

水质分析中绘制标准曲线应 建立直线回归方程

裴 钟 源

(水利系)

提 要: 标准曲线在环境监测水质分析中应用相当普遍。本文认为绘制标准曲线应用最小二乘法原理建立直线回归方程,以它所表达的直线是最佳的直线,误差最小。并举例说明建立回归方程运算的程序、方法,以及回归方程的相关分析。

关键词: 最小二乘法, 直线回归方程, 相关系数。

目前,在环境监测水质分析的国内外推荐方法中,光谱法的分析技术要占一半以上,而其中主要是以比耳定律为理论依据的比色——分光光度法和原子吸收分光光度法。在利用吸光度与浓度的关系借助分光光度计和原子吸收分光光度计等进行水质分析时,都需预先绘制标准曲线(或工作曲线,严格讲二者是有区别的),因为标准曲线法比“计算法”精确,它可以消除由于种种原因所引起的偏离吸收定律而造成的误差,并可判别待测溶液适用的测定浓度范围(线性范围)。虽然制作标准曲线比较费时,但对于象水质分析中成批样品的测定却有简便、省时的优点。该法使用方便、精确度高,是环境监测水质分析中质量控制的重要环节。

1 标准曲线的回归

习惯上,绘制标准曲线是在直角坐标纸上,以容易使人观察的方式,把纵横座标,分别表示仪器读数和标准溶液的浓度,选择合适的标度,用直线的形式画出,而最简单的方法是将相应的 (x_1, y_1) 、 (x_2, y_2) 、 (x_3, y_3) 、……、 (x_n, y_n) 点点在坐标纸上,凭视觉画一直线,使尽可能多的点在直线上和在直线两侧,但这样绘制的标准曲线即使在分析过程中各种条件控制非常严密,分析人员操作多么精确,实际上很难使一系列数据准确地成一直线,在很大程度上受主观判断力的影响,不可避免地存在着较大的误差,尤其对熟练程度较差的分析人员或操作不太精确的情况下,误差就更大。

本文1988年3月20日收到

为使标准曲线准确地表达两个相关因素（吸光度与浓度）间的相关关系，并且把误差减小到最低限度，必须进行一元线性回归分析，建立直线回归方程， $\hat{y} = ax + b$ ，以确定二变量间的定量关系式，准确地求算常数a和b值。常数a、b值的求法很多，最常用的有直线图解法、选点法（联立方程法），平均法、最小二乘法，其中以最小二乘法原理建立起来的回归方程所表达的直线是最佳的直线，它是一切直线中最接近所有样点的直线，它能使各实际观察值y与估计值 \hat{y} 之差的平方总和为最小，也就是说，以这条直线来代表x与y的相关关系其与样本数据的误差比任何其它直线都要小，即 $\sum (y - \hat{y})^2 = \sum (y - b - ax)^2$ 为最小。由微积分中求极值的方法推知，使上式平方和达到极小的回归线的a、b值分别为：

$$a = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad b = \bar{y} - a\bar{x}$$

而在实际运算中，a值多般应用下式：

$$a = \frac{\sum xy - (\sum x)(\sum y) / n}{\sum x^2 - (\sum x)^2 / n}$$

当x = 0时，y = b，故称b为y轴上的截距，a为直线的斜率，即当x变化一个单位时，y相应地变化的单位数，统计上称之为回归系数。

准确地绘制标准曲线是准确获得样品代表值的基础，也是环境水质监测中质量控制的重要环节。在直角坐标纸上确定自变量x（系列标准溶液浓度）与因变量y（仪器响应值——吸光度）的合适标度，尽量使标准曲线的几何斜率接近于1（与x座标成45度角），使两个轴上的读数误差相近。以最小二乘法原理建立起直线回归方程 $\hat{y} = ax + b$ ，找出x与y对应的两点 (\bar{x}, \bar{y}) 、 (x_0, y_0) ，联接这两点就得到精确的标准曲线。但直线的最高点为标准系列溶液的最高浓度所对应的仪器响应值（吸光度），不得将标准曲线任意外延。

具体运算举例如下：

1.1 用4-氨基安替比林（4-AAP）—氯仿萃取比色法测挥发酚时，得到下表所列数据：

x (μg/l)	2	4	8	16	24	32	40	60
y (A)	0.008	0.061	0.078	0.149	0.244	0.398	0.486	0.766

直线回归方程可由最小二乘法原理计算求出，计算时常编制成如下格式：

序号	x	y	x ²	y ²	xy
1	2	0.008	4	6.4 × 10 ⁻⁵	0.016
2	4	0.061	16	3.72 × 10 ⁻³	0.244
3	8	0.078	64	6.08 × 10 ⁻³	0.624
4	16	0.149	256	0.0222	2.384
5	24	0.244	576	0.0595	5.856
6	32	0.388	1024	0.1505	12.416
7	40	0.486	1600	0.2362	19.44
8	60	0.766	3600	0.5868	45.96

$$\begin{aligned} \Sigma x &= 186 & \Sigma y &= 2.18 & n &= 8 \\ \bar{x} &= 23.25 & \bar{y} &= 0.2725 \\ \Sigma x^2 &= 7140 & \Sigma y^2 &= 1.065 & \Sigma xy &= 86.94 \\ (\Sigma x)^2/n &= 4324.5 & (\Sigma y)^2/n &= 0.59405 & (\Sigma x)(\Sigma y)/n &= 50.685 \\ \Sigma x^2 - (\Sigma x)^2/n &= 2815.5 & \Sigma y^2 - (\Sigma y)^2/n &= 0.47095 \\ \Sigma xy - (\Sigma x)(\Sigma y)/n &= 3.6255 \\ a &= \frac{\Sigma xy - (\Sigma x)(\Sigma y)/n}{\Sigma x^2 - (\Sigma x)^2/n} = \frac{3.6255}{2815.5} = 0.0129 \\ b &= \bar{y} - a\bar{x} = 0.2725 - 0.0129 \times 23.25 = -0.0274 \end{aligned}$$

于是得直线回归方程：
 $\hat{y} = 0.0129x - 0.0274$

这条回归线一定通过 (\bar{x}, \bar{y}) 这一点，即 $(23.25, 0.2725)$ 点，再选方便的 x 值，令 $x_0 = 50$ 代入回归方程，得出相应的 $y_0 = 0.6176$ ，然后，联接 $\bar{x} = 23.25$ 、 $\bar{y} = 0.2725$ ， $x_0 = 50$ 、 $y_0 = 0.6176$ 两点，即得标准曲线图 (1)：

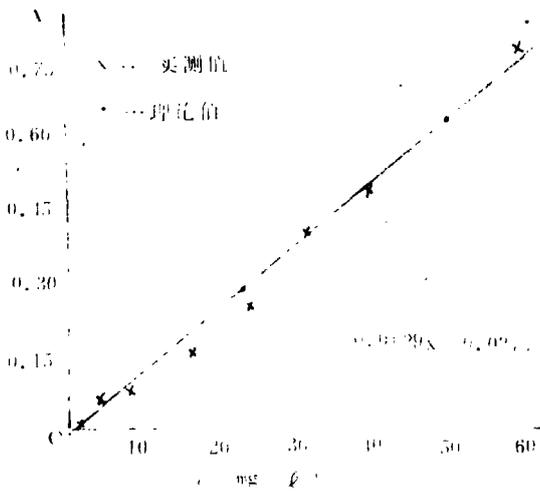


图 1 酚浓度和吸光度标准曲线

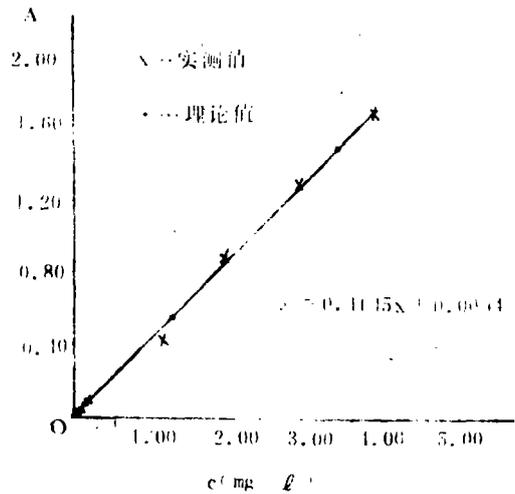


图 2 NO₃-N 浓度和吸光度标准曲线

1.2 用酚二磺酸比色法测 NO₃-N 时得到下表所列数据：

x (mg/l)	0.02	0.06	0.10	0.20	1.20	2.00	3.00	4.00
y (A)	0.008	0.020	0.045	0.085	0.420	0.870	1.240	1.650

直线回归方程计算如下：

$$\Sigma x = 10.38 \quad \Sigma y = 4.338 \quad n = 8$$

$$\bar{x} = 1.2975 \quad \bar{y} = 0.54225 \quad \Sigma x^2 = 30.054 \quad \Sigma y^2 = 5.20305 \quad \Sigma xy = 12.50272$$

$$(\Sigma x)^2 / n = 13.46805 \quad (\Sigma y)^2 / n = 2.35228 \quad (\Sigma x)(\Sigma y) / n = 5.62856$$

$$\Sigma x^2 - (\Sigma x)^2 / n = 16.58595 \quad \Sigma y^2 - (\Sigma y)^2 / n = 2.85077 \quad \Sigma xy - (\Sigma x)(\Sigma y) / n = 6.87416$$

$$a = \frac{\Sigma xy - (\Sigma x)(\Sigma y) / n}{\Sigma x^2 - (\Sigma x)^2 / n} = \frac{6.87416}{16.58595} = 0.4145$$

$$b = \bar{y} - a\bar{x} = 0.54225 - 0.4145 \times 1.2975 = 0.0044$$

于是得直线回归方程: $\hat{y} = 0.4145x + 0.0044$

联接 $\bar{x} = 1.2975$, $\bar{y} = 0.54225$ $x_0 = 3.5$, $y_0 = 1.45515$

两点, 即得标准曲线图 (2):

1.3 用纳氏试剂比色法测 $\text{NH}_3\text{-N}$ 时得到下表所列数据:

x (mg/l)	0.1	0.2	0.4	0.6	1.0	1.4	2.0
y (A)	0.02	0.04	0.088	0.120	0.200	0.246	0.366

直线回归方程计算如下:

$$\Sigma x = 5.7 \quad \Sigma y = 1.08 \quad n = 7 \quad \bar{x} = 0.8143 \quad \bar{y} = 0.1543$$

$$\Sigma x^2 = 7.53 \quad \Sigma y^2 = 0.258616 \quad \Sigma xy = 1.3936 \quad (\Sigma x)^2 / n = 4.6414$$

$$(\Sigma y)^2 / n = 0.16663 \quad (\Sigma x)(\Sigma y) / n = 0.87943 \quad \Sigma x^2 - (\Sigma x)^2 / n = 2.8886$$

$$\Sigma y^2 - (\Sigma y)^2 / n = 0.091986 \quad \Sigma xy - (\Sigma x)(\Sigma y) / n = 0.51417$$

$$\text{计算得出: } a = 0.178 \quad b = 0.009$$

于是得直线回归方程: $\hat{y} = 0.178x + 0.009$ 得标准曲线图 (3):

利用上述标准曲线, 以仪器读数 y (吸光度), 找出对应值 x , 即为水样待测成分的浓度值, 同时, 并可利用上述的回归方程, 以仪器读数值 y , 可以求出待测水样某成分的浓度值 \hat{x} , 该理论值 \hat{x} 比在直角坐标纸上查出的数字要精确得多。

以上运算可应用简单编程计算器来完成, 并将得到的直线回归方程编入程序, 这样每得到一个仪器读数值 y , 即可获知待测成分的浓度值 \hat{x} , 使用起来极为简便、迅速, 从而提高了工作效率。

2 回归方程相关分析

线性回归方程 $\hat{y} = ax + b$ 在一定程度上反映了仪器响应值 y (吸光度) 与样品所含某成分浓度值 (x) 之间的统计相关关系, 但变量 x 与 y 之间线性关系的密切程度如何, 还需对所建立的回归方程进行统计检验。

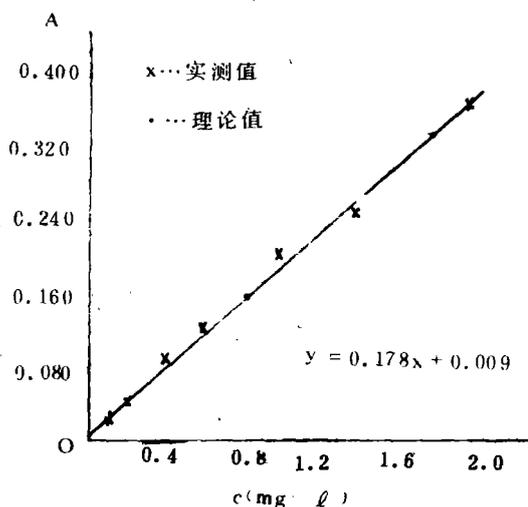


图3 $\text{NH}_3\text{-N}$ 浓度和吸光度标准曲线

2.1 相关系数 (r) :

相关系数是表示两种变量之间关系的性质和密切程度的指标, 它决定着标准曲线的质量和样品测量结果的准确度。公式为:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}}$$

在实际中应用较多的是下式:

$$r = \frac{\sum xy - (\sum x)(\sum y) / n}{\sqrt{[\sum x^2 - (\sum x)^2 / n] \cdot [\sum y^2 - (\sum y)^2 / n]}}$$

以上面三例为例, 计算得:

例(一)、用4-氨基安替比林-氯仿萃取比色法测挥发酚, 相关系数 $r=0.9956$

例(二)、用酚二磺酸比色法测 $\text{NO}_3^- - \text{N}$, 相关系数 $r=0.9997$

例(三)、用纳氏试剂比色法测 $\text{NH}_3 - \text{N}$, 相关系数 $r=0.9975$

从以上数据可知, x 与 y 几乎成完全正相关

2.2 显著性检验:

即指对回归方程的相关系数检验。

2.2.1 查相关系数检验表:

如例(一), 取显著水平 $\alpha=0.01$ 已知 $n=8$

查相关系数临界值表, 得 $r_{0.01(n-2)} = 0.8344$

显然, $r = 0.9956 \gg 0.8344 = r_{0.01}$

x 与 y 的相关系数达到 1% 显著水准, 说明用 4-氨基安替比林-氯仿萃取比色法测得挥发酚含量与吸光度间线性关系非常显著。

2.2.2 检验:

按公式:
$$t = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$$

如例(二), $r=0.9997$ $n=8$ 代入公式得:

$$t = \frac{0.9997}{\sqrt{1-0.9997^2}} \cdot \sqrt{8-2} = 99.95$$

查 t 值表得 $t_{0.01(6)} \text{ 单侧} = 3.14$

$$t = 99.95 \gg 3.14 = t_{0.01(6)}$$

说明用酚二磺酸比色法测得 $\text{NO}_3^- - \text{N}$ 含量与吸光度间线性关系有非常显著的意义。

3 结 语

标准曲线的绘制在环境监测水质分析中是通常迁到的一种方法, 使用相当普遍。为了准确、可靠地绘制最佳直线, 必须以最小二乘法原理建立直线回归方程。而相关系数则是两个变量之间密切程度的指标, 由它可知标准曲线的质量, 从而决定样品测量结果的准确度。通过对相关系数的检验, 进一步了解二变量间线性关系的显著程度。

参 考 文 献

- [1] 高玉堂, 环境监测常用统计方法, 原子能出版社 (1980)。
- [2] 李俊德, 应用数理统计方法, 河南科学技术出版社 (1985)。
- [3] 中国环境检测总站, 《环境水质监测质量保证手册》编写组编, 环境水质监测质量保证手册, 化学工业出版社 (1984)。
- [4] 冯师颜, 误差理论与实验数据处理, 科学出版社 (1964)。
- [5] 中国科学院数学研究所统计组, 常用数理统计方法, 科学出版社 (1979)。
- [6] APHA · AWWA, WPCF Joint Editorial Board, STANDARD METHODS For the Examination of water and wastewater, SIXTEENTH EDITION (1985) 23—24。

Plotting Standard Curve Must Build Line Regression Equation in the Water Analysis of Environmental Monitoring

Pei Zhongyuan

(Hydraulic Engineering Department)

Abstract: Standard curves are in common use in the water analysis of environmental monitoring. This paper suggests that when a standard curve is plotted, you'd better use the principle of least-square method to build line regression. The line expressed by the line regression equation is the best one and its error is the smallest.

In this paper, some living examples are given to show the calculating program and the method for building regression equation, and the relation analysis of regression equation.

Keywords: principle of least-square method, line regression equation, relation coefficient.