ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Государственное бюджетное образовательное учреждение
высшего профессионального образования
«Тихоокеанский государственный университет»

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Методические указания к самостоятельной
работе для студентов 2–го курса

Издание второе, стереотипное

Хабаровск 2012

УДК 519.281.2
Метод наименьших квадратов: Методические указания к лабораторной работе для студентов 2–го курса строительных специальностей / Сост. О. А. Романчук, Г. И. Целоусова. – Хабаровск : Хабар. политехн. ин – т, 1991. – 16 с.

Методические указания знакомят с одним из наиболее часто применяемых методов подбора эмпирических формул – методом наименьших квадратов. В указания включены вариант заданий, приведены примеры. При выполнении задания рекомендуется использовать настольную вычислительную технику.

Объём выполнения – 2 часа.

Печатается в соответствии решения кафедры высшей математики (С)
и методического совета строительного факультета.

Содержание

Цель работы: *Ознакомиться с одним из методов подбора эмпирических формул — методом наименьших квадратов.*

Задание: *Экспериментально получены* N *— значений величины* Y *при различных значениях величины* X*. Подобрать эмпирическую формулу, наиболее точно описывающую результаты эксперимента, отыскав еч̨ параметры по методу наименьших квадратов. Сделать чертёж, на котором в декартовой системе координат построить экспериментальные точки и график аппроксимирующег функции.*

Варианты заданий

Измерение температуры корпуса работающего агрегата, производимое с интервалом 5 минут, дало следующие результаты.

Вариант 1

X	0	5	10	15	20	25
Y	21	39	51	63	70	90

Вариант 2

X	1	6	11	16	21	26
Y	19	37	49	61	68	90

Значение удельной электропроводимости у стекла в зависимости от температуры X дается таблицей.

Вариант 3

X	14,5	30	64,5	74,5	86,7	94,5
Y	0	0,004	0,018	0,029	0,051	0,073

Вариант 4

X	42	53	61	74	83	90
Y	0,005	0,011	0,023	0,027	0,063	0,125

Количество Y вещества, (%), оставшегося в системе через X минут от начала химической реакции, даётся таблицей.

Вариант 5

X	7	12	17	22	27	32	37
Y	83,7	72,9	63,2	54,7	47,5	41,4	36,3

Вариант 6

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

На химическом производстве в течение пяти рабочих смен получены следующие данные зависимости выхода продукта Y (кг/ ч) от температуры реакции X.

Вариант 7

X	51	32	80	73	64	45	83
Y	52,7	15,2	89,5	94,8	76	39,3	114,8

Вариант 8

X	28	35	40	29	53	58
Y	5,3	20,7	21,7	9,2	55,4	64,3

Для исследования зависимости давления Y насыщенного пара (н/см) от удельного объёма X (m^3/кг) составлена таблица опытных данных.

Вариант 9

X	3,33	1,63	0,87	0,42	0,27	0,17
Y	0,48	1,03	2,02	4,25	7,16	11,5

Вариант 10

X	7,4	10,0	12,8	20,0	26,0	32,0
Y	16,3	15,2	14,5	13,5	13,1	12,8

Температура объекта Y зависит от температуры окружающей среды X. Для исследования зависимости Y от X составлена таблица.

Вариант 11

X	-12	29	0	4	6	8
Y	-3,75	0,92	1,87	9,5	5,21	6,2

Вариант 12

X	-20	-8	0	11	18	22
Y	9,5	10,5	13,2	23,1	57,6	168

Для исследования зависимости урожайности Y (ц/га) от количества внесённых удобрений X (т/га) проведены наблюдения над контрольными участками и составлена таблица.

Вариант 13

X	6	7	8	9	10	11
Y	28	30	31	32	32	33,5

Вариант 14

X	0,3	1,0	1,5	2,2	3,6	4,5
Y	5	10	13	16	17	18

Данные исследования количественных признаков: Y – среднемесячная выработка продукции на одного рабочего (тыс.р), X – стоимость основных производственных фондов (млн.р) собраны в таблицу.

Вариант 15

X	9,9	10,1	10,2	10,4	10,5	10,6
Y	0,8	0,95	1,05	1,2	1,25	1,25

Вариант 16

X	2,9	3,8	11,9	30,1	66,5	128
Y	0,3	1,1	1,9	3,2	4,1	5,2

Температура Y зависит от процентного содержания X компоненты A в теплоносителе. Для исследования зависимости Y от X составлена таблица.

Вариант 17

X	2	4	6	8	10	11
Y	9,5	7,8	5,2	3,1	-4	0,2

Вариант 18

X	0,5	1	2	4	8	12
Y	160	120	94	75	62	56

Данные исследования количественных признаков: X – длинна ампулы (мм) и Y – объём ампулы (cm^3), собраны в таблицу.

Вариант 19

X	18,8	19	19,5	20	21	23
Y	0,9	1,1	1	1	1,1	1,2

Вариант 20

X	39	40	45	51	55	56
Y	0,6	0,8	1	1,3	1,6	1,8

Данные исследования продолжительности решения систем линейных уравнений Y в минутах в зависимости от порядка системы X занесены в таблицу.

Вариант 21

X	2	4	6	8	10	12
Y	12	17	210	445	800	1010

Вариант 22

X	3	5	7	9	11	13
Y	13	130	315	600	950	1320

Для изучения распределения температуры в теплоизолированном тонком стержне дана таблица измеренных температур Y в соответствующих точках X стержня.

Вариант 23

X	0	2	6	8	10	14
Y	32	29,2	23,3	19,9	17,2	11,3

Вариант 24

X	0	5	10	15	20	22
Y	9,7	12.2	7,8	6,6	4,9	2,5

Данные исследования количественных признаков: X – стрела кривизны рельса (см), Y – количество дефектов рельса (см, на 25 – метровый рельс) собраны в таблицу.

Вариант 25

X	7	7,5	8	8,5	9	9,5
Y	0	5	10	15	20	22

Вариант 26

X	2,1	3,9	5,1	6,3	7,0	9,5
Y	0	5	10	15	20	25

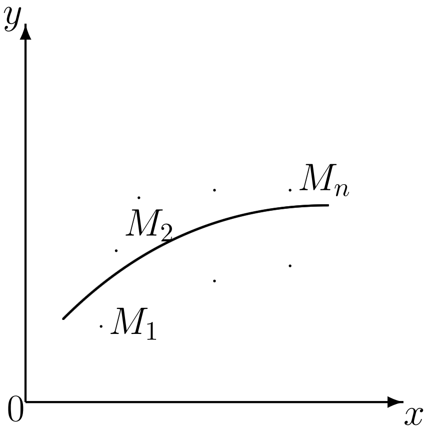
Постановка задачи

При эмпирическом (экспериментальном) изучении функциональной зависимости одной величины Y от другой величины X производят ряд измерений величины Y при различных значениях величины X. Результаты могут быть представлены в виде табл.1.

Таблица 1

X	X_1	X_2	$\dots$	X_{n-1}	X_n
Y	Y_1	Y_2	$\dots$	Y_{n-1}	Y_n

Возникает практически важная задача: установить формулу, дающую аналитическое выражение функциональной зависимости между исследуемыми величинами. Формулы, полученные по экспериментальным данным, называют **эмпирическими формулами**. Особенность задачи состоит в том, что наличие случайных ошибок измерения делает невозможным подбор такой формулы, которая точно бы описывала все опытные значения. Поэтому при подборе эмпирической формулы руководствуются требованиями: значения, вычисленные по эмпирической формуле, должны мало отличаться от опытных данных; эмпирическая формула должна быть более простой. Геометрическая задача построения эмпирической формулы состоит в проведении кривой, "возможно ближе"примыкающей к системе точек $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $\dots$, $M_n(x_n, y_n)$ *(рис.1)*. При этом необходимо условиться о математическом смысле понятия "близости"кривой и системе точек.



Построение эмпирической формулы производится по этапам

1) выбор формулы определенного типа (линейная, квадратичная, гиперболическая и так далее);
2) определение параметров формулы, то есть значений различных постоянных величин, входящих в формулу;
3) оценка результатов аппроксимации (приближения) экспериментальных данных эмпирической формулой.

Выбор типа эмпирической формулы

Нельзя указать общего метода для выбора наилучшего типа формулы, соответствующей опытным данным. Если неизвестен характер зависимости между величинами X и Y, то вид эмпирической зависимости является произвольным. Предпочтение отдаётся простым формулам, обладающим хорошей точностью.

Существует ряд приёмов, облегчающих выбор типа эмпирической формулы. Самым простым, но не самым надёжным является следующий. Опытные данные наносят на график, после чего на глаз, от руки проводят вблизи полученных точек наиболее правдоподобную кривую. При проведении кривой кроме экспериментальн точек можно использовать соображения общего характера: как должна вести себя кривая при значениях аргумента, весьма близких к нулю, при больших значениях аргумента, проходит ли кривая через начало координат, пересекает ли координатные оси и тому подобное. По виду построенной кривой выбирают тип эмпирической формулы.

Отыскание параметров эмпирических формул методом наименьших квадратов

Обозначим выбранную функциональную зависимость через

$$Y = f(x, a_1, a_2, \ldots, a_k) \tag{1}$$

где $a_1, a_2, \ldots, a_k$ – параметры, подлежащие определению. Согласно методу наименьших квадратов наилучшими значениями параметров $a_1, a_2, \ldots, a_k$ считаются те, для которых сумма квадратов отклонений измеренных значений Y_i от расчетных $f(x, a_1, a_2, \ldots, a$

то есть величина

$$S = \sum_{i=1}^{n} [f(x_i, a_1, a_2, \ldots, a_k) - Y_i]^2$$

будет минимальной.

Используем необходимые условия экстремума функции нескольких переменных. Отыскание тех значений параметров $a_1, a_2, \ldots, a_k$, которые доставляют наименьшие значения функции

$$S = S(a_1, a_2, \ldots, a_k)$$

сводится к решению системы уравнений

$$\frac{\partial s}{\partial a} = 0; \frac{\partial s}{\partial a_2} = 0; \ldots; \frac{\partial s}{\partial a_k} = 0$$

или в развёрнутом виде:

$$\left.\begin{aligned} \sum_{i=1}^{n} 2[f(x_i, a_1, a_2, \ldots, a_k) - Y_i]\frac{\partial f}{\partial a_1} = 0 \\ \sum_{i=1}^{n} 2[f(x_i, a_1, a_2, \ldots, a_k) - Y_i]\frac{\partial f}{\partial a_2} = 0 \\ \sum_{i=1}^{n} 2[f(x_i, a_1, a_2, \ldots, a_k) - Y_i]\frac{\partial f}{\partial a_k} = 0 \end{aligned}\right\} \qquad (2)$$

Если система имеет единственное решение, то оно и будет искомым. Если в эмпирическом формулу (1) параметры входят линейно, то система (2) будет линейной. Так как решение линейной системы значительно проще, чем нелинейной, следует по возможности преобразовывать выбираемую эмпиричную формулу к такому виду, чтобы определяемые параметры входили в неё линейно.

Оценка результатов аппроксимации

Оценка результатов аппроксимации может быть осуществлена путём сопоставления экспериментальных значений y_x с расчётными y_p, то есть вычисленными по эмпирической формуле. Результаты эксперимента более точно описывает та функция, для которой сумма квадратов отклонений экспериментальных значений от расчётных

$$S = \sum_{p=1}^{n} (y_x - y_p)^2,$$

будет наименьшей.

Метод наименьших квадратов обладает тем преимуществом, что если сумма S квадратов отклонений мала, то сами эти отклонения также малы по абсолютной величине. Недостатком этого метода является громоздкость вычислений. Поэтому к нему пребегают обычно, когда нужно получить весьма точные значения параметров.

Порядок выполнения работы

1. Вариант задания выдаётся для трёх – четырёх студентов.
2. В декартовой системе координат X_0Y на плоскости построить экспериментальные точки $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $\ldots$, $M_n(x_n, y_n)$. Соединить построенные точки кривой.
3. По виду полученной кривой подобрать 3–4 типа функциональных зависимостей. Рекомендуется выбирать функции с двумя линейно входящими параметрами. Например:

$$\begin{array}{l} y = ax + b; \quad y = a + b\ln x; \quad y = a + \dfrac{b}{x}; \\ y = ax^2 + b; \quad y = ax^2 + bx; \quad y = a + be^x \end{array} \quad \text{и тому подобное.}$$

4. Для каждых из подобранных функциональных зависимостей составить сумму квадратов отклонений экспериментальных значений y_x от расчётных $f(x_i, a, b)$ то есть величину

$$S(a, b) = \sum_{i=1}^{n} [f(x_i, a, b) - y_i]^2.$$

Найти частные производные $\dfrac{\partial S}{\partial a}$, $\dfrac{\partial S}{\partial b}$, составить систему уравнений $\dfrac{\partial S}{\partial a} = 0$, $\dfrac{\partial S}{\partial b} = 0$. Коэффициенты этой системы удобно вычислять с помощью вспомогательной расчётной таблицы. Решив систему, найти параметры a и b.
5. Оценить результаты аппроксимации.
Для каждого из полученных эмпирических формул вычислить сумму $S = \sum\limits_{p=1}^{n} (y_p - y_x)^2$. Сравнивая эти суммы, выбрать эмпирическую формулу, которая более точно описывает результаты эксперимента.
Замечание. Процесс вычисления расчётных значений y_p можно автоматизировать, используя программируемый микрокалькулятор. Для промежуточных вычислений

удобно составить вспомогательную расчётную таблицу 3.

6. На чертеже построить график полученной аппроксимирующей функции и сравнить его с расположением экспериментльных точек.

Пример вычисления параметров эмпирической формулы

Задание: Найти параметры a и b по методу наименьших квадратов эмпирической формулы $y = a + \dfrac{b}{x}$ по экспериментальным данным

x	2	4	6	12
y	8	5,25	3,5	3,25

Составим сумму квадратов отклонений экспериментальных значений y_i от расчётных $a + \dfrac{b}{x_i}$.

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

Найдем четыре производные:

$$\frac{\partial S}{\partial a} = 2\sum_{i=1}^{4}(a + \frac{b}{x_i} - y_i),$$

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{4}(a + \frac{b}{x_i} - y_i)\frac{1}{x_i}$$

Приравнивая их к нулю, получаем систему

$$\left.\begin{aligned} \sum_{i=1}^{4}(a + \frac{b}{x_i} - y_i) = 0 \\ \sum_{i=1}^{4}(a + \frac{b}{x_i} - y_i)\frac{1}{x_i} = 0 \end{aligned}\right\}$$

Преобразуем систему к виду:

$$\left.\begin{aligned} 4a + b\sum_{i=1}^{4}\frac{1}{x_i} = \sum_{i=1}^{4} y_i \\ a\sum_{i=1}^{4}\frac{1}{x_i} + b\sum_{i=1}^{4}\frac{1}{x_i^2} = \sum_{i=1}^{4}\frac{y_i}{x_i} \end{aligned}\right\} \tag{3}$$

Для вычисления коэффициентов системы составим расчётную таблицу 2.

Таблица 2

i	x_i	$1/x_i$	$1/x_i^2$	y_i	y_i/x_i
1	2	0,500	0,250	8,00	4,00
2	4	0,250	0,062	5,25	1,312
3	6	0,167	0,028	3,50	0,583
4	12	0,083	0,007	3,25	0,271
$\sum$		1,000	0,347	20,00	6,166

Система уравнений (3) примет вид

$$\left.\begin{aligned} 4a + b = 20 \\ a + 0,347b = 6,166 \end{aligned}\right\}$$

Решая её, получим: $a = 1,995 \approx 2$; $b = 12,021 \approx 12$.
Таким образом, искомая зависимость имеет вид: $y = 2 + \dfrac{12}{x}$

Пример оценки результатов аппроксимации

Задание: Оценить результаты аппроксимации экспериментальных данных

x	20	28	40	52	60	74
y	11	10	9	7	7	6

эмпирическими формулами:
$y_1 = 12,68 - 0,095x$,
$y_2 = 23,02 - 3,94\ln x$.

Вычисления удобно вести, фиксируя промежуточные результаты в таблице 3.

Таблица 3.

i	x	y_x	y_{1p}	$(y_{1p} - y_x)^2$	y_{2p}	$(y_{2p} - y_x)^2$
1	20	11	10,78	0,0484	11,22	0,0471
2	28	10	10,02	0,0004	9,89	0,0119
3	40	9	8,88	0,0144	8,49	0,2642
4	52	7	7,74	0,548	7,45	0,2040
5	60	7	6,89	0,0004	6,89	0,0123
6	74	6	5,65	0,1225	6,06	0,0038
$\sum$				0,7341		0,5433

Видим, что $\sum\limits_{i=1}^{6}(y_{1p} - y_x)^2 \approx 0,73$, $\sum\limits_{i=1}^{6}(y_{2p} - y_x)^2 \approx 0,54$. Следовательно, логарифмическая функция более точно описывает результаты эксперимента.

Замечание. Вычисления y_{1p} и y_{2p} велись на программируемом микрокалькуляторе по программе:
$12,68 \uparrow 0,095$ ПО $x-$ с/п
$23,02 \uparrow 3,94$ ПО $F \ln x-$с/п

Вопросы для защиты лабораторной работы

1. Какие формулы называются эмпирическими формулами?
2. Каким требованиям должна удовлетворять эмпирическая формула?
3. В чем заключается задача построения эмпирической формулы геометрически?
4. По каким этапам производится построение эмпирической формулы?
5. В чем суть метода наименьших квадратов?
6. Как оцениваются результаты аппроксимации?
7. Необходимые условия экстремума функции нескольких переменных?
8. Методы решения системы линейных уравнений?
9. Каковы достоинства и недостатки метода наименьших квадратов?

Список литературы

[1] Румшиский Л. З. Математическая обработка результатов эксперимента. – М. : Наука, 1971. – 192 с.

[2] Пискунов Н. С. Дифференциальное и интегральное исчисления.
Т. 1. – М. : Наука, 1978. – 456 с.

[3] Чернов М. В. Пакет программ решения строительно – технологических задач на микроЭВМ. – Хабаровск : ХПИ, 1988. – 32 с.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Методические указания к лабораторной работе для студентов 2 – го курса строительных специальностей.

Ольга Александровна Романчук
Галина Ивановна Целоусова

Редактор Т. Ф. Шейкина
Технический редактор Л. А. Ушакова
Н/К

Подписано в печать 15. 06. 90. Формат 60x84 1/16.
Бумага писчая. Офсетная печать. Усл. печ. л. 0. 9.
Уч. – изд.л 0. 8. Тираж 500 экз. Заказ 107. Бесплатно.

Редакционно – издательский отдел Хабаровского политехнического института.
680035, Хабаровск, ул. Тихоокеанская, 136.

Фотоофсетная лаборатория Хабаровского политехнического института. 680035, Хабаровск, ул. Тихоокеанская, 136.