

综述

浅谈一元线性加权回归在分析化学领域中的研究应用现状

成桂红¹, 成长玉¹, 张林², 邹丽¹, 陈洁³, 郑海峰⁴, 郑卫东¹

(1. 四川省食品检验研究院, 四川 成都 611731; 2. 四川工商职业技术学院, 四川 成都 611830;
3. 四川凯乐食品检测有限公司, 四川 成都 611730; 4. 成都产品质量检验研究院有限责任公司,
四川 成都 610199)

摘要: 本文主要阐述了一元线性回归中的普通最小二乘法和加权最小二乘法的原理、差异和选用原则, 介绍了异方差性检验、最佳权函数选择、回归方程检验、拟合质量比较等 4 个方面的应用方法, 剖析了加权最小二乘法在国内外的应用现状及其存在的问题, 并针对如何加快分析化学领域回归技术的应用研究提出建议, 以为分析化学和相关领域的研究人员提供借鉴和思考。

关键词: 分析化学; 一元线性回归; 加权回归; 回归分析; 最小二乘法

中图分类号: R155 **文献标识码:** A **文章编号:** 1004-8456(2023)02-0303-08

DOI: 10.13590/j.cjfh.2023.02.025

**Preliminary discussion on the research and application of univariate linear weighted regression
in analytical chemistry**

CHENG Guihong¹, CHENG Changyu¹, ZHANG Lin², ZOU Li¹, CHEN Jie³,
ZHENG Haifeng⁴, ZHENG Weidong¹

(1. Sichuan Institute of Food Inspection, Sichuan Chengdu 611731, China; 2. Sichuan Technology and
Business College, Sichuan Chengdu 611830, China; 3. Sichuan Kaile Food Testing Corporation Limited,
Sichuan Chengdu 611730, China; 4. Chengdu Quality Inspection Corporation Limited, Sichuan
Chengdu 610199, China)

Abstract: The paper mainly expounds the principles, differences and selection principles of ordinary least squares and weighted least squares in univariate linear regression, and introduces the application in four aspects, including heteroscedasticity test, optimal weight function selection, regression equation test and comparison of fitting quality. The application and existing problems of the weighted least squares method at home and abroad is analyzed. Suggestions on how to promote the application research of regression technology in analytical chemistry to provide reference and inspiration for researchers are put forward.

Key words: Analytical chemistry; univariate linear regression; weighted regression; regression analysis; least squares

回归分析(Regression analysis)一词来源于数理统计学,其定义是“评价响应变量与预测变量关系模型的技术”^[1-2]。在分析化学及其质量控制领域,回归分析通常被称为校准(calibration)或线性^[3-4]。此校准的含义是使用有证标准物质或已知准确含量的参考物质,确定目标分析物浓度与响应值的关系的一组操作^[4]。“校准”这一术语在分析化学领域、计

量学领域的定义是基本一致的^[5-6],只是计量学领域对校准定义的表述有些不同^[7-8]。传统分析化学领域使用的回归分析技术是最小二乘法(least squares),其定义是通过最小化观测值与界定模型的预测值的差值的平方和,进行参数估计的技术^[1]。从 1908 年德国数学家高斯(Gauss)提出最小二乘法以来,回归分析已经过 100 多年的发展,其研究和应用范围非常广泛,包含线性回归、回归诊断、回归变量选择、参数估计方法的改进、非线性回归、含有定性变量的回归等 6 大领域^[9]。本文重点关注线性回归在分析化学领域中的应用,如食品中添加剂的检测、食用农产品中农药残留的分析、环境监测中重金属的测定等。关于线性回归的定义,倪永年^[10]

收稿日期:2022-01-10

基金项目:四川省市场监督管理局科技计划项目(SCSJ2021016)

作者简介:成桂红 女 高级工程师 研究方向为食品检测与食品安全 E-mail:9085691@qq.com

通信作者:郑卫东 男 教授级高工 研究方向为食品安全研究与实验室质量控制 E-mail:542663298@qq.com

在1987年曾对此做过表述,即分析化学测试中常遇到各种变量,它们相互联系,互相依存,存在着一定的关系。假如两个变量存在线性关系,找出它们之间的关系便称之为线性回归。但是,可以从统计学建立回归模型的角度了解线性回归的意义。假设变量 x_1, x_2, \dots, x_n 与随机变量 y 之间存在相关关系,当 x_1, x_2, \dots, x_n 的值确定以后,对应 y 的取值应有相应的概率分布,可用如下模型表达这种关系:

$$y=f(x_1, x_2, \dots, x_n)+\varepsilon \quad \text{式(1)}$$

式(1)中, $f(x_1, x_2, \dots, x_n)$ 为变量 x_1, x_2, \dots, x_n 的确定关系及回归函数, ε 为随机误差项。

当其中的回归函数为线性函数时,即有:

$$y=a+b_1x_1+b_2x_2+\dots+b_nx_n+\varepsilon \quad \text{式(2)}$$

式(2)中 a, b_1, b_2, \dots, b_n 为未知参数,称为回归参数,式(2)就称为线性回归模型。

线性回归包含一元线性回归、多元线性回归、多个自变量与多个因变量的回归。此3种回归分析技术在分析化学领域都有广泛运用,尤其是一元线性回归,涵盖了分析化学的绝大多数领域。多元线性回归以及多个自变量与多个因变量的回归在化学计量学领域得到了广泛的应用,其中的多元校正和多元分辨是化学计量学研究的主要任务,也是该领域最有创造性的一部分内容^[11]。化学计量学自诞生以来,得益于物理学、数学和计算机科学的进步,经过了40多年的快速发展,取得了辉煌的成果。我国在这一领域也处于世界先进水平,取得了丰硕的成果^[12-15],如多元分辨校正模型在无机多组分分析中的运用^[12],二阶张量校正、三阶张量校正以及更高阶张量校正方法在复杂体系高阶仪器数据的定量解析^[13]等。多元校正食品分析领域也得到了较为成功的运用^[16]。可以说,多元线性回归以及多个自变量与多个因变量的回归技术是分析化学未来的方向,其相关理论和技术必将成为分析化学的主流。本文在此不多讨论化学计量学的内容,主要介绍一元线性回归中加权回归的理论和其在分析化学领域的应用问题。

1 一元线性回归简介

如前所述,式(2) $y=a+b_1x_1+b_2x_2+\dots+b_nx_n+\varepsilon$ 中,如果自变量 x_1, x_2, \dots, x_n 只有一项,则方程变为最简单的形式: $y=a+bx+\varepsilon$,这就是因变量 y 关于自变量 x 的一元线性回归。在分析化学领域,自变量可以是一系列已知准确值的样本(通常是含量已知并带有测量不确定度的标准溶液), y 是利用仪器设备检测已知值样本获得的响应信号,可以是单个样本检测结果,也可以是样本的独立重复检测结果的

平均值, ε 是回归方程的误差项。在实际运用中, y 和 x 所表达的含义可以互换。

1.1 普通最小二乘法

在传统的分析化学领域,通常用分析仪器检测有确定已知值(或参考值)的样本,以获得系列数据对 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$,用最小二乘法获得参数 a, b 的估计值 \hat{a} 和 \hat{b} ,进而得到回归方程 $\hat{y}=\hat{a}+\hat{b}x$,用此方程估计未知值样品中目标组分的含量。最小二乘法的原理是^[9]针对每一组样本观测值 (x_i, y_i) ,考虑观测值 y_i 与其回归值 $E(y_i)=a+bx_i$ 的离差越小越好,定义离差平方和为:

$$Q(a, b)=\sum_{i=1}^n [y_i - E(y_i)]^2 = \sum_{i=1}^n [y_i - a - bx_i]^2 \quad \text{式(3)}$$

最小二乘法是寻找参数 a 和 b 的估计值 \hat{a} 和 \hat{b} ,使式(3)定义的离差平方和达到最小,即满足:

$$Q(\hat{a}, \hat{b}) = \min_{a, b} \sum_{i=1}^n [y_i - \hat{a} - \hat{b}x_i]^2 \quad \text{(4)}$$

按照式(4)所限定的条件,求出的 \hat{a} 和 \hat{b} 称为回归参数 a, b 的最小二乘估计,称 $\hat{y}=\hat{a}+\hat{b}x$ 为 $y_i (i=1, 2, \dots, n)$ 的回归拟合值或拟合值,称 $e_i=y_i-\hat{y}_i$ 为 $y_i (i=1, 2, \dots, n)$ 的拟合残差。

这种方法称为普通最小二乘估计^[9],简称普通最小二乘法(Ordinary least square, OLS)。运用该方法必须确保误差项 ε 满足3个条件^[17]:误差项 ε 是一个随机变量,对于每个观测值 y_i 来说,误差项的期望值 $E(\varepsilon)=0$; ε 是一个服从正态分布的变量,即 $\varepsilon \sim N(0, \sigma^2)$; ε 的方差 $D(\varepsilon)=\sigma^2$,即对于所有自变量 x_i 来说,每个 x_i 的方差都相同。此三个基本条件是采用普通最小二乘法建立回归模型的先决条件。

1.2 加权最小二乘法

然而,在实际工作中遇到的情况大多是普通最小二乘法前两个条件能得到满足,第三个条件即所有观测值的方差相等这个条件难以得到满足。在分析化学领域,研究者们已经发现观测值的方差会随着自变量浓度的增加而变大,即所谓的“异方差(Heteroscedasticity)”问题^[18-19],有的文献也叫“异方差性”^[9,20]。术语名称虽有差异,含义是一致的,即因为 y_i (或误差 ε_i)的方差不恒定且随 x 的变化而变化,此时估算回归参数 \hat{a} 和 \hat{b} ,其方差不再最小化^[21],不能满足式(4)的要求,数理统计学术语中,这种非恒定方差被称为异方差。如果有异方差性存在,仍旧用普通最小二乘法估计回归参数,将会出现三个问题,一是参数估计虽然是无偏的,但不是最小方差线性无偏估计;二是参数显著性检验失效,即在模型整体显著性检验中所构造的 F 统计量不再服

从 F 分布,以及对于单个参数显著检验所构造的 t 统计量也不再服从 t 分布,模型的 F 检验、 t 检验都不能再使用^[22];三是回归方程应用效果不理想,预测值与真值(或参考值)偏离较大。为了避免出现此三个问题,需要运用加权最小二乘法(Weighted least square, WLS)。此方法的理论及计算推导如下:前述式(3)为 OLS 的离差平方和,在等方差的条件下,平方和中每一项的地位是相同的。但是在异方差的条件下,平方和中每一项的地位是不相同的,误差项方差 σ_i^2 在式(3)平方和中的作用就大,因而 OLS 估计的回归直线就被拉向方差大的项,而小方差项的拟合程度就差。WLS 的思路为在平方和中加入一个适当的权数项 w_i ,以调整各项在平方和中的作用。WLS 离差平方和表达如式(5):

$$Q_w(a, b) = \sum_{i=1}^n w_i [y_i - E(y_i)]^2 = \sum_{i=1}^n w_i [y_i - a - bx_i]^2 \quad \text{式(5)}$$

WLS 估计就是寻找参数 a 和 b 的估计值 \hat{a} 和 \hat{b} ,使式(5)定义的离差平方和达到最小。如果所有项的权数都相等,即 w_i 都等于某个常数,则 WLS 就成为 OLS,也可以说 OLS 是 WLS 的特例。统计学上不难证明以下二式成立:

$$\hat{a}_w = \bar{y}_w - \hat{b}_w \bar{x}_w \quad \text{式(6)}$$

$$\hat{b}_w = \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2} \quad \text{式(7)}$$

上述二式中, $\bar{x}_w = \frac{1}{\sum \omega_i} \sum \omega_i x_i$ 为自变量的加权平均; $\bar{y}_w = \frac{1}{\sum \omega_i} \sum \omega_i y_i$ 为因变量的加权平均。

本文不再给出具体的计算推导过程,详细内容可参考相关的书籍和文章^[9-10,22-25]。回归分析中若要运用 WLS,重要的过程包括识别是否有异方差性存在(异方差检验)、运用加权回归及选择最佳权函数、拟合质量评价三个方面。本文将围绕这三个重要方面的研究及运用展开讨论。

2 一元线性加权回归应用过程

2.1 异方差性检验

计算一元线性回归方程时,不得不考虑数据是否存在异方差性,这需要进行异方差检验。数理统计学给出了检验异方差的许多方法,白雪梅^[26]列出了在经济领域回归分析中检验异方差的 9 种方法及其简要过程,分别是图示检验、帕克(Park)检验、巴特利特(Bartlett)检验、格莱泽(Glejser)检验、斯皮尔曼(Spearman)等级相关检验、戈德菲尔特-匡特(Goldfeld-Quandt)检验、布劳特培干-戈弗雷

(Breuschpangan-Godfrey)检验、怀特(White)检验、自回归条件异方差(ARCH)检验。国内对异方差性的研究和检验运用比较多的是林业和计量经济学领域,在分析化学领域几乎不涉及。林业领域如使用图示检验法判断异方差性的存在^[19];在计量经济学领域,有人做了 Spearman 等级相关检验的实践^[27-28],利用专用软件 EVIEWS 进行异方差性 Goldfeld-Quandt 检验、Glejser 检验、White 检验^[29],取得了良好的效果。

2.2 最佳权函数选择

权被定义为因变量误差项的倒数^[9],即 $\omega_i = \frac{1}{\sigma_i^2}$, σ_i^2 为因变量第 i 个观测值误差项的方差。也有将权定义为 $\omega_i = \frac{\sigma_0^2}{\sigma_i^2}$ 的^[22,24](其中 σ_0^2 为任意选定的常数),并从理论上论述了权的有关性质,通过推导观测值函数的方差(方差传播律)和协因数(协因数传播律),从而获得权的计算方法。就其实质来说,这两种定义没有差别。权的定义表明误差项方差较大的观测值赋予较小的权数,相反误差项方差较小的观测值赋予较大的权数。在实际运用中, σ_i^2 往往是未知的,如果知道误差项方差随自变量变化而变化,也就是可以表达为 $\sigma_i^2 = \kappa x_i^2$,其中 κ 为比例常数,那么权数就可以表示为 $\omega_i = \frac{1}{\kappa x_i^2}$ 。参数估计中,比例系数是可以消去的,所以权数为 $\omega_i = \frac{1}{x_i^2}$ 。因为 σ_i^2 是模型 $y=f(x)+\varepsilon$ 中误差项 ε 的方差, σ_i^2 和模型中自变量、因变量都是有关的,所以权数也可以表达为 $\omega_i = \frac{1}{y_i^2}$ 。在实际应用中权数可以有三种选择,即 $\omega_i = \frac{1}{\sigma_i^2}$ 、 $\omega_i = \frac{1}{x_i^2}$ 、 $\omega_i = \frac{1}{y_i^2}$,还可以把权数表达为 $\omega_i = \frac{1}{\sigma_i^m}$ 、 $\omega_i = \frac{1}{x_i^m}$ 、 $\omega_i = \frac{1}{y_i^m}$, m 为实数,所以有文献在讨论药物分析中最佳权函数选择时,列出了多种权函数^[18]。

怎样确定权函数,目前还没有统一的标准和方法,某些情况下是凭经验^[19],还可以构造自变量 x 关于残差平方 e_i^2 的某个近似函数 $f(x)$,用 $f(x)$ 替代 σ_i^2 ,从而构造出可求的权重 $\omega_i = \frac{1}{f(x)}$,继而得到加权最小二乘估计量。怎么选择权函数才能达到最佳,这个问题较早就有相关研究^[30-31]。最合适的权函数是针对某一个模型而不是某一类模型^[17],作者给出了结论,即以模型本身构造的权函数就是要寻找的最佳权函数;并对此命题给以证明:假设模型为 $y=f(x)+\varepsilon$,两边同时除以 $f(x)$ 得新模型: $y'=y/f(x)=1+\varepsilon/$

$f(x)=1+\varepsilon'$ 。不难证明以模型本身构造的权函数进行加权回归,新模型误差项的期望值为0,其方差为常数,即对所有 x_i 来说,每个 ε'_i 的方差都相同,满足等方差的条件。在实际应用中,多数人没有刻意的针对某一具体模型,并没有深入研究有哪些可供选择的权函数,以及如何得到最佳权函数,而是假设误差项方差与变量存在函数关系,从而确定权函数^[17,31-32]。从理论上讲,最佳权函数是通过一系列迭代计算获得的^[33-34]。利用统计学软件SPSS可以确定幂指数权函数 $f(x)=x_i^m$ 中 m 的最优取值,效果良好^[28]。在林业方面,对加权回归中权函数的选择也做了比较深入的研究^[20]。还有研究结果认为,通过用原函数自身对模型做数据变换,能够使模型满足等方差的要求,而且对任一形式多个变量的回归模型都很容易实现^[19]。

除了关于权函数的选择在分析化学领域研究不多,关于权函数选择后所得模型预测值之间的比较,分析化学领域也不如林业、计量经济学等领域研究得透彻。国内较早的时候,有文献介绍了一元加权回归的理论知识^[10,25]。实际运用研究中,少有考虑并探讨如何选择权函数,也没有探讨如何获得最佳权函数^[25,35-37]。国外药物分析领域对一元线性加权回归的应用研究相对多一些,但是涉及如何选择最佳权函数的也很少,只是在回归计算中直接给出几种权函数然后比较效果^[18,38-45]。

2.3 回归方程检验

通常情况下,还应对拟合得到的回归方程进行显著性检验,以确定是否可用于预测和分析^[9]。最常用的检验有 t 检验、 F 检验和相关系数的显著性检验。 t 检验用于检验回归系数的显著性,也就是检验自变量 x 对因变量 y 的影响程度; F 检验也是用于检验回归方程的显著性;相关系数的显著性检验也是一种 t 检验,用于判断自变量 x 与因变量 y 是否高度相关。此三种检验实际上是等价的,在实际应用中选择一种即可。值得注意的有三点,一是不用每次检验,在实际的检测工作中,首次运用某一检测方法时,可能在方法确认(或验证)阶段需要进行回归方程检验,后续的检测方法运用,无须再做这些繁琐的检验;二是唯相关系数 r 为单一判定要求欠妥,在分析化学领域普遍认为相关系数 r 越大,回归方程的质量就越好,故有关著作^[46]、规范性文件^[47]都要求 r 要达到小数点以后两个9、三个9甚至四个9以上,在分析化学中沿袭了这一要求,目前分析化学界对此问题是有争议的^[48],诚然不能否认 r 中9越多越好,但是这只是表明自变量 x 与因变量 y 呈高度相关而已,至于回归方程的拟合质

量如何,回归系数 r 是不能反映的,这个问题在国外政府机构和权威组织均有论述和规定^[49-51],与 r 相关的另一个参数 r^2 还能在一定程度上说明拟合质量, r^2 称为决定系数,是相关系数 r 的平方,也称为判定系数、确定系数,反映了因变量的变异中能用自变量解释的部分所占比例;三是对于加权回归方程,残差(剩余)平方和或等价的回归平方和、剩余方差(或剩余标准差)、 F 统计量不能直接使用^[33]。

分析化学领域可以引用分析化学领域回归方程的拟合质量检验方法,如拟合残差分析和是否通过原点的检验。拟合残差 e_i 定义为自变量 x_i 处的响应值与该点拟合值之差,即 $e_i=y_i-\hat{y}_i$, e_i 是误差项 ε_i 的估计值。用自变量与拟合残差作直角坐标图,可以方便地观察残差随自变量变化的情况,从而验证所使用的拟合方法对与否则;利用残差分析还可以判断各点的拟合残差是否出现异常,一般认为 e_i 超过 $\pm 2\hat{\sigma}$ 或 $\pm 3\hat{\sigma}$ 为异常值^[9]。如何更好地判断异常残差,何晓群等^[9]给出了详细的方法供参考,然而在分析化学领域鲜见残差分析技术的应用。一些国家标准中也给出了其它的检验拟合残差方法^[52-55]。关于校准直线是否通过原点的检验,如果方法要求校准直线要通过原点,则应进行检验,前述国家标准也给出了检验方法及其运用实例。

2.4 OLS和WLS拟合质量比较

各个应用领域从不同角度对这个问题开展了较多的研究,一种是基于一个效果好的加权回归方程,所占权重较大的自变量 x 所对应的因变量 y 的预测方差,应该比OLS的预测方差要小得多,因此可据此比较对同一组数据分别使用OLS和WLS的拟合效果,只不过权重较小的点,可能会出现相反的结果,但这种方法只能得出一个较粗的结论^[33]。比较一致的看法是^[19-20,31]:OLS以残差平方和最小为原则,结果是模型拟合时,较多地考虑了样本值比较大的点,而较少考虑样本值较小的点。而WLS则是以加权残差平方和最小为原则,给予所有样本点以同等的重视程度。另外,基于残差平方和得到指标剩余标准差 s 和相关系数 r 理所应当是OLS优于WLS,所以在实际应用中用参数 s 和 r 来比较OLS和WLS的拟合质量是不恰当的。为此,研究者们构建了总相对误差、总系统误差、平均相对误差、平均相对误差绝对值和预估精度等指标,来比较OLS和WLS的优劣,值得借鉴。还有研究介绍了运用非参数性方法讨论模型误差及特性,既简单快捷又不用考虑误差分布情况,但不知是否可以在分析化学中加以运用^[56]。

在药物分析领域,OLS和WLS拟合质量的比较

也得到了广泛的研究。如分别计算 OLS 和 WLS 各个拟合点的相对误差(拟合误差),以此比较两者的优劣^[57];利用所得的 OLS 和 WLS 回归方程,通过计算各个点自变量 x 的预测置信区间加以比较两者的优劣^[25,35-37]。国外有研究人员做了比较深入细致的工作^[18],设计了一套评价指标,用于比较采用几种不同权重进行 WLS 拟合所得的回归方程的优劣,有一定的参考价值。还有研究人员选择了 WLS 拟合中常见的 4 种加权回归和双对数加权回归、稳健回归,共 6 种替代方法与 OLS 进行了比较,发现所有校准直线的决定系数和预测系数没有显著差异。所有替代方法获得的定量检测下限(Lower limit of quantitation, LLQ)值均比通过 OLS 获得的 LLQ 值低,权重为 $\frac{1}{x^2}$ 和 $\frac{1}{y^2}$ 时,其校准直线的 LLQ 比用 OLS 得到的校准直线的 LLQ 低 3~10 倍^[45]。

3 存在的问题

在涉及分析化学的食品检验、环境检验、农产品检验等领域,应用加权回归方法拟合校准曲线是比较少的。这表现在三个方面:一是对加权回归的研究比较少,现有的研究成果比较简单、孤立,研究不深入、不全面;二是注重加权回归的应用意识还欠缺,很多回归分析是存在异方差性的,但是在实际检测中很少去主动识别异方差,而默认用普通最小二乘法得到校准曲线;三是有应用加权回归的想法但不知从何着手,有的仪器设备带有分析软件,软件中也有是否进行加权回归以及选择什么权数的选项,但是使用者不会用,而提供技术支持的仪器设备供应商也说不清道不明。上述现象的出现,究其原因有以下四点:

第一,数理统计技术在分析化学的普及和推广应用不多。回归分析是数理统计理论运用中非常重要的一部分,也是分析化学应用中最基本最重要的内容。然而由于数理统计理论在分析化学领域应用中的缺失,导致很多分析测试工作者不能将完整的回归分析技术与分析方法相结合并应用。

第二,分析活动中的质量控制薄弱。严格的说,一个采用回归分析的检测方法要不要识别异方差,并根据识别的结果决定是否采用加权回归进行校准曲线的拟合,这项工作应该在建立方法之初或方法确认(或验证)阶段进行。方法确认(或验证)是实验室质量控制中重要的活动之一,是建立内部质量控制体系的基础,这项系统工程在我国实验室十分薄弱。近几年来,检测行业质量管理水平提升迅速,特别是具备 CNAS、CMA 等资质的实验室逐

步认识到建立内部质量控制体系意义非凡。但是如何建立一个适合实验室自身的内部质量控制体系,还有待人们不断探索,不断认识与实践。由于实验室内部质量控制薄弱,导致对回归分析技术掌握不够,加权回归技术得不到应用也不足为奇。

第三,对极高含量和极低含量等极端检测结果的准确度重视不够。分析化学家们早就注意到,校准直线的底端和高端的检测结果,其准确度可能不能满足质量要求。但是在选择解决这个问题的方法上,主要采用调整工作范围和布点等,而疏忽了使用加权回归的方法。

第四,分析化学应用基础研究薄弱。回归分析是数理统计理论在分析化学领域的应用,是数理统计理论及其技术与分析化学技术的有机结合,属于边缘结合科学,需要既掌握分析理论及其应用技术,同时又熟悉数理统计理论及其应用技术的科学家,对此展开广泛的应用研究。然而,数理统计理论和应用技术在分析化学领域的应用很少,由此带来的研究需求就少,钻研这类边缘科学的人就少,最终的结果就是应用基础研究薄弱。

此外,上述原因分析还不够全面,可能还存在其他很多原因共同导致了这一局面的出现。

4 建议

与国外发达国家和地区相比,我国在加权回归的研究与应用方面还存在较大差距;与别的应用领域如计量经济、林业等领域比较,加权回归在分析化学领域的应用还存在较大差距。根据前述分析的原因,为加快分析化学领域回归技术的应用基础研究,并将研究成果推广到实际应用中,给出以下建议。

第一,在分析化学界普及数理统计理论,进行应用推广。高校数理统计和试验设计相关课程除了在分析化学专业设置以外,还应同时在检验检测相关的学科中设为必修课,以加强与日后实际工作的衔接;另外应大力开展职业教育培训,让现有的分析化学工作者更多的了解这门与数学有关的理论和技术。

第二,落实对分析化学实验室建立质量控制体系的要求,加强对实验室建立内部质量控制体系的指导。不管是国内还是国外,相应的管理标准和法规规章,都要求实验室必须建立内部质量控制体系。然而这些要求往往不能得到很好地落实,究其原因还是建立有效的内部质量控制体系对技术管理的要求相当高,实验室现有的能力还达不到要求。现在国内大多数实验室还未充分意识到方法

验证(确认)是内部质量控制的重要基础,对其技术性能指标的验证与确认还停留在表面,包括线性与校准中回归分析的正确运用。因此,加强对实验室的指导就必须提上政府管理部门和社会权威机构的议事日程。欧美日等发达国家和地区组织在这方面做得比较严格,实验室相应的政府管理部门和权威的社会机构会根据相关标准和行业的要求,结合行业管理的技术特点,制定满足这些要求的指南文件,并要求实验室遵照执行。我国可以借鉴其成功的经验。

第三,加大科研支持力度。分析化学测试虽然是分析化学领域的小分支,但却是不可轻视的一支,有关管理部门和机构不但要引导做好分析化学已有成果在分析化学领域的应用,同时还应未雨绸缪,引导广大分析化学工作者积极投入到数理统计技术在分析化学中的应用研究中,并给予必要且充分的支持和帮助。

参考文献

- [1] 国家质量监督检验检疫总局,中国国家标准化管理委员会. 统计学词汇及符号 第2部分:应用统计: GB/T 3358.2—2009 [S]. 北京: 中国标准出版社, 2010.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Statistics—Vocabulary and symbols—Part 2: Applied statistics: GB/T 3358.2—2009[S]. Beijing: Standards Press of China, 2010.
- [2] ISO. Statistics-Vocabulary and symbols-Part 3: Design of experiments: ISO 3534-3: 2013 [S]. Genève: International Organization for Standardization, 2013.
- [3] 焦红. 食品化学实验室质量控制国际指南[M]. 北京: 科学出版社, 2011.
- JIAO H. International guidelines of quality control in the process of food chemical detecting[M]. Beijing: Science Press, 2011.
- [4] 国家质量监督检验检疫总局,中国国家标准化管理委员会. 化学分析方法验证确认和内部质量控制 术语及定义: GB/T 32467—2015[S]. 北京: 中国标准出版社, 2016.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Verification and validation of methods and internal quality control on chemical analysis—Terms and definitions: GB/T 32467—2015 [S]. Beijing: Standards Press of China, 2016.
- [5] 国家质量监督检验检疫总局. 分析化学术语: GB/T 14666—2003[S]. 北京: 中国标准出版社, 2004.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Terms for analytical chemistry: GB/T 14666—2003 [S]. Beijing: Standards Press of China, 2004.
- [6] DANZER K, CURRIE L A. Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998) [J]. Pure and Applied Chemistry, 1998, 70(4): 993-1014.
- [7] 中华人民共和国国家质量监督检验检疫总局. 通用计量术语及定义技术规范: JJF1001—2011[S]. 北京: 中国质检出版社, 2012.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. General terms in metrology and their definitions: JJF1001—2011 [S]. Beijing: China Quality Press, 2012.
- [8] BALAZS A. International vocabulary of metrology-Basic and general concepts and associated terms[J]. Chemistry International—Newsmagazine for IUPAC, 2008, 30(6): 21-22.
- [9] 何晓群, 刘文卿. 应用回归分析[M]. 5版. 北京: 中国人民大学出版社, 2019: 5-6, 19.
- HE X Q, LIU W Q. Applied regression analysis[M]. Beijing: China Renmin University Press, 2019: 5-6, 19.
- [10] 倪永年. 分析化学测试中的加权线性回归[J]. 南昌大学学报(理科版), 1987, 11(4): 55-58.
- NI Y N. Weighted linear regression in chemical analysis tests [J]. Journal of Nanchang University (Natural Science), 1987, 11(4): 55-58.
- [11] 梁逸曾, 吴海龙, 俞汝勤. 分析化学手册·10·化学计量学(第三版)[M]. 北京: 化学工业出版社, 2016: 294-302.
- LIANG Y Z, WU H L, YU R Q. Handbook of analytical chemistry·10·Stoichiometry (Version 3)[M]. Beijing: Chemical Industry Press, 2016: 294-302.
- [12] 陈展光, 吴海龙, 村松丫, 等. 多元分辨校正模型与方法研究-无机多组份分析[J]. 分析科学学报, 1997, 13(1): 11-16.
- CHEN Z G, WU H L, CUN S Y, et al. Study on models and methods for multivariate resolution and calibration determination of multicomponent inorganic ions [J]. Journal of Analytical Science, 1997, 13(1): 11-16.
- [13] 谷惠文. 高阶仪器结合化学多维校正用于复杂体系定量分析研究[D]. 长沙: 湖南大学, 2016: 4-18.
- GU H W. Research on the quantitative analysis of complex systems using high-order instruments coupled with chemical multiway calibration methods [D]. Changsha: Hunan University, 2016: 4-18.
- [14] 张晓华, 吕敏明, 郑晶晶, 等. 化学计量学二阶校正方法结合高效液相色谱用于蜂蜜中10种酚酸类物质的快速定量分析[J]. 色谱, 2020, 38(5): 572-580.
- ZHANG X H, LYU M M, ZHENG J J, et al. Fast high-performance liquid chromatography quantification of ten phenolic acids in honey using chemometric second-order calibration method[J]. Chinese Journal of Chromatography, 2020, 38(5): 572-580.
- [15] 熊智新. 基于小波变换的化学谱图数据处理[D]. 杭州: 浙江大学, 2004: 17-32.
- XIONG Z X. Chemograms data processing through wavelet transform[D]. Hangzhou: Zhejiang University, 2004: 17-32.
- [16] 杨清华, 李跑, 杜国荣, 等. 食品中多农药残留分析研究进展[J]. 农产品加工, 2020(1): 68-72.
- YANG Q H, LI P, DU G R, et al. Research progress in analysis of multi-pesticide residues in food[J]. Farm Products Processing,

- 2020(1): 68-72.
- [17] 曾伟生. 关于加权最小二乘法中权函数的选择问题[J]. 中南林业调查规划, 1996, 15(1): 54-55, 60.
- ZENG W S. On the choice of weighting functions in a weighted least squares [J]. *Contral South Forest Inventory and Planing*, 1996, 15(1): 54-55, 60.
- [18] JAIN R B. Comparison of three weighting schemes in weighted regression analysis for use in a chemistry laboratory[J]. *Clinica Chimica Acta*, 2010, 411(3-4): 270-279.
- [19] 张会儒, 唐守正, 胥辉. 关于生物量模型中的异方差问题[J]. 林业资源管理, 1999(1): 47-50.
- ZHANG H R, TANG S Z, XU H. On heteroscedasticity in biomass model [J]. *Forest Resources Management*, 1999 (1): 47-50.
- [20] 曾伟生, 骆期邦, 贺东北. 论加权回归与建模[J]. 林业科学, 1999, 35(5): 5-11.
- ZENG W S, LUO Q B, HE D B. Research on weighting regression and modelling[J]. *Scientia Silvae Sinicae*, 1999, 35(5): 5-11.
- [21] ZIEGEL E R, NETER J, WASSERMAN W, et al. Applied linear statistical models[J]. *Technometrics*, 1992, 34(1): 121.
- [22] 刘明. 基于一元线性回归模型异方差对加权最小二乘法的考察[J]. 统计与决策, 2012(19): 11-14.
- LIU M. Weighted least squares based on univariate linear regression model heteroscedasticity [J]. *Statistics & Decision*, 2012(19): 11-14.
- [23] 刘明. 线性回归模型基础应用问题研究[M]. 北京: 中国统计出版社, 2016.
- LIU M. Research on basic application of linear regression model [M]. Beijing: China Statistics Press, 2016.
- [24] 王仲锋. 加权线性回归及其应用[J]. 长春工程学院学报: 自然科学版, 2010, 11(3): 118-121, 150.
- WANG Z F. Weighted linear regression and its use [J]. *Journal of Changchun Institute of Technology: Natural Sciences Edition*, 2010, 11(3): 118-121, 150.
- [25] 毕瑞锋, 张发玲. 加权最小二乘法线性回归模型参数的理论推导与计算实例[J]. 计量与测试技术, 2016, 43(2): 67-68.
- BI R F, ZHANG F L. Theoretical derivation and calculation example of the parameters for weighted least square linear regression model [J]. *Metrology & Measurement Technique*, 2016, 43(2): 67-68.
- [26] 白雪梅. 异方差性的检验方法及评述[J]. 东北财经大学学报, 2002(6): 26-29.
- BAI X M. Methods and reviews of heteroscedasticity testing [J]. *Journal of Dongbei Unirersity of Finance and Economics*, 2002 (6): 26-29.
- [27] 曹玉茹. 基于SPSS加权回归的回归分析条件适用性研究[J]. 统计与决策, 2019, 35(4): 89-92.
- CAO Y R. Study on the Applicability of Regression Analysis Conditions Based on SPSS Weighted Regression [J]. *Statistics & Decision*, 2019, 35(4): 89-92.
- [28] 林天水, 陈佩树. 一元线性回归中异方差的处理[J]. 统计与决策, 2015(21): 86-88.
- LIN T S, CHEN P S. Treatment of Heteroscedasticity in Monovariate Linear Regression [J]. *Statistics & Decision*, 2015(21): 86-88.
- [29] 张振强, 韦兰英. 线性回归模型异方差的诊断与修正: 基于EViews软件的实例分析[J]. 中国集体经济, 2008(22): 82-84.
- ZHANG Z Q, WEI L Y. Diagnosis and Correction of Heteroscedasticity in Linear Regression Model: Case Analysis Based on EViews Software [J]. *China Collective Economy*, 2008(22): 82-84.
- [30] 曾伟生. 再论加权最小二乘法中权函数的选择[J]. 中南林业调查规划, 1998, 17(3): 9-11.
- ZENG W S. On the Choice of Weight Function in Weighted Least Squares [J]. *Contral South Forest Inventory and Planning*, 1998, 17(3): 9-11.
- [31] 郑春茂. 加权回归及权函数的变换在生物量建模中的应用[J]. 华东森林经理, 2012, 26(2): 77-79.
- ZHENG C M. Application of weighted regression and weight function transformation in biomass modeling [J]. *East China Forest Management*, 2012, 26(2): 77-79.
- [32] 周顺华. 线性回归模型异方差的诊断分析[J]. 学园, 2014(20): 167-168.
- ZHOU S H. Diagnostic Analysis of Heteroscedasticity in Linear Regression Model [J]. *Academy*, 2014(20): 167-168.
- [33] 冯士雍. 加权回归分析的适用场合, 加权回归方程效果的比较及权的估计[J]. 数理统计与管理, 1983, 2(6): 11-18.
- FENG S Y. Applicable occasions of weighted regression analysis, comparison of effects of weighted regression equations and estimation of weights [J]. *Application of Statistics and Management*, 1983, 2(6): 11-18.
- [34] ISO. Capability of detection — Part 2: Methodology in the linear calibration case: ISO11843-2: 2000 [S]. Genève: International Organization for Standardization, 2000.
- [35] 朱引保, 赵迎路. 应用加权回归方法分析特优汾酒中总酸总酯对应规律[J]. 酿酒, 2001, 28(5): 46-48.
- ZHU Y B, ZHAO Y L. Applying Weighted Regression Method to Analyze the Corresponding Law of Total Acids and Esters in Teyou Fen Liquor [J]. *Liquor-making*, 2001, 28(5): 46-48.
- [36] 胡新华. 测定室内空气甲醛的加权回归直线与等权回归直线的对比[J]. 实验室科学, 2011, 14(4): 113-117, 120.
- HU X H. A contrast about the weighted regression line and the isobar regression line of the measurement about formaldehyde of indoor air [J]. *Laboratory Science*, 2011, 14(4): 113-117, 120.
- [37] 颜志明. 不等精密度加权回归直线的计算及误差[J]. 环境监测管理与技术, 1995, 7(2): 35-36.
- YAN Z M. Calculation and Error of Unequal Precision Weighted Regression Straight Line [J]. *The Administration and Technique of Environmental Monitoring*, 1995, 7(2): 35-36.
- [38] SURVE D H, JINDAL A B. Development and validation of reverse-phase high-performance liquid chromatographic (RP-HPLC) method for quantification of Efavirenz in Efavirenz-Enfuvirtide co-loaded polymer-lipid hybrid nanoparticles [J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2019, 175: 112765.
- [39] POIRIER J M, ROBIDOU P, JAILLON P. Determination of indinavir in plasma by solid-phase extraction and column liquid chromatography [J]. *Therapeutic Drug Monitoring*, 1999, 21

- (4): 404-410.
- [40] DUBEY S K, DUDELLELY S, JANGALA H, et al. Rapid and sensitive reverse-phase high-performance liquid chromatography method for estimation of ketorolac in pharmaceuticals using weighted regression[J]. *Indian Journal of Pharmaceutical Sciences*, 2013, 75(1): 89-93.
- [41] ROCKE D M, LORENZATO S. A two-component model for measurement error in analytical chemistry[J]. *Technometrics*, 1995, 37(2): 176-184.
- [42] NAGARAJA N V, PALIWAL J K, GUPTA R C. Choosing the calibration model in assay validation[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 1999, 20(3): 433-438.
- [43] ALMEIDA A M, CASTEL-BRANCO M M, FALCÃO A C. Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods[J]. *Journal of Chromatography B*, 2002, 774(2): 215-222.
- [44] SADRAY S, REZAEI S, REZAKHAH S. Non-linear heteroscedastic regression model for determination of methotrexate in human plasma by high-performance liquid chromatography[J]. *Journal of Chromatography B*, 2003, 787(2): 293-302.
- [45] SZABO G K, BROWNE H K, AJAMI A, et al. Alternatives to least squares linear regression analysis for computation of standard curves for quantitation by high performance liquid chromatography: applications to clinical pharmacology[J]. *The Journal of Clinical Pharmacology*, 1994, 34(3): 242-249.
- [46] 赵静. 现代仪器在食品分析中的应用[M]. 北京: 化学工业出版社, 2013.
- ZHAO J. Application of modern instruments in food analysis[M]. Beijing: Chemical Industry Press, 2013.
- [47] 国家药典委员会. 中华人民共和国药典[M]. 北京: 中国医药科技出版社, 2015.
- Chinese Pharmacopoeia Commission. Chinese pharmacopoeia[M]. Beijing: China Medical Science Press, 2015.
- [48] MULHOLLAND M, HIBBERT D B. Linearity and the limitations of least squares calibration[J]. *Journal of Chromatography A*, 1997, 762(1-2): 73-82.
- [49] IRISH NATIONAL ACCREDITATION BOARD (INAB). Guide to method validation for quantitative analysis in chemical testing laboratories[J]. *INAB Officer*, 2012(3): 1-36.
- [50] National Association of Testing Authorities. Guidelines for the validation and verification of quantitative and qualitative test methods: Technical Note 17-March 2012 [Z/OL]. [2022-06-05]. <https://www.docin.com/p-375511692.html>.
- [51] HIBBERT D B. Further comments on the (miss-) use of r for testing the linearity of calibration functions[J]. *Accreditation and Quality Assurance*, 2005, 10(6): 300-301.
- [52] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 化学分析方法验证确认和内部质量控制要求: GB/T 32465—2015[S]. 北京: 中国标准出版社, 2016.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Requirements for verification & validation of detection methods and internal quality control on chemical analysis: GB/T 32465—2015[S]. Beijing: Standards Press of China, 2016.
- [53] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 化学分析方法验证确认和内部质量控制实施指南 色谱分析: GB/T 35655—2017[S]. 北京: 中国标准出版社, 2018.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Guideline for verification and validation of chemical analysis method and internal quality control—Chromatography: GB/T 35655—2017[S]. Beijing: Standards Press of China, 2018.
- [54] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 化学分析方法验证确认和内部质量控制实施指南 基于样品消解的金属组分析: GB/T 35657—2017[S]. 北京: 中国标准出版社, 2018.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Guideline for verification and validation of chemical analysis method and internal quality control—Metal component analysis based on sample digestion: GB/T 35657—2017[S]. Beijing: Standards Press of China, 2018.
- [55] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 海洋监测规范 第2部分: 数据处理与分析质量控制: GB 17378.2—2007[S]. 北京: 中国标准出版社, 2008.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. The specification for marine monitoring—Part 2: data processing and quality control of analysis: GB 17378.2—2007[S]. Beijing: Standards Press of China, 2008.
- [56] 钟义山, 燕爱玲, 钟云志. 模型误差几种估计方法的比较及应用[J]. *西北林学院学报*, 2003, 18(4): 145-147.
- ZHONG Y S, YAN A L, ZHONG Y Z. Comparisons and application of several estimation methods for model error[J]. *Journal of Northwest Forestry University*, 2003, 18(4): 145-147.
- [57] 钟大放. 以加权最小二乘法建立生物分析标准曲线的若干问题[J]. *药物分析杂志*, 1996, 16(5): 343-346.
- ZHONG D F. Some problems on establishing standard curve of biological analysis by weighted least square method[J]. *Chinese Journal of Pharmaceutical Analysis*, 1996, 16(5): 343-346.