

БОЧАРОВ Б.П.**ВОЕВОДИНА М.Ю.***г. Харьков, ХНАГХ*

Метод наименьших квадратов в нормировании библиотечных процессов

В 22 - ом (2007 г.) выпуске журнала была изложена методика нормирования трудовых процессов в библиотеке, а т.ч. использование циклового хронометража. Он применяется в тех случаях, если элементы операции имеют столь малую продолжительность, что отследить их практически невозможно. На наш взгляд, здесь лучше воспользоваться общепринятым для такого рода задач методом наименьших квадратов.

Считается, что известен общий вид функции:

$$Y = F(A_1, \dots, A_m, x_1, \dots, x_m).$$

Производится k наблюдений, в которых измеряются значения функции Y и переменных A_1, \dots, A_m .

Обозначим измерения значения функции как Y_1, \dots, Y_k , а измерения значения переменной A_i как a_{i1}, \dots, a_{ik} .

Необходимо определить значение переменных x_1, \dots, x_m таким образом, чтобы эти значения «наилучшим образом» соответствовали выбранной модели (функции F) и результатам наблюдений.

Считается, что «оптимальное» решение задачи достигается в точке минимума функции

$$L = \sum_{i=1}^k (y_i - F(a_{1i}, \dots, a_{mi}, x_1, \dots, x_m))^2$$

Если минимум функции L существует, то он определяется с помощью решения системы уравнений (эти уравнения называются нормальными уравнениями)

$$\frac{\partial L}{\partial x_j} = 0, j = 1, \dots, m.$$

Рассмотрим простой пример. Предположим, мы хотим определить время обслуживания читателя в некотором пункте выдачи. Это время выводится как сумма времени, затрачиваемого на выполнение двух операций - приёма и выдачи литературы.

Всего проведено три наблюдения и в каждом наблюдении измерены:

- продолжительность операции (время обслуживания),
- количество принятых книг,
- количество выданных книг.

Численные данные приведены в следующей таблице.

Номер измерения	Приём литературы x_1	Выдача литературы x_2	Продолжительность операции
1	$a_{11} = 2$	$a_{21} = 3$	$y_1 = 140$
2	$a_{12} = 3$	$a_{22} = 1$	$y_2 = 105$
3	$a_{13} = 2$	$a_{23} = 2$	$y_3 = 110$

Необходимо определить:

- время выдачи и приёма литературы;
- время приёма литературы.

В соответствии с методом наименьших квадратов, эти величины необходимо определять из соотношения:

$$L = \sum_{j=1}^3 (y_j - \sum_{i=1}^2 a_{ij} x_i)^2 = (y_1 - a_{11}x_1 - a_{21}x_2)^2 + (y_2 - a_{12}x_1 - a_{22}x_2)^2 + (y_3 - a_{13}x_1 - a_{23}x_2)^2 \rightarrow \min$$

Система нормальных уравнений имеет вид:

$$\begin{cases} \frac{\partial L}{\partial x_1} = (\sum_{j=1}^3 a_{1j} y_j) - (\sum_{j=1}^3 a_{1j}^2) x_1 - (\sum_{j=1}^3 a_{1j} a_{2j}) x_2 = 0 \\ \frac{\partial L}{\partial x_2} = (\sum_{j=1}^3 a_{2j} y_j) - (\sum_{j=1}^3 a_{2j}^2) x_2 - (\sum_{j=1}^3 a_{1j} a_{2j}) x_1 = 0 \end{cases}$$

или:

$$\begin{cases} p_{11}x_1 + p_{12}x_2 = q_1 \\ p_{21}x_1 + p_{22}x_2 = q_2 \end{cases}$$

где:

$$p_{11} = a_{11}^2 + a_{12}^2 + a_{13}^2 = 2^2 + 3^2 + 2^2 = 17$$

$$p_{22} = a_{21}^2 + a_{22}^2 + a_{23}^2 = 3^2 + 1^2 + 2^2 = 14$$

$$p_{12} = p_{21} = a_{11}a_{21} + a_{12}a_{22} + a_{13}a_{23} = 2 \times 3 + 3 \times 1 + 2 \times 2 = 13$$

$$q_1 = a_{11}y_1 + a_{12}y_2 + a_{13}y_3 = 2 \times 140 + 3 \times 105 + 2 \times 110 = 815$$

$$q_2 = a_{21}y_1 + a_{22}y_2 + a_{23}y_3 = 3 \times 140 + 1 \times 105 + 2 \times 110 = 745$$

Решение системы уравнений:

$$\begin{cases} x_1 = 25 \\ x_2 = 30 \end{cases}$$

В реальной жизни такие задачи уже давно не решают вручную. Мы заставим поработать Microsoft Excel. Все примеры, которые рассматриваются в статье, находятся в приложении (таблица Excel) на сайте журнала <http://lib-journal.ru>.

Предполагается, что читатель имеет элементарные навыки работы с программой Microsoft Excel.

На листе Ex1 приведены исходные данные для определения времени приёма и выдачи литературы из рассмотренного выше примера.

	A	B	C	D	E
1	1	1			
2					= СУММ(E3:E6)
3	2	3	140	\$A\$1*A3 + \$B\$1*B3	(C3 - D3)*(C3 - D3)
4	3	1	105	\$A\$1*A4 + \$B\$1*B4	(C4 - D4)*(C4 - D4)
5	2	2	110	\$A\$1*A5 + \$B\$1*B5	(C5 - D5)*(C5 - D5)

В ячейках A1 и B1 будет записано решение задачи, в начале их можно заполнять любыми значениями.

В ячейках A3:C5 находятся результаты наблюдений. Ячейки D3:E5 предназначены для вспомогательных вычислений.

В ячейке E2 находится значение функции, которую мы собираемся минимизировать.

Из меню «Сервис» выбираем пункт «Поиск решения...».

В диалоговом окне устанавливаем значения полей «Установить целевую ячейку» и «Изменяя ячейки»:

Далее нужно нажать кнопку «Параметры» и убедиться, что в окне «Параметры поиска решений» флажок «Линейная модель» не установлен.

Нажимаем кнопку «Выполнить» и получаем решение:

	A	B	C	D	E
1	25	30			
2					1.2E - 10
3	2	3	140	140	5.2E - 11
4	3	1	105	105	3.6E - 11
5	2	2	110	110	3.4E - 11

Числа в примере подобраны «точно», т.е. ошибки измерений не предусмотрены. Внесем некоторые изменения в исходные данные:

	A	B	C	D	E
1	25	30			
2					2.99999
3	2	3	141	140	1.00001
4	3	1	104	105	0.99999
5	2	2	109	110	0.99999

Ещё раз применяем процедуру «Поиск решения».

Нетрудно увидеть, что оптимальными становятся другие значения:

	A	B	C	D	E
1	24.39	30.57			
2					1.17391
3	2	3	141	140	0.27221
4	3	1	104	105	0.06806
5	2	2	109	110	0.83364

Применение Excel избавило нас от необходимости решать систему нормальных уравнений (кстати, он тоже не решает эту систему, минимум определяется другими методами) и, вообще, освободило от использования (и, к сожалению, понимания) каких-либо математических методов. Решение задачи свелось к правильному нажатию набора кнопок и получению результата.

Тут, однако, следует помнить, что Excel в качестве вычислительного инструмента способен решить любую, как корректную, так и некорректно поставленную задачу. Предположим, что нам необходимо по измерениям времени, затраченного на выдачу книг читателю, определить, какая часть этого времени приходится на действия библиотекаря и каково время, за которое читатель заполняет книжный формуляр. Пусть проведенные наблюдения дали следующие результаты.

Номер измерения	Количество книг		Продолжительность операции
	выданных (x1)	полученных (x2)	
1	a11 = 1	a21 = 1	y1 = 10
2	a12 = 2	a22 = 2	y2 = 20
3	a13 = 3	a23 = 3	y3 = 30

Понятно, что количество выданных библиотекарем и полученных читателем книг будет одинаково. В первом наблюдении оно составляет одну книгу, во втором - две, в третьем - три (цифры выбраны условно).

Процедура решения задачи в Excel ничем не отличается от предыдущего примера. Изменим цифровые значения и запустим поиск решения (лист Ex2). В результате получим:

	A	B	C	D	E
1	5	5			
2					0
3	1	1	10	10	0
4	2	2	20	20	0
5	3	3	30	30	0

На первый взгляд, вполне нормальное решение, целевая функция достигает абсолютного минимума. Казалось бы, всё в порядке. Однако, если изменить начальные значения изменяемых переменных (**запишем в ячейку A1 число 6**) и запустить поиск решения, то и в этом случае получаем другой, но вполне удовлетворительный ответ.

	A	B	C	D	E
1	7	3			
2					0
3	1	1	10	10	0
4	2	2	20	20	0
5	3	3	30	30	0

Два нормальных решения – это уже слишком много, нужно выяснить, где произошла ошибка. Опыт показывает, что первая и наиболее вероятная причина ошибки – нарушения в работа Microsoft Exce. Но в данном случае программа не виновата.

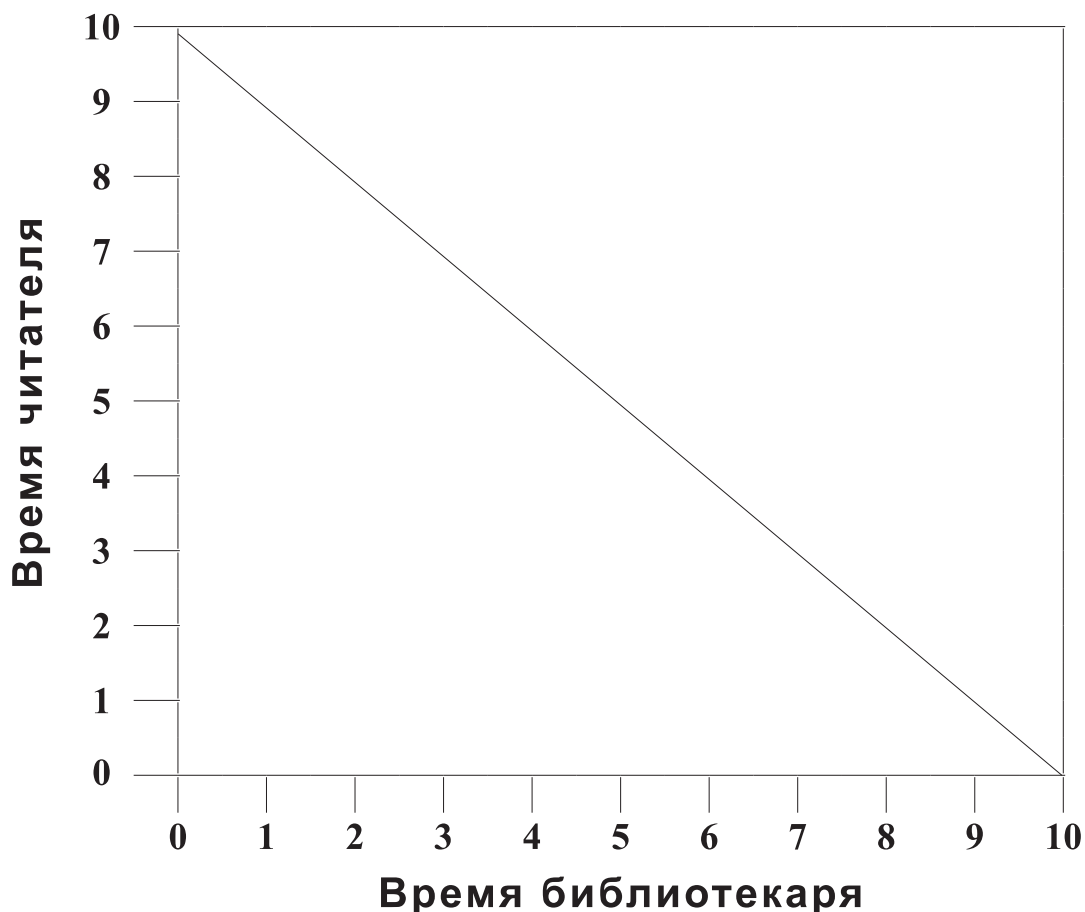
Запишем для наших данных систему нормальных уравнений метода наименьших квадратов.

$$\begin{cases} 14x_1 + 14x_2 = 140 \\ 14x_1 + 14x_2 = 140 \end{cases}$$

Очевидно, что такая система не имеет единственного решения. Решением может быть любая точка, лежащая на прямой

$$x_1 = 10 - x_2.$$

Графическое представления множества решений системы представлено на рисунке:



Более внимательное рассмотрение исходных данных показывает, что причина отсутствия единственного решения – равенство величин a_{1j} и a_{2j} для всех наблюдений. Действительно, библиотекарь не может обработать, к примеру, пять книг, а читателю дать расписаться за четыре.

Поэтому задача определения отдельно времени библиотекаря и времени читателя в принципе не имеет смысла. Можно определить только суммарное время на выдачу одной книги, с чем Excel успешно справился.

Все получаемые решения удовлетворяют условию:

$$X_1 + X_2 = 10$$

До сих пор нами рассматривались искусственные примеры. Перейдем теперь к практическому использованию метода наименьших квадратов. На листе Sol представлен соответствующий шаблон. Он предусматривает ввод результатов наблюдения при следующих ограничениях:

- до 200 наблюдений;
- не более 10 переменных.

Решение получается в ячейках C2:L2/

Значение целевой функции – в ячейке A1.

	A	B	C	D	E	F	G	H	I	J	K	L
1	1017328.177		1	2	3	4	5	6	7	8	9	10
2	Ras	Sum	9,9	117.7	13.1	40.3	35.2	11.4	47.3	89.1	28.5	86.1
3	14542	2772.59	59.6	941.23	39.2	322	35.2	11.4	426	623	142	172
4	13814	3129.53	79.5	705.92	105	161	282	79.6	142	802	171	602
5	182.59	2773.49	59.6	1176.5	91.5	121	282	45.5	94.6	445	199	258
6	686.52	2275.8	39.8	117.65	131	161	106	22.8	94.6	802	199	602
7	26.256	2517.12	99.4	352.96	91.5	40.3	247	79.6	378	713	85.5	430

Коэффициенты (значения a_{ij}) заносятся в ячейки N3:W202.

Общее время на выполнение операции для каждого измерения (y_j) заносится в ячейки N3:N202.

M	N	O	P	Q	R	S	T	U	V	W	X
											№
Sum	1	2	3	4	5	6	7	8	9	10	п/п
2652	6	8	3	8	1	1	9	7	5	2	1
3012	8	6	8	4	8	7	3	9	6	7	2
2787	6	10	7	3	8	4	2	5	7	3	3
2302	4	1	10	4	3	2	2	9	7	7	4
2512	10	3	7	1	7	7	8	8	3	5	5

В диалоговом окне «Поиск решения» необходимо заполнить поле ограничения, чтобы лишить Excel возможности получать отрицательные значения: $SC\$2:SL\$2 \geq 0$.

Данные для расчетов были сгенерированы с помощью датчика случайных чисел (лист Gen). Предоставляем читателю возможность самостоятельно разобраться в деталях.

Результаты расчетов представлены в таблице, в строке № 1 – определённые Excel значения, а в строке № 2 – «теоретические».

1	9.9	117.7	13.1	40.3	35.2	11.4	47.3	89.1	28.5	86.1
2	10	120	15	40	35	11	45	89	26	87

Метод наименьших квадратов тесно связан с методом наибольшего правдоподобия статистической оценки параметров по результатам наблюдений. Оценки, получаемые с помощью метода наибольшего правдоподобия, являются эффективными (т.е. ошибка такой оценки минимальна из всех возможных оценок). Приведем вкратце простой пример.

Пусть проводятся N измерений величины (параметра) x . Измеренные значения обозначим как y_1, \dots, y_N .

Необходимо определить эффективную оценку параметра x .

В данном случае $F(x) = x$ и оценку следует искать из соотношения:

$$L = \sum_{i=1}^N (y_i - x)^2 \rightarrow \min.$$

Нормальное уравнение (в данном случае всего одно) имеет вид:

$$\frac{\partial L}{\partial x} = -2 \sum_{i=1}^N (y_i - x) = 0.$$

Его решение:

$$x = \frac{1}{N} \sum_{i=1}^N y_i$$

Таким образом, эффективная оценка параметра по результатам наблюдений – среднее (арифметическое) значение измерений.

В завершение хотелось бы сказать несколько слов о том, как практически использовать данный материал библиотекам. В наше время прежде всего следует отказаться от ручных расчётов при нормировании труда и воспользоваться программными средствами.

Во - вторых, и это самое главное, важно помнить, что по некоторым операциям нормы могут быть получены не путём прямых замеров, а опосредствованно за счёт использования определённых, в т.ч. описанных выше, математических методов. Тем самым возможности нормирования существенно расширяются. Очевидно, что в этом случае библиотека должна в первую очередь выполнить постановку задачи, суметь грамотно сформулировать её.

Другое дело, что значительная часть библиотечных работников, скорее всего, не владеет в достаточной мере математическим аппаратом. Вот почему для проверки того, насколько корректно поставлена задача, имеет ли она решение, библиотека нуждается в помощи специалистов, имеющих на кафедрах вуза - математики, вычислительной техники и т.д.