International Journal of Advanced Research in Chemical Science (IJARCS)
Volume 4, Issue 11, 2017, PP 32-60
ISSN No. (Online) 2349-0403
DOI: http://dx.doi.org/10.20431/2349-0403.0411005
www.arcjournals.org

Multiple Linear Regression: An Overview with Analytical and Physico-Chemical Applications

Julia Martín[1, *], Ana María Jiménez[2], María José Navas[2], María Ángeles Fernández-Recamales[3], and Agustin G. Asuero[2]

[1] *Department of Analytical Chemistry, Escuela Politécnica Superior, The University of Seville, 41011-Seville (Spain)*

[2] *Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, 41012-Seville (Spain)*

[3] *Department of Chemistry and Materials Science, Faculty of Experimental Sciences, University of Huelva, Campus de El Carmen, Huelva 21007 (Spain)*

Corresponding Author: *Julia Martín, *Department of Analytical Chemistry, Escuela Politécnica Superior, The University of Seville, 41011-Seville (Spain)*

Abstract: *An overview on multiple linear regression (MLR) is envisaged in this paper. All but the final section is devoted to a discussion of the basic concepts of MLR. The corresponding MLR equations are derived and presented in a useful form for computing. However, the entirely general matrix approach to least squares applicable to any linear regression situation is also envisaged. In the final section selected analytical and physicochemical applications are shown in tabular form. MLR is one of the most widely used statistical tool and found applications on a number of areas such as quantitative structure property relationships (QSPR), quantitative structure retention relationships (QSRR), quantitative structure-transformation relationships (QSTR), molecular linear free energy relationships (MLFER) and quantitative structure activity relationships (QSAR), solvent polarity and solvatochromic effects, parameter estimation methods, correction of spectral (matrix) interferences, prediction, modelling and optimization, Fourier transform near infrared spectroscopy (NIR-FT) and multicomponent spectrophotometric analysis, and in many other areas.*

1. Introduction

We have focused traditionally our attention on fitting a straight line to data [1-10], because this problem is very common in applied science. Nevertheless, multiple linear regression (MLR) analysis is one of the most widely used [11-18] of all statistical tools. It is a prolongation of the linear regression where one response is linked to a number of independent variables, being used in a variety of circumstances: i) when it is known from theoretical considerations in the matter that the relationship follows that form; ii) or when the exact relationship connecting y and the x's, either is unknown or is too complicated to be used directly, being thus presumed than an approach of this kind is suitable.

The numerical calculations necessary to carry out the least squares analysis of multivariable relationships are often lengthy and tedious. Nevertheless, a mode of to arrange the work is shown in this contribution allowing the easy computation by using modern electronic computer or a spreadsheet. Basic equations are written both in traditional and in matrix form. MLR is an essential part of the model-dependent optimization techniques [19-20] and has extensive application in many subject areas, as we will have occasion to check at the end of this contribution, in which in tabular form selected applications of the MLR method in various areas are shown.

We follow in this contribution the same scheme as in previous [1-2, 6-7] reviews and book chapters. First, we review sequentially the various aspects that make up MLR. Subsequently, we collect in tabular form a hundred of selected applications, paying special attention to those of the analytical and physical-chemical nature.

2. The Basic Model

For simplicity we shall consider first the case of two independent variables only [21-33]. Suppose that there is a relationship between the true value of a response η_i (mean value) and the value of independent variables x_{1i} and x_{21} (regressors); this can be expressed as a plane in three dimensions

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{1}$$

The generally adopted practice of denoting random variables by Greek letters such as beta and their realization by small letter of the Latin alphabet is followed, in order to maintain a clear distinction [5] between parameter and their estimates. It is called first order model with two independent variables (linear in the parameters and linear in the independent variables). The first subscript, 1 or 2, describes [27] the independent variable. The second, *i*, makes reference [27] to the observational unit from which the observations on *y* and the independent variables were taken.

The parameter β_1 reveals the average change [21, 29] in η per unit increase in x_1 when x_2 is held constant. Likewise, β_2 reveals the average change in the η [21, 29] when x_2 is changed one unity, when x_1 is held fixed. When the effect of x_1 on the mean response does not depend on the level of x_2, and correspondingly the effect of x_2 does not depend on the level of x_1, the two independent variables do not interact having additive effects. Thus, the first order regression model is devised for independent variables whose effects on the mean response are additive or do not interact. The parameters β_1 and β_2 are commonly called partial regression coefficients because they reflect the partial effect of one independent variable when the other independent variable is included in the model and is held constant. Regression (Eqn. (1)) involves three dimensional, so that the complete picture cannot be represented on graph paper. For usual representation a three-dimensional model or diagram is required. Many computer packages for experimental design have the facility to produce three-dimensional diagrams.

The linear additive model can be extended to include any number of independent variables,. The generalization to more parameter involves merely the writing out of longer but precisely similar expressions, i.e. *p*, with *p+1* parameters β_j (j=0,1,...p) being estimated in those cases [27] in which the linear model includes the intercept β_0 . Instead of a regression line, one has to deal with a regression surface at *k*=2 (and with a regression hypersurface at *k* >2). In the general case, this is a response surface. In constructing a response surface, the numerical values of independent variables (factors) are laid off on the coordinate axis of a factor space. Experimental conditions (x_{1i}, x_{2i}) (i=1,2,...k) have been run yielding observations y_1, y_2, ...y_k... then adding the experimental error (random error) to the hypothetical model and by dropping the *i*-suffix throughout we have the approximate model

$$y = \eta_i + \varepsilon = b_0 + b_1 x_1 + b_2 x_2 \tag{2}$$

The random variability of *y*'s is explained by imposing the component ε_i on the "pure" linear function $\beta_0 + \beta_1 x_1 + \beta_2 x_2$.

We assume that the independent variables x_1 and x_2 can be manipulated or observed error-free, and only the dependent variable *y* is corrupted by measurement (all the errors in the *y*, and none in the *x*'s). We shall assume that the repressors vectors are not linearly dependent. For *k*=2 this say these are not constant k_1 and k_2 that $k_1x_1 + k_2x_2 = 0$, for j=1,2,...n: if this were not true, x_1 and x_2 would be perfectly related and the parameters, β_1, β_2, would be confounded. It would them not possible to estimate the parameters separately but only some linear combination of them. When the errors affecting the x'$_s$ are errors of control instead of errors of measurement, i.e., errors made in attempting to set x_1, x_2,... equal to their respective nominal values x_1', x_2',...then the methods of this contribution are pertinent, if the errors made in adjusting x_1, x_2,... to their respective nominal values are mutually independent (or uncorrelated at least).

The ε_i have in addition, the following [1, 4, 8, 34] properties:

a) The error ε_i is independent of x_{1i} and x_{2i}.

b) The expected value of ε_i is zero; our observed *y* is an unbiased estimate of η, that is $E(y)= \eta$.

c) The ε are statistically uncorrelated, i.e., the expected population value of $\varepsilon_i\ \varepsilon_j$ for any pair of points ($i \neq j$) is zero, cov(ε_i, ε_j) = 0.

d) The variance of ε is σ_y^2, which remains constant for all values of x_1 and x_2 (homocedasticity), or in any case is known, at least up to a common scalar factor in all variances (heterocedasticity).

e) ε is a normally distributed random variable; measurement errors follow a Gaussian distribution.

The unknown constants of proportionality β_p (p=0,1,2) are p variously termed as parameters, constants or coefficients. The x_{ji} (j=1,2) may be called independent variables; predictor variables or just predictors while y may be referred to as the dependent variable, the predicted variable, the outcome measurement or the criterion. However, as long as we deal with random variables the sense of the terms “dependent" and "independent” is strictly defined. Therefore, it seems reasonable, when it comes to regression relationships, to replace the term "independent variable” with an explanatory variable [21] and the term “dependent variable” with the variable being explained. The use of the concepts of dependence and independence requires great caution.

3. Approximation

We have the model (1) and must somehow estimate the unknown parameters β_0, β_1 and β_2 by statistics (i.e. functions of the data); we will obtain a fitted equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \tag{3}$$

The usage of small roman letters b_0, b_1 and b_2, to indicate estimates of the parameters given by Greek letters β_0, β_1, and β_2 is standard [5]. However, the representation $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ for the estimates is also usual. It is then meaningful to ask what the true value of the parameters is, though by the imprecise nature of measurements we can never hope to determine it with absolute certainty. It is relatively easy to agree that the “better” the selection of the values b_i, the smaller the differences

$$e_i = w^{1/2}\left(y_i - \hat{y}_i\right) \tag{4}$$

which are called weighted residuals and are estimates of the measuring error ε_i in the sample under consideration. Parameters of the approximating function [35] are frequently derived using least squares methodology. Assuming that the x_i’s are precisely known (i.e. σ_{xi}^2=0), selection of the parameters of the function describing the approximate model is made from the viewpoint of minimization of the sum of squares of the weighted residuals

$$Q(p) = \sum w_i e_i^2 = \sum w_i \left(y_i - \hat{y}_i\right)^2 \tag{5}$$

where the w’s are a priori fixed weighting coefficient [1, 5-7, 34] measuring the importance of particular observations in the sum. If the weighted errors are normally distributed, then with $w_i=\sigma_0^2/\sigma_i^2$ corresponds to the maximum likelihood objective function [26-27, 36]. Statistically the principal advantage of the least squares criterion is that it produces best linear unbiased estimates. That is, among all possible unbiased estimators of the β’s that are expressible as linear combinations of the scores, only least squares estimators have the lowest possible variance across successive samples. A statistic is considered to be an unbiased estimator of a parameter if the mean of the sample distribution of the statistics equals the parameter.

4. Centering

From the fact that the point $\left(\overline{y}, \overline{x}_1, \overline{x}_2\right)$ where $\overline{y}$, $\overline{x}_1$, and $\overline{x}_2$, are the weighted means of the variables for the set of k observations

$$\overline{y} = \frac{\sum w_i y_i}{\sum w_i} \tag{6}$$

$$\overline{x}_1 = \frac{\sum w_i x_{1i}}{\sum w_i} \tag{7}$$

$$\bar{x}_2 = \frac{\sum w_i x_{2i}}{\sum w_i} \tag{8}$$

lies on the plane, it follows that by centering the data the model can be converted [22, 24, 32] in

$$y - \bar{y} = \beta_1 \left(x_1 - \bar{x}_1 \right) + \beta_2 \left(x_2 - \bar{x}_2 \right) + \varepsilon \tag{9}$$

This form takes advantage of the fact that for any least squares fit the constant b_0 is always of the form

$$b_0 = \bar{y} - \sum_{i=1}^{p} b_i \bar{x}_i \tag{10}$$

for p constants fitted. In this case we need to find only the coefficients b_0, b_1 and b_2. In any case, centering is also a necessary preliminary for obtaining the correlation matrix of the variables.

5. Scaling

It is possible to simplify the process of finding the equation to the regression plane by normalizing all the variables x_1, x_2 and y, by the use of equations [22, 24]

$$Z_i = \frac{x_i - \bar{x}_i}{\sqrt{\sum \left(x_i - \bar{x}_i \right)^2}} = \frac{x_i - \bar{x}_i}{\sum S_{ii}} \qquad \left(i = 1, 2 \right) \tag{11}$$

$$G = \frac{y - \bar{y}}{\sqrt{\sum \left(y_i - \bar{y} \right)^2}} = \frac{y - \bar{y}}{\sqrt{S_{yy}}} \tag{12}$$

making each new variable have zero mean and unit sum of weighted squares

$$\bar{Z}_i = \bar{G} = 0 \tag{13}$$

$$\sum w Z_i^2 = \sum w G^2 = 1 \tag{14}$$

We note that the coded quantities Z_i (i=1,2) (and G) are simply convenient linear transformations of the, original x_i's and y_i, respectively, and so expressions containing the Z_i, can always be readily rewritten in terms of the x_i's (and y).

Another very useful version of autoscaling in common practice [37-39] in the chemometrics literature implies measurements to be normalized in the way (z-transformation, normalized variables)

$$Z_i = \frac{\left(x_i - \bar{x}_i \right)}{s_i} \tag{15}$$

being s_i the standard deviation of x_i measurements.

This kind of transformation constitutes an essential part [22] of a good computer routine. The major advantages of coding are:

a) it reduces by one the size of the matrix to be inverted later, being on this way helpful when fitting data via a pocket calculator. In larger matrices, say 5x5 and higher this may often lead to the occurrence of round off errors, when the matrix is inverted, even when the work is performed in an electronic computer.

b) the numerical values implied in matrix manipulation [30] are smaller, particularly the products and sum of products, and therefore are simpler to dealt with and do not suffer as much from round-off errors.

c) it allows to detect easily the dependence in the normal equations; an important property of the Z_i's values is that their covariance matrix is the same [38] as the correlation matrix of the x_i's.

It is to be emphasised, however, that the geometric interpretation of the parameter estimates evaluated by coding means is generally different [24] from the interpretation of those parameter estimates using uncoded factor levels.

6. The Normal Equations

By using Eqns. |11| y |12|, the centered data given by Eqn. |9| are transformed to a new scale

$$G\sqrt{S_{yy}} = \beta_1\sqrt{S_{11}}\,Z_1 + \beta_2\sqrt{S_{22}}\,Z_2 + \varepsilon \tag{16}$$

the new model to be adjusted being of the form

$$G = \alpha_1 Z_1 + \alpha_2 Z_2 + \delta \tag{17}$$

where

$$\alpha_1 = \beta_1\sqrt{\frac{S_{11}}{S_{yy}}} \tag{18}$$

$$\alpha_2 = \beta_2\sqrt{\frac{S_{22}}{S_{yy}}} \tag{19}$$

are new coefficients to be estimated from the transformed data (G, Z_1, Z_2) and represents scaled form of the original coefficients β_1 and β_2, and δ=ε/√Syy. Note that equations for normalized variables has no free term.

Thus the model (17) relates the observations [23] to the known transformed data (G_i, Z_{1i}, Z_{2i}) i=1,2,…n and the unknown α_1 and α_2 by n equations.

Expressed in matrix form [22, 30, 33, 40-42]

$$G = Z\alpha + \delta \tag{20}$$

where G is the vector of observations (n x 1), Z is the matrix of independent variables Z_1 and Z_2 of known form (n x p), α is the vector of parameters to be estimated (p x 1) and δ is a vector of errors (n x 1)

$$G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \\ \vdots & \vdots \\ Z & Z_{22} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} \tag{21}$$

The least squares estimator of α has to minimize the weighted sum of the squares of the residuals.

$$\begin{aligned} Q(\alpha) = \delta' w^{-1}\delta = \sum w_i\left(G_i - \hat{G}_i\right)^2 = \left(G - \hat{G}\right)' W^{-1}\left(G - \hat{G}\right) = \\ \left(G - Z\alpha\right)' W^{-1}\left(G - Z\alpha\right) = G'W^{-1}G - 2\alpha'Z'W^{-1}G + \alpha'Z'W^{-1}Z\alpha \end{aligned} \tag{22}$$

where the prime signifies the transpose of the matrix (i.e. rows and columns interchanged) , and W^{-1} is the inverse of the matrix of errors, W, which is diagonal with unequal diagonal elements, and G which is diagonal with unequal diagonal elements, and G is the column vector of predicted values of G for given values of Z_1 and Z_2

$$W = \begin{bmatrix} 1/w_1 & & & 0 \\ & 1/w_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & & & 1/w_n \end{bmatrix} \tag{23}$$

Note in this case, the use of 0 to denote [22] a large triangular block of zeros.

Note that the fitted values $\hat{y}_i$ are obtained by evaluating

$$\hat{G} = Z\alpha \tag{24}$$

By differentiating Eqn. (22) with respect to α we obtain [42]

$$\frac{dQ(\alpha)}{d\alpha} = -2Z'W^{-1}G + 2Z'W^{-1}Z\alpha \tag{25}$$

and by equalling to zero we obtain the normal equations

$$\left(Z'W^{-1}Z\right)A = Z'W^{-1}G \tag{26}$$

where *Z'* is the transpose of matrix *Z*, and *A* is the matrix of parameter estimates a_1 and a_2 of α_1 and α_2, respectively, giving de minimum sum of squares of (weighted) residuals

$$\underset{Z'}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & & Z_{2n} \end{bmatrix}} \underset{W^{-1}}{\begin{bmatrix} w_1 & & 0 \\ & w_2 & \vdots \\ \vdots & & \\ 0 & & w_n \end{bmatrix}} \underset{Z}{\begin{bmatrix} Z_{11} & Z_{21} \\ Z_{12} & Z_{22} \\ \vdots & \vdots \\ Z_{1n} & Z_{2n} \end{bmatrix}} \underset{A}{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}} =$$

$$= \underset{Z'}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \end{bmatrix}} \underset{W^{-1}}{\begin{bmatrix} w_1 & & 0 \\ & w_2 & \vdots \\ \vdots & & \\ 0 & & w_n \end{bmatrix}} \underset{G}{\begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{bmatrix}} \tag{27}$$

The product of two matrices exists if and only if the number of rows in the second matrix [22, 30, 42] is the same as the numbers of columns in the first matrix. By carrying out the products $Z'W^{-1}Z$ and $Z'W^{-1}$ we obtain

$$\underset{E}{\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}} \underset{A}{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}} = \underset{C}{\begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}} \tag{28}$$

Where $r_{12} = r_{21}$ is the correlation between Z_1 and Z_2 and r_{iy} is the correlation between Z_i (*i=1,2*) and, so that from Eqn. (14)

$$\sum wZ_1^2 = \sum wZ_2^2 = 1 \tag{29}$$

and

$$r_{12} = r_{21} = \sum wZ_1Z_2 = \frac{\sum w(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{\sqrt{\sum w(x_1-\bar{x}_1)^2}\sqrt{\sum w(x_2-\bar{x}_2)^2}} = \frac{S_{12}}{\sqrt{S_{11}\,S_{22}}} \tag{30}$$

$$r_{iy} = \sum wZ_iG = \frac{\sum w(x_i-\bar{x}_i)(y-\bar{y})}{\sqrt{\sum w(x_i-\bar{x}_i)^2}\sqrt{\sum w(y-\bar{y})^2}} = \frac{S_{iy}}{\sqrt{S_{ii}S_{yy}}} \tag{31}$$

Transforming the regression problem into a form in which it involves correlations is good in general because it makes all the numbers in the calculation lie between -1 and 1. When numbers are all of this order the adverse effects of roundoff error are minimized. While the dangers are slight when only two variables are involved this is very important when many predictor variables are being manipulated in a computer.

The matrix $Z'W^{-1}Z=E$ is symmetric, that is the element in the i_{th} row and j_{th} column [22, 30, 42] is the same. The transpose is identical in every element to the original matrix, that is if $A'=A$, then the matrix is called [22, 30, 42] a symmetrical matrix.

By multiplying the matrix E by A we obtain the set of normal equations

$$\begin{aligned} a_1 + r_{12}a_2 &= r_{1y} \\ r_{12}a_1 + a_2 &= r_{2y} \end{aligned} \tag{32a,b}$$

The use of matrices in statistics has many advantages [43] such as:

a) It summarizes expressions and equations very compactly.

b) It facilitates our memorizing these expressions.

c) One the problem is written and solved in matrix terms, the solution can be applied to any regression problem [22] not matter how many terms are in the regression equation.

d) It greatly simplifies the procedures for deriving solution to multivariate problems.

7. Solving the Normal Equations

To solve the set of linear equations (Eqns. (26) or (28)) it is necessary to take the inverse of the matrix $E=Z'W^{-1}Z$. Give this new matrix the name D and call its elements D_{ij}. Although the procedure that follows is laborious [30], the knowledge of matrix D is essential for the evaluation of the precision of the estimated parameter values. The E array can be eliminated from the left side of Eqn. (30) if both sides are premultiplied by its inverse E^{-1}

$$A = E^{-1}C = DC \tag{33}$$

so that the inverse of a non-singular square matrix (same number of rows and columns) has the property

$$E^{-1}E = EE^{-1} = I \tag{34}$$

Where I is the identity matrix of ones in the diagonal elements and zero on the off diagonal elements (of order p). So we get

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{35}$$

and thus we have

$$\begin{bmatrix} D_{11} + r_{12}D_{21} + r_{13}D_{31} & D_{12} + r_{12}D_{22} + r_{13}D_{32} \\ r_{12}D_{11} + D_{21} + r_{23}D_{31} & r_{12}D_{12} + D_{22} + r_{23}D_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{36}$$

Solving the resulting set of linear equations derived from (36) and taking into account that if every element of a matrix has a common factor [22, 30] it can be taken outside the matrix we get

$$D=\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}=\frac{1}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \tag{37}$$

(conversely, if a matrix is multiplied by a constant *c*, every element of the matrix [22, 30] must be multiplied by *c*).

By multiplying the matrix *D* by *C* we obtain the estimated parameters a_1 and a_2

$$a_1=\frac{r_{1y}-r_{12}r_{2y}}{1-r_{12}^2} \tag{38}$$

$$a_2=\frac{r_{2y}-r_{12}r_{1y}}{1-r_{12}^2} \tag{39}$$

Before Eqn. (2) can be used for practical purposes, one should change back to the natural scale, using the equations

$$b_i=a_i\sqrt{\frac{S_{yy}}{S_{ii}}} \qquad (i=1,2) \tag{40}$$

$$b_0=\bar{y}-\sum_{i=1}^{i=2}b_i\bar{x}_i \tag{41}$$

8. Alternative Ways of Deriving the Normal Equations

The coefficient parameters of Eqn. (17) are estimated from the condition

$$Q=\sum w\left(G-E(G)\right)^2=\sum w\left(G-\alpha_1 Z_1-\alpha_2 Z_2\right)^2=\min \tag{42}$$

The condition for the minimum of the function *Q* is defined in the same manner as in the case of a function of a single variable

$$\frac{\partial Q}{\partial \alpha_1}=-2\sum w\left(G-\alpha_1 Z_1-\alpha_2 Z_2\right)Z_2=0 \tag{43}$$

$$\frac{\partial Q}{\partial \alpha_2}=-2\sum w\left(G-\alpha_1 Z_1-\alpha_2 Z_2\right)Z_1=0 \tag{44}$$

On multiplying out these two expressions, on rearrangement we obtain the normal equations in the form

$$a_1\sum wZ_1^2+a_2\sum wZ_1Z_2=\sum wGZ_1 \tag{45}$$

$$a_1\sum wZ_1Z_2+a_2\sum wZ_1^2=\sum wGZ_2 \tag{46}$$

which according to Eqns. (29), (30) and (31) gives Eqn. (28).

In general, least squares estimates are always such that the vector of residuals from the fitted least squares equation is normal to each of the regressor vectors [24]. By multiplying Equation (17) by $\sqrt{w}$ the weighted regression model has the form

$$G\sqrt{w}=\alpha_1\sqrt{w}\,Z_1+\alpha_2\sqrt{w}\,Z_2+f^* \tag{47}$$

where

$$f^*=\delta\sqrt{w} \tag{48}$$

Under (47) the data have to satisfy the following equation

$$G^{*}=\alpha_1 Z_1^{*}+\alpha_2 Z_2^{*}+f^{*} \tag{49}$$

Where $G^{*}$and Z_i^{*} are pseudovariates [24]

$$G^{*}=G\sqrt{w} \tag{50}$$

$$Z_i^{*}=Z_i\sqrt{w} \tag{51}$$

The variance of f^{*} is given by

$$Var\left(f^{*}\right)=Var\left(\delta\sqrt{w}\right)=wVar\left(\delta\right)=wVar\left(\frac{\varepsilon}{\sqrt{S_{yy}}}\right)=\frac{w}{S_{yy}}Var\left(\varepsilon\right)=\frac{w}{S_{yy}}\frac{\sigma^2}{w}=cte \tag{52}$$

It is possible to perform the weighted analysis by carrying out an ordinary unweighted least squares analysis using the pseudovariates.

9. The Variance-Covariance Matrix

Let α designate the column vector of the regression coefficients α_1 and α_2, and let the expected value of A be. Then

$$E\left[\left(A-\alpha\right)\left(A-\alpha\right)'\right]=E\left[\begin{pmatrix} a_1-\alpha_1 \\ a_2-\alpha_2 \end{pmatrix}\begin{pmatrix} a_1-\alpha_1 & a_2-\alpha_2 \end{pmatrix}\right]= E\left[\begin{pmatrix} \left(a_1-\alpha_1\right)^2 & \left(a_1-\alpha_1\right)\left(a_2-\alpha_2\right) \\ \left(a_2-\alpha_2\right)\left(a_1-\alpha_1\right) & \left(a_2-\alpha_2\right)^2 \end{pmatrix}\right] \tag{53}$$

The covariance between two random variables, a_i and a_j (with a joint distribution) is defined [11, 30, 33, 41] as the expectation value of the product of deviations of a_i and a_j from their expected values (true or population means) α_i and α_j, respectively. On the other hand, the variance is a special case [11] of the covariance of a random variable with itself. On this way, the variance-covariance matrix V(a) is given by

$$V\left(a\right)=E\left[\left(A-\alpha\right)\left(A-\alpha\right)'\right]=\begin{pmatrix} \sigma_{a_1}^2 & \operatorname{cov}\left(a_1,a_2\right) \\ \operatorname{cov}\left(a_2,a_1\right) & \sigma_{a_2}^2 \end{pmatrix} \tag{54}$$

Taking into account that $A=D\,C=DZ'W^{-1}$ and $\alpha=E(A)=DZ'W^{-1}E(G)$, we get

$$V\left(a\right)=E\left[\left(DZ'W^{-1}G-DZ'W^{-1}E\left(G\right)\right)\left(DZ'W^{-1}G-DZ'W^{-1}E\left(G\right)\right)'\right]= E\left[\left(DZ'W^{-1}\left(G-E\left(G\right)\right)\right)\left(DZ'W^{-1}\left(G-E\left(G\right)\right)\right)'\right]=E\left[DZ'W^{-1}G^0\left(DZ'W^{-1}G^0\right)'\right] \tag{55}$$

where $G^0$is a random vector with independent elements

$$G^0=G-E\left(G\right)=\begin{bmatrix} G_1-E\left(G_1\right) \\ G_2-E\left(G_2\right) \\ \ldots \\ G_n-E\left(G_n\right) \end{bmatrix} \tag{56}$$

Using the transpose rule and nothing that D and W^{-1} are symmetric

$$V(a)=E\left[\left(DZ'W^{-1}G^{0}G^{0\prime}W^{-1}ZD'\right)\right]=DZ'W^{-1}E\left[G^{0}G^{0\prime}\right]W^{-1}ZD' \tag{57}$$

On the other hand

$$E\left[G^{0}G^{0\prime}\right]=E\left[\begin{pmatrix} G_1-E(G_1)\\ G_2-E(G_2)\\ \cdots\\ G_n-E(G_n)\end{pmatrix}\begin{pmatrix} G_1-E(G_1) & G_2-E(G_2) & \cdots & G_n-E(G_n)\end{pmatrix}\right]=$$
$$\begin{pmatrix} \sigma^2_{G_1} & \mathrm{cov}(G_1,G_2) & \cdots & \mathrm{cov}(G_1,G_n)\\ \mathrm{cov}(G_2,G_1) & \sigma^2_{G_2} & \cdots & \mathrm{cov}(G_2,G_n)\\ \vdots & \vdots & \vdots & \vdots\\ \mathrm{cov}(G_n,G_1) & \mathrm{cov}(G_n,G_2) & \cdots & \sigma^2_{G_n}\end{pmatrix} \tag{58}$$

$\sigma_{Gi}{}^{2}$ is the variance of observation i and cov(G_i,G_j) is the covariance of observations i and j. In many circumstances the G_i may expect to be independent if they come from separate isolated [41], non-interfering measurements. When the G_i's are independent, uncorrelated, cov(G_i, G_j) for all $i \neq j$ is equal to zero.

In those cases in which the variances are not equal, the observations G_1, G_2,…G_k have variances σ^2/w_1, σ^2/w_2, …σ^2/w_k where σ^2 is unknown but the constants w_1, w_2,…, w_k that determine the relative accuracy of observations are known, and thus, we have

$$E\left[G^{0}G^{0\prime}\right]=\begin{pmatrix} \sigma^2/w_1 & & & \\ & \sigma^2/w_2 & & \\ & & \ddots & \\ & & & \sigma^2/w_n\end{pmatrix}=\begin{pmatrix} 1/w_1 & & & \\ & 1/w_2 & & \\ & & \ddots & \\ & & & 1/w_n\end{pmatrix}\sigma^2=w\sigma^2 \tag{59}$$

Then

$$V(a)=DZ'W^{-1}WW^{-1}ZD\sigma^2=D\sigma^2 \tag{60}$$

$W^{-1}W=I$. The matrix D is called an error matrix.

10. Sample Estimates of the Population Variances σG/Z^2 and σY/X^2

The sum of the squares of the residuals divided by its associated degrees of freedom may be taken as the sample estimate $s_{G/Z}{}^{2}$ of the population variance of residuals, $\sigma_{G/Z}{}^{2}$

$$s^2_{G/Z}=\frac{\sum w\left(G-\hat{G}\right)^2}{N-3} \tag{61}$$

One parameter is lost with a corresponding lost also in the data as the differences $y_i-\bar{y}$, i=1,2,…n denote only (n-1) separate pieces of information given that their sum is zero, whereas y_1, y_2,…y_n denote n separate pieces of information. The residual mean square estimates $\sigma_{G/Z}{}^{2}$ assuming our chemical model is adequate, but no otherwise.

Substituting the estimated value of $\hat{G}$ in Eqn. (72) and making operations we have

$$s^2_{G/Z}=\frac{\sum w\left(G-a_1Z_1-a_2Z_2\right)^2}{n-3}=\frac{1-2a_1r_{1y}-2a_2r_{2y}+a_1^2+a_2^2+2a_1a_2r_{12}}{n-3} \tag{62}$$

So that $\sum wZ_i^2 = \sum wG_i^2 = 1$ and $\sum wZ_1Z_2 = r_{12}$. Taking into account the set of normal equations (28) we get finally

$$s_{G/Z}^2 = \frac{1 - a_1r_{1y} - a_2r_{2y}}{n-3} \tag{63}$$

On the other hand, by applying Eqn. (12) to an experimental and a fitted G value we get

$$G_i - \hat{G}_i = \frac{y_i - \hat{y}_i}{\sqrt{S_{yy}}} \tag{64}$$

since the average of the $\hat{y}_i$ is the same of the $y's$

$$\hat{\bar{y}} = \frac{\sum w\hat{y}}{\sum w} = \frac{\sum (b_0 + b_1x_1 + bx_2)}{\sum w} = b_0 + b_1\frac{\sum wx_1}{\sum w} + b_2\frac{\sum wx_2}{\sum w} = b_0 + b_1\bar{x}_1 + b_2\bar{x}_2 = \bar{y} \tag{65}$$

By multiplying Eqn. (75) through $w^{1/2}$, summing and squaring, on rearrangement we get

$$\sum w_i\left(y - \hat{y}\right)^2 = S_{yy}\sum w_i\left(G_i - \hat{G}_i\right)^2 \tag{66}$$

$$s_{y/x}^2 = S_{yy}s_{G/Z}^2 \tag{67}$$

11. Precision of the Estimated Parameters Values A^1 and A^2

The information concerning the precision of the estimated parameter values is comprised in the variance-covariance matrix *V*(*a*). The product of $s_{G/Z}^2$ and the *D* matrix provides the estimated variance-covariance matrix

$$V(a) = \begin{pmatrix} s_{a_1}^2 & \text{cov}(a_1, a_2) \\ \text{cov}(a_2, a_1) & s_{a_2}^2 \end{pmatrix} = Ds_{G/Z}^2 = \frac{1}{1 - r_{12}^2}\begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix} s_{G/Z}^2 \tag{68}$$

Each of the upper left to corner right diagonal elements of *V*(*a*) is an estimated variance of the parameter estimates s_{ai}^2; those elements correspond to the parameter as they appears in the model from the left to right. The off diagonal of the matrix represent the covariance between a_1 and a_2 (or identically between a_2 and a_1):

$$s_{a_1}^2 = s_{a_2}^2 = \frac{1}{1 - r_{12}^2}s_{G/Z}^2 \tag{69}$$

$$\text{cov}\left(a_1, a_2\right) = \text{cov}\left(a_2, a_1\right) = \frac{-r_{12}}{1 - r_{12}^2}s_{G/Z}^2 \tag{70}$$

12. Precision of the Parameters in the Original Model: Collinearity

By applying the random error propagation law [11] to Eqn. (41) we get

$$s_{b_i}^2 = s_{a_i}^2\frac{S_{yy}}{S_{ii}} \qquad (i = 1, 2) \tag{71}$$

and taking into account Eqns. (69) and (67)

$$s_{b_1}^2 = s_{a_1}^2\frac{S_{yy}}{S_{11}} = \frac{1}{1 - r_{12}^2}\frac{S_{yy}}{S_{11}}s_{G/Z}^2 = \frac{1}{1 - r_{12}^2}\frac{s_{y/x}^2}{S_{11}} \tag{72}$$

$$s_{b_2}^2 = s_{a_2}^2 \frac{S_{yy}}{S_{22}} = \frac{1}{1-r_{12}^2} \frac{S_{yy}}{S_{22}} s_{G/Z}^2 = \frac{1}{1-r_{12}^2} \frac{s_{y/x}^2}{S_{22}} \qquad (73)$$

r_{12}^2 is the degree of non-orthogonality between x_1 and x_2. It measures the fraction of the variation in one independent variable that is accounted for by the variation in one independent variable that is accounted for by the other in an equation of the form $x_1 = A + B\, x_2$; when $r_{12}^2=0$ we have complete (linear) independence of x_1 and x_2 or "orthogonality". When $r_{12}^2 \neq 0$, we have some degree of dependence of x_1 and x_2 determine how much the variance of b_i (i=1,2) is inflated. The term $1/(1-r_{12}^2)$ is named the variance inflation factor (VIF), an indicator of collinearity [44-47], also called multicollinearity (the best known remedial procedure to dealt with collinearity is ridge regression). A VIF larger than 5 or 10 is generally considered large and is an indication that the corresponding coefficient is poorly estimated. Two key problems arise under collinearity: variable effects cannot be separated and extrapolation is likely to be seriously erroneous.

Extreme non-orthogonality has several undesirable consequences in least squares regression. The columns of the design matrix Z are nearly dependent, $Z'W^{-1}Z$ is nearly singular and the estimation equation for the regression parameters is ill-conditioned. Parameter estimates may be unstable (small changes in the data causes large changes), which may be unreasonable large (in absolute value) or have the wrong sign. Standard errors on estimates are inflated (reflected in their large variances), magnifying the effects of errors in the regression variables, leading easily to unreliable predictions. Though exact collinearity seldom occurs in real experimental situations [44], near-collinearity is a frequent occurrence in real life data.

Once the variances and covariances of a set of quantities are known, they may be used to evaluate [48] the variances and covariances of other quantities. The equation (18)

$$\operatorname{cov}(f_k, f_l) = \sum \frac{\partial f_k}{\partial x_i} \frac{\partial f_l}{\partial x_j} \operatorname{cov}(x_i, x_j) \qquad (74)$$

allows the evaluation of the covariance of functions f_k and f_l from the covariance of x_i and x_j; the summation is over all i and j.

If we apply Eqn. (74) to Eqn. (40) (i=1,2) we obtain

$$\operatorname{cov}(b_1, b_2) = \left(\frac{\partial b_1}{\partial a_1}\right)\left(\frac{\partial b_2}{\partial a_2}\right) \operatorname{cov}(a_1, a_2) = \frac{S_{yy}}{\sqrt{S_{11}S_{22}}} \operatorname{cov}(a_1, a_2) \qquad (75)$$

and then by combining Eqns. (75) and (70)

$$\operatorname{cov}(b_1, b_2) = \frac{-r_{12}}{1-r_{12}^2} \frac{S_{yy}}{\sqrt{S_{11}S_{22}}} s_{G/Z}^2 = \frac{-r_{12}}{1-r_{12}^2} \frac{s_{y/x}^2}{\sqrt{S_{11}S_{22}}} \qquad (76)$$

To obtain the variance of the constant b_0, from Eqn. (41)

$$s_{b_0}^2 = s_{\bar{y}/x}^2 + \bar{x}_1^2 s_{b_1}^2 + \bar{x}_2^2 s_{b_2}^2 + 2\bar{x}_1\bar{x}_2 \operatorname{cov}(b_1, b_2) \qquad (77)$$

Substituting estimates for $s_{\bar{y}/x}^2$

$$s_{\bar{y}/x}^2 = \frac{s_{y/x}^2}{\sum w} \qquad (78)$$

and taking into account that $s_{y/x}^2 = s_{G/Z}^2 S_{yy}$, we obtain

$$s_{b_0}^2 = s_{G/Z}^2 S_{yy}\left[\frac{1}{\sum w} + \frac{1}{1-r_{12}^2}\left[\frac{\overline{x}_1^2}{S_{11}} + \frac{\overline{x}_2^2}{S_{22}} - 2\overline{x}_1\overline{x}_2\frac{S_{12}}{S_{11}S_{22}}\right]\right] = s_{G/Z}^2 S_{yy}\left[\frac{1}{\sum w} + \frac{1}{1-r_{12}^2}\left[z_1^2 + z_2^2 - 2z_1z_2\,r_{12}\right]\right] \tag{79}$$

In most regression applications, a t value (t=b_i/s_{bi}) is calculated fro each independent variable, being often used, e.g., for testing significance. However, a joint confidence region (confidence ellipse) for β_1 and β_2 may be bounded. Please see Box et al. [24] for details. Residuals should be plotted in various ways in order to detect possible anomalies. The topic has also been treated [3, 22, 31] previously with detail by some authors.

13. Calculation of the Variance of any Fitted Value in the Scaled and Original Models

To obtain the variance of any fitted $\hat{g}$ value, since $\hat{g}$ is a linear combination of the random variables a_1 and a_2 $(\hat{G} = a_1Z + a_2Z_2)$ it follows that

$$s_g^2 = Z_1^2 s_{a_1}^2 + Z_2^2 s_{a_2}^2 + 2Z_1Z_2\operatorname{cov}(a_1, a_2) \tag{80}$$

and taking into account Eqns. (69) and (70)

$$s_g^2 = \frac{1}{1-r_{12}^2}\left[Z_1^2 + Z_2^2 - 2Z_1Z_2r_{12}\right]s_{G/Z}^2 \tag{81}$$

To obtain the variance of any fitted $\hat{y}$ value, s_y^2, we look at the Eqn. (11) in a slightly different form, namely

$$\hat{y} = \overline{y} + \hat{G}\sqrt{S_{yy}} \tag{82}$$

$$s_{\hat{y}}^2 = \frac{s_{y/x}^2}{\sum w} + S_{yy}s_G^2 \tag{83}$$

so that $\overline{y}$ and $\hat{G}$ are uncorrelated.

Thus the variance of the predicted mean value of y, $\hat{y}_0$ at a specified value of x_1 and x_2

$$s_{y_0}^2 = s_{G/Z}^2 S_{yy}\left[\frac{1}{\sum w} + \frac{1}{1-r_{12}^2}\left[Z_1^2 + Z_2^2 - 2Z_1Z_2r_{12}\right]\right] \tag{84}$$

This equation is identical to Eqn. (79). We may find the variance of b_0 as a particular case of variance of any mean estimated value of $\hat{y}_i$, in which x_i=0 and $Z_i = \overline{x}_i / \sqrt{S_{ii}}$.

If we are not concerned with the mean value $E_i(Y)$ which can be obtained for given values x_{1i} and x_{2i} but with the average deviation of a single measurement y_i from the mean $E(Y)$, then the variance of the difference

$$\Delta_i = y_i - E(y_i) \tag{85}$$

to be determined, is expressed as

$$s_{\Delta_i}^2 = s_{y_i}^2 + s_{\hat{y}_i}^2 \tag{86}$$

and then the variance of predicting a single new value of response at given x_1 and x_2 is equal to plus the variance of estimating the mean response at that point, that is

$$s_{\Delta}^2 = s_{\hat{y}}^2 \left[1 + \frac{1}{\sum w} + \frac{1}{1-r_{12}^2} \left[Z_1^2 + Z_2^2 - 2Z_1Z_2r_{12} \right] \right] s_{y/x}^2 \tag{87}$$

14. The Multiple Coefficient of Determination and the Coefficient of Multiple Correlation

The adequacy of the regression model in terms of fit is usually assessed by the magnitude of a summary statistics knows as the multiple coefficient of determination or R^2 value, which is defined as [22, 49-52]

$$R^2 = \frac{\sum w_i \left(\hat{y}_i - \bar{y} \right)^2}{\sum w_i \left(y_i - \bar{y} \right)^2} \tag{88}$$

The total sum of squares may be divided in two parts: a) the sum of squares due to the fitted equation; and b) the residual sum of squares. In terms of the original model

$$\sqrt{w_i}\left(y_i - \bar{y} \right) = \sqrt{w_i}\left(\hat{y}_i - \bar{y} \right) + \sqrt{w_i}\left(y_i - \hat{y}_i \right) \tag{89}$$

Squaring and summation over i gives

$$\sum w_i \left(y_i - \bar{y} \right)^2 = \sum w_i \left(\hat{y}_i - \bar{y} \right)^2 + \sum w_i \left(y_i - \hat{y}_i \right)^2 \tag{90}$$

the cross product vanish on the summation $2\sum w_i \left(\hat{y}_i - \bar{y} \right)\left(y_i - \hat{y}_i \right) = 0$

Thus

$$R^2 = 1 - \frac{\sum w_i \left(y_i - \hat{y} \right)^2}{\sum w_i \left(y_i - \bar{y} \right)^2} \tag{91}$$

It is a measure of the strength of correlation. The range of R^2 values is $0 \le R^2 \le 1$. The R^2 value will be close to one if the fitted regression model represents a good fit of the data. Similarly, a poor fit of the fitted regression model will result in a R^2 value near zero.

From Eqn. (12) and (75)-(76)

$$y_i - \bar{y} = G_i \sqrt{S_{yy}}$$

(92)

$$\hat{y}_i - \bar{y} = \hat{G}_i \sqrt{S_{yy}} \tag{93}$$

$$R^2 = \frac{\sum w_i \left(\hat{y}_i - \bar{y} \right)^2}{\sum w_i \left(y_i - \bar{y} \right)^2} = \frac{\sum w_i \hat{G}^2}{\sum w_i G^2} \tag{94}$$

but $\sum w_i G^2 = 1$ (Eqn. 14)

$$R^2 = \sum w_i \hat{G}_i^2 = \sum w_i \left(a_1 Z_1 + a_2 Z_2 \right)^2 = a_1^2 + 2a_1a_2r_{12} + a_2^2 = a_1\left(a_1 + a_2r_{12} \right) + a_2\left(a_1r_{12} + a_2 \right) \tag{95}$$

and taking into account the normal equations (33) and (34)

$$R^2 = \sum w_i \hat{G}_i^2 = a_1 r_{1y} + a_2 r_{2y} \tag{96}$$

The R^2 is invariant under non-singular transformation [53] of the original variables. This property implies that the same correlation will be obtained from the matrix of correlation as from the covariance matrix. The R^2 measure “the proportion of total variation about the mean of y_i values explained by the regression”. R^2 may adopt value as high as 1 (or 100 %) when all the (x_i) values are different. When data are replicated (different results being obtained), however, the value of R^2 cannot achieve 1 with independence of how well the model fits. As a matter of fact no model, however good, may account for the variation in the data [22-23] due to pure error. The square root of the coefficient of multiple determination is the coefficient of multiple correlation, R. From a strictly point of view, correlation is defined only for random variables and as the x_i’s values are predetermined, this name is not totally correct.

The sample value of R^2 is a biased estimate of the corresponding population coefficient. With samples of small size, the value of R must be corrected for the systematic error. The fewer the degrees of freedom, $\nu = n$-1, the more the degree of correlation given by the multiple correlation coefficient is overestimated. Unbiased estimates of R^2 is available namely, the correlated multiple coefficient of determination which is defined as a ratio of variances

$$\bar{R}^2 = 1 - \frac{residual\ variance}{y\ variance} = 1 - \frac{\dfrac{\sum w_i\left(y_i - \hat{y}\right)^2}{n-l-1}}{\dfrac{\sum w_i\left(y_i - \bar{y}\right)^2}{n-1}} = 1 - \frac{n-1}{n-l-1}\frac{\sum w_i\left(y_i - \hat{y}\right)^2}{\sum w_i\left(y_i - \bar{y}\right)^2} \tag{97}$$

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-l-1} \tag{98}$$

l is the number of coefficients in the regression equation. Note that the coefficient of determination does not indicate if the lack of (perfect prediction) fit is due to a wrong (inadequate) model used or to the purely experimental uncertainty.

15. Adequacy of the Model

Using an F test sometimes checks the adequacy of the model

$$F = \frac{sum\ of\ squares\ due\ to\ regression / \left(l-1\right)}{sum\ of\ residuals\ squares / \left(n-l\right)} \tag{99}$$

and taking into account Eqn. (91), the F value as a function of R^2 will be given by

$$F = \frac{R^2 / \left(l-1\right)}{\left(1 - R^2\right) / \left(n-l\right)} \tag{100}$$

Greater the value of R^2, greater the calculated value of F: when R^2=1, F=∞; and when R^2=0 then F=0.

16. The Three Parameter Model

For a model of the kind

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{101}$$

we get

$$D = \left(Z'W^{-1}Z\right)^{-1} = \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix} = \frac{\begin{pmatrix} 1 - r_{23}^2 & r_{13}r_{23} - r_{12} & r_{12}r_{23} - r_{13} \\ r_{13}r_{23} - r_{12} & 1 - r_{13}^2 & r_{12}r_{13} - r_{23} \\ r_{12}r_{23} - r_{13} & r_{12}r_{13} - r_{23} & 1 - r_{12}^2 \end{pmatrix}}{1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2\, r_{12}\, r_{13}\, r_{23}} \tag{102}$$

and then

$$\begin{aligned} a_1 &= r_{1y}D_{11} + r_{2y}D_{12} + r_{3y}D_{13} \\ a_2 &= r_{1y}D_{12} + r_{2y}D_{22} + r_{3y}D_{23} \\ a_3 &= r_{1y}D_{13} + r_{2y}D_{23} + r_{3y}D_{33} \end{aligned} \tag{103a,b,c}$$

Eqns. (40) and (41) being now valid with i=1,2,3.

17. Higher Order Models: Solving Simultaneous Over Determined (Least Squares) Equations System

The matrix approach to the solution of a set of simultaneous linear equations is entirely general. By expressing the regression problem in matrix notation a solution is obtained that is applicable to any linear regression situation, including the simple straight line. For an over determined set of equations *ZA*=*G*, the normal equations are given by Eqns. (26) or (27), $(Z'W^{-1}Z)A=Z'W^{-1}G$. By introducing [36] the matrix

$$Z_w = W^{-\frac{1}{2}}Z \tag{104}$$

and the corresponding vector

$$G_w = W^{-\frac{1}{2}}G \tag{105}$$

Eqn. (26) or (27) reduces to

$$Z'_w Z_w A = Z'_w G_w \tag{106}$$

leading to

$$A = \left(Z'_w Z_w\right)^{-1} Z'_w G_w \tag{107}$$

The normal equations -Eqn. (106)- can be solved by means [36, 54-57] of three basic methods. The first bases the calculation on the $Z_w'Z_w$ matrix: Gaussian elimination (and related methods, e.g. Gauss-Jordan), the sweep operator or the Cholesky decomposition. The second method avoiding forming the $Z_w'Z_w$ matrix works directly with Z_w by using *QR* decomposition (modified Gram-Schmidt, Househölder reflection or Givens rotation algorithms). The third method involves [39, 58-62] singular value decomposition (*SVD*).

Note that the inverse of a matrix developed in terms of its determinant offers an exact connection between elements of a matrix and those of its inverse (when it is exists, i.e. when it is non singular). The process of matrix inversion, however, is a relatively complicated one, as we have previously indicate in section 5, because using determinants is computationally tedious even for matrices of small order such as 4x4 or 5x5. Direct solution of the normal equations is not generally the best way to find the least squares solution. Product matrices of the form *Z'Z*, however, have unpleasant roundoff properties. Fortunately, more efficient methods used in computer program exist to produce readily the pseudo inverse, without inverting matrices.

The *SVD* is intimately related to the familiar theory of diagonalizing a symmetric matrix. The operation performed in *SVD* is sometimes referred to as eigenanalysis, principal component analysis, or factor analysis. It is also used for modern computations of principal partial least squares regression (PLSR). The major benefit from using SVD is that it can handle ill-conditioned matrices much better than the approach expressed in Eqn. (107). This is the most computationally extensive approach, but also it is the most numerically stable

18. Polynomials Models

MLR can also be used to solve polynomial regression problems

$$\eta = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m \tag{108}$$

By setting $x=x_1$, $x^2=x_2,\ldots,x^m=x_m$, a MLR model of order *m*+1 is obtained, which can be estimated as previously described.

19. Applications

MLR was first applied [63-69] in the field of medicinal and pharmaceutical chemistry in order to study QSAR, and in correlation analysis of organic reactivity, i.e. prediction of n-octanol water from structures, some interesting monographs being published on this respect [65-69], and an in the search of solute-solvent interactions of various kinds [63]. A survey of selected (mainly) analytical and physico-chemical applications [70-163] of MLR is given in Table 1. Papers are summarized in Table 1 in chronological order from 2017 downwards, as in other previous papers and contributions [2, 3, 6-7] from our laboratory. Emphasis is put on the most recent applications, 54 of the 96 applications selected date from 2000 onwards. The number of applications, however, seems to be unlimited.

We may indicate at first various kinds of functional relationships based on structure and response depending on the topic subject of research, e.g. property, retention, transformation, free linear energy, or activity:

i) Quantitative structure property relationships (QSPR), e.g. those dealing with solution kinetics [73], melting points [79], physico-chemical properties [84] and drug n-octanol water partition [103].

ii) Quantitative structure retention relationships (QSRR), e.g. those related with steroid [97], essential oil compounds [101], local anaesthetics [127], and apparent volume of distribution [108].

iii) Quantitative structure-transformation relationships (QSTR), e.g. papers [115, 118] on sulfonylurea derivatives.

iv) Molecular linear free energy relationships (MLFER), e.g. studies on prediction partition coefficients [122], solute-solvent interactions [125], specific descriptors [126], and empirical parameters of solvent potency [156].

v) Quantitative structure activity relationships (QSAR), e.g. those that cyclooxygenase inhibitor [112], effect of OATP1B1 transport [80], chromatographic studies [150], steric considerations [154], chance factors involved [157], carminative activity [154, 158], parameters in connection with correlation analysis [161], drug design [162], and mathematical aspects of the topic [163].

Another group of papers dealt with solvent polarity parameter and solvatochromic effects and solvatochromic parameters and solvent polarity scale [86, 92, 102, 110, 114, 135, 137, 144, 151]. Papers on parameter estimation methods have also been selected, e.g. redox potential of cytochrome C [105], apparent dissociation constants [116], overlapping acidity constants [134, 149], physico-chemical parameters [113, 119, 141], neutralization enthalpies [148], position and confidence limits of an extreme [155, 160], counting measurements [159], linear and non linear calibration response [133], and multiparameter models [153]. The field of prediction is also an important area, e.g. chromatographic retention indices [96, 99, 100, 120, 135], acute toxicity of phenol derivatives [74], some physico-chemical properties [84], apparent volume of distribution [108], partition coefficients [122], acidity constant values [123] fuel ignition quality by NMR [76], models for ginsenosides [100], and performance of MLR [139]. Solution kinetics [73, 142], modelling and optimization on ionisable compounds [111] and acid-base behaviour of solvent effects [117], analysis of IR and NIR-FT spectra [77, 85, 90], UV absorption spectra [81], correction of spectral interferences (matrix effect) [87, 93, 129] on molecular absorption and atomic emission techniques, and multicomponent spectrophotometric analysis of mixtures [94, 95, 107, 146] have also been the subject of study.

Table 1. *Some selected applications of Multiple Linear Regression*

[Ref.]	Chemical Problem	Authors
[70]	Setting water quality criteria for copper using MLR models: A complementary approach to the biotic ligand model	Brix et al., 2017
[71]	Soil organic carbon distribution in Mediterranean areas under a climate change scenario via MLR analysis	Olaya-Abril et al., 2017
[72]	A MLR model for the estimation of blue ballpoint pen ink dating by measuring the fading of ink with respect to time using UV–Vis spectrophotometry	Sharma and Kumar, 2017
[73]	Use of Monte Carlo method in addition to functional and individual weighting to overcome multicollinearity problems in MLR equations	Elvas-Leitao, 2016

	applied as QSPR. The method was applied to rate constants for the Menschutkin reaction between Et_3N and EtI in mono- and di-alcohols	
[74]	Prediction of acute toxicity of phenol derivatives using MLR approach for *tetrahymena pyriformis* contaminant identification in a median-size database	Dieguez-Santana et al., 2016
[75]	A note on the use of MLR in molecular ecology to assess the relative effects of genetic characteristics on individual fitness or traits, or how environmental characteristics influence patterns of genetic differentiation	Frasier, 2016
[76]	An improved model for the prediction of ignition quality of hydrocarbon fuels using ^1H nuclear magnetic resonance spectroscopy and MLR modelling. Cetane number and derived cetane number of 71 pure hydrocarbons and 54 hydrocarbon blends were utilized as a data set	Jameel et al., 2016
[77]	A new method named “consensus successive projections algorithm – MLR method” is proposed in order to make a full use of the useful information in the spectra	Chen et al., 2015
[78]	A method for the quantification of the sum of short chain chlorinated paraffins by gas chromatography-mass spectrometry. The method is suited to the special demands of environmental sediment analysis using an approach of sum determination by MLR	Geiβ et al., 2015
[79]	QSPR study on melting point of carbocyclic nitroaromatic compounds by MLR and artificial neural network	Wang et al, 2015
[80]	QSAR analysis of the effects of OATP1B1 transporter by structurally diverse natural products using a particle swarm optimization-combined MLR approach	Cao et al., 2014
[81]	Determination of thiophanate-methyl using UV absorption spectra based on MLR	Jiao et al., 2014
[82]	A method for estimating multivariate functional relationships between sets of measured oceanographic, meteorological, and other field data using MLR approach	Richter and Stavn, 2014
[83]	A data set of 1-adamantylthiopyridine analogues (1-19) with antioxidant activity was used for constructing QSAR models. MLR was employed for the development of QSAR models	Worachartcheevan et al, 2014
[84]	Develop of QSPR to predict characteristic properties of a series of 62 new glycerol derivatives, relevant to solvent classification and substitution uses using structural descriptor variables by MLR analysis	García et al., 2013
[85]	A high-throughput screening technology based on near-infrared spectroscopy for the rapid and accurate determination of algal biomass composition using MLR and multivariate linear regression analysis	Laurens and Wolfrum, 2013
[86]	The feasibility of colorimetric measurements coupled with multivariate data analysis to determine the empirical solvent polarity parameter $E_T(30)$	Shakerizadeh-Shirazi et al., 2013
[87]	Internal correction of spectral interferences and mass bias for selenium metabolism studies using enriched stable isotopes in combination with MLR	Lunne et al., 2012
[88]	Mathematical procedure of MLR in combination with on-line liquid chromatography applied for the measurement of Sr and Nd isotope ratios by multicollector ICP-MS in the presence of isobaric isotopes. The separation of Rb and Sr and the separation of Nd and Sm were accomplished by ion exchange chromatography with large volume injection	Rodriguez-Castrillón et al., 2012
[89]	Effects of E-beam irradiation on several food compositional parameters such as protein, fat, moisture, nitrate and nitrite content, as well as free amino acids and some of their decomposition products. To evaluate food modifications, principal component analysis and MLR statistical tools were used	Guillén-Casla et al., 2011
[90]	Fourier-transform infrared spectroscopy, followed by multivariate treatment of spectral data, is proposed to evaluate the oxidised fatty acid (OFA) concentration in virgin olive oil samples characterised by different oxidative status	Lerma-García et al., 2011
[91]	A new procedure using a student-friendly least-squares MLR technique utilizing a function within Microsoft Excel is described that enables students to calculate molecular constants from the vibronic spectrum of iodine	Cooper, 2010

[92]	Statistical analysis is applied to study the solvatochromic effects using the solvent parameters (regressors) influencing the spectral shifts in the electronic spectra	Dorohoi, 2010
[93]	Control of matrix interferences by MLR models in the determination of arsenic and lead concentrations in fly ashes by inductively coupled plasma optical emission spectrometry	Ilander and Väisänen, 2010
[94]	An exercise is described for the analysis of a mixture containing acetaminophen, aspirin, and caffeine using UV spectroscopy and HPLC. The concentrations of various components in a mixture are determined directly from the UV spectra using MLR, and then the concentrations of various components in a mixture are determined using HPLC data	Smith et al., 2010
[95]	A novel approach for the estimation of the concentration of chemical components through spectrophotometric measurements. It is based on the exploitation of the whole spectral information available in the original spectral data space by means of a MLR system	Benoudjit et al., 2009
[96]	QSRR for the prediction of Kováts retention indices of 180 alkylphenols and their derivatives using the MLR and support vector machine	Fatemi et al., 2009
[97]	Gas chromatographic QSRR of trimethylsilylated anabolic androgenic steroids by MLR and partial least squares	Fragkaki et al., 2009
[98]	A rapid and non-destructive method to evaluate the advanced oxidation of virgin olive oils (VOOs). An electronic nose based on an array of six metal oxide semiconductor sensors was used, jointly with MLR, to predict the oxidized fatty acid concentration in VOO samples characterized by different oxidative status	Lema-García et al. 2009
[99]	Uses of support vector machines, radial basis function neural networks and MLR methods to investigate the correlation between gas chromatography retention indexes and physicochemical descriptors for diverse organic compounds	Chen et al., 2008
[100]	MLR and artificial neural network retention prediction models for ginsenosides on a polyamine-bonded stationary phase in hydrophilic interaction chromatography	Quiming et al., 2008
[101]	QSRR for components of the essential oil of the plant *Bidens pilosa* Linn. var. A suitable set of molecular descriptors was calculated and the best-fitting descriptors were selected by using stepwise MLR and a genetic algorithm the selection of variables	Riahi et al., 2008
[102]	Pyridinium-N-phenolate betaine dyes as empirical indicators of solvent polarity	Reichardt, 2008
[103]	A QSPR study of n-octanol-water partition coefficients of some of diverse drugs using MLR	Ghasemi and Saaidpour, 2007
[104]	The usefulness of robust MLR techniques implemented in the expectation maximization framework in order to model successfully data containing missing elements and outlying objects	Stanimirova et al., 2007
[105]	Spectroelectrochemical determination of the redox potential of cytochrome C via MLR: An undergraduate instrumental analysis or biochemistry laboratory exercise	Whitaker et al., 2007
[106]	Some of the earlier proposed empirical equations used for retention modeling are tested in micellar liquid chromatography	Boichenko et al., 2006
[107]	Precision in multi-wavelength spectroscopic analysis with classical-least squares regression	Cabezon and Oliveri, 2006
[108]	Apparent volume of distribution for drug entities belonging to different chemical classes was studied using a quantitative structure pharmacokinetic relationship approach	Ghafourian et al, 2006
[109]	Micellar liquid chromatography for the determination of drug materials in pharmaceutical preparations and biological samples	Esteve-Romero et al., 2006
[110]	Empirically determination of the polarity of room-temperature ionic liquids by means of solvatochromic pyridinium N-phenolate betaine dyes	Reichardt, 2005
[111]	Considerations of the modelling and optimization of resolution of ionizable compounds in extended pH-range columns	Torres-Lapacio et al., 2005
[112]	Selection of the most important descriptors (taken as independent variables) to build QSAR models with MLR method	Lü et al., 2004
[113]	Determining the relative contribution of structural properties of aminoacids	Madden et al., 2004

	to the formation of beta sheets in proteins and predicting the properties of a molecule using parameters derived from IR spectroscopy by using graphing calculators	
[114]	A study was made to correlate an overall solute polarity descriptor (p) with several molecular parameters: excess molar refraction, dipolarity/polarizability, effective hydrogen-bond acidity and basicity, and McGowan volume, through the linear solvation model	Torres-Lapacio et al., 2004
[115]	Model development to predict transformation of sulfonylureas in different matrices using MLR	Berger, Muller and Eing, 2002
[116]	A mathematical model for calculating apparent acid dissociation constants (pK$_a$) in hydroorganic mixtures with respect to the concentration of organic solvent in a binary mixture	Jouyban et al., 2002
[117]	Expressions of Kamlet-Taft equations obtained by MLR applied to pK$_a$ values of adenine (pK$_1$, pK$_2$) and adenosine 3’, 5’-cyclic monophosphate (sodium salt)	Marqués et al., 2002
[118]	Developments of QSAR models between the structure of phenylurea herbicides and their transformation in different matrices	Berger et al., 2001
[119]	Comparison and description of high throughput methods to measure the properties: solubility, permeability, lipophilicity, pKa, stability and integrity fpr drugs discovery	Kerns, 2001
[120]	Prediction of retention factors of phenolic and nitrogen-containing compounds in reverse-phase liquid chromatography based on logP and pK$_a$ obtained by computational chemical calculation	Hani et al., 2000
[121]	Development of a method of ion-interaction chromatography for the simultaneous separation of 21 polar aromatic sulphonates using a Box-Cox transformation	Marengo, Gennaro and Gianotti, 2000
[122]	A previously published method for the prediction of MLFER descriptors is tested against experimentally determined partition coefficients in various solvent systems. Modified solvation equations for water−octanol and water−cyclohexane partition are presented, and their implications discussed	Platts et al., 2000
[123]	Application of the Hammett and Taft one-parameter model and Drago dual-parameter model to a very wide series of dissociation constants in methanol of carboxylic aliphatic acids, benzoic acid derivatives, phenols, protonated amines, anilinium and pyridinium derivatives	Bosch et al., 1999
[124]	Application of a general growth curve model with different covariance structures to assess the similarity of dissolution rates of several drug lots	Lee et al., 1999
[125]	Solute-solvent interactions in normal-phase liquid chromatography: a MLFER study	Oumada et al., 1999
[126]	Estimation of molecular free energy relation descriptors using a group contribution approach	Platts et al., 1999
[127]	A novel retention model that includes the hydrophobicity of compounds and the molar fraction of the charged form of compounds by means of MLR	Escuder-Gilabert et al., 1998
[128]	The technique of MLR is applied instead of the Yates’ method on factorial and fractional designs in experimental science according to Response Surface methodology. Several examples taken from the literature are analysed	González, 1997
[129]	An undergraduate laboratory experiment is designed for the simultaneous determination of both Co(II) and Cr(III) in unknown liquid mixtures based upon a bilinear least squares regression analysis of measured absorbance data	Pandey et al., 1998
[130]	A computer program for searching the best model for describing different experimental systems	Bohanec and Moder, 1997
[131]	Correct and incorrect use of MLR. A set of criteria useful to judge the quality of an experimental plan, before carrying out any experiment	Sergent et al., 1995
[132]	Study of the relationship between *pH*(s) and solvent composition, expressed as a fraction in order to assessing the presence of preferential solvation effects. The linear solvation energy relationships method has been applied	Barbosa and Sanz-Nebot, 1994
[133]	A calibration routine based upon the curve $y = ax\ln x + bx + c$ is presented, which describes the non linear behaviour, including electron capture, nitrogen-phosphorus and UV photometric detectors. The method gives comparable results to weighted linear regression with assays showing linear	Burrows and Watson, 1994

	concentration versus response relationship	
[134]	Develop of a new methodology for calculation of acid-base dissociation constants of monoprotic and diprotic compounds from absorptiometric data and pH measurements by MLR	Cladera et al., 1994
[135]	Correlation between the retention (log *k*′) values (using different columns and mobile phases) with the solute and mobile phase solvatochromic parameters for fifteen phenols	Rosés and Bosch, 1993
[136]	Relationship of partition coefficients of mono-substituted diazines and pyridines in different partioning systems	Yamagami et al., 1993
[137]	Relationships between E_T polarity and composition in binary solvent mixtures	Bosh and Roses, 1992
[138]	Use of MLR in order to reduce the number of absorbance measurements and provide a means for evaluating the precision of the results of the simultaneous determination of cobalt, copper, and nickel in solution by UV-vis spectroscopy	Dado and Rosenthal, 1990
[139]	The predictive performance of three commonly used biased regression methods and MLR (ordinary least squares) using classical model selection techniques are evaluated on five data sets published in the chemical and statistical literature	Kowalski, 1990
[140]	A method to evaluate the composition of a mixture of two different proteins from the amino acid composition by MLR with the aid of a computer program written in BASIC	Antoni and Presentini, 1989
[141]	Examination of the least squares method applied to the evaluation of physicochemical parameters with linearized equations	Ramos and Alvarez-Coque, 1989
[142]	Analysis of spectrally resolved kinetic data and time-resolved spectra by bilinear regression	Roman and Gonzalez, 1989
[143]	MLR approach by recasting the analysis of variance in order to compare the effects of several different treatments on a continuous variable of interest, applicable with missing data	Slinker and Glantz, 198
[144]	The Py scale of solvent polarities. The relative intensities I_1/I_3 of the vibronic bands of pyrene fluorescence in 94 solvents and the vapor phase are reported	Dong and Winnik, 1984
[145]	The octanol-water partition coefficient of aromatic solutes: the effect of electronic interactions, alkyl chains, hydrogen bonds, and ortho-substitution	Leo, 1983
[146]	A commonly used undergraduate experiment in multicomponent analysis which has been modified to include the standard addition method, thereby eliminating the need for MLR	Raymond et al., 1983
[147]	X-ray fluorescence determination of zinc, including the relative and absolute errors	Adam and Suchonel, 1982
[148]	The application of thermometric titrimetries to the determination of enthalpies of neutralization of diprotic and triprotic acids using a linear least-squares method	Mongay et al., 1982
[149]	A critical study of the linear least-squares method applied to the spectrophotometric determination of protonation constants of diprotic acids	Mongay et al., 1982
[150]	The rational bases, experimental techniques and conditions required for the chromatographic determination of the structural data of importance for studies on quantitative relationships between chemical structure and biological activity of drugs	Kaliszan, 1981
[151]	An empirical relationship between the eluant strength parameter ε^o and solvent Lewis acidity and basicity	Krygowski et al, 1981
[152]	Quantitative trace gas analysis by infrared spectroscopy	Haaland and Esterling, 1980
[153]	Joint parametric uncertainty intervals for parameters of a MLR model	Schwartz, 1980
[154]	The unexplained variation in the relationships between carminative activities and octanol-water distribution coefficients of various classes of compounds	Evans et al., 1979
[155]	Determination of the absorption maximum in wide bands in the spectrum of dimethylphtalate	Heilbronner, 1979
[156]	Possibilities of establishing reaction and absorption series using solvent-dependent standard reactions or standard absorptions of organic compounds. Particular attention is merited by the summary of the 24 most	Reichardt, 1979

	important empirical parameters of solvent polarity and the table of $E_T(30)$ values for 151 solvents	
[157]	Chance factors in studies of QSAR. In this regard, a critical distinction must be made between the number of variables screened for possible correlation and the number which actually appear in the regression equation	Topliss and Edwards, 1979
[158]	The unexplained variation in the relationships between carminative activities and octanol-water distribution coefficients of various classes of compounds	Evans et al., 1978
[159]	Methods are developed for calculating statistical uncertainties in the form of approximate confidence limits for analyses determined by calibration of counting experiments for which the calibration curve is linear	Schwartz, 1978
[160]	A precise method of determining absorption maxima where Gaussian functions occur based on a logarithmic transformation of the Gaussian equation	De la Zerda et al., 1975
[161]	Chromatographic parameters in correlation analysis of QSAR	Tomlinson, 1975
[162]	Utilization of operation schemes for analog synthesis in drug design	Topliss, 1972
[163]	A mathematical technique is suggested as a means of describing QSAR of a series of chemical analogs	Free and Wilson, 1964

20. Final Comments

An account to give sequentially a clear description of the MLR topic has been attended in this paper. This paper presents formulas useful in computing MLR, which led to an efficient method of computing. The entirely general matrix approach to least squares applicable to any regression situation has also been envisaged. The subject of MLR has enough importance as to devote a paper of this nature. Note that simple linear regression based on straight line has been reviewed and go on reviewing many times, but in spite of this, MLR usually receives minor attention. A number of selected references have been compiled in tabular form in order to advise the importance and wide range of application of the subject. References selected belong mainly to the fields of QSAR and related topics, solvent and solvatochromic effect and parameters, prediction in a variety of ways, parameter estimation methods and multicomponent analysis of mixtures from diverse analytical techniques.

References

[1] J. Martin, A. R. Gracia, and A. G. Asuero, Weighting and transforming data in linear regression, In Linear Regression: Models, Analysis and Applications, V. L. Beck (Ed.), New York: Nova Science Publishers, 2017, Ch. 1, pp. 1-68.

[2] J. Martin, and A. G. Asuero, Regression through the origin, In Linear Regression: Models, Analysis and Applications, V.L. Beck (Ed.), New York: Nova Science Publishers, 2017, Ch. 2, pp. 69-115.

[3] J. Martin, D. D. R. de Adana, and A. G. Asuero, Fitting models to data: residual analysis, a primer, In Uncertainty Quantification and Model Calibration, J. P. Hessling (Ed.), Rijeka: InTech, 2017, Ch. 7, pp.133-173.

[4] Sayago A., Boccio M. and Asuero A. G., Fitting straight lines with replicated observations by linear regression: The least squares postulates, Crit. Rev. Anal. Chem. 34(1), 39-50 (2004).

[5] Sayago A. and Asuero A. G., Fitting straight lines with replicated observations by linear regression. Part II. Testing for homogeneity of variances, Crit. Rev. Anal. Chem. 34(2), 133-146 (2004).

[6] Asuero A. G. and Gonzalez G., Fitting straight lines with replicated observations by linear regression. Part III. Weighting data, Crit. Rev. Anal. Chem. 37, 143-172 (2007).

[7] Asuero A. G. and Bueno J. M., Fitting straight lines with replicated observations by linear regression. Part IV. Transforming data, Crit. Rev. Anal. Chem. 41(1), 36-69, (2011).

[8] Asuero A. G. and Gonzalez A. G., Some observations on fitting a straight line to data, Microchem. J. 40, 216-225 (1989).

[9] Asuero A. G., Sayago A. and González A. G., The correlation coefficient: an overview, Crit. Rev. Anal. Chem. 36(1), 41-59 (2006).

[10] Gonzalez A. G., Herrador M. A., Asuero A. G. and Sayago A., The correlation coefficient attacks again, Accred. Qual. Assur. 11(5), 256-258 (2006).

[11] Asuero A. G., Gonzalez G. G., de Pablos F. and Gomez-Ariza J. L., Determination of the optimum working range in spectrophotometric procedures, Talanta 35, 531-537 (1988).

[12] Giacomino A., Abolino O., Malandrino M. and Mentasti E., The role of chemometrics in single and sequential extraction assays: a Review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques, Anal. Chim. Acta. 688, 122-139 (2011).

[13] Guo Y., Recent progress in the fundamental understanding of hydrophilic interaction chromatography (HILIC), Analyst 140, 6452-6466 (2015).

[14] Rajalahti T. and Kvalheim O. M., Multivariate data analysis in pharmaceutics: A tutorial review, Int. J. Pharm. 417, 280-290 (2011).

[15] Andrade J. M., Cal-Prieto M. J., Gómez-Carracedo M. P., Carlosena A. and Prada D., A tutorial on multivariate calibration in atomic spectrometry techniques, J. Anal. At. Spectrom. 23, 15-28 (2008).

[16] Reinholds I., Bartkevics V., Silvis I. C. J., van Ruth S. M. and Esslinger S., Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review, J. Food Compos. Anal. 44, 56-72 (2015).

[17] Krzywinski M. and Altman N., Multiple linear regression, Nature Methods 12(12), 1103-1104 (2015).

[18] Pirhadi S., Shiri F. and Ghasemi J. B., Multivariate statistical analysis methods in QSAR, RSC Adv. 5, 104635-104665 (2015).

[19] Montgomery D. C., Design and Analysis of Experiments, 9th ed., Wiley, New York (2017).

[20] Myers R. H., Montgomery D. C. and Anderson-Cook C. M., Response Surface Methodology. Process and Product Optimization Using Designed Experiments, 4th ed., Wiley, New York (2016).

[21] Kutner M., Nachtsheim C. and Neter C., Applied Linear Regression Models, 4th ed., McGraw-Hill, New York (1996).

[22] Draper N. R. and Smith H., Applied Regression Analysis, 3th ed., Wiley, New York (1998).

[23] Box G. E. P. and Draper N. R., Empirical Model-Building and Response Surfaces, Wiley, New York (1987).

[24] Box G. E. P., Hunter W. G. and Hunter J. S., Statistics for Experimenters, an Introduction to Design, In Data Analysis and Model Building, Wiley, New York (1978).

[25] De Levie R., Advanced Excel for Scientist Data Analysis, Cambridge University Press, Cambridge (2008).

[26] Cook D. and Weisberg S., Applied Regression including Computing and Graphics, Wiley, New York (1999).

[27] Rawling J. O., Pantula S. G. and Dickey D. A., Applied Regression Analysis. A Research Tool, 2nd ed., Springer-Verlag, New York (1998).

[28] Czermimki J., Iwasiewicz A., Paszek Z. and Sikorski A., Statistics in Applied Chemistry, Elsevier, Amsterdam (1990) pp. 291-297.

[29] Chatterjee S., Hadi A. and Price B., Regression Analysis by Example, 5th ed., Wiley, New York (2012).

[30] Deming S. N. and Morgan S. L., Experimental Design: a Chemometric Approach, 2nd ed., Elsevier, Amsterdam (1993).

[31] Tomassone R., Lesguoy E. and Miller C., La Régression, nouveaux regards sur une ancienne méthode statistique, Masson, Paris (1983).

[32] Kennedy J. B. and Neville A. M., Basic Statistical Methods for Engineers and Scientists, 2nd ed., Harper International Edition, New York (1976).

[33] Akhnazarova S. and Kafarov V., Experiment Optimization in Chemistry and Chemical Engineering, Mir, Moscu (1982).

[34] Bates D. M. and Watts D. G., Non Linear Regression Analysis and its Applications, Wiley, New York (1988).

[35] S. N. Deming, Linear models and matrix least squares in clinical chemistry, In Chemometrics, Mathematical and Statistics in Chemistry Ed. B. R. Kowalski, Dordrecht: Reidel, 1984, pp. 267-394.

[36] Seber G. A. F. and Lee A. J., Linear Regression Analysis, 2nd ed., Wiley, New York (2003).

[37] Sharaf M. A., Illman D. L. and Kowalski B. R., Chemometrics, Wiley, New York (1986).

[38] Massart D. L. and Kaufman L., The Interpretation of Analytical Chemical Data by the use of Cluster Analysis, Wiley, New York (1983).

[39] Mandel J., Use of the singular value decomposition in regression analysis, Am. Stat. 36, 15-24 (1982).

[40] Garden J. S., Mitchell D. G. and Mills W. N., Nonconstant variance regression techniques for calibration curve based analysis, Anal. Chem. 52, 2310-2315 (1980).

[41] Pattengill M. D. and Sands D. E., Statistical significance of linear least squares parameters, J. Chem. Educ. 56, 244-247 (1979).

[42] Wise B. M. and Gallagher N. B., An introduction to linear algebra, Crit. Rev. Anal. Chem. 28, 1-20 (1998).

[43] Harris R. J., A Primer of Multivariate Statistics, 2nd ed., Academic Press, Orlando, Fla., (1985), p. 56.

[44] Mandel J., The regression analysis of collinear data, J. Res. Nat. Bur. Stand. 90, 465-477 (1985).

[45] Stewart G. W., Collinearity and least squares regression, Stat. Sci. 2, 68-100 (1987).

[46] Korkhin A. S., Method for the estimation of regression parameters in the case of multicollinearity, Ind. Lab. (English Translation) 55, 1198-1205 (1989).

[47] Mansfield E. R. and Helms B. P., Detecting multicollinearity, Am. Stat. 36(3), 158-160 (1981).

[48] Sands D. E., Correlation and covariance, J. Chem. Educ. 54, 90-94 (1977).

[49] Kvalseth T., Cautionary note about R^2 , Am. Stat. 39, 279-285 (1985).

[50] Sheather S. J., A Modern Approach to Regression with R, Springer, New York (2009).

[51] H. Abdi, Multiple correlation coefficient, In Encyclopedia of Measurement and Statistics, Ed. N. Salkind, Ca, Sage: Thousand Oaks, 2007.

[52] Sonnergaard J. M., On the misinterpretation of the correlation coefficient in pharmaceutical sciences, Int. J. Pharm. 321, 12-17 (2006).

[53] Morrison D. F., Multivariate Statistical Methods, 2nd ed., McGraw-Hill, New York (1976).

[54] de Levie R., Linear least squares, the spreadshet, and Filip, Am. J. Physics 75(7), 619-629 (2007).

[55] Mark H. and Workman J., Chemometrics in Spectroscopy, Elsevier, Amsterdam (2007).

[56] Natrella M. G., Experimental Statistics, NBS Handbook 91, National Bureau of Standards, Washington DC (1963, reprinted 1966).

[57] Massart D. L., Vandeginste B. G. M., Buidens L. M. C., de Jong S., Lewi P. J. and Smeyers-Verbeke J., Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam (1997).

[58] Hendler R. W. and Shrager R. I., Deconvolution based on singular value decomposition and the pseudoinverse: a guide for beginners, J. Biochem. Biophys. Methods 28, 1-33 (1994).

[59] Kunral M. and Dowd P. A., Singular vale decomposition as an equation solver in co-kriging matrices, J. South. Afr. Inst. Min. Metall. 112, 853-858 (2012).

[60] Desa R. J. and Matheson I. B. C., A practical approach to interpretation of singular value decomposition results, Methods Enzymol. 384, 1-8 (2004).

[61] Shaw R., Debsarma S. and Kundu S., An algorithm for removing stoichiometric discrepancies in biochemical reaction databases, Current Sci. 103(11), 1328-1334 (2012).

[62] Jaumot J., Vives M. and Gargallo R., Application of multivariate resolution methods to the study of biochemical and biophysical processes, Anal. Biochem. 327, 1-13 (2004).

[63] Abraham M. H., Scales of solute hydrogen bonding –their construction and application to physicochemical and biochemical processes, Chem. Soc. Rev. 22, 73-83 (1993).

[64] Leo A. J., Calculating log P_{oct} from structures, Chem. Rev. 93, 1281-1306 (1993).

[65] Hansch C. and Leo A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley-Interscience, New York (1987).

[66] Kaliszan R., Quantitative Structure-Activity Chromatographic Relationships, Wiley, New York (1987).

[67] Bergman S. W. and Gittins J. C., Statistical Methods for Pharmaceutics Research Planning, Marcel Dekker, New York (1985).

[68] Shorter J., Correlation Analysis of Organic Reactivity with Particular Reference to Multiple Regression, Wiley, Chichester (1982).

[69] Hansch C., Quantitative SAR in Drug Design 1, In Medical Chemical Series, Ed. E. J. Ariens, 1971.

[70] Brix K. V., DeForest D. K., Tear L., Grosell M. and Adams W. J., Use of multiple linear regression models for setting water quality criteria for copper: a complementary approach to the biotic ligand model, Environ. Sci. Technol. 51, 5182-5192 (2017).

[71] Olaya-Abril A., Parras-Alcántara L., Lozano-García B. and Obregón-Romero R., Soil organic carbon distribution in Mediterranean areas under a climate change scenario via multiple linear regression analysis, Sci. Total Environ. 592, 134-143 (2017).

[72] Sharma V. and Kumar R., Dating of ballpoint pen writing inks via spectroscopic and multiple linear regression analysis: a novel approach, Microchem. J. 134, 104-113 (2017).

[73] Elvas-Leitao R. A., Fighting collinearity in QSPR equations for solution kinetics with the Monte Carlo method and total weighting, J. Braz. Chem. Soc. 27, 2070-2075 (2016).

[74] Dieguez-Santana K., Pham-The H., Villegas-Aguilar P. J., Le-Thi-Thu H., Castillo-Garit J. A. and Casañola-Martin G. M., Prediction of acute toxicity of phenol derivatives using multiple linear regression approach for Tetrahymena pyriformis contaminat identification in a median-size database, Chemosphere 165, 434-441 (2016).

[75] Frasier T. R., A note on the use of multiple linear regression in molecular ecology, Mol. Ecol. Resour. 16(2), 382-387 (2016).

[76] Jameel A. G. A., Naser N., Emwas A-H., Dooley S. and Sarathy S. M., Predicting fuel ignition quality using 1H NMR spectroscopy and multiple linear regression, Energy Fuels 30, 9819-9835 (2016).

[77] Chen K. L. X., Li L., Chen H., Ruan X. and Liu W., A consensus successive projections algorithm – multiple linear regression method for analysing near infrared spectra, Anal. Chim. Acta 858, 16-23 (2015).

[78] Geiβ S., Löffer D., Körner B., Engelke M., Sawal G. and Bachhausen P., Determination of the sum of short chain chlorinated n-alkanes with a chlorine content between 50% and 67% in sediment samples by GC-ECNI-MS and quantification by multiple linear regression, Microchem. J. 119, 30-39 (2015).

[79] Wang D., Yuan Y., Duan S., Liu R., Gu S., Zhao S., Liu L. and Xu J., QSPR study on melting point of carbocyclic nitroaromatic compounds by multiple linear regression and artificial neural network, Chemometr. Intell. Lab. Systems 143, 7-15 (2015).

[80] Cao D-S., Liu S., Fan L. and Liang Y-Z., QASR analysis of the effects of OATP1B1 transporter by structurally diverse natural products using a particle swarm optimization-combined multiple linear regression approach, Chemometr. Intell. Lab. Systems 130, 84-90 (2014).

[81] Jiao L., Dong D., Zheng W., Zhao X., Zhang S. and Shen C., Determination of thiphanate-methyl using UV absorption spectra based on multiple linear regression, Int. J. Light Electron Optics 125, 183-185 (2014).

[82] Richter S. J. and Stavn R. H., Determining functional relations in multivariate oceanographic systems: model II multiple linear regression, J. Atmospheric Ocean Technol. 31, 1663-1672 (2014).

[83] Worachartcheewan A., Nantasenamat C., Owasirikul W., Monnor T., Naruepantawart O., Janyapaisarn S., Prachayasittikul S. and Prachayasittikul V., Insights into antioxidant activity of 1-adamantylthiopyridine analogs using multiple linear regression, Eur. J. Med. Chem. 73, 258-264 (2014).

[84] García J. I., García-Marín H., Mayoral J. A. and Pérez P., Quantitative structure-property relationships prediction of some physic-chemical properties of glycerol based solvents, Green Chem. 15, 2283-2293 (2013).

[85] Laurens L. M. L. and Wolfrum E.J., High-throughput quantitative biochemical characterization of algal biomass by NIR spectroscopy; multiple linear regression and multivariate linear regression analysis, J. Agric. Food Chem. 61(50), 12307-12314 (2013).

[86] Shakerizadeh-Shirazi F., Hemmateenejad B. and Mehranpour A. M., Determination of the empirical solvent polarity parameter ET(30) by multivariate image analysis, Anal. Methods 5, 891-986 (2013).

[87] Lunne K., Martínez-Sierra J. G., Gammelgaard B. and Alonso J. I. G., Internal correction of spectral interferences and mass bias for selenium metabolism studies using enriched stable isotopes in combination with multiple linear regression, Anal. Bioanal. Chem. 402, 2749-2763 (2012).

[88] Rodríguez-Castrillón J. A., García-Ruiz S., Moldovan M. and Alonso J. I. G., Multiple linear regression and on-line ion exchange chromatography for alternative Rb-Sr and Nd-Sm MC-ICP-MS isotopic measurements, J. Anal. At. Spectrom. 27, 611-618 (2012).

[89] Guillén-Casla V., Rosales-Conrado N., León-González M. E., Pérez-Arribas L. V. and M. Polo-Díez L., Principal component analysis (PCA) and multiple linear regression (MLR) statistical tools to evaluate the effect of E-beam irradiation on ready-to-eat food, J. Food Compos. Anal. 24, 456-464 (2011).

[90] Lerma-García M. J., Simó-Alfonso E. F., Bendini A. and Cerretani L., Rapid evaluation of oxidized fatty acid concentration in virgin olive oil using Fourier-transform infrared spectroscopy and multiple linear regression, Food Chem. 124, 679-684 (2011).

[91] Cooper P. D., A simple and convenient method of multiple linear regression to calculate iodine molecular constants, J. Chem. Educ. 97, 7, 687-690 (2010).

[92] Dorohoi D-O, About the multiple linear regressions applied in studying the solvatochromic effects, Spectrochim. Acta A 75, 1030-1035 (2010).

[93] Ilander A. and Väisänen A., Control of matrix interferences by multiple linear regression models in the determination of arsenic and lead concentrations in fly ashes by inductively coupled plasma optical emission spectrometry, J. Anal. At. Spectr. 25, 1581-1587 (2010).

[94] Smith E. T., Belogay E. A. and Hoim T., Using multiple linear regression to analyze mixtures: an excel spreadsheet exercise for undergraduates, Chem. Educ. 15, 103-107 (2010).

[95] Benoudjit N., Melgani F. and Bouzgou H., Multiple regression system for spectrophotometric data analysis, Chemometr. Intell. Lab. Systems 95, 144-149 (2009).

[96] Fatemi M. H., Baher E. and Ghorbanzade'h M., Predictions of chromatographic retention indices of alkylphenols with support vector machines and multiple linear regression, J. Sep. Sci. 32, 4133-4142 (2009).

[97] Fragkaki A. G., Tsantili-Kakoulidou A., Angelis Y. S., Koupparis M. and Georgakopoulos C., Gas chromatographic quantitative structure-retention relationships of trimethylsilylated anabolic androgenic steroids by multiple linear regression and partial least squares, J. Chromatogr. A 1216, 8404-8420 (2009).

[98] Lerma-García M. J., Simó-Alfonso E. F., Bendini A. and Cerretani L., Rapid evaluation of oxidized fatti acid concentration in virgin olive oils using metal oxide semiconductor sensors and multiple linear regression, J. Agric. Food Chem. 57, 9365-9369 (2009).

[99] Chen H-F., Quantitative predictions of gas chromatography retention indexes with support vector machines, radial basis neural networks and multiple linear regression, Anal. Chim. Acta 609, 24-36 (2008).

[100] Quiming N. S., Denola N. L., Saito Y. and Jinno K., Multiple linear regression and artificial neural network retention prediction models for ginsenosides on a polyamine-bonded stationary phase in hydrophilic interaction chromatography, J. Sep. Sci. 31, 1550-1563 (2008).

[101] Riahi S., Ganjali M. R., Pourbasheer E. and Norouzi P., QSRR study of GC retention indices of essential-oil compounds by multiple linear regression with a genetic algorithm, Chromatographia 67(11/12), 917-922 (2008).

[102] Reichardt C., Pyridinium-N-phenolate betaine dyes as empirical indicators of solvent polarity: some new findings, Pure Appl. Chem. 80, 7, 1415-1432 (2008).

[103] Ghasemi J. and Saaidpour S., Quantitative structure-property relationship study of n-octanol-water partition coefficients of some of diverse drugs using multiple linear regression, Anal. Chim. Acta 604, 99-106 (2007).

[104] Stanimirova I., Serneels S., Van Espen P. J. and Walczak B., How to construct a multiple regression model for data with missing elements and outlying objects, Anal. Chim. Acta 581, 324-332 (2007).

[105] Whitaker G. T., Belogay E. A. and Smith E. T., Spectroelectrochemical determination of the redox potential of cytochrome c via multiple regression. An undergraduate instrumental analysis of biochemistry laboratory exercise, Chem. Educ. 12, 392-395 (2007).

[106] Boichenko A. P., Iwashchenko A. L., Loginova L. P. and Kulikov A. U., Heterocedasticity of retention factor and adequate modeling in micellar liquid chromatography, Anal. Chim. Acta 576, 229-238 (2006).

[107] Cabezon M. A. and Oliveri A. C., Precision in multi-wavelength spectroscopic analysis with classical-least squares regression, Chem. Educ.11, 394-401 (2006).

[108] Ghafourian T., Barzegar-Jalali M., Dastmalchi S., Khavari-Khorasani T., Hakimiha N. and Nokhodchi A., QSPR models for the prediction of apparent volume of distribution, Int. J. Pharm. 319, 82-97 (2006).

[109] Esteve-Romero J., Carda-Broch S., Gil-Agusti M., Carella-Pero M. E. and Bose D., Micellar liquid chromatography for the determination of drug materials in pharmaceutical preparations and biological samples, Trends Anal. Chem. 24, 75-91 (2005).

[110] Reichardt C., Polarity of ionic liquids determined empirically by means of solvatochromic pydridinium N-phenolate betaine dyes, Green Chem. 7, 339-351 (2005).

[111] Torres-Lapasio J. R., Alvarez-Coque M. C. G., Bosch E. and Rosés M., Considerations of the modelling and optimization of resolution of ionizable compounds in extended pH-range columns, J. Chromatogr. A, 1089, 170-186 (2005).

[112] Lü J. X., Shen Q., Jiang J. H., Shen G. L. and Yu R. Q., QSAR analysis of cyclooxygenase inhibitor using particle swarm optimization and multiple linear regression, J. Pharm. Biomed. Anal. 35, 679-687 (2004).

[113] Madden S. P., Wilson W., Dong A., Geiger L. and Mecklin C. J., Multiple linear regression using a graphing calculator. Applications in biochemistry and physical chemistry, J. Chem. Educ. 81, 903-907 (2004).

[114] Torres-Lapasio J. R., Alvarez-Coque M. C. G., Rosés M., Bosch E., Zissimos A. M. and Abrahm M. H., Analyses of a solute polarity parameter in reversed-phase liquid chromatography on a linear solvatation relationships basis, Anal. Chim. Acta 515, 209-227 (2004).

[115] Berger B. M., Müller M. and Eing A., Quantitative structure-transformation relationships of sulfonylurea herbicides, Pest. Manag. Sci. 58, 724-735 (2002).

[116] Jouyban A., Chan H. K., Clark B. J. and Acree Jr. W. E., Mathematical representation of apparent dissociation constants in aqueous-organic mixtures, Int. J. Pharm. 246, 135-142 (2002).

[117] Marqués I., Fonrodona G., Barço A., Guiteras J. and Beltran J. L., Study of solvent effects on the acid-base behavior of adenine, adenosine 3'5'-cyclic monophosphate and poly(adenylic) acid in acetonitrile-water mixtures using hard-modelling and soft-modelling approaches, Anal. Chim. Acta 471, 145-148 (2002).

[118] Berger B. M., Müller M. and Eing A., Quantitative structure-transformation relationships of phenylurea herbicides, Pest. Manag. Sci. 57, 1043-1054 (2001).

[119] Kerns E. H., High throughput physicochemical profiling for drug discovery, J. Pharm. Sci. 90, 1838-1858 (2001).

[120] Hani T., Kouizumi K. and Kinoshita T., Prediction of retention factors of phenolic and nitrogen-containing compounds in reverse-phase liquid chromatography based on log P and pKa obtained by computational chemical calculation, J. Liquid Chromatogr. Relat. Technol. 23, 363-385 (2000).

[121] Marengo E., Gennaro M. C. and Gianotti V., Chemometrically assisted simultaneous separation of 21 aromatic sulfonates in ion-interaction RP-HPCL, Chemometr. Intell. Lab. Systems 53, 57-67 (2000).

[122] Platts J. A., Abraham M. H., Butina D. and Hersey A., Estimation of molecular linear free energy relation descriptors using a group contribution approach. 2. Prediction of partition coefficients, J. Chem. Inf. Comput. Sci. 40, 71-80 (2000).

[123] Bosch E., Riked F., Rosés M. and Sales J., Hammett-Taft and Drago models in the prediction of acidity constant values of neutral and cationic acids in methanol, J. Chem. Soc., Perkin Transactions, 2, 1953-1958 (1999).

[124] Lee J. C., Chen D. T., Hung H. N. and Chen J. J., Analysis of drug dissolution data, Stat. Med. 18, 799-814 (1999).

[125] Oumada F. Z., Roses M., Bosch E. and Abraham M. H., Solute-solvent interactions in normal-phase liquid chromatography: a linear free-energy relationship study, Anal. Chim. Acta 382, 301-308 (1999).

[126] Platts J. A., Butina D., Abraham M. H. and Hersey A., Estimation of molecular free energy relation descriptors using a group contribution approach, J. Chem. Inf. Comput. Sci. 39, 835-845 (1999).

[127] Escuder-Gilaber L., Sagrado S., Villanueva-Camañas R. M. and Medina-Hernandez M. J., Quantitatie retention-structure and retention-activity relationship studies of local anesthetics by micellar liquid chromatography, Anal. Chem. 70, 28-34 (1998).

[128] González A. G., Two level factorial experiment designs based on multiple linear regression models: a tutorial digest illustrated by case studies, Anal. Chim. Acta 360, 227-241 (1998).

[129] Pandey S., McHale M. E. R., Coym K. S. and Acree Jr. W. E., Bilinear regression analysis as a means to reduce matrix effects in simultaneous spectrophotometric determination of CrIII and CoII. A quantitative analysis laboratory experiment, J. Chem. Educ. 75, 878-880 (1998).

[130] Bohanec S. and Moder M., A computer program for searching the best model for describing different experimental systems, Anal. Chim. Acta 340, 267-275, (1997).

[131] Sergent M., Mathieu D., Phan-Tan-Luu R. and Drava, G. Correct and incorrect use of multilinear regression, Chemometr. Intell. Lab. Systems 27, 153-162 (1995).

[132] Barbosa J. and Sanz-Nebot V., Preferential solvatation in acetonitrile-water mixtures. Relationship between solvatochromic parameters and standard pH values, J. Chem. Soc. Faraday Transact. 90, 3287-3292 (1994).

[133] Burrows J. L. and Watson K. V., Development and application of a calibration regression routine in conjunction with linear and non linear chromatographic detector responses, J. Pharm. Biomed. Anal. 12, 523-531 (1994).

[134] Cladera A., Tomás C., Estela J. M. and Cerdà V., New methodology for calculation of acid-base dissociation constants of monoprotic and diprotic acids with close pK values from absorptiometric data and pH measurements, Anal. Chim. Acta 286, 253-263 (1994).

[135] Rosés M. and Bosch E., Linear solvatation energy relationships in reverse-phase liquid chromatography. Prediction of retention from a single solvent and a single solute parameter, Anal. Chim. Acta 274, 147-162 (1993).

[136] Yamagami C., Takao N. and Fujita T., Hydrophobicity parameter of diazines III. Relationship of partition coefficients of monosubstituted diazines and pyridines in different portioning systems, J. Pharm. Sci. 82, 155-161 (1993).

[137] Bosh E. and Roses M., Relationships between E_T polarity and composition in binary solvent mixtures, J. Chem. Soc. 88, 3541-3546 (1992).

[138] Dado G. and Rosenthal J., Simultaneous determination of cobalt, copper, and nickel by multivariate linear regression, J. Chem. Educ. 67, 797-800 (1990).

[139] Kowalski K. G., On the predictive performance of biased regression methods and multiple linear regression, Chemometr. Intell. Lab. Systems, 9, 177-184 (1990).

[140] Antoni G. and Presentini R., A least-squares computer method for the determination of the molecular ratio of conjugates between two different proteins from the results of the amino acid analysis, Anal. Biochem. 179, 158-161 (1989).

[141] Ramos G. R. and Alvarez-Coque M. C. F., Examination of the least squares method applied to the evaluation of physicochemical parameters with linearized equations, Anal. Chim. Acta 220, 145-153 (1989).

[142] Roman E. S. and González M. C., Analysis of spectrally resolved kinetic data and time-resolved spectra by bilinear regression, J. Phys. Chem. 93, 3532-3536 (1989).

[143] Slinker B. K. and Glantz S. A., Multiple linear regression is a useful alternative to traditional analyses of variance, Am. J. Physiol. 255, R353-R367 (1988).

[144] Dong D. C. and Winnik M. A., The Py scale of solvent polarities, Canadian J. Chem. 62(11), 2560-2565 (1984).

[145] Leo A., The octanol-water partition coefficient of aromatic solutes: the effect of electronic interactions, alkyl chains, hydrogen bonds, and ortho-substitution, J. Chem. Soc. Perkins Transactions II 1(6), 825-838 (1983).

[146] Raymond M., Jochum C. and Kowalski B. R., Optimal multicomponent analysis using the standard addition method, J. Chem. Educ. 60, 1072-1073 (1983).

[147] Adam K. and Suchomel J., Über die absolute und relative Regression. Fres. Z. Anal. Chem. 310, 121-123 (1982).

[148] Mongay C., Garcia M. C. and Ramis G., Calculation of neutralization enthalpies from thermometric titrations of diprotic and triprotic acids using a linear least-squares method, Thermochim. Acta 56, 307-323 (1982).

[149] Mongay C., Ramis G. and Garcia M. C., A critical study of the linear least-squares method applied to the spectrophotometric determination of protonation constants of diprotic acids, Spectrochim. Acta A 38, 247-252 (1982).

[150] Kaliszan R., Chromatography in studies of quantitative structure activity relationship, J. Chromatogr. 220, 71-83 (1981).

[151] Krygowski T. M., Radomski J. P., Rzeszowiah A. and Wrona P. K., An emprirical relationship between the eluant strength parameter ε^o snd solvent Lewis acidity and basicity, Tetrahedron 37(1), 119-125 (1981).

[152] Haaland D. M. and Easterling R. G., Improved sensitivity of infrared spectroscopy by the application of least squares methods, Appl. Spectrosc. 34, 539-548 (1980).

[153] Schwartz L. M., Multiparameter models and statistical uncertainties, Anal. Chim. Acta 122, 291-301 (1980).

[154] Evans B., James K. C. and Luscombe D. K., Quantitative structure-activity relationships and carminative activity II. Steric considerations, J. Pharm. Sci. 68(3), 370-371 (1979).

[155] Heilbronner E., Position and confidence limits of an extremum. The determination of the absorption maximum in wide bands, J. Chem. Educ. 56, 240-243 (1979).

[156] Reichardt C., Empirical parameters of solvent polarity as linear free-energy relationships, Angewandte Chem. Int. Edit. 18(2), 98-100 (1979).

[157] Topliss J. G. and Edwards R. P., Chance factors in studies of quantitative structure-activity relationships, J. Med. Chem. 22(10), 1238-1244 (1979).

[158] Evans B., James K. C. and Luscombe D. K., Quantitative structure-activity relationships and carminative activity, J. Pharm. Sci. 67(2), 277-278 (1978).

[159] Schwartz L. M., Statistical uncertainties of analyses by calibration of counting measurements, Anal. Chem. 50, 980-985 (1978).

[160] de la Zerda J., de Milleri P. and Villaveces J. L., Precise determination of the absorption maxima in wide bands, J. Chem. Educ. 52, 415 (1975).

[161] Tomlinson E., Chromatographic parameters in correlation analysis of structure-activity relationships, J. Chromatogr. 113, 1-45 (1975).

[162] Topliss J. G., Utilization of operation schemes for analog synthesis in drug design, J. Med. Chem. 15, 1006-1011 (1972).

[163] Free S. M. and Wilson J. W., A mathematical contribution to structure activity studies, J. Med. Chem. 7, 395-399 (1964).

Citation: *J. Martin et al., "Multiple Linear Regression: An Overview with Analytical and Physico-Chemical Applications", International Journal of Advanced Research in Chemical Science (IJARCS), vol. 4, no. 11, pp. 32-60, 2017. http://dx.doi.org/10.20431/2349-0403.0411005*