Усовершенствованная версия градуировочного графика

Для версий:
Microsoft Office 2013
LibreOffice 6.4

Данную заметку можно рассматривать как продолжение пункта «Построение градуировочных графиков» (*Пособие*, с. 57). Описанное там можно дополнительно усовершенствовать, причём потребность в этом может быть связана со следующей ситуацией, возникающей в лабораторной практике.

Градуировочная зависимость имеет довольно ограниченный «срок годности» и её регулярно нужно переснимать. На моей памяти никогда не было такого, чтобы при этом все заново измеренные значения аналитического сигнала у стандартных образцов в точности совпали со всеми предыдущими. Это приводит к тому, что новая градуировочная линия отличается от старой – либо слегка меняется угол её наклона, либо вся линия смещается чуть выше или ниже. Поэтому чтобы правильно выполнить расчёт концентрации определяемого вещества в исследуемом образце, нужно в ячейках “G5” и “G6” листа указать обновлённые значения коэффициентов a и b уравнения прямой. При замене данных в ячейках со значениями концентраций определяемого вещества x и соответствующих им величин аналитического сигнала y, табличный процессор отобразит на графике пересчитанные коэффициенты (обновлённый вид уравнения линии) – их-то и надо будет не забыть заново вписать в указанные выше ячейки.

Если же при этом учесть влияние человеческого фактора, то не исключена возможность, что после ввода значений x – y новой градуировки ручная замена коэффициентов можно быть запамятована, в итоге получится неправильное значение концентрации, поскольку она окажется рассчитанной фактически по старой калибровке.

Чтобы этого избежать, удобнее и надёжнее сделать так, чтобы коэффициенты a и b в “G5” и “G6” вычислялись сами. Вот именно этот расчёт и предлагается реализовать при «апгрейде» градуировочного графика.

Уравнение прямой линии табличный процессор вычисляет с использованием метода наименьших квадратов (МНК), поэтому сначала придётся позанудствовать и дать краткое описание этого метода.

Итак, пусть у нас есть n стандартных образцов с концентрациями определяемого компонента x_1, x_2, … , x_n, а измеренные соответствующие значения аналитического сигнала равны y_1, y_2, … , y_n. Если зависимость $y=f(x)$ носит линейный характер, то оптимальной аппроксимирующей линией будет такая, коэффициенты которой являются решением следующей системы уравнений (под оптимальной в МНК понимается линия, для которой минимальна сумма квадратов всех отклонений каждого измеренного значения y_i от «теоретического» значения $ax_i + b$):

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^2\right)\cdot a + \left(\sum_{i=1}^{n} x_i\right)\cdot b = \sum_{i=1}^{n} x_i \cdot y_i \\ \left(\sum_{i=1}^{n} x_i\right)\cdot a + n\cdot b = \sum_{i=1}^{n} y_i \end{cases} \tag{1}$$

Решение системы «на лету» удобно выполнять с использованием метода Крамера, вскользь упоминавшегося в пункте «Реализация метода Монте-Карло» (*Пособие*, с. 85). Рассмотрим систему из двух линейных уравнений с двумя неизвестными a и b:

$$\begin{cases} k_{11} \cdot a + k_{12} \cdot b = k_{13} \\ k_{21} \cdot a + k_{22} \cdot b = k_{23} \end{cases}$$

Сначала вычисляется так называемый основной определитель системы Δ (он состоит из коэффициентов, стоящих перед неизвестными переменными в левых частях уравнений):

$$\Delta = \begin{vmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{vmatrix} = k_{11} \cdot k_{22} - k_{21} \cdot k_{12} ,$$

затем – определители Δ_a и Δ_b (они составляются аналогично Δ, только в них коэффициенты перед соответствующей неизвестной переменной заменяются числом из правой части уравнения):

$$\Delta_a = \begin{vmatrix} k_{13} & k_{12} \\ k_{23} & k_{22} \end{vmatrix} = k_{13} \cdot k_{22} - k_{23} \cdot k_{12} ; \; \Delta_b = \begin{vmatrix} k_{11} & k_{13} \\ k_{21} & k_{23} \end{vmatrix} = k_{11} \cdot k_{23} - k_{21} \cdot k_{13} .$$

После этого можно вычислить сами a и b:

$$a = \frac{\Delta_a}{\Delta} ; \; b = \frac{\Delta_b}{\Delta} .$$

Допустим у нас имеются такие значения концентраций x и аналитического сигнала y:

	A	B	C	D	E	F	G
1							
2		№ п/п	x	y			
3		1	10	7,14			
4		2	20	14,37			y=ax+b
5		3	40	25,19		a=	1
6		4	60	34,45		b=	1
7		5	80	48,82			
8		6					
9		7				y=	1
10		8				x=	0
11		9					
12		10					
13		11					
14		12					
15		13					
16		14					
17		15					
18							

Прежде чем, начать решать систему уравнений, нужно посчитать значения коэффициентов в (1), а для этого необходимо знать суммы по x_i , y_i , $x_i \cdot y_i$, $x^2{}_i$, а также n – общее количество пар значений $x - y$. Сделать это можно таким образом:

	A	B	C	D	E	F	G	H	I	J
1										
2		№ п/п	x	y					x^2	x*y
3		1	10	7,14					100	71,4
4		2	20	14,37			y=ax+b		400	287,4
5		3	40	25,19		a=	1		1600	1007,6
6		4	60	34,45		b=	1		3600	2067
7		5	80	48,82					6400	3905,6
8		6							0	0
9		7				y=	1		0	0
10		8				x=	0		0	0
11		9							0	0
12		10							0	0
13		11							0	0
14		12							0	0
15		13							0	0
16		14							0	0
17		15							0	0
18										
19		Сумма:	210	129,97		n=	5		12100	7339
20										

В диапазонах “I3:I17” и “J3:J17” вычисляются значения x^2_i и $x_i{\cdot}y_i$ по формулам (показано на примере ячеек “I3” и “J3” соответственно):

```
=C3^2
=C3*D3
```

В ячейке “G19” подсчитывается n по формуле:

```
=СЧЁТ(C3:C17)
```

```
=COUNT(C3:C17)
```

При помощи функции

```
=СУММ(   )
```

```
=SUM(   )
```

в ячейках “C19”, “D19”, “I19” и “J19” вычисляются необходимые суммы по диапазонам “C3:C17” (x_i), “D3:D17” (y_i), “I3:I17” (x^2_i) и “J3:J17” ($x_i{\cdot}y_i$) соответственно.

Теперь можно переходить к вычислению определителей Δ, Δ_a и Δ_b. Для этого можно разместить элементы каждого из определителей в отдельном блоке ячеек, введя в них формулы, которые просто дублируют уже вычисленные перед этим значения сумм (а также n):

	A	B	C
18			
19		Сумма:	210
20			
21		D=	
22		=I19	=C19
23		=C19	=G19
24			
25		D(a)=	
26		=J19	=C19
27		=D19	=G19
28			
29		D(b)=	
30		=I19	=J19
31		=C19	=D19

Следующий шаг – нахождение значений определителей. Для этого сперва в ячейку “C21” нужно ввести формулу:

```
=МОПРЕД(B22:C23)
```

```
=MDETERM(B22:C23)
```

Использованная в ней функция вычисляет величину определителя, который указан в качестве её аргумента, представляющего собой квадратный массив (диапазон) ячеек.

Чтобы посчитать два остальных определителя, достаточно скопировать формулу из “C21” в “C25” и “C29”. Последний этап – собственно нахождение коэффициентов a и b уравнения линии выполняется вводом формул в ячейки “G5” (a) и “G6” (b) соответственно:

```
=C25/C21
=C29/C21
```

Хотелось бы напомнить, что при наличии в ячейке “G10” формулы вида

```
=(G9-G6)/G5
```

будет выполняться расчёт концентрации вещества в исследуемом образце на основании значения аналитического сигнала и градуировки. В связи с этим выходит, что в получившейся программе ячейки листа, в которые должны вводится эмпирические данные – это “G10” и “C3:D17”, остальные же ячейки (прежде всего те, которые содержат формулы) трогать не следует, поэтому для большей надёжности при желании можно установить защиту листа (*Пособие*, с. 55).

В учебных целях приведённые расчёты выполнены достаточно развёрнуто для большей наглядности и понятности последовательности действий при вычислениях. Разумеется, их можно сократить, действуя по одному из следующих вариантов.

В первом варианте в ячейке “I19” можно указать формулу вида:

```
=СУММКВ(C3:C17)
```

```
=SUMSQ(C3:C17)
```

Использованная в ней функция вычисляет сумму квадратов значений указанного в качестве аргумента диапазона ячеек, в данном случае – $\sum_{i=1}^{n} x_i^2$, в связи с чем отпадает необходимость сначала вычислять каждое значение x^2_i по отдельности в диапазоне “I3:I17”, а затем складывать их. Далее, в ячейке “J19” следует разместить вот такую формулу:

```
=СУММПРОИЗВ(C3:C17;D3:D17)
```

```
=SUMPRODUCT(C3:C17;D3:D17)
```

Функция в ней может содержать несколько аргументов (в данном случае – два), каждый из которых представляет собой диапазон ячеек. Работает она так: перемножает числа из соответствующих ячеек разных диапазонов (здесь – значения x_i с соответствующими им значениями y_i) и складывает их между собой. Таким образом, приведённая формула считает $\sum_{i=1}^{n} x_i \cdot y_i$, избавляя от потребности делать эти вычисления раздельно в ячейках “J3:J17”.

Второй вариант сокращения значительно более «экстремальный», поскольку в нём все вычисления сводятся всего к двум формулам. Из МНК и метода Крамера следует, что коэффициенты уравнения аппроксимирующей линии выражаются так:

$$a=\frac{\Delta_a}{\Delta}=\frac{\begin{vmatrix}\sum_{i=1}^{n} x_i\cdot y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n\end{vmatrix}}{\begin{vmatrix}\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n\end{vmatrix}}=\frac{n\cdot\sum_{i=1}^{n} x_i\cdot y_i-\sum_{i=1}^{n} x_i\cdot\sum_{i=1}^{n} y_i}{n\cdot\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2},$$

$$b=\frac{\Delta_b}{\Delta}=\frac{\begin{vmatrix}\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i\cdot y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i\end{vmatrix}}{\begin{vmatrix}\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n\end{vmatrix}}=\frac{\sum_{i=1}^{n} x_i^2\cdot\sum_{i=1}^{n} y_i-\sum_{i=1}^{n} x_i\cdot\sum_{i=1}^{n} x_i\cdot y_i}{n\cdot\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2}.$$

Эти равенства, записанные в нотации табличного процессора, можно поместить в ячейках “G5” (коэффициент *a*):

```
=(СЧЁТ(C3:C17)*СУММПРОИЗВ(C3:C17;D3:D17)-
СУММ(C3:C17)*СУММ(D3:D17))/(СЧЁТ(C3:C17)*СУММКВ(C3:C17)-
СУММ(C3:C17)^2)
```

```
=(COUNT(C3:C17)*SUMPRODUCT(C3:C17;D3:D17)-
SUM(C3:C17)*SUM(D3:D17))/(COUNT(C3:C17)*SUMSQ(C3:C17)-
SUM(C3:C17)^2)
```

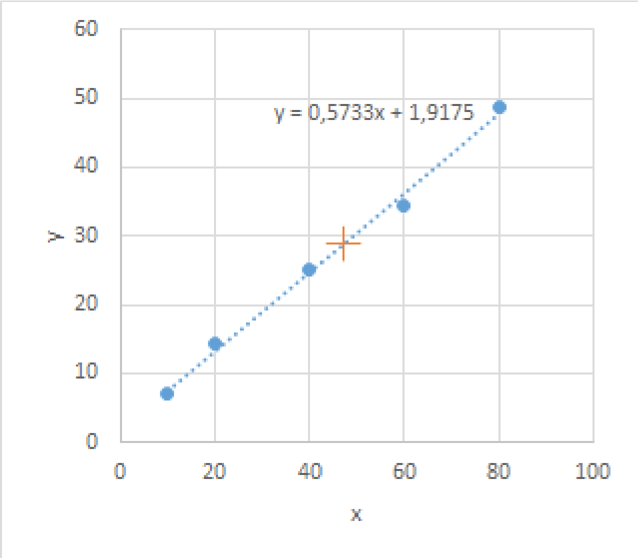
и “G6” (коэффициент *b*):

```
=(СУММКВ(C3:C17)*СУММ(D3:D17)-
СУММ(C3:C17)*СУММПРОИЗВ(C3:C17;D3:D17))/
(СЧЁТ(C3:C17)*СУММКВ(C3:C17)-СУММ(C3:C17)^2)
```

```
=(SUMSQ(C3:C17)*SUM(D3:D17)-
SUM(C3:C17)*SUMPRODUCT(C3:C17;D3:D17))/
(COUNT(C3:C17)*SUMSQ(C3:C17)-SUM(C3:C17)^2)
```

Как видно, все необходимые расчёты размещены весьма компактно и не «размазаны» по листу, но вместе с тем подобные формулы в силу своей громоздкости трудны для восприятия, особенно для тех, кто только начал осваивать работу с электронными таблицами.

В качестве финального оформительского штриха рекомендуется добавить на график с градуировочной линией ещё один ряд данных (*Пособие*, с. 78), состоящий лишь из одной пары значений *x* и *y* (то есть из одной точки) – значений ячеек “G10” и “G9”. Это позволит наглядно отмечать место расположения на графике (в серединной области, в районе одного из краёв или за его пределами) измеренной величины аналитического сигнала у исследуемого образца. После настроек параметров маркера такого дополнительного ряда результат может выглядеть примерно следующим образом:



© Широков Александр, 06.09.2020