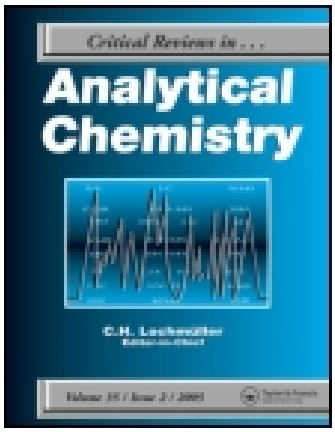

Critical Reviews in Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/batc20

Fitting Straight Lines with Replicated Observations by Linear Regression: The Least Squares Postulates

Ana Sayago [a], Maravillas Boccio [a] & Agustin Asuero [a]
[a] Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, Seville, Spain
Published online: 10 Aug 2010.

To cite this article: Ana Sayago, Maravillas Boccio & Agustin Asuero (2004) Fitting Straight Lines with Replicated Observations by Linear Regression: The Least Squares Postulates, Critical Reviews in Analytical Chemistry, 34:1, 39-50, DOI: 10.1080/10408340490273744

To link to this article: http://dx.doi.org/10.1080/10408340490273744

Critical Reviews in Analytical Chemistry, 34:39–50, 2004

ISSN: 1040-8347 print
DOI: 10.1080/10408340490273744

Fitting Straight Lines with Replicated Observations by Linear Regression: The Least Squares Postulates

Ana Sayago, Maravillas Boccio, and Agustin G. Asuero
Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, Seville, Spain

The results obtained by statistical techniques are valid if the assumed conditions are satisfied. Fitting straight lines with replicated observations by linear regression is considered in this article paying special attention to the compliance of the least squares postulates. Normality, robustness, independence, abscissa free from error, and proper weights are contemplated sequentially in this article. A detailed consideration of multiple measurements at one or more points is included, with the importance of genuine replicates as well the number of necessary replications being emphasized. The authors expect the results of this review to be of value to investigators and also in the teaching.

Keywords least squares postulates, replicated observations, straight lines

In 1923, Uhler (1) stated, "The method of least squares is so old and well known that the publication of a new article on this subject may well require some preliminary remarks in behalf of its justification." Anyway, several aspects of least squares, particularly in regard to (2–5) the use of replication, error analysis, and weighting and data transformations, appears to be poorly understood by a number of experimenters.

Statistical techniques are always based on assumptions, and the validity of results obtained through their use in practice always depends, sometimes critically, on the assumed conditions being met (6–8), at least to a sufficient degree of approximation. If statistically significant results are to be obtained, it is very important (9) that experiments be properly designed. Unfortunately, a percentage of analytical papers are married (10, 11) by inept or incorrect use of basic statistical methods or by failure to apply them to suspect data. Examples are given by Exner (12) from older and more recent literature where experimental data were processed in an incorrect way from the point-of-view of statistics. Our final aim is to describe the fitting of a straight line to a set of bivariate data in those cases in which replicate observations have been carried out, introducing the concepts without use of matrix algebra and without losing sight of the field of physical and chemical measurements. An easy understanding by analysts interested in improving the quality of their data processing may be obtained on this way.

Address correspondence to Agustin G. Asuero, Professor of Analytical Chemistry, Faculty of Pharmacy, University of Seville, 41012-Seville, Spain. E-mail: asuero@us.es

The first article of the series is devoted to the topic of the least squares postulates, paying special attention to multiple measurements at one or more points. Testing for homogeneity of variances and weighting and data transformation will be the subject of future articles. However, the authors expect the results of this critical review to be of value to investigators making use of these methods and also in teaching.

More often than not, simple models facilitate interpretation of complex physicochemical phenomena. If there is any competitor to the *t*-test in the application popularity stakes then it is probably (13) regression. In addition, perhaps the most common plea for assistance that the statistician receives from nonstatistical colleagues concerns (14) the fitting of a linear relationship to a set of bivariate data. As a matter of fact, however, the statistical methods most commonly misapplied by analytical chemists are (11) correlation and regression. There is no doubt about the importance of the regression topic, which is closely related to very basic operations, for example, calibration (15, 16) and the comparison of two analytical methods (17, 18) applied to a range of test materials. On the other hand, the introduction of regulations to control the production of foods and pharmaceutical products (19–22) and for environmental monitoring (23) has led to the great interest in method validation (24, 25) and error estimation (26, 27).

THE LEAST SQUARES POSTULATES

In analytical chemistry (28, 2) as well as in other quantitative sciences, it is often necessary to fit a mathematical equation or model to experimental data. Common situations that may be

described by functional relationships include calibration curves relating measured value of response to a property of material (21, 29–30), comparison of analytical procedures (17, 31–35), relationships in which time is the x-variate (36, 37), and parameter estimation methods (38–40). In this respect Deming (41) emphasized that some researchers still "fit the data to a model," which suggests a lack of scientific integrity. What is mean of course is that they fit a model to the data.

Models that take into account the possibility of uncertainty are called probabilistic or stochastic models, whereas those that do not allow for uncertainty are known as deterministic ones (4). The mechanistic model is a type of probabilistic model that is based on some known or presumed transformation between factors (variables chosen for study in an experimental plan) and responses. In this case, there is an exact mathematical formula (y as a function of x) relating the two variables and the only reason that the observations do not fit this equation is because of disturbances or errors in measurements of the observed value of one or both variables.

If mere estimates of parameters are necessary, any criterion (42) according to one's aesthetic preferences (e.g., the least sum of modules or the criterion of minimax) can be employed. The meaningful problem appears if reliability of parameters inferred and/or reliability of the model is needed. Parameters of the approximating function, however, are frequently derived using the least squares methodology, which dates from Gauss (43) and Legendre (44), priority being claimed by the first (45).

Under ideal conditions (46, 47), the method of least squares is the preferred method for fitting theoretical equations to experimental data, the weighted (30, 48) sum of squares of deviations of the observations from the fitted function (residuals) being as small as possible. It can be argued that the main advantage of least squares analysis is that it provides estimates of the uncertainties of the parameters. The statistical fitting of a straight line is generally referred to as linear regression, where the word "regression" has only (28, 49–50) historical meaning. However, straight-line least squares fitting and linear least-squares fitting are not necessarily synonymous. Linear least squares properly means that the parameters of the fitting model appear only linearly in it, no matter whether the model (51) is linear or nonlinear in its variables.

Least squares estimates can be determined using closed forms, and then it is straightforward to computationally determine this minimum. Least squares solutions (52) are unique (there is only a minimum), unbiased (parameter estimates do not contain any systematic error), consistent (when the sample size increases the solution converges toward the true population values), and efficient (their variance is finite and there is no solution with smaller variance among all unbiased linear estimators). Under certain constraints, the least squares parameters are those parameter estimates that have the highest probability to be correct. Then, least-squares parameters are (53) the maximum likelihood parameters.

To demonstrate that a least squares criterion is valid, it is necessary (54–56) to assume: (i) that the errors, ε_i, are random rather than systematic, with mean cero and variances $\sigma_i^2 = \sigma^2/w_i$ (where σ is a constant and w_i is the weight of point i) and follow a Gaussian distribution (this distribution is so common that is also referred to as the normal one); (ii) that the independent variable, that is, x, the abscissa, is known exactly or can be set by the experimenter either; (iii) the observations, y_i, are in an effective sense uncorrelated and statistically independent, that is, for $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, with means equal to their respective expectations or true values, $E\{y_i\} = \eta_i$; and (iv) that the correct weights, w_i, positive numbers, are known (this requires that the functional form of the dependence on x of the variance of y be known). The least squares criterion gives indeed poor results, however, if the observations are incorrectly weighted or if the data contain "outliers," that is, very poor observations at higher frequency than allowed (57) for by the normal distribution.

The general case of nonuniform precision or heteroscedasticity is assumed, where the measuring quantity is determinable nor with constant (i.e., homoscedasticity condition), but with different variance dependent (58) on the size of the measuring quantity. In fact in many phenomena, as the level (the value of a factor in an experimental plan) of the "signal" increases, the level of the noise increases. A much more consistent approach can be obtained on this way, thus making (59) in addition the use of a variance model to define the weights possible. Although the weighted approach is quite well known, the applications in analytical chemistry are not widespread, presumably since these types of statistics and the way they are treated in statistical software (5, 60) are not attractive to most analytical chemistry.

Many researchers are not even aware that their published data have grossly violated at least one of the assumptions inherent in regression modeling. Unfortunately, deviations from the conventional assumptions are the rule in chemical and physical analysis. Not only do we rarely have exact information concerning the functional relationship, but systematic error (61) is generally present, and there is evidence that observations follow distributions which are longer tailed than normal. In fact one characteristic of most chemical analysis, which counteracts the effectiveness of many of the classical methods of assumption testing, is (62) the small number of samples that may be processed. It is well known that the power of statistical method for detecting small discrepancies is severely limited by the number of observations.

There is no way to unequivocally prove that the requirements assumed in least squares calculations are met in any given situation. The only thing that can be done is to look for departures from the necessary requirements. This should always be done before applying statistical treatment to any data set. However, if no significant violations are found, there is no reason to believe that are violations (note that this is different from saying that the requirements have been met). The statistician calls this the test of the null hypothesis. In other words, is the concept (63, p. 16) "innocent until proven guilty." Fisher stated (64) in this context: "Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis."

When the conditions are met, the parameter estimates found by minimization of a least squares criterion are (39, 65–66) best unbiased linear estimates of the regression parameters. The term *best estimates* means that any linear combination of these estimates has the smallest variance of all linear unbiased estimates. An estimate a of a parameter α is said to be unbiased if its expected value is equal to the population value at any sample size, $E\{a\} = \alpha$. The term *linear estimates* means that they can be written as a weighted linear combination of measurements y_i.

NORMALITY

Observed data possess (67) an empirical distribution which may conform to a variety of probability distributions (e.g., normal, Poisson, binomial, lognormal). The least squares estimator is the optimal one for data with normal error distribution. If this assumption is violated, this estimator can perform very poorly. The concern in testing for distributional assumption should be (68) whether or not it is reasonable to approximate the data with the model, not whether the data came from the hypothesized distribution. A process is denoted as parametric when we use an appropriate distribution to describe (69) empirical data. However, one aspect of data analysis in which current practice is inadequately supported by experimental evidence is the assumption that observations are normally distributed. This assumption is critical because departures from it can cause the least squares method to give poor results, even with correct weighting. The assumption that responses are distributed as a normal distribution is frequently made (51) to calculate confidence intervals, tests for significant effects, or make additional data comparisons. Tests for normality exist (70, 71), for example, skewness, chi-square, Anderson-Darling, Shapiro-Wilk, but they require more observations than are generally made in usual experiments. Note that normality is particularly difficult to diagnose in small samples. Even when ample data are available, it is most usual for any tests of normality to be applied. Thus the condition of normality is generally assumed (72) by the analyst without checking, unless there are (73) theoretical objections or empirical indications to the contrary.

The most commonly used probability distribution in medicine is (74) the Gaussian, and much biological data are adequately represented by it. However, small deviations from the Gaussian distribution may be expected in the majority of real distribution of analytical results. Tukey and McLaughin (75) have suggested that the normal distribution is actually so rare that it might be more instructively termed the pathological distribution; error distributions in which extreme deviations occur more frequently than with the normal distribution often arise. Such distributions are said to be (76) heavy tailed or *leptokurtic* (meaning finely or thinly curved). "Many if not most routine analyses may have a leptokurtic error system" already, indicated Student (77) in the 1920s. In fact, any distribution is a mathematical concept. Geary suggested in 1947 (68, p. 5) that in front of all statistical texts should be printed: "Normality is a myth." There never was and never will be a normal distribution. However, there need be no cause for fear that a serious error is being made by the use of statistical methods based on the normal distribution.

Assumption of normality, nevertheless, is a plausible assumption (78) as an error term is made up of the combination of a large number of small chance effects (independent random errors) arising from several sources. Such a combination tends to produce (79) a normal distribution, regardless of the distribution of the separate errors (the Central Limit Theorem) if its variance is finite. Since most experiments involve many operations to set up and measure the results, it is reasonable to assume, at least tentatively, that the disturbances will be normally distributed. Then, normality is a very frequent assumption in regression analysis. Model parameters are mostly estimated by least squares, since it is efficient if the errors are normally distributed.

Extensive statistical methods and accurate tables have been derived for the normal distribution, many normal results hold for nonnormal populations and nonnormal responses may be transformed (80) to induce normality. It is often true that a variance-stabilizing transformation is also effective in transforming (81) a skew, nonnormal variable into a reasonably symmetric and approximate normal one. Thus, the transformation of scale often achieves a dual beneficial purpose. If the original variable is normally distributed, then the transformed variable cannot be. Frequently, however, the lack of constancy of variance (heteroscedasticity) is simultaneously associated with a nonnormality, and the transformation that gives a constant variance (homoscedasticity) also simultaneously gives a distribution (52) closer to normal.

ROBUSTNESS

One or few observations which do not follow the same model as the rest of the data can strongly influence the regression model. Real data are often subject to problems (82) that make the use of classical statistics, based on the normal distribution, difficult. The main practical problem probably is the occurrence of outliers. Another difficulty can be that the distribution of the data is not normal. One can attempt to correct this is with outlier tests or with procedures that make distribution normal. Outlier detection and rejection in particular often is not evident (83), and a possible alternative is then to use statistical procedures that are robust to outliers or deviations from normality.

When no assumptions are made at all about the distribution of data, a process is described as "nonparametric" or "distribution free." Nonparametric methods are generally robust, that is, insensitive to departures from an assumed statistical model (relatively immune to the form of the error distribution) or to the presence of outliers (resistant) though robust methods are not necessarily nonparametric. Although parametric tests are perhaps more commonly used, there are certain circumstances where nonparametric tests (84) should be used in preference.

A variety of robust regression methods with different degrees of robustness that safeguard against violation of the classical

assumptions (85–88) have been described and there has been considerable interest in recent years in the robust method of estimating, because of the usual absence of genuine knowledge about the error structure of data. Such tools require only simple assumptions such as (62) randomness, independence, and symmetry. The calculation of robust regressions is very time consuming and is designed for use (89–93) with various error distributions which can arise in practice, as well as normal data contaminated with wild observations. These are methods that are a little less accurate than least squares estimation when the least squares assumptions are true, but are more accurate—sometimes much more accurate—when they are false (i.e., they are insensitive to violations of these assumptions). On the other hand computer intensive methods of inference, for example, bootstrap (94–97) or jacknife (98–99), are recognized as powerful tools in judging data in complicated, nonstandard situations with only a few distributional assumptions.

INDEPENDENCE

The implication of the independence assumptions is that the disturbances in different experiments are independent of one another, that is, the disturbances on separate runs are not systematically related, an assumption which can usually be made to be more appropriate by randomization. It should also be confirmed that the errors are independent or, in other words, not correlated with some variable. This assumption is, in fact, quite reasonable to expect the y_i to be independent in many situations if they are the results (100) of separate isolated, noninterfering measurements. However, correlation among error terms in calibration work may appear if, for example, sampling is carried out inappropriately or analytical conditions vary (101) over time, such as temperature, degradation, evaporation, drifting devices.

The most common cause of nonindependence responses is that they are collected sequentially in time or in a systematic manner, which usually introduces (80) positive correlation among the observations. Sampling is one of the more important factors that influences (101) correlation among the error terms and is generally problem dependent. Serial correlation or autocorrelation in data represents in this way a violation concerning independence of measurement errors. Positive correlation among the observations causes too many significant results if either a t-test or an F-test in analysis of variance is used to test for treatment differences. Autocorrelation may be found (39) in problems involving data concerned with time dependencies, wherein the residual at one point (in time) r_k is in some way dependent on the residual at the previous point r_{k-1}, as well as on some purely random measurement error, ε_k, as in $r_k = y_k^{\text{exp}} - y_k^{\text{pred}} = f(r_{k-1}) + \varepsilon_k$. On the other hand, we are often faced with cumulative errors which appear as a consequence of a sampling technique used when all experiments are carried out, for example, on a single solution.

Correlation often goes undetected because its presence is difficult to detect by inspection of the data. Time series analysis, a special area of statistics, incorporates (102) correlation structure into the model used to analyze the data. The Durbin and Watson or the Durbin test criteria may be applied in order to check the independence postulate, as shown by Baumann (101) and Draper and Smith (51). If, however, the y_i are obtained through some functional combination of a number of measured values, they usually will not be independent and a full formalism (100) must be employed. For the majority of experiments run in the chemistry laboratory, the most effective precaution against correlation is randomization. The requirement for independence is important since it determines the degrees of freedom associated with statistical parameters.

The simplest forms of variation of a distribution parameter are shift and trend. A *shift* is a jump which persists, a *trend* (more specifically, a monotone trend) is a series of unidirectional jumps. Shifts or trends are possible alternatives to the hypotheses of randomness. Therefore, any test of randomness can be used as test against trend. *Randomness* of a given set of stochastic variables means that these are statistically independent and should all be described by the same probability distribution. A number of tests of randomness (63) that are sensitive against possible trends have been proposed. The run test analyzes the sequence of residual signs (17, 101). Nonindependent disturbances can be treated by generalized least squares but, as in the case where there is nonconstant variance, modifications to the model must be made (79) either through information gained from the data or by additional assumptions as to the nature of the interdependence.

ABSCISSA FREE FROM ERROR

It will be assumed that all the errors occur in the measured y values and that the errors in the measurements needed for the x values are negligible relative to y. This is not a restriction in practice. Case in which the errors have the same magnitude seldom occur (103); if the errors x are larger, x and y can be interchanged. This is to say that it is the structure of an experiment, rather than the convenience or comfort of the programmer, that determines (104) which is the independent variable and which is the dependent one.

Least squares linear regression is often applied to determine (105) a mathematical calibration model that approximates the relation between concentration and response. Making up standards always produces (106, 107) error. Thus, the assumption that x is error free is always formally invalid. However, errors in x have no consequences if they are less than one-tenth of the errors in y. If the error in x is greater, then the overall error is significantly increased. Moreover, regression parameters and confidence intervals of a calibration curve are then biased using (ordinary) weighted least squares (108).

In the context of most calibration problems the assumption relative to the abscissa variable is reasonable because (14) the analyte concentrations (x values) are precise enough. The assumption that errors only occur in the y-direction, on the other hand, is effectively valid in many experiments; errors in

instrument signals are often at least 2–3% (relative standard deviation [*RSD*], the ratio of the standard deviation to the mean), whereas the errors in making up the standards should be not more than one-tenth of this.

In order to use the full benefits of precise instrumental equipment, the error in making up the standards should be reduced to less than 0.05%. This is always possible using precise scales and precise volumetric equipment. The precision of volumetric operations like dilutions can often be improved using scales instead of pipettes and measuring flasks for volume determinations, or by using internal standards (109). However, modern automatic techniques are dramatically improving the precision (110) of many instrumental methods (gas chromatography (GL), liquid chromatography (LC), and capillary electrophoresis (CE) provide signal repeatabilities of 0.5% to 1%), spectroscopic techniques are equal or better, and flow injection analysis shows many examples of *RSD* of 0.5% or less. In such cases, it may be necessary either to abandon the assumption that x is free from error or to maintain the validity of the assumption by making up the standards gravimetrically rather than volumetrically (i.e., with an even greater accuracy than usual).

In the case of linear titration curves, the abscissa value is the added volume of titrant which may be considered (104) free from random errors. By using an adequate syringe or microburette with an automatic device, the measurement of volume can be made with high precision.

In kinetic studies, y is some function that represents the concentration of starting materials (9, 55, 111) or products, while x represents a timescale. A common practice in determining a kinetic mechanism is to determine the rate constants at a large number of different concentrations. It is generally assumed in those cases that the concentrations are known without error, and thus only the rates contain error terms. The same considerations are applied in enzyme kinetics (112), the precision with which the substrate concentrations are known depends on the accuracy of the pipetting in making up the reaction mixtures and, if different dilutions of one stock solution are used and care is taken, the resultant random errors in the substrate concentrations will be small. Thus, it is nearly always reasonable to assume that substrate concentrations are known much more accurately than rates, so that it is not too big an assumption to treat (113) all the error as applying to the rate and this is what is normally done.

The accuracy and precision of the timer used when the optical absorbance of a transient species is measured as a function of time (114), for example, will usually be much superior to that of the absorbance, in which case we are also justified in considering random errors in the absorbance only. For some analytical methods, however, such as X-ray fluorescence (XRF), certified reference materials (CRM) are often used as calibration standards because real samples (i.e., geological materials) are too complex. For this reason, uncertainties are associated (115) with both CRM concentration values and instrumental responses.

Although there may be some experiments where it is reasonable to assume that one variable is largely free from errors, there are others where such an assumption is manifestly absurd (12, 114, 116), as in cases in which both variables are calculated from the same observation. Particular attention must been given to equations in which one variable is involved on both sides. Then an error in this quantity appears in both coordinates mutually correlated in both conditions, that is, the independent variable x is not an exact quantity and the independence of errors is not fulfilled.

In method comparison studies a number of samples are examined by each of the two methods under study, and the two sets of results obtained are plotted on the x- and y-axes. Each point on the graph therefore represents a single sample examined by the two methods. In this instance, it is obvious that measurement errors must be expected in both the x- and y-directions (117, 118). In general, when the values for y and x are obtained by measurements

$$y_i = \eta_i + \varepsilon_i \quad [1]$$

$$x_i = \xi + \delta_i \quad [2]$$

where y_i and x_i are the measured values of the variables. When both variables contain errors, any distinction between dependent and independent variables is ambiguous, although one usually attempts to control one of them x, and observe the other y (119). Their true values are η_i and ξ_i and their respective errors ε_i and δ_i. In the usual least squares, one assumes that the error $\delta = 0$ or that $\delta \ll \varepsilon$. Note that when the weighted least squares method is used in the comparison of two methods, it is biased on the assumption that the x-values are known without error. In such a situation, wrong conclusions may be drawn when weighted least squares is used for the computation of regression coefficients.

If for any reason the precision with which the x_i values are known is not considerably better than the precision of measurement of the y_i values, the statistical analysis based on the (ordinary) weighted least squares is not valid and a more general approach is necessary (17, 120–124). The fundamental problem that arises if deviations are measured in any direction other than parallel with one or other axis is that such deviations do not have properly defined dimensions (except in the unusual case where x and y have the same dimensions and fall naturally in the same scales). The practical consequence of this is that the fit obtained in such a case will depend arbitrarily (113) on the scales chosen for plotting. Strategies concerning the topic of orthogonal regression have been (125–130) reviewed and treated.

If the consequence of the error in x is not clear beforehand, a simple but laborious test (101) meaningful for a sufficient number of data, can reveal it. The standard deviation of a number of repetitive measurements of the same sample is compared to the standard deviation of measurements of individually prepared samples of the same concentration. If these standard deviations are not found to be significantly different (e.g., F-test), no major error in x is assumed. Ideally, this test should be performed for concentrations at the upper and lower end of the range because

signal and sample preparation errors usually depend on the concentration.

PROPER WEIGHTS

In analytical chemistry, the assumption of homoscedastic distribution will generally turn out to be justificied within the calibration range. When the abscissa range (e.g., concentration) spans several orders of magnitude, however, as occurs with calibration in those cases in which (131) drug concentrations in urine or other body fluids are investigated, the precision of the y values varies greatly over the range of the x values. This condition contravenes the homoscedastic requirement of normal unweighted regression. Peculiar cases of heteroscedasticity occur when constant relative standard deviation or constant relative variance (counts: Poisson distribution) are involved. Photometric absorbances obeying the Lambert-Beer law over a wide range may also tend to heteroscedastivity as well as chromatographic analysis under certain conditions. In these cases, the introduction of weighting factors may be considered. With inductively coupled plasma mass spectrometry (ICPMS), even when calibration is over a relatively narrow range, weighted least squares estimates (132) are required. One should carry out a noise analysis (133) and always consider a weighted regression as (58) the general mode, the unweighted regression as a special mode. Nearly always one will find that (134) the absolute precision of the determination (i.e., the standard deviation σ_i) increases with concentration x_i, whereas the relative precision (i.e., the relative standard deviation $RSC = \sigma_i/x_i$) decreases with concentration. A common type of heteroscedasticity occurs in practice when the errors are of a constant relative size (30, 135–136). As developed chemical methods are tested at different concentrations, it is possible to seek a relationship between precision and concentration (137, 138) over the range of concentration tested. A variety of authors have suggested relationships and ISO 5725-1994 (139, 140) provides guidelines for establishing the existence of a given relationship.

More reliable data (smaller variability) are given greater emphasis, or weight. In this way one does not have to refit response versus concentration, since (30) the original data remain unchanged. The method of least squares is a powerful tool for the treatment of data, but its advantages may be vitiated by failure to include (141) proper weights. The problem is made particularly acute by the fact that the least squares criterion is highly sensitive to outliers, and often produces a paradoxical situation where the observations recognized to be the worst make the greatest contribution to the estimates of parameters. Although replication can be a severe constraint, it also has the advantage of providing (48) a form of "robust" regression. The most common way to perform weighted regression is simply to use (142) reciprocal variance for weights, that is, $w_i = 1/s_i^2$, where s_i^2 is the experimental estimate of σ_i^2. This relationship ensures that, if replication is employed, outlier values of y_i will be given low weights.

GOODNESS-OF-FIT OF THE MODEL

It is known that if the linear model is true the residuals (differences between the experimental and the fitted values) from the least squares regression line may be used to estimate (41) the error variances, and replicates are not necessary. However, if the linear model is not applicable, the residuals estimate a sum of an error of fit and experimental error. It is obvious that estimates of error variances independent both of the assumed model and the method of fitting can be obtained only from replicates at each point (143). Furthermore a comparison of a sum of squares of residuals with the error sum of squares obtained from the replicates provides (41, 144, 145) a test of goodness of fit of the model.

MULTIPLE MEASUREMENTS AT ONE OR MORE POINTS

It is frequently useful to employ experimental arrangements in which two or more runs are made at an identical set of values of the input variable x. The term *set* itself (70) is defined as referring to a number of independent replicate measurements of the same property. Measurements which are repeated so that they are subject to all the sources of random error in the experiment (146) are called *replicates*. In those cases in which these runs are made in such a way that they are subject to all the sources of error that beset runs made at different conditions (76), we call them *genuine replicates*. *Replication* is thus defined (147) as the independent performance of two or more experiments at the same level of all controlled factors. To avoid ambiguity (148), however, the term "replicate" or "replication" used alone should be employed only in the context of measurement (analysis) and not in the sense of "separation of multiple units" or collect "replicates" unless the usage is explicit. If the same factors are not present (149), our so-called replicates are pseudoreplicates and do not serve to estimate the variance of the distribution from which the observations were drawn.

Suppose that there are k specimens (samples) from a single variate (normal population) to be analyzed. Remember that a variable can take any physically admissible value whereas a variate is a variable that must also satisfy (46) a frequency distribution. For each specimen, a different number of replicates are made:

$$y_{11}\, y_{12} \cdots\cdots y_{1n1} \text{ are } n_1 \text{ repeat observations at } x_1$$
$$y_{21}\, y_{22} \cdots\cdots\cdots y_{2n2} \text{ are } n_2 \text{ repeat observations at } x_2$$
$$\cdots\cdots$$
$$y_{k1}\, y_{k2} \cdots\cdots\cdots y_{knk} \text{ are } n_k \text{ repeat observations at } x_k$$

To avoid ambiguity, subscripts greater than 9 can be written in parentheses or separated with commas (147). Any number of actual replicate measurements, be it one or more, is considered to be a random sample from this hypothetical infinite population. The size of the (statistical) sample is the number of measurements constituting it. Thus a set of three replicate measurements is (81) a sample of size 3 (not three samples).

Attention must be paid to the unfortunate, but unavoidable, dual use of the word "sample" with two distinctly different meanings: the chemical sample and the statistical sample.

The replicate results for each sample are scattered about their mean value, y_i, because of the random error (29) of measurement: the fluctuations around the mean value following the normal distribution. Most experimental determinations (150) are subject to fluctuations of an unpredictable kind, usually on account of experimental factors which are not under rigid control and the inherent mechanical limitations of the experimental apparatus, and sometimes on account of the inherent variability in the phenomena under investigation. We call "homogeneous" any set of values (81) that should, apart from experimental error, be alike. Thus, any set of replicate measurements is homogeneous. For such sets, there exist rules for the rejection of outliers based on the theory of the normal distribution.

The response of interest depends in addition to the property measured, also on other factors, such as (151) the surrounding temperature and relative humidity, the leveling of the instrument, the impurities in the chemical reagents, the correctness of a balance, or the skill of the operator. Attempts are always made to exercise control over such environmental factors, though to achieve total control is humanly impossible as well as is being aware of all factors that could affect the measuring process.

It might happen even in a well-designed experiment that some results are lost due to (152) gross error, breakdown of analytical equipment, or some other reasons. The loss of the results cannot always be recovered by means of repeated assays, although it that is sometimes possible in simple experiments. One also cannot exclude the case in which unequal number of repetition in objects results (5, 153) from the design of the experiment.

GENUINE REPLICATES

It is important to understate that repeated runs (51) must be genuine repeats and not just repetitions of the same reading. In chemical experiments, a succession of readings made during steady-state running does not provide (154) genuine repeat points. When genuine run replicates are made under a given set of experimental conditions, the variation between their associated observations (155) may be used to estimate the standard deviation of the effects.

By *genuine run replicates* we mean that variation between runs made at the same experimental condition is (51, 76) a reflection of the total variability afflicting runs made at different experimental conditions. This point requires careful consideration. In particular, several chemical analyses from a single run would provide only an estimate of analytical variance, usually only a small part of the run-to-run variance. Similarly, several samples from the same run could provide only a small part of the run-to-run variance. Generally, this problem of wrongly assessing experimental error variance has been particularly troublesome.

However, if a certain set of conditions was reset anew, after intermediate runs at other x-levels have been made, and provided that drifts in the response level had not occurred, genuine repeat runs (51) would be obtained. An obvious benefit of replication is improved reliability of the results. Other benefit is the easy of testing the goodness-of-fit of the model. The purely experimental uncertainty can be obtained only (147) by setting all the controlled factors at fixed levels and replicating the experiment.

In the preparation and analysis of a control sample involving drying, weighting, dilution, and subsequent two injections of a single sample prepared in this way onto an high-performance liquid chromatography (HPLC) column (156), the only factors that can produce difference between the measured results are those that operate from the injection stage onward, viz., injection, separation, and detection. Accordingly so, two genuine replicates in this case not only involve differences due to all these factors, but also differences due to drying error, to weighting error, and to dilution error. Accordingly, the partial replicates could seriously underestimate the size of the random error present in the measurement.

Repetition beginning at any later stage (e.g., aliquots from the same dissolved test portion) does not provide (148) an estimate of repeatability since the variability introduced by the omitted steps is not included in the final measurement. Presenting a test solution repeatedly to an instrument provides an estimate of instrumental precision only.

NUMBER OF REPLICATES (NORMALITY)

Replicates allow the magnitudes of the random variations to be estimated and the mean of replicates is expected (57), if systematic errors are absent or have been corrected, to be closer to the true value of the measurand under study than the individual reading.

The question of how many replicate measurements to take (57, 157–160) must include consideration of the magnitude of variability, availability of the test material and reagent, the time required, the cost of each measurement, and the variability required in the final result. This is a question (161) to which there is no simple, clear-cut answer. Even within the concepts that are based on the construction of a calibration curve (162, 163), there is no consensus about the choice of calibration samples and the number of replicates. For many situations, duplicates or triplicates are quite adequate. Five replicates in each group has been proposed by Jacquez et al. (164). Too much replication wastes our effort (165), while too little (166) fails to give us the sensitivity we need. Three or four replicates are not sufficient to properly estimate a variance (48); at least 8–10 are usually required. Three replicates (experimental replicates) for each of five concentration values equally spaced have been recommended by the ACS Committee on Environmental Improvement (167). However, the reduction in experimental work (a calibration design with fewer experimental and instrumental replicates) does not necessarily imply (168) a loss of analytical information.

As the number of replicate increases, the estimate of the total variance obviously improves. However, superficially one might be tempted to answer "the more the better," concerning the number of replicates that it is useful to be run, by referring to the theorem on the standard error of the mean. Such reasoning, however, is a fallacy because replication error (151) is only a portion, sometimes quite small, of the total error. Statistical and practical significances should always be remembered (161) in this respect. Systematic errors are not reduced (81) through the process of averaging.

As the number of replicates increases, however, the central limit theorem states that the frequency distribution for the mean value approaches normality (very rapidly indeed, especially if the parent distribution is symmetric). This fortunate circumstance provides a very important, but little recognized, basis for replication of analyses. In effect, replication (62) permits us to assume normality (for the mean value not for individual values), an assumption which is quite difficult to substantiate otherwise; the mean of four observations is already very close to normal. Good analytical practice takes this fact into account when establishing the minimum desirable number of observations. In particular, uniform and binomial distributions provide a simple but dramatic illustration of the approach to normality. In fact, random numbers with normal (Gaussian) distribution, with given mean μ and standard deviation σ_x can be conveniently generated (169–171) using the following equation

$$x = \mu + \sigma_x \left(\frac{\sum_{i=1}^{n} r_i - \frac{n}{2}}{\left(\frac{n}{12}\right)^{\frac{1}{2}}} \right) \qquad [3]$$

where r_i is a uniformly distributed random number whose value lies between 0 and 1, provided by most high-level programming languages. The value of the variable n is frequently selected to be 12. A certain computational advantage is achieved in this way (172–174) because the preceding equation is reduced to

$$z = \sum_{i=1}^{12} (r_i - 6) \qquad [4]$$

and despite this value of n restricts the distribution of the limits $\pm$ 6, it leads to total valid values of z up to three times the standard deviation.

A test regarding normal distribution of the y_{iv} measuring values may be carried out (58, 175). At least three values ($n_i \geq 3$) should be measured for any value of the independent variable to carry out the following test. The critical value D is the ratio of the spread R_i and the standard deviation s_i

$$R_i = (y_{iv})_{\max} - (y_{iv})_{\min} \qquad [5]$$

$$s_i = \sqrt{\frac{\sum (y_{iv} - \bar{y}_i)^2}{n_i - 1}} \qquad [6]$$

$$D_i = \frac{R_i}{s_i} \qquad [7]$$

If D does not lie between the k-dependent limits compiles in statistical tables, it may be concluded with 90% probability that the data are not normally distributed and in the following less precise, so-called nonparametric procedures have to be applied. However, if the David test is passed, normality of the measuring data can be assumed and the parametrical procedure may subsequently be applied.

Some case is needed to achieve genuine replicate runs. In particular, a group of such runs should normally be run consecutively but should be randomly ordered. Replicate runs must be subject to all the usual setup errors, sampling errors, and analytical errors which affect runs made at different conditions. Failure to achieve this will typically cause underestimation of the error and will invalidate the analysis.

CONCLUSIONS

The regression topic is of vital importance because it is closely related to calibration, comparison of two analytical methods, method validation, and error estimation. Under ideal conditions, the least squares method is the preferred choice implying a variety of assumptions (i.e., normality, independence, abscissa free from error, and proper weights). However, some published data have grossly violated at least one of the assumptions inherent in regression modeling. If the experimental deviations may essentially be confined to the dependent variable y, a major simplification can be made. When measurements are obtained over a wide range of the x variable, the assumption of uniformity in the variance of y is not valid. It sometimes happens that some observations used in a regression analysis are more reliable than others, and so the direct application of conventional least squares may produce gross errors. Although the assumption of homoscedasticity is valid for some analytical procedures, there are others for which it is not, including these methods based on counting measurements either photon or radioactive and also photometric and chromatographic analysis under certain conditions. There are two main solutions to the problem of nonconstant variance: Transform the data, or perform a weighted least squares regression analysis as several authors have pointed out is a better solution. However, weighting and data transformation will be the subject of a future article. Compared with the problem of nonconstant variance, chemists have paid less attention to the normality assumption underlying the least squares method. Replication is associated with the standard deviations of the effects. In this respect, replicates must be genuine replicates. Too much replication wastes our effort, while too little fails to give the information found.

REFERENCES

1. H. S. Uhler, Method of least squares and curve fitting. *J. Opt. Soc.* 7 (1923):1043–1066.
2. R. de Levie, Curve fitting with least squares. *CRC Crit. Rev. Anal. Chem.* 30 (2000):59–74.
3. M. J. Cardone, Detection and determination of error in analytical methodology. Part I. In the method verification program. *J. Ass. Off. Anal. Chem. Int.* 66 (1983):1257–1282.

4. A. G. Asuero and A. G. González, Some observations of fitting a straight line to data. *Microchem. J.* 40 (1989):216–225.
5. E. Desimoni, A Program for the weighted least squares regression of unbalanced response arrays. *Analyst* 124 (1999):1191–1195.
6. K. J. Ellis and R. G. Duggleby, What happens when data are fitted to the wrong equation. *Biochem. J.* 171 (1978):513–517.
7. J. R. J. Belloto and T. D. Sokolovski, Residual analysis in regresión. *Am. J. Pharm. Educ.* 49 (1985):295–303.
8. M. Meloun, J. Militki, K. Kupla, and R. G. Brereton, The effect of influential data, model and method on the prescision of univariate calibration. *Talanta* 57 (2002):721–740.
9. S. Bolton, *Pharmaceutical Statistics: Practical and Clinical Applications*, 3rd ed. (New York: Marcel Dekker, 1994).
10. M. Thompson, Why analysts need chemometrics. *Anal. Proc.* 25 (1988):380.
11. M. Thompson, Abuse of statistics software package. *Anal. Proc.* 27 (1990):142–144.
12. O. Exner, How to get wrong results from good experimental data: a survey of incorrect applications of regression. *J. Phys. Org. Chem.* 10 (1997):797–813.
13. A. Unwin, Using your eyes—making statistics more visible with computers. *Comput. Stat. Data. Anal.* 32 (2002):303–312.
14. J. S. Hunter, Calibration and the straight line: current statistical practices. *J. Ass. Off. Anal. Chem.* 64 (1981):574–583.
15. W. Penninckx, D. Hartmann, D. L. Massart, and J. Smeyers-Verbeke, Validation of the calibration procedure in atomic absorption spectrometric methods. *J. Anal. Atom. Spectrom.* 11 (1996):237–246.
16. K. Danzer and L. A. Currie, IUPAC, Guidelines for calibration in analytical chemistry. Part I. Fundamental and single component calibration. *Pure Appl. Chem.* 70 (1998):993–1014.
17. M. Thompson, Regression methods in the comparison of accuracy. *Analyst* 107 (1982):1169.
18. C. Hartmann, J. Smeyers-Verbeke, W. Pennickx, and D. L. Massart, Detection of bias in method comparison by regression analysis. *Anal. Chim. Acta* 338 (1997):19–40.
19. S. Braggio, R. J. Barnaby, P. Grossi, and M. Cugola, A strategy for validation of bioanalytical methods. *J. Pharm. Biom. Anal.* 14 (1996):375–378.
20. J. Wieling, G. Hendriks, W. J. Tamminga, J. Hempenius, C. K. Mensink, B. Ooesterhuis, and J. H. G. Jonkman, A Rational experimental design for bioanalytical method validation illustration using an assay method for total captocril in plasma. *J. Chromatogr.* A730 (1996):381–394.
21. H. T. Karnes and C. C. March, Calibration and validation of linearity in chromatographic biopharmaceutical analysis. *J. Pharm. Biomed. Anal.* 9 (1991):911–918.
22. M. Feinberg and N. Raguenes, Development and application of a standardized validation procedure for food chemistry laboratories. *Anal. Chim. Acta* 391 (1999):239–252.
23. International Standard, *Water Quality: Calibration and Evaluation of Analytical Methods and Estimation of the Performance Characteristics: Part 1. Statistical Evaluation of the Linear Calibration Function*, ISO 8466-1-1990 (Geneva: International Organization for Standradization, 1990).
24. C. Hartmann, J. Smeyers-Verbeke, D. L. Massart, R. D. Mc Dowal, Validation of bioanalytical chromatographic methods. *J. Pharm. Biomed. Anal.* 17 (1998):193–218.
25. N. V. Nagaraja, J. K. Paliwal, and R. C. Gupta, Choosing the calibration model in assay validation. *J. Pharm. Biomed. Anal.* 20 (1999):433–438.
26. L. Brüggemann and R. Wennrich, Evaluation of measurements uncertainty for analytical procedures using a linear calibration function. *Accred. Qual. Assur.* 7 (2002):269–273.
27. K. Heydorn and T. Anglov, Calibration uncertainty. *Accred. Qual. Assur.* 7 (2002):153–158.
28. J. D. Hwang and J. D. Winefordner, Regression methods in analytical chemistry. *Prog. Anal. Spectrosc.* 11 (1988):209–249.
29. Analytical Methods Committee, Is my calibration linear. *Analyst* 119 (1994):2363–2366.
30. M. E. Zorn, R. D. Gibbons, and W. C. Sonzogni, Weighted least squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. *Anal. Chem.* 69 (1997):3069–3075.
31. C. Hartmann, J. Smeyers-Verbeke, and D. L. Massart, Problems in method comparison studies. *Analysis* 21 (1993):125–132.
32. K. Linnet, Evaluation of regression procedures for methods comparison studies. *Clin. Chem.* 39 (1993):424–432.
33. K. Linnet, Performance of Deming regression analysis in case of misspecified analytical error rario in method comparison studies. *Clin. Chem.* 44 (1998):1024–1031.
34. D. Stöck, K. Dewitte, and L. M. Thienpont, Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data? *Clin. Chem.* 44 (1998):2340–2346.
35. J. O. Westgard, Points of care in using statistics in method comparison studies. *Clin. Chem.* 44 (1998):2240–2242.
36. R. J. Cvetanovic, D. L. Singleton, and G. Paraskevopoulos, Evaluation of the mean values and standard errors of rate constants and their temperature coefficients. *J. Phys. Chem.* 83 (1979):50–60.
37. L. M. Schwartz, Linearization methods for first-order kinetic analysis. *Anal. Chem.* 53 (1981):206–213.
38. O. Exner, Calculating equilibrium constants from spectral data: reliability of the Benesi-Hildebrand method and its modifications. *Chem. Intell. Lab. Syst.* 39 (1997):85–93.
39. M. Meloun, J. Militki, and F. Forina, *Chemometrics for Analytical Chemistry: Vol. 2. PC-Aided Regression and Related Methods* (New York: Ellis Horwood, 1994).
40. R. Sundberg, Statistical aspects on fitting the Arrhenius equation. *Chem. Intell. Lab. Syst.* 41 (1998):249–252.
41. S. N. Deming, Linear models and matrix least squares in clinical chemistry, in *Chemometrics: Mathematics and Statistics in Chemistry*, ed. B. R. Kowalski (Dordrecht: Reidel, 1984), 267–304.
42. E. B. Rudnyi, Statistical model of systematic errors: lineal error model. *Chem. Intell. Lab. Syst.* 34 (1996):41–54.
43. C. F. Gauss, *Theoria Motus Corporum Coelestium* (Hamburg: Perthes, 1809), Ar. 179.
44. A. M. Legendre, Appendice sur la méthode des moindres quarrés, in *Nouvelles Méthodes pour la Determination des Orbites des Comètes* (Paris: Courcier, 1805), 71–80.
45. S. M. Stigler, Gauss and the invention of least squares. *Ann. Stat.* 5 (1981):465–474.
46. K. A. Connors, *Binding Constants: The Measurements of Molecular Complex Stability* (New York: Wiley, 1987).

47. P. Valko and S. Vajda, *Advanced Scientific Computing in BASIC with Applications in Chemistry, Biology and Pharmacology* (Amsterdam: Elsevier, 1989).
48. D. L. MacTaggart and S. O. Farwell, Analytical use of linear regression. Part I. Regression procedures for calibration and quantitation. *J. Ass. Off. Anal. Chem.* 75 (1992):594–607.
49. F. Galton, Family likeness in stature. *Proc. Royal. Soc. London* 40 (1888):42–72.
50. R. L. Plackett, Studies in the history of probability and statistics XXIX: The discovery of the method of least squares. *Biometrika* 59 (1972):239–251.
51. N. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. (New York: Wiley, 1998).
52. K. A. Brownlee, *Statistical Methodology in Science and Engineering*, 3rd ed. (Malabar, FL.: Robert G. Krieger, 1984).
53. G. Meinrath, C. Ekberg, A. Landgren, and J. A. Liljenzin, Assessment of uncertainty in parameter estimation and prediction. *Talanta* 5 (2000):231–246.
54. *Design and analysis of Enzyme and Pharmacokinetics Experiments*, ed. L. Endreny (London: Plenum Press, 1981).
55. K. A. Connors, *Chemical Kinetics: The Study of Reaction Rates in solution* (New York: VCH, 1990), Chapter 2, 41–49.
56. P. C. Meier and R. E. Zünd, *Statistical Methods in Analytical Chemistry*, 2nd ed. (New York: Wiley, 2000).
57. J. N. Miller, Outliers in experimental data and their treatment *Analyst*. 118 (1993):455–461.
58. H. Bubert and R. Klockenkämper, Precision-dependent calibration in instrumental analysis. *Fresenius J. Anal. Chem.* 316 (1983):186–193.
59. M. Davidian and P. D. Haaland, Regression and calibration with non constant error variance. *Chem. Intell. Lab. Syst.* 9 (1990):231–248.
60. S. M. Gort and R. Hoogerbrugge, A user-friendly spreadsheet program for calibration using weighted regression. *Chem. Intell. Lab. Syst.* 28 (1995):193–199.
61. R. C. Castells and M. A. Castillo, Systematic errors: detection and correction by means of standard calibration. Youden calibration and standard additions method in conjunction with a method response model. *Anal. Chim. Acta* 423 (2000):179–185.
62. L. Currie, Approach to accuracy in analytical chemistry, in *Treatise on Analytical Chemistry, Part 1*, Vol. 1, 2nd ed., eds. I. M. Kolthoff, P. J. Elving (New York: Wiley, 1978), 95–242.
63. J. K. Taylor, *Statistical Techniques for Data Analysis* (Chelsea, Mich: Lewis, 1990).
64. R. A. Fisher, *The Design of Experiments*, 8th ed. (New York: Hafner, 1966), p 16.
65. M. Meloun and J. Militki, Detection of single influential points in OLS regression model building. *Anal. Chim. Acta* 439 (2001):169–191.
66. M. Meloun, J. Militki, M. Hil, and R. G. Brereton, Crucial problems in regression modelling and their solutions. *Analyst* 127 (2002):433.
67. W. R. Ott, *Environmental Statistics and Data Analysis* (Boca Raton, FL.: Lewis/CRC Press, 1995).
68. S. S. Shapiro, *How to Test Normality and Other Distributional Assumptions* (Wisc.: American. Society for Quality Control, 1990), p 5.
69. N. I. Fisher, Graphical methods in nonparametric statistics: a review and annotated bibliography. *Int. Stat. Rev.* 51 (1983):25–28.
70. G. Kateman and L. Buidens, *Quality Control in Analytical Chemistry* (New York: Wiley, 1993).
71. International Standard, *Statistical Interpretation of Data: Tests for Departure from the Normal Distribution*, ISO 5479-1997 (Geneva: International Organization for Standardization, 1997).
72. V. Clancey, Statistical methods in chemical analysis. *Nature* 159 (1947):339–340.
73. D. J. Finney, The choice of a response metameter in bio-assay. *Biometrics* 5 (1949):261–272.
74. A. R. Henderson, Chemistry with confidence: should clinical chemistry require confidence intervals for analytical and other data. *Clin Chem.* 39 (1993):929–935.
75. J. W. Tukey and D. H. McLauguin, Less vulnerable confidence and significance procedures for location based on a single sample: trimming/winsorization I. *Sankhya Ser A* 25 (1963):331–352.
76. G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces* (New York: Wiley, 1987).
77. Student, Errors of routine analysis. *Biometrika* 19 (1927):151–164.
78. M. Thompson and R. J. Howarth, The frequency distribution of analytical error. *Analyst* 105 (1980):1188–1195.
79. D. Bates and D. G. Watts, *Nonlinear Regression Analysis and Its Applications* (New York: Wiley, 1988).
80. C. K. Bayne and I. B. Rubin, *Practical Experimental Designs and Optimization Methods for Chemistry* (Deerfield Beach, FL.: VCR, 1986), 61–62.
81. J. Mandel, Accuracy and precision: evaluation and interpretation of analytical results, in *Treatise on Analytical Chemistry*, Part I, Vol. 1, 2nd ed., eds. I. M. Kolthoff and P. J. Elving (New York: Wiley, 1978), 243–298.
82. P. Vankeerberghen, C. Vandenbesh, J. Smeyers-Verbeke, and D. L. Massart, Some robust statistical procedures applied to the analysis of chemical data. *Chem. Intell. Lab. Syst.* 12 (1991):3–13.
83. B. Walczak and D. L. Massart, Multiple outlier detection revisited. *Chem. Intell. Lab. Syst.* 41 (1998):1–15.
84. C. P. Wheater and P. A. Cook, *Using Statistics to Understand the Environment* (London: Routledge, 2000).
85. M. Thompson, Robust statistics and functional relationship estimation for comparing bias of analytical procedures over extended concentration ranges. *Anal. Chem.* 61 (1989):1942–1945.
86. K. Danzer, Robuste Statistik in de analytischen Chemie. *Fresenius J. Anal. Chem.* 335 (1989):869–875.
87. P. J. Rousseew, Tutorial to robust statistics, *J. Chemomet.* 5 (1991):1–20.
88. M. C. Ortiz-Fernández and A. Herrero-Gutierrez, Regression by least median squares, a methodological contribution to titration analysis. *Chem. Intell. Lab. Syst.* 27 (1995):231–243.
89. D. L. Massart, L. Kaufman, P. Rousseeuw, and A. Leroy, Least medium of squares: a robust method for outlier and model error detection in regression and calibration. *Anal. Chim. Acta* 187 (1986):171–179.
90. G. Caviglioli, G. Drava, S. Lafaggi, B. Parodi, and G. Bignardi, Median-based robust regression methods in prediction of Drug Stability. *J. Pharm. Sci.* 85 (1996):1096–1104.
91. E. J. Dietz, A comparison of robust estimators in simple linear regression. *Commun. Statist-Simula.* 16 (1987):1209–1227.
92. Y. Hu, J. Smeyers-Verbeke, and D. L. Massart, Exploratory study on median based methods in robust regression for routine

calibration in atomic absorption analysis. *J. Anal. At. Spectr.* 4 (1989):605–611.
93. H. Yuzhu, J. Smeyers-Verbeke, and D. L. Massart, Outlier detection in calibration. *Chem. Intell. Lab. Syst*. 9 (1990):31–44.
94. P. Bonate, Approximate confidence intervals in calibration using the Boostrap. *Anal. Chem*. 65 (1993):1367–1372.
95. G. Jones, M. Wortberg, S. B. Kreissig, B. D. Hammock, and D. M. Rocke, Application of the bootstrap to calibration experiments. *Anal. Chem*. 68 (1996):763–770.
96. G. Meinrath, Computer-intensive methods for uncertainty estimation in complex situations. *Chem. Intell. Lab. Syst*. 51 (2000):175–187.
97. R. Wehrens, H. Putter, and L. M. C. Buydens, The bootstrap: a Tutorial. *Chem. Intell. Lab. Syst*. 54 (2000):35–52.
98. E. K. Kimanani and J. Lavigne, Bioanalytical calibration curves: variability of optimal powers between and within analytical methods. *J. Pharm. Biomed. Anal*. 16 (1998):1107–1115.
99. E. K. Kimanani, Bioanalytical calibration curves: proposal for statistical criteria. *J. Pharm. Biomed. Anal*. 16 (1998):1117–1124.
100. M. D. Pattengil and D. E. Sands, Statistical significance of linear least squares parameters. *J. Chem. Educ*. 56 (1979):224–247.
101. K. Baumann, Regression and calibration for analytical separation techniques. Part II. Validation, weighted and robust regression. *Process, Control. Qual*. 10 (1997):75–112.
102. G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control* (San Francisco: Holden-Day, 1970).
103. J. Kragten, Least squares polynomial curve-fiting for calibration purposes (STATCAL-CALIBRA). *Anal. Chim. Acta* 241 (1990):1–13.
104. G. Kateman, H. C. Smith, and L. Meites, Weighting in the interpretation of data for potentiometric acid-base titration by non linear regression. *Anal. Chim. Acta* 152 (1983):61–72.
105. L. C. Rodríguez, A. M. Campaña, C. J. Linares, and M. R. Ceba, Estimation of the performance characteristics of an analytical method using the data set of the calibration experiment. *Anal. Lett*. 26 (1993):1243–1258.
106. R. J. Carroll and C. H. Spiegelman, The effect of ignoring small measurement errors in precision instrument calibration. *J. Qual. Technol*. 18 (1986):170–173.
107. C. H. Spiegelman, R. L. Watters, and L. Hungwu, A statistical method for calibrating flame emission spectrometry which takes account of errors in the calibration standards. *Chem. Intell. Lab. Syst*. 11 (1991):121–130.
108. A. Martinez, F. J. del Rio, I. Riu, and F. X. Rius, Detecting proportional and constant bias in method comparison studies using linear regression with errors in both axes. *Chem. Intell. Lab. Syst*. 49 (1999):179–193.
109. K. Baumann and H. Wätzig, Regression and calibration for analytical separation techniques. Part I. Design considerations. *Process. Control Qual*. 10 (1997):59–73.
110. J. N. Miller, Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review. *Analyst* 116 (1991):3–14.
111. D. T. Elmore, A. E. Kingston, and D. B. Shields, The computation of velocities and kinetic constants of reactions with particular reference to enzime-catalysed processes. *J. Chem. Soc*. (1963):2070.
112. W. Cleland, The statistical analysis of enzyme kinetic data. *Advances in Enzymology* 29 (1967):1–32.
113. A. Cornish-Bowden, *Analysis of Enzyme Kinetic Data* (Oxford: Oxford University Press, 1995).
114. R. de Levie, When, why, and how to use weighted least squares. *J. Chem. Educ*. 63 (1986):10–15.
115. A. Martinez, *Calibració Lineal i Comparació de Mètodes Analítics mitjançant Tècniques de Regressió que Incorporen Errors en Totes les Variables* (Ph.D. Dissertation Universitat Rovira i Virgili, Tarragona, 2001).
116. A. Cornish-Bowden and R. Eisenthal, Statistical considerations in the estimation of enzyme kinetic parameters by the direct linear plot and other methods. *Biochem. J*. 139 (1974):721–730.
117. D. L. MacTaggart, S. O. Farwell, Analytical use of linear regression. Part I. Regression procedures for calibration and quantitation. *J. Ass. Off. Anal. Chem*. 75 (1992):608–614.
118. R. F. Martin, General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clin. Chem*. 46 (2000):100–104.
119. J. R. Macdonald and W. J. Thompson, Least-squares fitting when both variables contain errors: pitfalls and possibilities. *Am. J. Phys*. 60 (1992):66–73.
120. B. D. Ripley and M. Thompson, Regression techniques for the detection of analytical bias. *Analyst* 112 (1987):277–383.
121. W. J. Thompson, *Computing for Scientist and Engineers: A Workbook of Analysis Numerics, and Applications* (New York. Wiley, 1992), 190–208.
122. A. G. González and A. G. Asuero, Computational program for validating analytical results. *Fresenius J. Anal. Chem*. 346 (1993):885–887.
123. J. Riu and F. X. Rius, Univariate regression models with errors in both axes. *J. Chemometrics* 9 (1995):343–362.
124. H. W. Zwanziger and C. Sarbu, Validation of analytical methods using a regression procedure. *Anal. Chem*. 70 (1998):1277–1280.
125. K. Danzer, M. Wagner, and C. Fischbacher, Calibration by orthogonal and common least squares. Theoretical and practical aspects. *Fresenius J. Anal. Chem*. 352 (1995):407–412.
126. J. Riu and F. X. Rius, Assessing the accuracy of analytical methods using linear regression with error in both axes. *Anal. Chem*. 68 (1996):1851–1857.
127. J. Riu and F. Rius, Method comparison using regression linear with uncertainties in both axes. *TrAC* 16 (1997):211–216.
128. A. Martinez, J. Riu, and F. X. Rius, Lack offit in linear regression considering error in both axes. *Chem. Intell. Lab. Syst*. 54 (2000):61–73.
129. A. Martinez, J. Riu, and F. X. Rius, Evaluating bias in method comparison studies using linear regression with errors in both axes. *J. Chemometrics* 16 (2002):41–53.
130. F. J. del Rio, J. Riu, and F. X. Rius, Prediction intervals in linear regression taking into account errors on both axes. *J. Chemometrics* 15 (2001):773–788.
131. K. Baumann and H. Wätzig, Appropriate calibration functions for capillary electrophoresis. II. Heterocedasticity and its consequences. *J. Chromatog*. A700 (1995):9–20.
132. S. N. Ketkar and T. J. Bzik, Calibration of analytical instruments. Impact of nonconstat variance in calibration data. *Anal. Chem*. 72 (2000):4762–4765.
133. R. Matsuda, Y. Hayashi, K. Hayashi, Y. Saito, K. Fwaki, H. Harakawa, M. Satoh, Y. Ishizuki, and T. Kato, Deductive prediction of precision in measurement, calibration and standard addition method in atomic absorption spectrometry for cadmium. *Anal. Chem*. 70 (1998):319–327.

134. M. Thompson, Variation of precision with concentration in an analytical system. *Analyst* 112 (1988):1579–1587.
135. K. P. Anderson and R. L. Snow, A relative deviation, least squares method of data treatment. *J. Chem. Educ.* 44 (1967):756–757.
136. E. D. Smith and D. M. Mathews, Least square regression lines. Calculations assuming a constant percent error. *J. Chem. Educ.* 44 (1967):757–759.
137. H. Hughes and P. W. Hurley, Precision and accuracy of test methods and the concept of K-factor in chemical analysis. *Analyst* 112 (1987):1445–1449.
138. R. Albert and W. Horwitz, A heuristic derivation of the Horwitz curve. *Anal. Chem.* 69 (1997):789–790.
139. International Standard, *Accuracy (Trueness and Precision) of Measurement Methods and Results*, Part 1 to 6, ISO 5725-1994 (Geneva: International Organization for Strandardization, 1994).
140. Y. V. Heyden, The ruggedness of analytical methods. *Analusis* 22 (1994):M27–M29.
141. D. E. Sands, Weighting factors in least squares. *J. Chem. Educ.* 51 (1974):473–474.
142. J. S. Garden, D. G. Mitchell, and W. N. Mills, Non-constant variance techniques for calibration curve-based analysis. *Anal. Chem.* 52 (1980):2310–2315.
143. M. Feinberg, *La Validation des M'ethodes d' Analyse. Une approache chimiometrique de l'assurance qualit'e au laboratoire* (Paris: Masson, 1996).
144. S. N. Deming and S. L. Morgan, Teaching the fundamentals of experimental design. *Anal. Chim. Acta* 150 (1983):183–198.
145. S. N. Deming, Chemometrics: an overview. *Clin. Chem.* 32 (1986):1702–1706.
146. J. C. Miller and J. N. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 4th ed. (Chichester: Ellis Horwood, 2000).
147. S. N. Deming and S. L. Morgan, *Experimental Design: A Chemometric Approach*, 2nd ed. (Amsterdam: Elsevier, 1993).
148. L. A. Currie and G. Svehla, Nomenclature for the presentation of results in chemical analysis. *Pure Appl. Chem.* 66 (1994):595–608.
149. J. R. Green and D. Margerison, *Statistical Treatment of Experimental Data* (Amsterdam: Elsevier, 1977).
150. G. W. Wilkinson, Statistical estimation in enzyme kinetics, *Biochem. J.* 80 (1961):324–332.
151. J. Mandel, *The Statistical Analysis of Experimental Data* (New York: Wiley-Interscience, 1964).
152. J. Cserminski, A. Iwasiewicz, Z. Paszek, and A. Sikorski, *Statistical Methods in Applied Chemistry* (Amsterdam: Elsevier, 1990), 425.
153. E. Desimoni, A. Daghetta, and S. Valsecchi, Evaluation of uncertainty by ordinary least squares regression of replicated data. A revised formulation to deal with unbalanced data set. *Ann. Chim. (Roma)* 88 (1998):601–607.
154. G. R. Phillips, J. M. Harris, and E. M. Eyring, Treatment of replicate measurements. *Anal. Chem.* 54 (1982):2053–2056.
155. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenter: An Introduction to Design, Data Analysis and Model Building* (New York: Wiley, 1978), 319.
156. E. Mullins, Introduction to control charts in the analytical laboratory. Tutorial review. *Analyst* 119 (1994):369–375.
157. D. G. Mitchell, W. N. Mills, J. S. Garden, and M. Zdeb, Multiple-curve procedure for improving precision with calibration-curve based analyses. *Anal. Chem.* 49 (1977):1655–1660.
158. J. Agterdenbos, Calibration in quantitative analysis. Part 1. General considerations. *Anal. Chim. Acta* 1058 (1979):315–323.
159. G. J. Kemp, The susceptibility of calibration methods to errors in the analytical signal. *Anal. Chim. Acta* 176 (1985):229–247.
160. K. G. Owens, C. F. Bauer, and C. L. Grant, Effects of Analytical Calibration Models on Detection Limit Estimates, in *Detection Limits in Analytical Chemistry, Importance, Theory and Practice: ACS Symposium Series, 361*, ed. L. A. Currie (Washington, D.C.: American Chemical Society, 1988), ch. 10, 194–207.
161. L. Davies, *Efficiency in Research, Development and Production: The Statistical Design and Analysis of Chemical Experiments* (Cambridge: RSC, 1993), 51.
162. International Union of Pure and Applied Chemistry (IUPAC), Harmonized protocols for the adoption of standardized methods and for the presentation of their performance characteristics (W. D. Pocklington). *Pure Appl. Chem.* 62 (1990):149–162.
163. International Union of Pure and Applied Chemistry (IUPAC), International Standarization Organization (ISO), and Association of Official and Analytical Chemist (AOAC) Guidelines for collaborative study procedure to validate characteristics of a method of analysis. *J. Ass. Off. Anal. Chem.* 72 (1989):694–704.
164. J. A. Jacquez, F. J. Mather, and C. R. Crawford, Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics.* 24 (1968):607–626.
165. M. A. Castillo and R. C. Castells, Initial evaluation of quantitative performance of chromatographic methods using replicates at multiple concentrations. *J. Chromatogr.* A921 (2001):121–133.
166. G. T. Wernimont, *Use of Statistics to Develop and Evaluate Analytical Methods*, ed. W. Spendley (Arlington, Va: Association of Official and Analytical Chemist (A.O.A.C.), 1985), 64–78.
167. American Chemical Society Committee on Environmental Chemistry, Guidelines for data adquisition and data quality evaluation in environmental chemistry. *Anal. Chem.* 52 (1980):2242–2249.
168. A. González-Casado, A. M. G. Campaña, L. C. Rodríguez, J. C. Vilchez, and M. R. B. Garcia, A comparative study of different calibration data subsets in an attempt to minimise experimental work. *LC-GC Int.* 11 (1998):726, 728, 730, 732–735.
169. B. S. Gottfried, *Programming with Basic*, Schaom's Outline Series (New York: McGraw-Hill, 1975), p 172.
170. O. A. Güell and J. A. Holcombe, Analytical applications of Monte Carlo techniques. *Anal. Chem.* 62 (1990):529A–534A, 536A, 538A, 540A–542A.
171. S. Bi, A simple model for predicting the pH values of acidic natural waters. *Anal. Chim. Acta* 314 (1995):111–119.
172. C. Mongay, G. Ramis, and M. C. García, A critical study of the linear least squares method applied to the spectrophotometric determination of protonation constants of diprotic acids. *Spectrochim. Acta* A38 (1982):247–252.
173. G. R. Ramos and M. C. G. Alvarez-Coque, Examination of the least squares method applied to the evaluation of physicochemical parameters with linearized equations. *Anal. Chim. Acta* 220 (1989):145–153.
174. R. Klicka and L. Kubaceta, Statistical properties of linearization of the Arrhenius equation via the logarithmic transformation. *Chem. Intell. Lab. Syst.* 39 (1997):69–75.
175. H. A. David, H. O. Hartley, and E. S. Pearson, The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika* 41 (1954):482–483.