

用 EXCEL 进行线性回归分析 及测量不确定度的计算

董夫银^①

(深圳出入境检验检疫局 深圳市福田区福强路 1011 号大厦 15 楼 518045)

摘 要

用 EXCEL 提供的函数 LINEST 和回归分析工具进行线性回归分析,并利用 EXCEL 的插入图表功能做线性回归拟合图。也示例说明了线性回归产生的测量不确定度的计算。

关键词 EXCEL, LINEST, 回归, 线性拟合图, 测量不确定度。

中图分类号: O652; TP3 **文献标识码**: B **文章编号**: 1004-8138(2005)06-1234-05

1 前言

在日常检验中,尤其是化学分析中,经常要用到一元线性回归方程。这是由于被分析物的浓度与仪器响应值之间通常是一个线性关系。仪器分析就是利用这个特性,先测得一组已知浓度的标准溶液的仪器响应值,然后求出该标准溶液浓度与仪器响应值之间的线性关系,即一元线性回归方程,利用这个方程就可通过待测溶液的仪器响应值求得待测溶液的浓度。上述线性回归过程尽管复杂,但在现代仪器分析中,都不要自己计算,仪器已自动完成。只不过仪器计算完后给出的有关该一元线性回归方程的信息较少,给计算测量不确定度以及其他一些需要用到一元线性回归特性参数的场合带来很大的不便。尽管市面上有许多专门的统计计算软件,但对于简单的统计计算都不如 EXCEL 经济、易得和实用。在 EXCEL 中集成很多统计计算功能,它的回归计算功能很强,不但可以进行一元线性回归计算,而且可以进行多重线性回归计算,也能画线性拟合图。因此非常有用。下面以一元线性回归为例,介绍具体的计算过程,该计算过程也同样适用于多重线性回归。

2 测量方法

假设,为了得到一元线性回归方程,用已知浓度 C_i 的不同标准溶液进行 n 次测量,得到 n 个仪器响应值 A_i ,然后再按最小二乘法就可求得的一元线性回归方程为 $A=C \cdot B_1+B_0$,其中 A 为仪器响应值; C :待测溶液的浓度; B_1 :斜率; B_0 :截距。

由于这是一条拟合直线,按照该线性回归方程求得的待测浓度就不可避免地带来测量不确定度。该测量不确定度按下式进行计算^[1]:

$$u(C) = \frac{S}{B_1} \sqrt{\frac{1}{P} + \frac{1}{n} + \frac{(C-\bar{C})^2}{S_{xx}}}$$

其中: $u(C)$ ——待测溶液浓度 C 的测量不确定度; P ——重复测试的次数。

S 和 S_{xx} 值由下式求得:

① 联系人,电话:(0755)83390293;(0755)83886163;手机:013600191964;传真:(0755)83396455;E-mail:don868@yahoo.com.cn

作者简介:董夫银(1964—),男,浙江省泰顺县人,高级工程师,从事化学及分析化学中测量不确定度的评估的研究。

收稿日期:2005-07-05

$$S^2 = \frac{\sum_{j=1}^n [A_j - (B_0 + B_1 \cdot C_j)]^2}{n-2}$$

$$S_{xx} = \sum_{j=1}^n (C_j - \bar{C})^2$$

\bar{C} :不同标准溶液浓度的平均值(共测量 n 次)。

下面将测量分成单次测量与重复测量两种情况分别进行介绍。

2.1 单次测量

对于单次测量即每个标准溶液浓度只测量一次的情形。例如,校准标准溶液的浓度与仪器的响应值见表 1。

表 1 单次测量

标准溶液浓度	0.00	1.00	2.00	5.00	10.00
仪器响应值	1	125	246	619	1250

2.1.1 用 EXCEL 求一元线性回归方程及其参数

打开 EXCEL,在 EXCEL 的 A 和 B 列分别输入标准溶液浓度和仪器响应值(见图 1)。在 EXCEL 中有两种方法求一元线性回归方程,即 LINEST 函数法和回归法。

(1) LINEST 函数法

先选一个放计算结果的区间,如 C1:D5,然后输入=LINEST,会出现 LINEST(known_y's, known_x's,const,stats),设浓度为 X 轴,仪器响应值为 Y 轴,而 const 和 stats 均用 true 代入,则在括号内分别输入(B1:B5,A1:A5,true,true),然后同时按 Ctrl+Shift+Enter,则出现下图 1。

C1	* {=LINEST(B1:B5,A1:A5,TRUE,TRUE)}					
	A	B	C	D	E	F
1	0.00	1	124.9141	-1.4908		
2	1.00	125	0.430906	2.197199		
3	2.00	246	0.999964	3.479417		
4	5.00	619	84034.52	3		
5	10.00	1250	1017350	36.31902		
6						

图 1 Linest 法计算示意图

在 C1:D5 区域就出现了一元线性回归计算的结果,按照 EXCEL 的规定,C1 单元格为斜率;D1 单元格为截距;C2 单元格为斜率的标准差;D2 单元格为截距的标准差;C3 单元格为 R^2 值;D3 单元格为仪器响应值 A 的标准差;C4 单元格为 F 统计值,D4 单元格为自由度,C5 单元格为回归平方和,D5 单元格为残差平方和。

所以,从图 1 可得斜率 $B_1=124.9141$;截距 $B_0=-1.4908$,一元线性回归方程为:

$$A=124.9141 \times C - 1.4908$$

(2) 用“数据分析”中的“回归”工具

在“工具”菜单上,单击“数据分析”。如果没有“数据分析”,则要加载“分析工具库”加载宏。具体做法如下:

① 在“工具”菜单上,单击“加载宏”。

② 在“可用加载宏”列表中,选中“分析工具库”框,再单击“确定”。

单击“数据分析”后,在所出现的分析工具列表中选中“回归”,单击确定后出现图 2。在输入 X

值区域时,可先将“X 值输入区域”的框清空,然后在框上点上鼠标指针,涂黑放 X 数值的区域即 A1:A5。Y 也按同样的方式输入。然后在需要显示的内容前打上相应的勾后,按确定,得图 3 的数据(仅列出所产生的部分数据,没有列出所产生的线性拟合图)。

在表中,每个数据代表什么,一清二楚,不像 LINEST 函数计算的结果,需要记住 EXCEL 的规定才能知道每个数据的含义。另外,在图 2 的列表中有许多选项,可以做线性拟合图、正态概率图等。利用所给出的 F 值、t 值及置信概率,还可以对线性回归方程进行显著性检验。图 4 的数据表明, F 值大于查表所得的 $F_{1,3}(0.05)=10.13$ 或 t 值大于查表所得的 $t_3(0.025)=3.1824$,因此 X 与 Y 之间的线性相关关系显著。

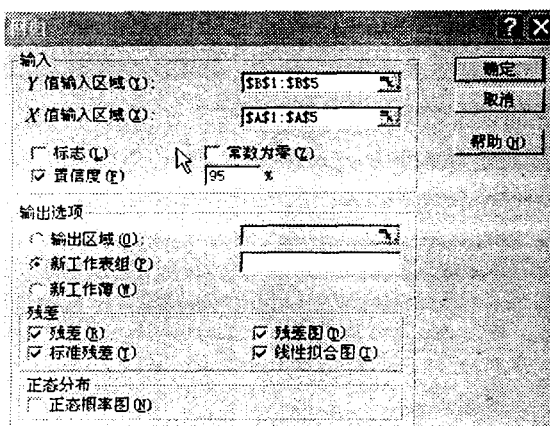


图 2 数据分析法中的回归工具示意图

SUMMARY OUTPUT						
	A	B	C	D	E	F
4	Multiple	0.999982				
5	R Square	0.999964				
6	Adjusted	0.999952				
7	标准误差	3.479417				
8	观测值	5				
9						
10	方差分析					
11		df	SS	MS	F	Significance F
12	回归分析	1	1017350	1017350	84034.52	9.05244E-08
13	残差	3	36.31902	12.10634		
14	总计	4	1017387			
15						
16		Coefficient	标准误差	t Stat	P-value	Lower 95%
17	Intercept	-1.4908	2.197199	-0.6785	0.546106	-8.483270452
18	X Variabl	124.9141	0.430906	289.8871	9.05E-08	123.5427737
19						

图 3 利用数据分析法中的回归工具的计算示意图

2.1.2 画线性拟合图

上述所介绍的“数据分析”中的“回归”工具的功能较多,有画线性回归拟合图的功能。除此之外,线性拟合图还可以通过以下方式获得:

在“插入”菜单上,单击“图表”,在“标准类型”列表上选中“XY 散点图”,点击“下一步”,在“系列”列表上分别输入 X 和 Y 的数值的区域(与图 2 回归图表中的 X 和 Y 值的区域一样)。或在“数据区域”列表上输入 X 与 Y 的总的数值区域即 A1:B5,并在“系列产生在”上选中“列”。点击“下一步”在“标题”列表上在“图表标题”中填入“线性回归拟合图”或认为合适的其他名称;在数值(X)和数值(Y)上分别填入“X”与“Y”。点击“下一步”,然后在出现的列表选中“作为其中的对象插入”,点击“完成”,则会看到一个线性回归图。点击该图表的任何位置,则在工具栏中会出现“图表”菜单,单击“添加趋势线”,在“类型”列表中选择“线性”,点击“确定”就得到图 4。

2.1.3 测量不确定度的计算

计算测量不确定度 $u(C)$ 时,可利用上述计算所得的 B1 结果,即图 1 中 C1 单元格的数值。对于 S,可直接利用 D3 单元格的标准差数据,也可利用图 1 D5 单元格的残差平方和数值,这是由于 $S^2 = \text{残差平方和}/(n-2)$,故 $S = \sqrt{36.31902/3} = 3.479417$ 。对于 S_{xx} ,则需要利用 EXCEL 的标准偏差函数 stdev,因为 S_{xx} 等于标准偏差的平方乘以 $(n-1)$ 。为求标准偏差,任选一个单元格 F1,输入 =stdev,在出现的括号内输入标准溶液浓度的区间 A1:A5,得 =stdev(A1:A5),按回车键,得标准偏差值为 4.037326,故 $S_{xx} = (4.037326)^2 \times 4 = 65.2$ 。待测浓度 C 可通过待测物的仪器响应值,再按上述所求得的回归方程求出浓度即可(仪器一般自动产生),设所求得的待测浓度 $C = 2.06$,并假设测量次数为一次,即 $P = 1$,本例中标准溶液测量 5 次故 $n = 5$, \bar{C} 用 AVERAGE 函数计算,结果为 3.60,将上述参数代入 $u(C)$ 的计算公式则得 $u(C) = 0.030972$,该测量不确定度计算过程也可通过 EXCEL 来进行,具体见图 5 第一行 f_6 框中所列出的式子。

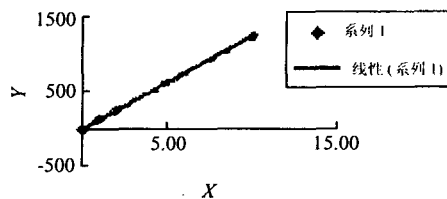


图 4 线性回归拟合图

X——浓度:μg/L;Y——仪器响应值。

F6		f ₆ =F3/C1*SQRT(1/1+1/5+(F5-F2)*(F5-F2)/F4)				
	A	B	C	D	E	F
1	0.00	1	124.9141	-1.4908	标准偏差	4.037326
2	1.00	125	0.430906	2.197199	平均值	3.60
3	2.00	246	0.999964	3.479417	S值	3.479417
4	5.00	619	84034.52	3	S _{xx}	65.2
5	10.00	1250	1017350	36.31902	C值	2.06
6					u(C)值	0.030972

图 5 测量不确定度的计算示意图

2.2 重复测量

对于重复测量即每个标准浓度测量多次的情形,见表 2。

表 2 重复测量

标准溶液浓度(μg/L)	仪器响应值 1	仪器响应值 2	仪器响应值 3
0.1	0.028	0.029	0.029
0.3	0.084	0.083	0.081
0.5	0.135	0.131	0.133
0.7	0.180	0.181	0.183
0.9	0.215	0.230	0.216

(本例选自参考文献[1]P75 页的例子)

用 LINEST 进行计算时,要注意标准溶液的浓度的数量为 15 个而不是 5 个。结果如下:

从图 6 可得 $S = 0.005486$;斜率 $B_1 = 0.241$;斜率的标准差为 0.005008;截距 $B_0 = 0.0087$;截距的标准差为 0.002877; $R^2 = 0.994418$ (即 $R = 0.997205$);另按上述 2.1.3 的步骤可计算得 $S_{xx} = 1.2$ (要按 15 个浓度计算)。这些数值与参考文献给出的数值完全一致。

另外,上述 2.1.1(2)所介绍的“回归”工具不适合于重复测量的情况。对于上述 2.1.2 所介绍的画线性拟合图,除了能添加趋势线外,也能进行拟图。对于本例要将数据分成三个系列,即 A 列对 D 列,B 列对 E 列,C 列对 F 列,然后按照上述方法进行即可。

= {=LINEST (D1:F5, A1:C5, TRUE, TRUE)}							
A	B	C	D	E	F	G	H
0.1	0.1	0.1	0.028	0.029	0.029		
0.3	0.3	0.3	0.084	0.083	0.081		
0.5	0.5	0.5	0.135	0.131	0.133		
0.7	0.7	0.7	0.18	0.181	0.183		
0.9	0.9	0.9	0.215	0.23	0.216		

图 6 重复测量的计算示意图

参考文献

[1] 中国实验室认可委员会. 化学分析中不确定度的评估指南[M]. 北京: 中国计量出版社出版, 2002. 78.

Linear Regression Analysis and Calculation of Uncertainty with EXCEL Software

DONG Fu-Yin

(Shenzhen Entry-Exit Inspection and Quarantine Bureau, Shenzhen, Guangdong 518045, P. R. China)

Abstract

The LINEST technique, the regression tool and chart tool built in the EXCEL were used to generate a best-fit line and calculate relevant parameters as well as uncertainty.

Key words EXCEL, LINEST, Regression, Linear Fit Chart, Uncertainty Calculation.

《光谱实验室》2005 年售价实际继续下降

由于《光谱实验室》投稿数量不断增加, 为了保证出版周期, 从 2005 年第 1 期开始, 在 2004 年的基础上每册正文增加页码 16 页, 而售价保持不变:

2003 年售价: 20 元/册, 页码为 160 页/册, 平均 0.125 元/页;

2004 年售价: 25 元/册, 页码为 208 页/册, 平均 0.120 元/页。

2005 年售价: 25 元/册, 页码为 224 页/册, 平均 0.112 元/页。

因此, 售价实际继续下降。

《光谱实验室》编辑部