

最小平方距离法在比色分析中的应用

唐万龙 刘光崧
(中国科学院南京土壤研究所)

比色分析(例如吸收光度法)结果的计算,一般沿用绘图法^[1],即先用标准系列溶液作自变量 x ,测得的吸收值作因变量 y ,绘制出工作曲线,然后再依据工作曲线上的 y 预测待测液的含量。为了进行定量分析,有些学者提出用回归分析的方法进行吸收光度法的计算^[1,2]。回归分析的数学原理是最小二乘法^[3],其中自变量 x 与因变量 y 的地位不是对等的,用 x 来预报 y 和用 y 来预报 x ,其模型是不同的,自然两条直线也是不同的。在吸收光度法的计算中, x 和 y 的地位是对等的。先用 x 来预报 y ,然后用 y 来预报 x ,其数学模型应是同一的。如果用最小二乘法来建立回归方程,对换 x 、 y 的位置,其模型则不同,因此,预报结果的精度是不高的。为了解决这一类问题,下面介绍最小平方距离法^[4]。

现以具体实例来说明最小平方距离法的数学原理和应用以及与最小二乘法的差别。

某分析工作者用钒钼黄法测定待测液中的磷时,标准系列溶液的浓度(x)及其吸收值(y)的数据如下表所示。

标准溶液 X(mg/L)	0	4	8	12	16	24
吸收值 Y	0.020	0.115	0.205	0.293	0.375	0.540

用最小二乘法求回归方程的方法如下:

$$\begin{aligned}\bar{x} &= \sum_{i=1}^n x_i \\ \bar{y} &= \sum_{i=1}^n y_i \\ l_{xx} &= \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \\ l_{yy} &= \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \\ l_{xy} &= \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) \\ b &= \frac{l_{xy}}{l_{xx}} \\ a &= \bar{y} - b\bar{x}\end{aligned}$$

以标准溶液为自变量 x ,吸收值为因变量 y ,其回归方程为

$$\begin{aligned}y &= a + bx \\ y &= 0.0276 + 0.0216x\end{aligned}\tag{1}$$

以吸收值 y 为自变量 x' , 标准溶液为因变量 y' , 其回归方程为

$$y' = c + dx'$$

$$y' = -1.267 + 46254x' \quad (2)$$

把(1)转化(2)

$$x = \frac{y-a}{b}$$

$$y' = \frac{x'-a}{b}$$

$$y' = -1.278 + 46.296x' \quad (3)$$

同理, 把(2)转化为(1)

$$x' = \frac{y'-c}{d}$$

$$y = \frac{x-c}{d}$$

$$y = 0.0274 + 0.0216x \quad (4)$$

方程(2)和方程(3)来自于同一种观测数据, 又表示相同的含义, 理应是相同的, 可事实是不同的。方程(1)和(4)同理。方程(1)和(2)之所以不等价, 是由于最小二乘法的数学原理本身造成的。方程(1)选定标准溶液为自变量 x , 求目标函数吸收值 y 与直线的纵向误差平方和

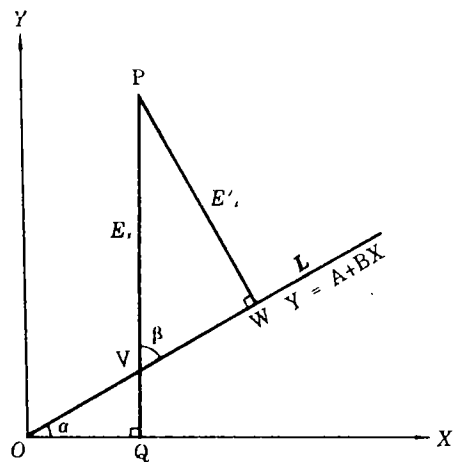
$$\sum_{i=1}^n (y_i - bx_i - a)^2 \quad (5)$$

达到最小, 方程(2)选定吸收值 y 作为自变量, 求目标函数 x 与直线横向误差平方和

$$\sum_{i=1}^n (x_i - dy_i - c)^2 \quad (6)$$

达到最小。由于目标函数(5)与(6)不一样, 自然得出的方程(1)与(2)就不同了。因此, 对 x 、 y 的地位是对等的、可以互换的这类问题, 不能采用最小二乘法, 而应采用最小平方距离法。

最小平方距离法既不要求纵向, 也不要求横向, 而是要求观测数据点到直线距离的平方和



$$\sum_{i=1}^n \left(\frac{y_i - Bx_i - A}{\sqrt{1+B^2}} \right)^2 \quad (7)$$

达到最小, 这就是最小平方距离法。下面来推导(7)式的来历:

设上图直线 L 为 $y = A + Bx$
 B 为斜率, 显然 $B = \tan \alpha$ 设点 $P(x_i, y_i)$, E_i 为 P 到 x 轴交于直线 L 的距离, E'_i 为 P 到直线 L 的距离。

$$PQ = y_i$$

$$VQ = A + Bx_i$$

$$E_i = PQ - VQ = y_i - Bx_i - A$$

显然,最小二乘法是求 $\min \sum_{i=1}^n E_i^2$,而最小平方距离法是求 $\min \sum_{i=1}^n E_i'^2$,这就是两种方法的本质区别。

在 $\triangle PVW$ 中, $E_i' = E_i \cdot \sin\beta$

$$= (y_i - Bx_i - A) \cos\alpha$$

$$\cos\alpha = \frac{1}{\sqrt{1+\tan^2\alpha}} = \frac{1}{\sqrt{1+B^2}} \quad \therefore E_i' = \frac{y_i - Bx_i - A}{\sqrt{1+B^2}}$$

$$\min \sum_{i=1}^n E_i'^2 = \min \sum_{i=1}^n \left(\frac{y_i - Bx_i - A}{\sqrt{1+B^2}} \right)^2 \quad (8)$$

在(8)式中, B 、 A 是两个待定系数,我们仍采用最小二乘法中的记号,当 $l_{xy} \neq 0$ 时,

$$B = \frac{(l_{yy} - l_{xx}) + \sqrt{(l_{yy} - l_{xx})^2 + 4l_{xy}^2}}{2l_{xy}}$$

$$A = \bar{y} - B\bar{x}$$

在最小平方距离法中, x 、 y 的地位是对等的,互换其位置之后仍是同一条直线,在数学上是可以证明的。这里仍用前表的数据进行验证:

以标准溶液为自变量 x ,以吸收值为因变量 y ,其直线方程为:

$$y = A + Bx$$

$$y = 0.0276 + 0.0216x \quad (9)$$

以吸收值 y 为自变量,以标准溶液 x 为因变量,其直线方程为

$$x = -1.278 + 46.296y \quad (10)$$

把(9)化为(10)

$$x = \frac{y - A}{B}$$

$$x = -1.278 + 46.296y \quad (11)$$

(11)同(10);同理,把(10)化为(9)

$$y = 0.0276 + 0.0216x \quad (12)$$

(12)与(9)同。因此方程(9)与(10)是同一直线。用最小平方距离法和最小二乘法建立的直线方程都是正确的,但是两者的适用范围不一样。当 x 、 y 的地位对称时,或两者都有误差时,特别是要求双向预测时,宜采用最小平方距离法。

参 考 文 献

- [1] 中国土壤学会农业化学专业委员会编,土地农业化学常规分析方法,科学出版社,1983。
- [2] Bauer, Edward L., A Statirtical Manual For Chemists, Academic Press Inc., New york, London, 1971.
- [3] 冯士雍编,回归分析方法,科学出版社,1985。
- [4] 王柱,最小平方距离法介绍,数理统计与管理,第4期,1986。