

Приравнивая их нулю, имеем

$$\left.\begin{aligned} & an+b\sum_{i=1}^{n}\frac{1}{x_i}=\sum_{i=1}^{n}y_i, \\ & a\sum_{i=1}^{n}\frac{1}{x_i}+b\sum_{i=1}^{n}\frac{1}{x_i^2}=\sum_{i=1}^{n}\frac{y_i}{x_i}. \end{aligned}\right\} \qquad (7.18)$$

При вычислении коэффициентов данные удобно свести в таблицу 7.4.

Таблица 7.4 Расчет коэффициентов уравнения регрессии

	x_i	$\frac{1}{X_i}$	$\frac{1}{X_i^2}$	y_i	$\frac{y_i}{X_i}$
1	x_1	$\frac{1}{X_1}$	$\frac{1}{X_1^2}$	y_1	$\frac{y_1}{X_1}$
2	x_2	$\frac{1}{X_2}$	$\frac{1}{X_2^2}$	y_2	$\frac{y_2}{X_2}$
…	…	…	…	…	…
k	x_k	$\frac{1}{X_k}$	$\frac{1}{X_k^2}$	y_k	$\frac{y_k}{X_k}$
…	…	…	…	…	…
n	x_n	$\frac{1}{X_n}$	$\frac{1}{X_n^2}$	y_n	$\frac{y_n}{X_n}$
$\sum$	$\sum x_i$	$\sum\frac{1}{X_i}$	$\sum\frac{1}{X_k^2}$	$\sum y_i$	$\sum\frac{y_i}{X_i}$

Числовые значения параметров **a** и **b** найдем путем решения системы уравнений (7.18).

7.4 Показательная функция

Данные опыта могут быть аппроксимированы показательной кривой

$$\mathbf{y = a \cdot b^x}. \tag{7.19}$$

Для получения параметров **a** и **b** прологарифмируем обе части функции. При этом вспомним, что логарифм произведения равен сумме логарифмов сомножителей, а логарифм степени равен произведению показателя степени на логарифм ее основания. Затем следует найти величины **lg(a)** и **lg(b)**.

$$\mathbf{\lg y = \lg a + x \lg b}, \tag{7.20}$$

Сумма квадратов отклонений:

$$\mathbf{S = \sum_{i=1}^{n} (x_i \lg b + \lg a - \lg y_i)^2}. \tag{7.21}$$

Частные производные по **lg(b)** и **lg(a)**

$$\mathbf{\frac{dS}{d(\lg b)} = 2\sum_{i=1}^{n} (x_i \lg b + \lg a - \lg y_i) x_i}, \tag{7.22}$$

$$\mathbf{\frac{dS}{d(\lg a)} = 2\sum_{i=1}^{n} (x_i \lg b + \lg a - \lg y_i)}. \tag{7.23}$$

Приравнивая частные производные нулю, имеем

$$\mathbf{\lg b \sum_{i=1}^{n} x_i^2 + \lg a \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \lg y_i}, \tag{7.24}$$

$$\mathbf{\lg b \sum_{i=1}^{n} x_i + n \lg a = \sum_{i=1}^{n} \lg y_i}. \tag{7.25}$$

Для решения системы удобно составить таблицу 7.5.

Таблица 7.5 Расчет коэффициентов уравнения регрессии

i	x_i	x_i^2	y_i	$\lg y_i$	$x_i \lg y_i$
1	x_1	x_1^2	y_1	$\lg y_1$	$x_1 \lg y_1$
2	x_2	x_2^2	y_2	$\lg y_2$	$x_2 \lg y_2$
…	…	…	…	…	…
k	x_k	x_k^2	y_k	$\lg y_k$	$x_k \lg y_k$
…	…	…	…	…	…
n	x_n	x_n^2	y_n	$\lg y_n$	$x_n \lg y_n$
$\sum$	$\sum\limits_{i=1}^{n} x_i$	$\sum\limits_{i=1}^{n} x_i^2$	$\sum\limits_{i=1}^{n} y_i$	$\sum\limits_{i=1}^{n} \lg y_i$	$\sum\limits_{i=1}^{n} x_i \lg y_i$

Найденные из нормальной системы значения **lg(a)** и **lg(b)** с помощью таблиц антилогарифмов позволяют легко определить **a** и **b**.

Экспериментальные данные могут быть также аппроксимированы показательной кривой вида

$$y = a \cdot e^{bx}. \tag{7.26}$$

Прологарифмируем это выражение по основанию **e**

$$\ln y = \ln a + bx. \tag{7.27}$$

Сумма квадратов отклонений

$$S = \sum_{i=1}^{n} (\ln a + bx_i - \ln y_i)^2. \tag{7.28}$$

Частные производные по основаниям **b** и **ln(a)**

$$\frac{dS}{db} = 2\sum_{i=1}^{n} (\ln a + bx_i - \ln y_i) x_i, \tag{7.29}$$

$$\frac{dS}{d(\ln a)} = 2\sum_{i=1}^{n} (\ln a + bx_i - \ln y_i). \tag{7.30}$$

Приравняв частные производные нулю, имеем

$$\ln a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \ln y_i\,, \tag{7.31}$$

$$n\ln a + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \ln y_i\,. \tag{7.32}$$

Для решения системы уравнений удобно свести данные в таблицу 7.6.

Таблица 7.6 Расчет коэффициентов уравнения регрессии

i	x_i	x_i^2	y_i	$\ln y_i$	$x_i \ln y_i$
1	x_1	x_1^2	y_1	$\ln y_1$	$x_1 \ln y_1$
2	x_2	x_2^2	y_2	$\ln y_2$	$x_2 \ln y_2$
…	…	…	…	…	…
n	x_n	x_n^2	y_n	$\ln y_n$	$x_n \ln y_n$
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} \ln y_i$	$\sum_{i=1}^{n} x_i \ln y_i$

В результате решения нормальной системы можно найти **b** и **ln(a)**, а по таблице антилогарифмов найти **a**. Так как экспериментальные данные получаются с определенной погрешностью, то следует стараться получить достаточно большое их число. При этом случайные ошибки отдельных измерений погашают друг друга и решение, найденное по методу наименьших квадратов, становится более достоверным.