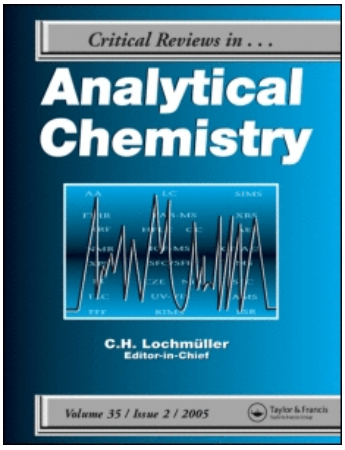

Critical Reviews in Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713400837

Fitting Straight Lines with Replicated Observations by Linear Regression: Part II. Testing for Homogeneity of Variances

Ana Sayago[a]; Agustin G. Asuero[a]
[a] Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, Seville, Spain

Online publication date: 10 August 2010

To cite this Article Sayago, Ana and Asuero, Agustin G.(2004) 'Fitting Straight Lines with Replicated Observations by Linear Regression: Part II. Testing for Homogeneity of Variances', Critical Reviews in Analytical Chemistry, 34: 3, 133 — 146
To link to this Article: DOI: 10.1080/10408340490888599
URL: http://dx.doi.org/10.1080/10408340490888599

Critical Reviews in Analytical Chemistry, 34:133–146, 2004

ISSN: 1040-8347 print
DOI: 10.1080/10408340490888599

Fitting Straight Lines with Replicated Observations by Linear Regression: Part II. Testing for Homogeneity of Variances

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

The purpose of the previous, this, and subsequent articles in the series is to deal with the fitting of a straight line to experimental data in those cases in which replicate observations are realized. Special attention has been paid in this occasion to the straight line model, to the determination of the variance from a number of small samples, and to the problem of testing for homogeneity of variances. Ideally, weights should be determined by replicate measurements. However, little attention has been paid to the effect of replicating the response measurements in the context of regression analysis. Several methods that compare variances at different factor levels, such as the *F*, Bartlett, Cochran, Burr and Foster, Hartley, and Levene tests have been envisaged. Weighting and data transformation and variance analysis will be the subject of a future report. The authors expect the results of this critical review to be of value to investigators making use of these methods and also in the teaching.

Keywords least squares method, replicated observations, straight line model, straight lines, testing for homogeneity of variances

THE STRAIGHT LINE MODEL

The most immediate goal of scientific or industrial experimentation is (1–4) to find relationships among manipulated, x, and observed, y, variables, or to validate such relationships coming from some underlying theory, $y = f(x, p)$, where p is the number of parameters in the model. In many applications the model is based on a theoretical relationship that governs (5–6) the system, and its parameters have some well-defined physical meaning. In this context, "simple problems—as indicated by Deming (7 p. 14)—afford nearly as much opportunity for thought in the field of statistical inference as the more complicated ones do." Regression analysis has extremely wide use (6, 8–10) in many fields of science and technology (i.e., in chemistry and pharmacy). The 1993 International Conference on Harmonization (ICH) guidelines on Stability Testing of New Drug Substances and Products establish, for example, the use regression analysis to define the shelf lives of drug products (11–12). We will label an individual measurement on a given sample as y_{iv} this means that we are dealing with the vth measurement on the ith sample. If all $v_i = 1$, so that there is only one observation on y at each x, we have a special, and important, case for which the analysis, both theoretical and arithmetical, is simpler (13). For the purpose of the present exposition we will assume that the y_{iv}values are linearly (first-order model) related to the x_i values (14), and that no error is involved in the x_i's. We are concerned with a linear function relationship $\eta = f(\xi)$

$$\eta = \alpha_0 + \alpha_1 \xi \qquad [1]$$

The true variables η and ξ are unobservable, only experimental variables x_i, y_{iv} can be observed. Thus, the linear relation that constitutes the statistical model with replicates for our data (15–16) is

$$y_{iv} = \alpha_0 + \alpha_1 x_i + \varepsilon_{iv} \qquad [2]$$

where (x_i, y_{iv}) is the ith pair of associated x_i and y_{iv} values, α_0 and α_1 are the model parameters to be estimated on the basis of the finite test, and ε_{iv} is the error associated with the measurement y_{iv}. In fact ε_{iv}would be difficult to discover since it changes for each observation y_{iv} which will be assumed

Received 15 March 2004; accepted 27 July 2004.
Address correspondence to Agustin G. Asuero, Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, 41012 Seville, Spain. E-mail: asuero@us.es

uncorrelated (mutually independent) random variables, with mean zero and unknown variances σ^2/w_i, where σ is a constant and w_i is the weight of points iv. Variances can depend on the response, on the predictor, or possibly on other factors. The variance is the square of the standard deviation of the y-value, of any particular value of x. Each measurement y_{iv} is the sum of two terms, the expected value $E\{y_{iv}\} = \alpha_0 + \alpha_1 x_i$ (deterministic part) of the parameters (e.g., we assume that the model is correct and there are no systematic measurement errors), and the error term or disturbance which perturbs the response for that case (stochastic part); $\varepsilon_{iv} = y_{iv} - \alpha_0 - \alpha_1 x_i$, the difference between the observed value and the true value η_{iv}. For n_i genuine replicates, in which x_i is fixed, the model can be written

$$\bar{y}_i = \alpha_0 + \alpha_1 x_i + \bar{\varepsilon}_i \quad [3]$$

where

$$\bar{y}_i = \frac{\sum y_{iv}}{n_i} \quad [4]$$

$$\bar{\varepsilon}_i = \frac{\sum \varepsilon_{in_i}}{n_i} \quad [5]$$

We assume then that the values of the independent variable x are known without error or that the observed values of the independent variable can be chosen and set by the experimenter. In the latter case, we have the controlled independent variable model of Berkson (17) which reduces to the above model. 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_{iv}values, the statistical analysis given here is not valid and a more general approach is necessary (18–22). Now, the true values of the regression parameters, α_0, called the intercept, offset, or blank value (calibration), and α_1, called the regression coefficient, slope, or sensitivity (calibration), are unknown in Equation 3. However, x_i remain fixed, and although we cannot find them exactly without examining (23–24) all possible occurrences of y and x, we can use the information provided by the different observations y_{iv} to give us estimates a_0 and a_1 of α_0 and α_1; thus, we can write

$$\hat{y}_i = a_0 + a_1 x_i \quad [6]$$

where $\hat{y}$, read "y hat," denotes the predicted value of $\bar{y}_i$ for a given x_i, when a_0 and a_1 are determined. Equation (6) could then be used as a predictive equation; substitution for a value of x_i would provide a prediction of the true value of $\bar{y}_i$ for that x_i. The residuals, e_{iv}, are the differences between the responses actually measured, y_{iv}, and those predicted by the model, $\hat{y}_i$, $e_{iv} = y_{iv} - \hat{y}_i$. After the performance of the experiments, one should first evaluate the data graphically. Often this permits an easy detection of important problems. A good way to do this is by examining the residuals (23). In spite of the importance of this topic, it is not included in the International Organization for Standardization (ISO) recommendations (25–26). Residual plots can be used to give information about the error structure of the data, and hence the appropriate weighting factors to be used in the analysis (27). The residuals, $\bar{e}_i = (\Sigma e_{iv})/n_i$, are estimates of the error terms $\bar{\varepsilon}_i$ and should have approximately the same properties. Accordingly, they should show no systematic trends or outliers; if they do, they call into question the assumption underlying the regression model (1, 2, 11). The use of small roman letters a_0 and a_1 to denote estimates of the parameters given by Greek letters α_0 and α_1 is standard. However, the notation $\hat{\alpha}_0$ and $\hat{\alpha}_1$for the estimates is also frequently seen. We use the first type of notation in this article. Since Equation 6 is algebraically equivalent to $x = a_0' + a_1' y$, one might expect $a_1 = 1/a_1'$ and $a_0 = -a_0'/a_1'$ (28), but the best values do not satisfy these equalities because the equations are not statistically equivalent (the two lines are not the same except in the hypothetical situation when all the points lie exactly on a straight line). The discrepancy cannot be eliminated (29) by weighting. Moreover, measurements of y are subject to random experimental errors so we cannot replace y and x by the values of y_i and x_i to write a simple equation for the line (24).

The purpose of weighting factors is to make constant the variance of the weighted $\bar{y}_i$ values. The weighted disturbances $\sqrt{w_{\bar{y}_i}} \cdot \bar{\varepsilon}_i$, should be normally distributed with constant variance. The notation $w_{\bar{y}_i}$ for the weight of $\bar{y}_i$ is due to Deming (7). In principle, weighting factors $w_{\bar{y}_i}$ are based on the information error of y at position x_i. Only relative weights are needed, and for replicate data with constant variances all weights are usually equal to n_i (7) and in absence of replication equal to 1 (30). However, in the weighted model we set $\sqrt{w_{\bar{y}_i}}\ \bar{y}_i$ as the weighted observation. According to Equation 3 we get

$$\sqrt{w_{\bar{y}_i}} \cdot \bar{y}_i = \sqrt{w_{\bar{y}_i}} \cdot \alpha_0 + \sqrt{w_{\bar{y}_i}} \cdot \alpha_1 x_i + \sqrt{w_{\bar{y}_i}} \cdot \bar{\varepsilon}_i \quad [7]$$

the terms $\sqrt{w_{\bar{y}_i}} \cdot \alpha_0$ and $\sqrt{w_{\bar{y}_i}} \cdot \alpha_1 x_i$ remain fixed (free from errors) and then the variance of the weighted observations can be shown to be constant (31), by using the formula $\mathrm{Var}\,(ay) = a^2 \mathrm{Var}\,(y)$, where a is a constant and $w_{\bar{y}_i} = n_i w_i$ (7)

$$\mathrm{Var}\,(\sqrt{w_{\bar{y}_i}} \cdot \bar{y}_i) = w_{\bar{y}_i} \mathrm{Var}\,(\bar{y}_i) = n_i w_i \frac{\mathrm{Var}\,(y_i)}{n_i} = w_i \sigma_i^2 = \sigma^2 \quad [8]$$

($\sigma^2 =$ constant). Ordinarily, the values of the adjustable parameters which lead to a minimum in the sum of weighted squares deviation from the true line

$$S(\alpha_0, \alpha_1) = \sum w_{\bar{y}_i} (\bar{y}_i - \alpha_0 - \alpha_1 x_i)^2 \quad [9]$$

are found by differentiating the sum with respect to each parameter in turn (α_0 and α_1), and equating the derivatives to zero, the coefficients of which, the estimates a_0 and a_1, may be calculated. Note that for linear models, derivatives with respect to any of the parameters are independent of all the parameters.

EVALUATION OF SLOPE AND INTERCEPT

$\partial S/\partial\alpha_0 = 0$ (we substitute α_0 and α_1 for a_0 and a_1 when we equate $\partial S/\partial\alpha_i = 0;\ i = 0, 1$) gives

$$a_0 = \bar{\bar{y}} - a_1\bar{x} \qquad [10]$$

where $\bar{x}$ and $\bar{\bar{y}}$ are the weighted means of the x_i and $\bar{y}_i$ values, respectively.

$$\bar{x} = \frac{\sum w_{\bar{y}_i} x_i}{\sum w_{\bar{y}_i}} \qquad [11]$$

$$\bar{\bar{y}} = \frac{\sum w_{\bar{y}_i} \bar{y}_i}{\sum w_{\bar{y}_i}} \qquad [12]$$

Remember that $\bar{y}_i$ is the mean of the replicate measurements carried out in the run i. The double overbar for the grand mean is used in ISO 5725-2 (32). This means that the weighted centroidal point $(\bar{x},\ \bar{\bar{y}})$ lies on the fitted line. An error in the value of $\bar{\bar{y}}$ leads to a constant error in $\bar{y}$ for all points on the line, the line being translated up or down without a change in slope. Equations 10–12 are also valid when there is an error in both x and y coordinates at some or all of the observed points, but in those cases (33–34) the weight in each point is a reciprocal function of both variances of x and y values as well as the squared slope ($w_i = 1/(\mathrm{Var}\ y_i + a_1^2 \mathrm{Var}\ x_i)$).

On the other hand, Equations 9–10 lead to

$$\begin{aligned} S(\alpha_0, \alpha_1) &= \sum w_{\bar{y}_i} (\bar{y}_i - \bar{\bar{y}} - \alpha_1(x - \bar{x}))^2 \\ &= S_{YY} - 2\alpha_1 S_{XY} + \alpha_1^2 S_{XX} \end{aligned} \qquad [13]$$

where S_{XX} and S_{YY} are the sums of the squares of deviations from the mean for the two variables (x and $\bar{y}$) and S_{XY} is the corresponding sum of the cross-products

$$\begin{aligned} S_{XX} &= \sum w_{\bar{y}_i} (x_i - \bar{x})^2 \\ &= \sum w_{\bar{y}_i} x_i^2 - \frac{\left(\sum w_{\bar{y}_i} x_i\right)^2}{\sum w_{\bar{y}_i}} \end{aligned} \qquad [14]$$

$$\begin{aligned} S_{XY} &= \sum w_{\bar{y}_i} (x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}}) \\ &= \sum w_{\bar{y}_i} x_i \bar{y}_i - \frac{\left(\sum w_{\bar{y}_i} x_i\right)\left(\sum w_{\bar{y}_i} \bar{y}_i\right)}{\sum w_{\bar{y}_i}} \end{aligned} \qquad [15]$$

$$S_{YY} = \sum w_{\bar{y}_i} (y_i - \bar{\bar{y}})^2 = \sum w_{\bar{y}_i} \bar{y}_i^2 - \frac{\left(\sum w_{\bar{y}_i} \bar{y}_i\right)^2}{\sum w_{\bar{y}_i}} \qquad [16]$$

This notation is similar to that used in unweighted linear regression (23). Similar shortcuts were due to Gauss (7). The first forms in Equations 14–16 require the computation of $(x - \bar{x})$ and $(y - \bar{\bar{y}})$ for all observations. This may be tedious if x and $\bar{y}$ involve several decimal places, and the computation of products and squares becomes more laborious than the direct operation with the second form in Equations 14–16, which is normally performed (23, 35–37) in pocket calculators. It is worth noting that the first and second parts of the right member of the second form of Equations 14–16, involve the subtraction of two large positive quantities from each other. Therefore, these equations are potentially inaccurate if only a limited number of significant figures are carried out on calculators and computers (36–37). As a matter of fact, the forms of Equations 14–16 are unstable. They have the property that they suffer (38–39) from subtractive cancellation for data sets with small coefficients of variation (the ratio of standard deviation s to the mean $\bar{x}$). To avoid rounding errors, it is best to carry as many significant figures as possible in these computations. Rounding is best done at the reporting, not intermediate, stage of a calculation. Most digital computers, because of their round off characteristics (23) provide more accurate answers using the first forms in Equations 14–16.

Then, $\partial S/\partial\alpha_1$ gives

$$a_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum w_{\bar{y}_i}(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})}{S_{XX}} = \frac{\sum w_{\bar{y}_i}(x_i - \bar{x})\bar{y}_i}{S_{XX}} \qquad [17]$$

since the other term removed from the numerator is

$$\sum w_{\bar{y}_i}(x_i - \bar{x})\,\bar{\bar{y}} = \bar{\bar{y}} \sum w_{\bar{y}_i}(\bar{x}_i - \bar{x}) = 0 \qquad [18]$$

Equation 17 and

$$\begin{aligned} a_0 = \bar{\bar{y}} - a_1\bar{x} &= \frac{\sum w_{\bar{y}_i}\bar{y}_i}{\sum w_{\bar{y}_i}} - \frac{\sum w_{\bar{y}_i}(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})}{\sum w_{\bar{y}_i}(x_i - \bar{x})^2} \frac{\sum w_{\bar{y}_i}\bar{x}_i}{\sum w_{\bar{y}_i}} \\ &= \frac{\sum w_{\bar{y}_i}\bar{y}_i \cdot \sum w_{\bar{y}_i}(x_i - \bar{x})^2 - \sum w_{\bar{y}_i}(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}}) \cdot \sum w_{\bar{y}_i}\bar{x}_i}{\sum w_{\bar{y}_i} \sum w_{\bar{y}_i}(x_i - \bar{x})^2} \end{aligned} \qquad [19]$$

should always be calculated on a computer (35, 40) in preference. Fitting a straight line through the origin represents a strong assumption, which in general is not justified. Even when the model is known to pass through the origin, it does not mean that a straight line fit through the origin is necessarily appropriate (23). The question of whether to force the regression line through the origin has been discussed in detail (13). Though in most analytical situations $a_0 = 0$ could be justified by theory, reality is rarely as simple as theory, for example, the lack of selectivity, or unexpected interaction between chemical species (41).

Equation 17 shows that the parameter estimates are weighted linear combinations of all $\bar{y}_i$, the magnitudes of the weight coefficient depending on the location of experimental data. This important conclusion tells us that when a value of x_i is far from the mean $\bar{x}$, the weight is large so that the point $(x_i, \bar{y}_i)$ has a great "weight coefficient," and is more significant in the estimate (23).

VARIANCE OF THE ESTIMATED PARAMETERS AND PREDICTED VALUES

By applying the random error propagation law (42) and recognizing that x has a negligible error, from Equation 17 we

get

$$\mathrm{Var}\,(a_1) = \frac{1}{S_{XX}^2}\mathrm{Var}\left(\sum\left(\sqrt{w_{\bar{y}_i}}\,(x_i-\bar{x})\sqrt{w_{\bar{y}_i}}\,\bar{y}\right)\right) = \frac{1}{S_{XX}^2}(S_{XX}\,\sigma^2) = \frac{\sigma^2}{S_{XX}} \quad [20]$$

The variance of the weighted mean $\bar{\bar{y}}$ is given by

$$\mathrm{Var}\,(\bar{\bar{y}}) = \mathrm{Var}\left(\frac{\sum w_{\bar{y}_i}\bar{y}_i}{\sum w_{\bar{y}_i}}\right) = \frac{\mathrm{Var}\left(\sum\left(\sqrt{w_{\bar{y}_i}}\left(\sqrt{w_{\bar{y}_i}}\bar{y}_i\right)\right)\right)}{\left(\sum w_{\bar{y}_i}\right)^2} = \frac{\sigma^2}{\sum w_{\bar{y}_i}} \quad [21]$$

Taking into account that $\bar{\bar{y}} = \sum a_i\sqrt{w_{\bar{y}_i}}\bar{y}_i$ and $a_1 = \sum c_i\sqrt{w_{\bar{y}_i}}\bar{y}_i$ where

$$a_i = \frac{\sqrt{w_{\bar{y}_i}}}{\sum w_{\bar{y}_i}} \quad c_i = \frac{\sqrt{w_{\bar{y}_i}}\,(x_i-\bar{x})}{S_{XX}} \quad [22]$$

we get

$$\mathrm{Cov}(a,c) = \left(\sum a_i c_i\right)\mathrm{Var}\left(\sqrt{w_{\bar{y}_i}}\cdot\bar{y}_i\right) = \left(\frac{\sum w_{\bar{y}_i}(x_i-\bar{x})}{S_{XX}\sum w_{\bar{y}_i}}\right)\sigma^2 = 0 \quad [23]$$

and so, $\bar{\bar{y}}$ and a_1 are uncorrelated random variables. Then, from Equation 19 we have

$$\mathrm{Var}\,(a_0) = \mathrm{Var}\,(\bar{\bar{y}}) + \bar{x}^2\,\mathrm{Var}\,(a_1) \quad [24]$$

and taking into account Equations 21, 20, and 24, the variance of intercept is given by

$$\mathrm{Var}\,(a_0) = \left(\frac{1}{\sum w_{\bar{y}_i}} + \frac{\bar{x}^2}{S_{XX}}\right)\sigma^2 = \left(\frac{\sum w_{\bar{y}_i}\,x_i^2}{S_{XX}\sum w_{\bar{y}_i}}\right)\sigma^2 \quad [25]$$

If we now apply the propagation variance law to Equation 10 one rewritten as $\bar{\bar{y}} = a_0 + a_1\bar{x}$, we obtain

$$\mathrm{Var}\,(\bar{\bar{y}}) = \mathrm{Var}\,(a_0) + \bar{x}^2\,\mathrm{Var}\,(a_1) + 2\bar{x}\,\mathrm{Cov}\,(a_0,\,a_1) \quad [26]$$

and on rearrangement and taking into account Equations 21, 25, and 20,

$$\mathrm{Cov}\,(a_0,a_1) = \frac{\mathrm{Var}\,(\bar{\bar{y}}) - \mathrm{Var}\,(a_0) - \bar{x}^2\mathrm{Var}\,(a_1)}{2\bar{x}} = -\bar{x}\frac{\sigma^2}{S_{XX}} = -\bar{x}\mathrm{Var}\,(a_1) \quad [27]$$

Estimation of the parameters and their associated errors is not the end-all of linear regression. Usually, these results are used to "predict" values of one of the data variables based on a value of the other variable (15). If these values are within the range of the x data, it is termed "interpolation." If they are not, it is termed "extrapolation." Extrapolation is strongly discouraged because it cannot always be correctly assumed that the model on which the experiment is based will hold outside the limits of the actual data.

The fitted regression equation is

$$\hat{y}_i = \bar{\bar{y}} + a_1(x_i-\bar{x}) \quad [28]$$

where both $\bar{\bar{y}}$ and a_1 are subject to error, which will influence $\hat{y}_i$. Let x_i be a selected value of x. The predicted mean value of $\hat{y}$ for this value of x is

$$\mathrm{Var}\,(\hat{y}_i) = \mathrm{Var}\,(\bar{\bar{y}}) + (x_i-\bar{x})^2\mathrm{Var}\,(a_1) = \left(\frac{1}{\sum w_{\bar{y}_i}} + \frac{(x_i-\bar{x})^2}{\sum w_{\bar{y}_i}(x_i-\bar{x})^2}\right)\sigma^2 \quad [29]$$

This formula applies only to values of $\bar{y}$ predicted from values of x that were included in the x data set. We might expect to make our test prediction in the middle of our observed range of x and would expect our prediction to be less good away from the middle. The expected value of $\hat{y}$ is $E\,[\hat{y}] = \alpha_0+\alpha_1 x = \eta$. Thus, $\hat{y}$ is a linear function of a_0 and a_1 and the latter are too linear functions on $\bar{y}_i$ which is normally distributed, and therefore $\hat{y}$ is also normally distributed. A new observation $\bar{y}_0$ at x_0 will be distributed about η_0 with a variance $\sigma/w_{\bar{y}_0}$ independently of $\hat{y}_0$, so the expectation $E[\bar{y}_0 - \hat{y}_0] = 0$ and the variance is

$$\mathrm{Var}\,(\bar{y}_0-\hat{y}_0) = \frac{\sigma^2}{w_{\bar{y}_0}} + \mathrm{Var}\,(\hat{y}_0) = \left(\frac{1}{w_{\bar{y}_0}} + \frac{1}{\sum w_{\bar{y}_0}} + \frac{(x_0-\bar{x})^2}{\sum w_{\bar{y}_0}(x_0-\bar{x})^2}\right)\sigma^2 \quad [30]$$

where $w_{\bar{y}_0}$ is a weighting factor appropriate to the value of $\bar{y}_0$. Thus, for values of y, $\hat{y}_0$, predicted from new values of x, x_0, an extra component of y variance must be added to the variance of Equation 29 to obtain the variance of $\hat{y}_0$.

The weighted residuals are defined as the k differences

$$\sqrt{w_{\bar{y}_i}}e_i = \sqrt{w_{\bar{y}_i}}(\bar{y}_i - \hat{y}_i) \quad [31]$$

$i = 1, 2, k$, where $\bar{y}_i$ is the mean of the y_{iv} values and $\hat{y}_i$ is the corresponding fitted value obtained by use of the fitted regression equation (the observed errors in the model are correct). In order to examine the residuals in the weighted model, the weighted residuals may be plotted versus predicted values, chronological order of measurements, observed values, and design variables (23).

The estimate of variance provided by the weighted squared residuals is given by (3, 5, 7, 15)

$$s_{\bar{y}/x} = \frac{\sum w_{\bar{y}_i}(\bar{y}_i-\hat{y}_i)^2}{k-2} = \frac{S_{YY} - a_1^2 S_{XX}}{k-2} \quad [32]$$

The estimated variances of a_0, a_1, and $\hat{y}$ are obtained simply by replacing σ^2 by $s^2_{\bar{y}/x}$ in the above equations. For the weighted case, the size of $s^2_{\bar{y}/x}$ is directly affected by the relative size of the weights, and thus is not usually an estimate of σ^2 (15). However, it is needed for calculating estimates for the variances of the intercept and slope.

Interval estimates (prediction intervals) for the parameters a_0, a_1, and η are given by their statistic $\pm\, t_{f,\alpha/2}\cdot$ (estimated variance of the statistics)$^{1/2}$, where $t_{f,\alpha/2}$ is that value of Student's t statistic (with $f = k - 2$ degrees of freedom) such that the probability of $t_f > t_{f,\alpha/2}$ equals $\alpha/2$. To obtain simultaneous confidence curves appropriate for the whole regression function over its entire range, it would be necessary to replace $t_{(v;1-1/2\alpha)}$ by $\sqrt{2F}_{(2,k-1,1-\alpha)}$. The confidence region for the entire line, the Working-Hotelling region is given by (14)

$$a_0 + a_1 x \pm \left(2F_{2,f,\alpha}\left[\frac{1}{\sum w_{\bar{y}_i}} + \frac{(x-\bar{x})^2}{S_{XX}}\right]s^2_{\bar{y}/x}\right)^{1/2} \qquad [33]$$

WEIGHTED VERSUS UNWEIGHTED LINEAR REGRESSION

Contrary to the effect of adding in x-errors (43), the weighted adjustment causes no severe impediments to obtain formulas for the slope and intercept. If a weighted least squares analysis was called for but an ordinary least squares analysis was performed, the estimates obtained would still be unbiased but would not have minimum variance, since the minimum variance estimates are obtained from the correct weighted least squares analysis (1). The confidence limits for weighted regression lines show the weighted centroid $(\bar{x}, \bar{\bar{y}})$ closer to the origin than its unweighted counterpart (44–46). The precision varies with the position on the graph. A given graph shows maximum absolute precision at an x-value equal to the weighted mean $\bar{x}$-value. Variance is greater at both higher and lower x-values. It increase rapidly with x-values at x-values above the highest x-level.

The incorporation of the heteroscedasticity in the calibration procedures is recommended/prescribed by several international (ISO 9169, ISO/CD 13-752) standards (25–26). The International Union of Pure and Applied Chemistry (IUPAC) has incorporated the issue of "heteroscedasticity" or nonconstant variance into their recommendations for calculating limits of detection and quantification (47). Calculations of weighted regression lines with replication are evidently more complex than unweighted regression computations (48–49). Many commercial software packages for data analysis include weighted least squares (usually without replication). Unfortunately spreadsheets do not (yet) do so, although macros are available to remedy this deficiency (2, 50). The necessity to retain many extra figures in the calculations is characteristic of this approach to the least squares. Before proceeding with use of weighted least squares, however, the analyst should consider the possibility (31) that large variations in variance could due to nonoptimum instrument operating conditions. Correction of problem such as a noisy lamp in fluorimetry will usually yield greater improvements than better statistical techniques.

DETERMINING THE VARIANCE FROM A NUMBER OF SMALL SAMPLES: PURELY EXPERIMENTAL UNCERTAINTY

Linear least squares regression is widely used to analyze the results of calibration curves or standard addition methods (51), but it requires repeated observations of standard solutions in order to properly verify homogeneity of variance and linearity. For a statistical evaluation of the results, one should check whether the data are free from outliers. Outlying points may disturb the normality of the data, which is required by most of the tests used to examine the behavior of the variance and the goodness-of-fit (52). Tests to trace single outliers, namely the Dixon (53) and the Grubbs (54–55) tests may be applied. We now consider a number of populations of different means $\bar{y}_i$ but with common variance σ^2. The problem is to find the best estimate of this common variance σ^2. Since the individual variances are found from (genuine) replicate reading, it follows (if the y_{iv} are independent random variables with variance σ^2) that each of the sample variances are given (56) by

$$s_i^2 = \frac{\sum (y_{iv} - \bar{y}_i)^2}{n_i - 1} = \frac{Q_i}{f_i} \qquad [34]$$

$(i = 1, 2, \ldots k)$ where y_{iv} is the vth observation $(v = 1, 2, \ldots n_i)$ at x_i and $\bar{y}_i$ is the mean of the y values at x_i, Q_i is the contribution to the pure error sum of squares from the n_i observations at x_i (internal sum of squares of the y_{iv} about their average $\bar{y}_i$), and f_i is the respective number of degrees of freedom. To avoid difficulties with false (single) outliers, the number of replicate measurements (52) at each x level (i.e., concentration in calibration) should be sufficiently high (at least >4). On the other hand, the minimum number of observations necessary to obtain a useful estimate of standard deviation is six (57). However, this is barely realistic in a routine operation in order to avoid the increase in cost and time of analyses.

For those experiments that were not replicated, the mean response is simply the single value of response. Each of the k calculated sample variance values s_i^2 are estimates of the population variance σ^2 of the parent distribution. Thus being so, all s_i^2 values may be combined in order to yield a better value for σ^2. The pooled replication variance for all experiments is equal to (58) the weighted mean of the individual variances; the weights being the degrees of freedom

$$\bar{s}^2 = \frac{\sum_1^k Q_i}{\sum_1^k f_i} = \frac{\sum_1^k f_i s_i^2}{\sum_1^k f_i} \qquad [35]$$

The number of degrees of freedom of the pooled variance is equal to the number of observations minus the number of

constraints used to determine the k means (56)

$$\sum_{1}^{k} f_i = \sum_{1}^{k} (n_i - 1) = \sum_{1}^{k} n_i - k \qquad [36]$$

From Equations 34–36, we get

$$\bar{s}^2 = \frac{\sum_{i=1}^{k} \sum_{v=1}^{n_i} (y_{iv} - \bar{y}_i)^2}{\sum_{i=1}^{k} n_i - k} \qquad [37]$$

the numerator of the right hand of Equation 37 being the total pure error sum of squares. The total number of degrees of freedom for the total reproducibility variance, $\bar{s}^2$, is markedly greater than for each variance taken separately. Therefore, the total replication variance gives a far better estimate of the population variance σ^2. Note that the pure error mean square $\bar{s}^2$ contains only deviation from the group averages that are independent of the model used; it consists of a within group sum of squares not containing the parameters of the model, and therefore is an estimate of σ^2 irrespective of whether the model being fitted is correct or not (59).

Complete symmetry of replication makes the arithmetic computation easier, allows dealing with missing values in a logical way, and makes it simpler to judge outliers (60). If the sample variance are obtained from samples of the same size, $n_1 = n_2, \ldots, n_k, \ldots = n$, the expression for the replicate variance is greatly simplified

$$\bar{s}^2 = \frac{\sum f_i s_i^2}{\sum f_i} = \frac{\sum (n-1) s_i^2}{\sum (n-1)} = \frac{\sum s_i^2}{k} \qquad [38]$$

because the number of degrees of freedom for the pooled sample variance is equal to

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} (n-1) = \sum_{i=1}^{k} n - k = k(n-1) \qquad [39]$$

The estimated reproducibility variance in this case is given by

$$\bar{s}^2 = \frac{\sum_{i=1}^{k} \sum_{v=1}^{n_i} (y_{iv} - \bar{y}_i)^2}{k(n-1)} \qquad [40]$$

If there are only two observations y_{i1}, y_{i2} at the point x_i, then

$$\sum_{v=1}^{v=2} (y_{iv} - \bar{y}_i)^2 = \frac{1}{2}(y_{i1} - y_{i2})^2 \qquad [41]$$

and this is an easier form to compute (23). This sum of squares has one degree of freedom (61). Thus, with only $n_i = 2$ replicates at each of the k set of conditions, the formula for the ith variance (Equation 35) reduces (62) to

$$s_i^2 = \frac{d_i^2}{2} \qquad [42]$$

where d_i is the difference between the duplicate observations for the ith set of conditions.

Thus, Equation 35 yields

$$\bar{s}_i^2 = \frac{\sum d_i^2}{k} \qquad [43]$$

An often-used relationship (63) that can be derived from Equation 37 is

$$\bar{s}^2 = \frac{\sum_{i=1}^{k} \left(\sum_{v=1}^{n_i} y_{iv}^2 - \frac{(\sum_{v=1}^{n_i} y_{iv})^2}{n_i} \right)}{\sum_{i=1}^{k} n_i - k} = \frac{\sum_{i=1}^{k} \sum_{v=1}^{n_i} y_{iv}^2 - \sum_{i=1}^{k} \left(\frac{(\sum_{v=1}^{n_i} y_{iv})^2}{n_i} \right)}{\sum_{i=1}^{k} n_i - k} \qquad [44]$$

However, the result obtained by means of Equation 44 is less accurate (40) than that obtained by using Equation 37. As a matter of fact, if no prior estimate of σ^2 is available, but repeat measurements of y (i.e., two or more measurements) have been made at the same value of x, we can use these repeat measurements to obtain an estimate of σ^2. Such an estimate is said to represent "pure error" because, if the setting of x is identical for several observations, only the random variation can influence the results and provide differences between them. Such differences will usually provide an estimate of σ^2 which is more reliable than we can obtain from any other source. For this reason, it is sensible when designing experiments to arrange for repeat observations. The lack of replicates also hampers the application of outlier tests at the different levels, as well as the evaluation of tests that compare variance estimates at different levels (52).

The estimate of σ^2 made from the internal consistency of the data (i.e., from the consistency of observations within samples) is only possible if there are points in which there is more than one observation. When there is but one observation at each point, the estimate of σ^2 by internal consistency (7) is not a possibility. It should be emphasized that the pooling of variance estimates in accordance with Equation 35 is justified only if we can confidently assume that all the populations involved have a common variance σ^2. Should this not be the case, then the pooling procedure degenerates into a mere exercise in arithmetics and leads to a quantity which at best is of nebulous meaning (64) and at worst has no meaning at all.

TESTING FOR HOMOGENEITY OF VARIANCES

Modern measurement systems generally involve some kind of instrumental response (peak area, peak height, electric current, etc.) which is an indirect reflection of the magnitude of the quantity being measured (11). Many analytical data show nonuniform variance. In most cases, knowledge of variance must be gained experimentally. One of the most evident ways of

achieving this is by measuring a sufficient number of replicates at each concentration x_i (65–66). Knowledge of the variance of experimental data is fundamental to optimal design and proper analysis in many areas of investigation. The validation of the calibration procedure, For example, involves an examination of the behavior of the variance and of goodness-of-fit of the selected model (11, 52). For many reasons, most of which are of a statistical nature, it is desirable to deal with homoscedastic data (i.e., homogeneous variance, standard structure) whenever this is possible. The responses $y_1 \ldots y_n$, are independently sampled, or for a lesser restriction have zero covariances, and come from population (67) with a common variance σ^2. This we will call the standard structure. The variance of y at each point x_i should be equal (i.e., constant over the whole working range x); in other words, the errors in measuring y_{iv} are independent of the values of x. However, in the experimental sciences, more often than not the error increases as the value of the measured variable increases (68–70), that is, the variance is a function of the factor point associated with the response. This we will call the heteroscedastic structure (68, 71) or nonuniform variance condition (69). However, weights are often neglected in regression methods. Nonconstant variance or heteroscedasticity often appears in chromatography and capillary electrophoretic applications if a certain range of concentration is exceeded (71–72). In spectrophotometric measurements, the uncertainty in the absorbance depends strongly on absorbance and on other parameters of the experiment as well. Spectrophotometric absorbances are strongly heteroscedastic (73–74). Variance function of the response is not uniform over the concentration range in bioanalysis (75). According to some recent (71) results, weighted least squares saves us from nonnegligible errors when the ratio of standard deviations s_i within the x range (i.e., concentration) explored is higher than 5, while it is mandatory if it exceeds 50. If sufficient replicate measurements in $x_1, x_2, \ldots x_k$ are available it is possible to estimate $\sigma_1^2\ \sigma_2^2\ \sigma_k^2$ separately. In weighted regression, the observations are weigthed by a factor inversely proportional to the variance or to the deviation standard of the y value, for example,though other sophisticated weighting schemes can also be proposed (76). In other words, the regression takes less notice of points with greater absolute variation, and provides a line that passes closer to the points with smaller variance; hence, providing a more realistic fit (19, 77). Not only does the error in many experimental techniques vary with the values of the independent variable but it often varies (78) from day to day and from technician to technician. Indeed so many factors other than direct instrumental error contribute to the overall error of any experimental technique that it has become a basic precept of the experimental method that whenever possible an experiment should be designed to include a sufficient number of complete replicates so as to provide an estimate of experimental error from the data within any run of an experiment.

Testing for homogeneity of variances in linear regression then requires (13, 79) us to arrange repeat runs at fixed values of the predictor variable x, in order to obtain estimates of the variance of the observations at the different values of the x variate (when the functional relationship between x and σ is not known, or if no relationship of this kind exists). In the absence of sufficient number of replicates, a functional relationship between variance and concentration can be assumed (65, 80–83). The random errors which are obtained by repeatedly measuring the same sample are caused by noise which may be due to several factors (65, 68, 84). Often it is not clear by inspection whether variances are heterogeneous (i.e., the variance of the y_i observations is or is not dependent on the x variate). As there is no justification for assuming that this spread is necessarily the same at all x levels, it is for ease of detection of this effect that the samples are arranged in increasing order of magnitude of the measured characteristic, being instructive to examine the relationship between s_i and $\bar{y}_i$(or s_i and x_i) by plotting the first quantity versus the second and fitting an equation through the data points (65). This plot may suggest (69, 65) a proportional relationship. The variance is often assumed to be linearly related to the logarithm of the estimate or to the expected value of the variable (85). A diagnostic procedure particularly designed for regression, and avoiding the need of repeated measurements, has been developed by Cook and Weisberg (86).

THE F TEST

The F test, so named by George Snedecor (62) in honor of R. A. Fischer, is used for comparing two variances of the two normal populations (two independent variates each distributed as chi square, χ^2). The null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ is tested by using a statistic that is in the form of a ratio of estimates, s_1^2/s_2^2 (where s_1^2 is the larger variance estimate and s_2^2 is the smaller variance estimate), which follows a theoretical distribution known as an F distribution (one-sided test when the alternative to the null hypothesis is $s_1^2 > s_2^2$, and a two-sided test when the alternative is simply $s_1^2 \neq s_2^2$). If the computed F exceeds the tabulated $F_{1-\alpha}(f_1, f_2)$ value, the probability that the difference between the two variances is caused by chance alone is smaller than the specified probability, and we are justified in rejecting the null hypothesis, with the given probability (28) of committing a type I error, that is, the critical level has a specifically defined false positive, unlike an undefined false negative or type II error.

Intuitively if the variances are equal, this ratio should be approximately equal to one, so values that differ greatly from one indicate inequality. The larger the sample the more accurately the variance is determined the lower value of F at which the null hypothesis is rejected with a given probability of a correct decision. Another formal test based on an F test has been published by Franke et al. (87) and is proposed by the ISO (88–89). Since the variance ordinarily changes smoothly with the response in analytical chemistry applications, ISO proposes to carry out a one-tailed F test checking whether the variance at

the higher concentration level is significantly larger than at the lowest concentration level.

F tests are robust with respect to departures from homogeneity, that is, moderate departures from equality of variance do not greatly affect the F statistic. Apparently the F test, however, is more sensitive (90) to unequal variances than to nonnormality. When we wish to compare more than two variances to see if they differ significantly, we could apply the F test repeatedly to all possible pairs of the set of variances. If we have a set of k variances, the F test could be applied to $C_k^2 = k(k-1)/2$ variance pairs, and if we select the 95% confidence level for all comparisons, by the time the $k(k-1)/2$ pairs were compared, the confidence level for making the correct decision on all $k(k-1)/2$ would have dropped to only $(0.95^{k(k-1)/2}) \cdot 100\%$. Note, however, that the confidence level for each comparison is still 95%. Moreover the F test, by comparing variances of replicates at the lowest and highest x value, does not save us from oscillations of variance at intermediate x level and implies large increases of cost and time of analysis (91).

To avoid this situation, there are several tests for comparing more than two variances if the experimenter fears a great departure from homogeneity (testing for homogeneity of variances), though unfortunately most of these rely on the assumption of normality. The most popular test is that suggested by Bartlett (92–94). For samples varying in size, the null hypothesis that the respective population variances are equal can be tested by the Bartlett test. However, if the sample variances are obtained from samples of the same size, resort is made to the more convenient and accurate Cochran's test (95–98). Bartlett's test has the disadvantage of being more complicated to perform and being very sensitive to departures from normality of distribution. A very simple test which does not show those disadvantages has been reported by Burr and Foster (99–100). The calculations involved in the Burr and Foster's test (101) are very simple, and, in addition the test is not affected if a sample variance is equal to zero, unlike other simple tests such as those due to Hartley (101). Another approach is to use methods which are insensitive to the unequal variances. Levene's test (81) is a robust test in that it is not sensitive to the violation of normality assumption.

BARTLETT'S TEST

The Bartlett test is a combination of a sensitive test of normality, more precisely the "long-tailedness" of a distribution, with a less sensitive test of equality of the variances (103). It is the only one (of the common tests) which can be used to compare more than two variances whose degrees of freedom are not equal; this means that it can manage unbalanced data sets. Bartlett's test is a special application of the χ^2 test, in which we compare (104–108) the difference between the total number of degrees of freedom times the natural logarithm of the pooled estimate of variance, and the sum, extended over all samples, of the product of the degrees of freedom and the natural logarithm of the estimate of variance

$$B = \ln \bar{s}^2 \sum_{i=1}^{k} f_i - \sum_{i=1}^{k} \left(f_i \ln s_i^2 \right) \qquad [45]$$

This enables us to test the hypothesis that all the variances are homogeneous, with $k-1$ degrees of freedom, if all f_i are greater than two (28). The null hypothesis for a statistical test would be $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$; the alternative hypothesis is H_1: at least two of the population variances are not equal. If the calculated value of $\chi^2\,[B]$ is greater than the tabulated value at a specified level of significance, we conclude that the variances are not homogeneous (the level of significance represents, as always, the probability of our having reached the wrong conclusion). Nevertheless, the value of χ^2 as calculated from Equation 45 is biased toward the high side so that we may wrongly reject the null hypothesis. If χ^2 indicates acceptance, then after correction for bias the hypothesis would be even more likely to be correct so that we need not worry about the bias. However, rejection at the 5% level has to be checked by calculating a corrected value of χ^2, say $\chi^2[C]$. This is given by

$$\chi_{[C]} = \frac{B}{C} \quad \frac{B}{C} \leq \chi_{1-\alpha}^2 \qquad [46]$$

where

$$C = 1 + \frac{1}{3(k-1)} \left(\sum \left(\frac{1}{f_i} \right) - \frac{1}{\sum f_i} \right) \qquad [47]$$

Since $C > 1$ always, the null hypothesis must be accepted if $B \leq \chi_{1-\alpha}^2$. If $B > \chi_{1-\alpha}^2$ Bartlett's test is applied in full. Hartley (109) has shown that the approximation involved in using the χ^2 distribution for the distribution of B/C is poor for small n_i and has given a modification with the necessary tables.

Often a considerable computational effort can be saved by applying an F test to the largest and smallest variances before Bartlett's test. If an F test indicates that the largest variance is not significantly different from the smallest one, then one can reasonably assume (60) that the variances can be regarded as homogeneous. When applied to data whose distributions are not normal, Barlett's test may give an erroneous verdict of nonhomogeneity (110). Because of this sensitivity to distribution, some analysts do use it when the distributions are at least approximately normal and because other tests for homogeneity of more than two variances whose degrees of freedom are not equal are not very popular. Bartlett's test was recommended by Danzer and Currie (111) in the recent guidelines for calibration in analytical chemistry. One should always keep in mind that these tests are very sensitive when the number of replicates is small. It can happen that nonuniformity of variance might not be detected due to a lack of replicates (28).

With an equal number of replicate determinations in each experiment, $n_1 = n_2 = \cdots = n_k = n$, the total variance is equal to the mean of the individual variances, and in these cases

Equation (45) reduces to

$$B = (n-1)\left(k \ln \bar{s}^2 - \sum \ln s_i^2\right) \quad [48]$$

The rationale for the statistic $\chi^2[C]$ can be clearly noticed in this case (102). Note that the numerator of $\chi^2[C]$, (i.e., B) is $\ln 10(n-1)k$ times (log of the arithmetic mean of the s_i^2 minus log of the geometric mean of the s_i^2). When the s_i^2 are all close together, the arithmetic mean and the geometric mean will be close. When the s_i^2 are discrepant, the arithmetic mean will be much larger than the geometric mean, and hence will lead to a large value of B, and so, of the $\chi^2[C]$.

In the Bartlett test modified for kurtosis, the statistic B of Equation 45 is multiplied by $d = 2/(g_2 - 1)$, where g_2 estimates the kurtosis γ_2 of the sets of repeats (112–113). In normally distributed data, the true γ_2 would be 3 and d would typically be close to 1

$$g_2 = \frac{k \sum_{i=1}^{k} \sum_{v=1}^{n_i} (y_{iv} - \bar{y}_i)^4}{\sum_{i=1}^{k} \sum_{v=1}^{n_i} (y_{iv} - \bar{y}_i)^2} \quad [49]$$

COCHRAN'S TEST

Cochran's test (56, 62, 114–115) is designed for use in situations in which the numbers of repeat determinations are equal. It is one of several significative tests which are known as homogeneity of variance tests, appropriate if we suspect a single one of the variances of being appreciably greater than the remaining $k-1$, and which share the null hypothesis that several variances are equally variable. Bartlett's test is sensitive to nonnormality (56, 106) so that it cannot be relied on when the normality of the data is seriously in doubt. Cochran's test is particularly sensitive to the case where all the variances σ_k^2 are expected to be equal, except for one variance which could be larger.

Like the F test, the Cochran test uses a ratio of variances, but in a different manner (115). This ratio is called the G statistic. It compares the largest of the set of variances with the other variances in the set

$$G = \frac{s_{\max}^2}{\sum_{i=1}^{k} s_i^2} = \frac{\sum_{v=1}^{n} (y_{iv} - \bar{y}_i)_{\max}^2}{\sum_{i=1}^{k} \sum_{v=1}^{n} (y_{iv} - \bar{y}_i)^2} \quad [50]$$

The G distribution depends only on the number k of variances and the degrees of freedom f with which each variance is determined: $f = n - 1$. If the value of the Cochran statistic as found from the sample variance is smaller than the tabulated one $G < G_{1-\alpha}(n-1, k)$ where k is the number of variances, the difference between the variances at the selected significance level (114) is explained by chance. Note that the test is inherently one-sided. Cochran's maximum variance test is recommended by ISO 5725-2 for use in the analysis of precision experiments (31). Theoretically this test (likely Hartley's test) requires an equal number of measurement points at each x level. However, ISO indicates that for the Cochran's test small differences in the number of points can be ignored, and applies the Cochran criterion for the number of measurements that occur at most x levels (i.e., concentration level). The Association of Official Analytical Chemists (AOAC) guidelines recommend in interlaboratory collaborative trials (11) a significance level of 0.025 (i.e., the critical value leaves an area of 0.025 in the right-hand tail of the sampling distribution). Sometimes, a different number of observations is available or retained. The error introduced with unbalanced data arrays, those containing a different number of observations (e.g., one result had been discarded) would be small and would not have affected our decision. To overcome this difficulty, we could have used a method for replacing missing values. These methods can be very useful when the data from a carefully balanced experiment is spoiled by a small number of missing or rejected determinations (51). When the sample sizes are not too different, one computes their harmonic mean $\bar{y}_H$ and interpolates in tables for $v = \bar{y}_H - 1$ (103). The harmonic mean of a set of numbers is the reciprocal of the arithmetic mean of the reciprocal of the numbers.

In most situations, Cochran's test is equivalent to Hartley's. Since the Cochran test utilizes more information, it is somewhat more sensitive (106). A very similar test (79), which is however based on the ranges of the individual samples, is described by Bliss, Cochran, and Tukey (116); examples and the upper 5% bounds can be found in the original paper.

BURR AND FOSTER'S TEST

For samples varying in size, the statistics q_c is built as (99, 117)

$$q_c = \bar{f} \frac{f_1 s_1^4 + \cdots + f_k s_k^4}{\left(f_1 s_1^2 + \cdots + f_k s_k^2\right)^2} = \bar{f} \frac{\sum f_i s_i^4}{\left(\sum f_i s_i^2\right)^2} \quad [51]$$

$f_i = n_i - 1$ being the degrees of freedom and $\bar{f}$ the harmonic mean of the degrees of freedom

$$\bar{f} = \frac{k}{\sum_{i=1}^{k} \left(\frac{1}{f_i}\right)} \quad [52]$$

If $q_c > q$, being q the corresponding tabulated value, the null hypothesis is rejected, the variances being regarded as heterogeneous.

For samples of the same size, $n_1 = n_2 = \cdots = n_k$, the q_c statistics is given by

$$q_c = \bar{f} \frac{f_i \sum s_i^4}{f_i^2 \left(\sum s_i^2\right)^2} = \frac{\sum s_i^4}{\left(\sum s_i^2\right)^2} \quad [53]$$

the harmonic mean of the degrees of freedom being equal in this case to $\bar{f} = n - 1$.

HARTLEY'S F_{max} TEST

Hartley's test is one of the simplest tests; it may be used when all treatment groups are of the same size and involves

comparing the largest sample variance with the smallest sample variance. This is a quick test that sometimes may be used with advantage (79). If the samples which serve as the basis for computing the values s_i^2 have equal sizes the H_0 hypothesis ($\sigma_j^2 = \sigma^2$; $j = 1, \ldots, k$) can be tested by means of the Hartley's $F_{\max}$ test (101) normality of the data being assumed.

The statistic test for the null hypothesis has the following form:

$$F_{\max} = \frac{\max\left[s_1^2 \ldots s_i^2 \ldots s_k^2\right]}{\min\left[s_1^2 \ldots s_i^2 \ldots s_k^2\right]} \qquad [54]$$

The random variable $F_{\max}$ has a probability distribution which depends on k and $f = n - 1$ degrees of freedom. If the computed value $F_{\max}$ satisfies the condition $F_{\max} \geq F_{\max,k,f,\alpha}$ we then reject the null hypothesis H_0 in favour of the alternative hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_k^2$ and the probability of making a type I error is α. If however, $F_{\max} < F_{\max,k,f,\alpha}$, there are no grounds for rejecting the null hypothesis H_0. Because of the sensitivity of this test to departures from normality, if $F_{\max}$ is significant it indicates either unequal variances or a lack of normality.

The values of $F_{\max}, k, f, \alpha$ such that $P\ (F_{\max} \geq F_{\max,k,f,\alpha,})$ are given in Tables (e.g., 62, 118). Tables of critical values of $F_{\max}$ are available only for the case of equal number of degrees of freedom. The tabulation applies for up to 12 variances being compared with 5% and 1% significance levels. The Hartley's $F_{\max}$ test is less powerful than the Bartlett's or Levene's tests, and tables of critical values (119) are not as readily available as, for example, those for F. Nevertheless the simplicity of the test makes it worth consideration.

If the appropriate tables of critical values are not available, or if the degrees of freedom are unequal, we may compare our test criterion with $F_{f(\text{num}),f(\text{den})}(1-\alpha)$, where f_{num} and f_{den} are the degrees of freedom of the numerator and denominator of $F_{\max}$. If $t < F_{\text{f(num),f(den)}}(1-\alpha)$, we may conclude that our value is nonsignificant at the α level of significance since the correct critical value always exceeds that for the ordinary F test of two variances. Such a nonsignificant result eliminates the need for the special tables and also the computation required for the Bartlett's and Levene's (102) tests. However, a significant result here does not necessarily stand. Of the three tests (Cochran, Hartley, Bartlett), Bartlett's test is the most powerful when the alternative hypothesis is unknown and normality can be relied on. However, all three tests are sensitive to nonnormality. The tests of Hartley and Cochran lead to the same decisions in most cases.

LEVENE'S *S* TEST

Since result of the F test can be strongly influenced even by small deviations from the normal distribution, Levene (120) has proposed an approximate and ingenious nonparametric procedure for the equality of group variances. Levene's test is the one-way analysis of variace F test on $|y_{iv} - \bar{y}_i|$, the absolute deviations of the y_{iv} from their group mean $\bar{y}_i$. This essentially converts a test of variances into a test of means, which is relatively unaffected by nonnormality. This test has been shown to be reasonably robust to nonnormality when the sample sizes are equal (120–121). Consider each observation's absolute deviation from its group mean$|y_{iv} - \bar{y}_i|$. In a group with small variances observations will be clustered about group means, and therefore the $|y_{iv} - \bar{y}_i|$ will be small. In contrast, in a group with large variances, the $|y_{iv} - \bar{y}_i|$ will be large. Therefore, to test for equality of group variances, Levene suggested performing an ANOVA on the $|y_{iv} - \bar{y}_i|$ (122).

For each set, we define the quantities Y_{iv} by

$$Y_{iv} = |y_{iv} - \bar{y}_i| \qquad [55]$$

where $\bar{y}_i = \Sigma y_{iv}/n_i$, the mean of the ith set and $(v = 1, 2, \ldots, n_i)$. We test whether or not the means of k sets of Y's are significantly different, since these are proportional to the sample variances. Compare the "between groups" mean square with the "within groups" mean square via an F test. The appropriate F statistic, T, is then (23)

$$T = \frac{\sum_{i=1}^{k} n_i(\bar{Y}_i - \bar{Y})^2/(k-1)}{\sum_{i=1}^{k}\sum_{v=1}^{n_i}(Y_{iv} - \bar{Y}_i)^2 \Big/ \sum_{i=1}^{k} n_i - k} \qquad [56]$$

where

$$\bar{Y}_i = \sum_{i=1}^{n_i} Y_{iv}/n_i \qquad [57]$$

$$\bar{Y} = \sum_{i=1}^{k}\sum_{v=1}^{n_i} Y_{iv} \Big/ \sum_{i=1}^{k} n_i - k \qquad [58]$$

Under H_0, T is approximately distributed as $F_{k-1,\,\Sigma n_i - k}$. Hence we reject H_0, the hypothesis that the variances are equal, at the significance level α if $T \geq F_{k-1,\,\Sigma n_i - k,\,1-\alpha}$. Here, we are performing a one-way analysis of variance on the Y's. One difficulty, however, is that Y's are not normally distributed, but this is not too important since the analysis of variance test is robust to nonnormality. Another difficulty is that the Y values within a set are not independent because each is derived using the same set mean. However, it has been shown that this departure from the proper conditions for one-way analysis of variance causes only a very small disturbance unless the number of observations in the various set is very small.

Various modifications of Levene's test have been proposed and investigated. Brown and Forsythe (123) suggested using the absolute deviations from the group medians, which is a more robust estimation of location, as an alternative to the mean in computing the absolute deviations. O'Brien proposed a test that is basically a modification of Levene's Y_{iv}^2 using a dispersion variable (124). Levene's test using group medians is more robust than the Levene test when groups are unequal in size and the absolute deviation scores (deviations from the group means)

are highly skewed, causing violation of the normality assumption. It has been shown that, asymptotically, Levene's test has the correct level whenever the estimate of group "center" is an estimate of group median (125–126). Seven tests of equality of variances have been compared in terms of robustness and power in a simulation experiment with small-to-moderate sample sizes. It was found that the Levene test and the Bartlett test with kurtosis adjustment are robust. Among them, the bootstrap version of the Levene test tends to have the higher power (127).

TEST BASED ON THE COMPARISON OF THE RANGES

This is one of the simplest tests (128–130). It involves the calculation of the range between the highest and lowest responses y_i reported for each x_i and calculating the ratio

$$C = \frac{R_{\max}}{R_1 + R_2 + \ldots + R_k} \qquad [59]$$

where $R_1, R_2, \ldots R_k$ are the ranges for each of k samples and $R_{\max}$ is the largest range. This test can be applied with as few as two replicates at each x_i value, but is more reliable and sensitive when a greater amount of replication is available. When the calculated ratio exceeds the tabular value for a 0.05 significance level, we reject the hypothesis that the variances are homogeneous.

A rapid test based on the quotient of the largest and the smallest ranges was introduced by Leslie and Brown (131). The upper critical limits for four significance levels can be found in the original work.

CONCLUDING REMARKS

The analytical literature, surprisingly, rarely describes important statistical details related to the processing of data in regression analysis with replicate measurements. Linear least squares, however, require repeated observations in order to properly verify homogeneity of variances and linearity. Weighted regression is a way of preserving the conceptual simplicity of linear models. As a matter of fact, the idea behind the weighted least squares is to attach the most importance to the data that are measured with the greatest precision. If enough large samples with replication are available, then within group variances may be used to provide approximate weights. An obvious weighting factor, therefore, is simply inversely proportional to the variance of the response at each x_i. Nevertheless, this approach can be unreliable if the number of observations n_i at each x_i are small because these estimated weights can be poor characterizations of the truth. Anyway, this danger can be avoided if a function can be found which models the variance. Our ability to detect departures from random scatter will be much improved if, instead of plotting the raw data, we use residual plots for model validation.

Formal tests for testing homogeneity of variances of replicate measurements of y at each x_i exist, but all have drawbacks. As Box (110) noted, "To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" It can be difficult in practice to tell whether significant F test values are due to heterogeneity of variance or to the fact that the populations are not normal. Cochran's test is recommended by ISO even in case of unbalanced data sets. The Bartlett test is dependent on meeting the assumption of normality and therefore Levene's test has now largely replaced it. Levene's test is the usual F test for equality of means computed on pseudo-observations, which one defines as the absolute deviations of the data from an estimate of the group "center." A more robust test very similar to the Levene test has been proposed by Brown and Forsythe (123). Instead of performing the t-test for mean differences on the deviations from the mean, one can perform the analysis on the deviations from the group medians. Most software packages can give you measures for homogeneity of the variance, though they are not included in linear regression analysis with replicated data.

The slope and intercept of a weighted regression line are often very similar to those obtained when unweighted calculations are applied to the same data; an analyst who ignores nonconstant variance will not sacrifice much statistical reliability in the mid-range of the curve. However, only the weighted line provides proper estimates of standard deviations and confidence limits when the weights vary significantly with x_i. An increased number of measurements at the low end of the concentration range can offset both the effects of nonlinearity and heteroscedasticity as they will weight the least squares function in favor of reducing the residuals at this extreme of the curve. Weighted regression calculations on the other hand must also be used when curvilinear data are converted into rectilinear data by a suitable algebraic transformation.

REFERENCES

1. A. Sayago, M. Boccio, and A. G. Asuero, Fitting straight lines with replicated observations by linear regression: the least squares postulates. *CRC Crit. Rev. Anal. Chem*. 34 (2004):39–50.
2. R. de Levie, Curve fitting least squares. *CRC Crit. Rev. Anal. Chem*. 30 (2000):59–74.
3. P. Valko and S. Vajda, *Advanced Scientific Computing in BASIC, with Applications in Chemistry, Biology and Pharmacology* (Amsterdam: Elsevier, 1989):139–144.
4. M. Meloun, J. Militki, K. Kupla, and R. G. Brereton, The effect of influential data, model and method on the precision of unvariate calibration. *Talanta* 57 (2002):721–740.
5. K. A. Connors, Statistical Treatment of Data in *Binding Constants: The Measurement of Molecular Complex Stability*, Chapter 3. (New York: Wiley, 1987):103–138.
6. M. Davidian and P. D. Haaland, Regression and calibration with non constant error variance. *Chem. Intel. Lab. Syst.* 9 (1990):231–248.
7. W. E. Deming, *Statistical Adjustment of Data* (New York: Dover, 1964):14.
8. J. R. J. Belloto and T. D. Sokolovski, Residual analysis in regression. *Am. J. Pharm. Educ*. 49 (1985):295–303.

9. S. Bolton, *Pharmaceutical Statistics: Practical and Clinical Applications*, 3rd ed. (New York: Marcel Dekker, 1994).
10. P. P. Mager, *Design Statistics in Pharmacochemistry* (New York: Wiley, 1991).
11. E. Mullins, *Statistics for the Quality Control Laboratory* (Cambridge: RSC, 2003).
12. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), *Guidelines for Industry: Validation of Analytical Procedures*, Q2A (Geneva: ICH, 1994).
13. K. A. Brownlee, *Statistical Methodology in Science and Engineering*, 3rd ed. (Malabar, Fla.: R. G. Krieger, 1984).
14. J. S. Hunter, Calibration and the straight line: current statistical practices. *J. Assoc. Off. Anal. Chem*. 64 (1981):574–583.
15. D. L. Mac Taggart and S. O. Farwell, Analytical use of linear regression. Part I. Regression procedures for calibration and quantitation. *J. Assoc. Off. Anal. Chem*. 75 (1992):594–607.
16. B. Jülicher, P. Gowik, and S. Uhlig, Assessment of detection methods in trace analysis by mean of a statistically based in-house validation concept. *Analyst* 123 (1998):173–179.
17. J. Berkson, Are there two regressions? *J. Am. Stat. Ass*. 45 (1950):164–180.
18. M. Thompson, Regression methods in the comparison of accuracy. *Analyst* 107 (1982):1169–1180.
19. B. D. Ripley and M. Thompson, Regression techniques for the detection of analytical bias. *Analyst* 112 (1987):277–383.
20. F. J. Del Rio Bocio, J. Riu, R. Boqué, and F. X. Rius, Limit of detection in linear regression with errors in the concentration. *J. Chemometr*. 17 (2003):413–421.
21. J. M. Lisy, A. Cholvadova, and J. Kutej, Multiple straight-line least-squares analysis with uncertainties in all variables. *Computers Chem*. 14 (1990):189–192.
22. R. F. Martin, General Deming regression for estimating systematic bias and its confidence interval in the method-comparison studies. *Clin. Chem.* 46 (2000):100–104.
23. N. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. (New York: Wiley, 1998).
24. J. W. Einax, H. W. Zwanziger, and S. Geiß, *Chemometrics Environmental Analysis*. (Weinheim: VCH, 1997).
25. International Standard, *Air Quality—Assessment of the Accuracy of the Calibration Function of a Measurement Method under Field Conditions*, ISO 9169 (Geneva: International Organization for Standardization, 1994).
26. International Standard, *Air Quality—Assessment of the Accuracy of the Calibration Function of a Measurement Method under Field Conditions*, ISO/CD 13752 (Geneva: International Organization for Standardization, 1994).
27. J. Keith and G. Duggleby, What happens when the data are fitted to the wrong equation? *Biochem. J*. 171 (1978):513–517.
28. J. B. Kennedy and A. M. Neville, *Basic Statistical Methods for Engineers and Scientists*, 2nd ed. (New York: Harper, 1976).
29. 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.
30. A. Olin, Some statistical properties of the results obtained by the standard addition method demonstrated by computer simulation. *Chem. Anal*. (*Warsaw*) 40 (1995):939–951.
31. J. S. Garden, D. G. Mitchell, and W. N. Mills, Non constant variance techniques for calibration curve-based analysis. *Anal. Chem*. 52 (1980):3210–2315.
32. International Standard, *Accuracy (Trueness and Precision) of Measurement Methods and Results: Part 2. Basic Method for the Determination of Repeatability of a Standard Measurement Methods*, ISO 5725-2 (Geneva: International Organization for Standardization, 1994).
33. D. L. MacTaggart and S. O. Farwell, Analytical use of linear regression. Part II: Statistical error in both variables. *J. Ass. Off. Anal. Chem.* 75 (1992):608–614.
34. A. G. González and A. G. Asuero, Computational program for validating analytical results. *Fresenius Z' Anal. Chem*. 346 (1993):885–887.
35. J. D. Lee and T. D. Lee, *Statistics and Computer Methods in BASIC* (New York, Van Nostrand-Reinhold, 1982):108–109.
36. P. M. Wanek, R. E. Whippe, T. E. Fickies, and P. M. Grant, Inaccuracies in the calculation of standard deviation with electronic calculators. *Anal. Chem*. 54 (1982):1877–1878.
37. H. E. Solberg, Inaccuracies in computer calculation of standard deviation. *Anal. Chem*. 55 (1983):1611.
38. B. D. McCullough and B. Wilson, On the accuracy of statistical procedures in Microsoft Excel 97. *Comp. Stat. Data Anal.* 31 (1999):27–37.
39. M. G. Cox and P. M. Harris, Design and use of reference data sets for testing scientific software. *Anal. Chim. Acta* 380 (1999):339–351.
40. A. G. Asuero and A. G. González, Some observations of fitting a straight line to data. *Microchem. J*. 40 (1989):216–225.
41. P. C. Meier and R. E. Zünd, *Statistical Methods in Analytical Chemistry*, 2nd ed. (New York: Wiley, 2000).
42. A. G. Asuero, G. Gonzalez, F. de Pablos, and J. L. Gomez Ariza, Determination of the optimum working range in spectrophotometric procedures, *Talanta* 35 (1988):531–537.
43. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed. (Cambridge: Cambridge University Press, 1999).
44. J. N. Miller, Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review. *Analyst* 116 (1991):3–14.
45. M. Mulholland and D. B. Hibert, Linearity and the limitations of least squares calibration. *J. Chromatogr.* A762 (1997):73–82.
46. J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 4th ed. (Essex: Ellis Horwood, 2000).
47. L. A. Currie, Recommendations in evaluation of analytical methods including detection and quantification capabilities. *Pure Appl. Chem*. 67 (1995):1699–1723.
48. S. M. Gort and R. Hoogerbrugge, A user-friendly spreadsheet program for calibration using weighted regression. *Chem. Intell. Lab. System*, 28 (1995):193–199.
49. M. E. Zorn, R. D. Gibbons, and W. C. Sonzogni, Weighted least squares approach to calculating limits of detection and quantification by modelling variability as a function of concentration. *Anal. Chem.* 69 (1997):3069–3075.
50. R. de Levie, *How to Use Excel in Analytical Chemistry and in General Scientific Data Analysis* (Cambridge: Cambridge University Press, 2001).
51. E. Desimoni, A. Daghetta, and S. Valsecchi, Evaluation of uncertainty by ordinary linear least squares regression of replicated

data. A revised formulation to deal with unbalanced data set. *Ann. Chim*. (Roma) 88 (1998):601–609.
52. W. Penninckk, C. Hartmann, and D. L. Massart, J. Smeyer-Verbeke, Validation of the calibration procedure in atomic spectrometric methods. *J. Anal. At. Spectrosc*. 11 (1996):237–246.
53. W. J. Dixon, Processing data outlier. *Biometrics* 9 (1953):74–89.
54. F. E. Grubbs, Procedures for detecting outlying observations in samples. *Technometrics* 11 (1969):1–21.
55. F. E. Grubbs and G. Beck, Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics* 14 (1972):847–854.
56. S. Aknazarova and V. Kafarov, *Experiment Optimization in Chemistry and Chemical Engineering* (Moscow: Mir Publishers, 1982):50–52.
57. L. A. Currie and G. Svehla, Nomenclature for the presentation of results of chemical analysis. *Pure Appl. Chem*. 66 (1994):595–608.
58. M. J. Cardone, Detection and determination of error in analytical methodology. Part I. In the method verification program. *J. Ass. Off. Anal. Chem*. 66 (1983):1257–1282.
59. S. N. Deming and S. L. Morgan, *Experimental Design: A Chemometric Approach*, 2nd ed. (Amsterdam: Elsevier, 1993).
60. G. T. Wernimont, *Use of Statistics to Develop and Evaluate Analytical Methods* (ed. W. Spendley) (Arlington, VA: AOAC, 1985).
61. G. E. P. Box and N. R. Draper, *Empirical-Model-Building and Response Surfaces* (New York: Wiley, 1987).
62. R. L. Anderson, *Practical Statistics for Analytical Chemistry* (New York: Van Nostrand, 1987), 56, 109.
63. J. Cserminski, A. Iwasiewicz, Z. Paszek, and A. Sikorski, *Statistical Methods in Applied Chemistry* (Amsterdam: Elsevier, 1990), 425.
64. J. Mandel, *The Statistical Analysis of Experimental Data* (New York: Wiley-Interscience, 1964): 68, 352–353,
65. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: A Textbook* (Amsterdam: Elsevier, 1988).
66. J. H. Noggle, *Practical Curve Fiting and Data Analysis. Software and Self-Instruction for Scientists and Engineers* (Chichester: Ellis Horwood, 1993).
67. D. McCormick and A. Roach, *Measurement Statistics and Computation* (New York: Wiley, 1987), 303–321.
68. H. Bubert and R. Klockenkämper, Precision-dependent calibration in instrumental analysis. *Fresenius Z' Anal. Chem.* 316 (1983):186–193.
69. L. M. Schwarz, Calibration curve with non uniform variance. *Anal. Chem.* 51 (1979):723–727.
70. M. Thompson, Variation of precision with concentration in an analytical system. *Analyst* 112 (1988):1579–1587.
71. K. Baumann and H. Wätzig, Appropriate calibration functions for capillary electrophoresis. II. Heterocedasticity and its consequences. *J. Chromatog.* A700 (1995):9–20.
72. K. Baumann, Regression and calibration for analytical separation techniques. Part II. Validation, weighted and robust regression. *Process and Control Quality* 10 (1997):75–112.
73. S. N. Ketkar and T. J. Bzik, Calibration of analytical instruments. Impact of nonconstant variance in calibration data. *Anal. Chem.* 72 (2000):4762–4765.
74. J. K. G. Watson, Robust weighting in least squares fits. *J. Mol. Spectr*. 219 (2003):326–328.
75. J. Tellinghuisen, Statistical error calibration in UV-visible spectrophotometry. *Appl. Spectrosc*. 54 (2000):431–437.
76. J. Tellinghuisen, On the role of statistical weighting in the least squares analysis of UV-visible spectrophotometric data. *Appl. Spectrosc*. 54 (2000):1208–1213.
77. N. V. Nagaraja, J. K. Paliwal, and R. C. Gupta, Choosing the calibration model in assay validation. *J. Pharm. Biom. Anal*. 20 (1999):433–438.
78. Y. V. Heyden, The ruggedness of analytical methods. *Analysis* 22(5) (1994):M27–M29.
79. S. Dowy and S. Wearden, *Statistics for Research*, 2nd ed. (New York: Wiley, 1991), 348–356.
80. E. Desimoni, S. Mannino, and B. Brunetti, On the assessment of compliance with legal limits. Part 1: Signal concentration domains. *Accred. Qual. Assur*. 6 (2001):452–458.
81. J.-C. Lee, and M. H. Ramsey, Modelling measurement uncertainty as a function of concentration: an example from a contaminated land investigation. *Analyst* 126 (2001):1784–1791.
82. R. L. Watters, R. J. Carroll, and C. H. Spiegelman, Error modelling and confidence interval estimation for inductively coupled plasma calibration curves. *Anal. Chem.* 59 (1987):1639–1643.
83. L.-J. Hwang, Impact of variance function estimation in regression and calibration. *Meth. Enzymol.* 240 (1994):150–170.
84. E. D. Prudnikov and Y. S. Shapling, Random errors in analytical methods. *Analyst* 109 (1984):305–307.
85. J. H. Ottoway, Normalization in the fitting data by iterative methods. Application to tracer kinetics and enzyme kinetics. *Biochem. J*. 134 (1973):729–736.
86. R. D. Cook and S. Weisberg, Diagnostics for heteroscedasticity in regression. *Biometrika* 70 (1983):1–10.
87. J. P. Franke, R. A. de Zeeuw, and R. Hakkert, Evaluation and optimization of the standard addition method for absorption spectrometry and anodic stripping voltammetry. *Anal. Chem.* 50 (1978):1374–1380
88. 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 (Geneva: International Organization for Standardization, 1990).
89. International Standard, *Water Quality—Calibration and Evaluation of Analytical Methods and Estimation of the Performance Characteristics: Part 2. Calibration Strategy for Non Linear Second Order Function*, ISO, 8466-2 (Geneva: International Organization for Standardization, 1990).
90. H. Scheffe, *The Analysis of Variance* (New York: Wiley, 1959).
91. E. Desimoni, A program for the weighted least squares regression of unbalanced response arrays. *Analyst* 124 (1999):1191–1196.
92. M. S. Bartlett, Some examples of statistical methods of research in agriculture and applied biology. *J. Roy. Statist. Soc.* 4(2) (1937, suppl.):137–170.
93. M. S. Bartlett, Properties of sufficiency and statistical tests, *Proc. Roy. Soc. Ser.* A160 (1937):268–282.
94. M. S. Bartlett, A note on the multiplying factors for various approximations. *J. Roy. Statist. Soc. Ser*. B16 (1954):296–298.
95. W. G. Cochran, Some difficulties in the statistical analysis of replicated experiments, *Empire Journal of Experimental Agriculture* 6 (22) (1938):157–175.

96. W. G. Cochran, The distribution of the largest of a set of estimated variances as a fraction of the total. *Ann. Eugenics* 11 (1941):47–53.
97. W. G. Cochran, Some consequences when the assumptions for the analysis of the variance are not satisfied. *Biometrics* 3 (1947):22–38.
98. W. G. Cochran, The combination of estimates from different experiments, *Biometrics* 10 (1954):101–129.
99. I. W. Burr and L. A. Foster, *A Test of Equality of Variances*, Department of Statistics Mimeo Series No. 282 (Lafayette, Ind.: Purdue University, 1972).
100. E. Carbonell, J. B. Dewis, R. Calvo, F. Gonzalez, and W. Pruñozosa, *Regresión lineal, un enfoque conceptual y práctico* (Madrid: INIA, Ministerio de Agricultura, Pesca y Alimentación 1983), 97–99, 175–176.
101. H. O. Hartley, The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37 (1950):308–312.
102. K. E. Peace, *Biopharmaceutical Statistics for Drug Development* (New York: Marcel Dekker, 1988).
103. L. Sach, *Applied Statistics: A Handbook of Techniques*, 2nd ed. (New York: Springer-Verlag, 1982).
104. S. D. Roof, An IBM 704 Computer Program for computing Bartlett's test of homogeneity of variance. *Biometrics* 18 (1962):251–252.
105. J. Philippe, *Les Methodes Statistiques en Pharmacie et en Chimie (applications a la recherche, a la production et au controle)* (Paris: Masson, 1967), 125–126.
106. D. J. Finney, *Statistical Methods in Biological Assay*, 3rd ed. (London: Griffin, 1978), 55–57.
107. S. Ebel and M. Lorz, Weyandt-Spangenberg, Optimierte kalibrierung. *Fresenius Z. Anal. Chem.* 335 (1989):960–965.
108. S. C. Gad, *Statistics and Experimental Design for Toxicologists*, 3th ed. (Boca Raton, Fla.: CRC Press, 1999), 42–44.
109. H. O. Hartley, Testing the homogeneity of a set of variances. *Biometrika* 31 (1940):249–255.
110. G. E. P. Box, Non-normality and tests on variances. *Biometrika* 40 (1953):318–335.
111. K. Danzer and L. A. Currie, IUPAC Guidelines for calibration in analytical chemistry. Part1. Fundamentals and single component calibration. *Pure Appl. Chem.* 70 (1998):993–1014.
112. M. Meloun and J. Militki, Computer-assisted data treatment in analytical chemistry. IV. Classical estimates of parameters of location, scale and shape. *Chem. Papers* 49 (1995):68–73.
113. W. R. Ott, *Environmental Statistics and Data Analysis* (Boca Raton, Fla.: CRC Press, 1995).
114. V. P. Spiridonov and A. A. Lopatkin, *Tratamiento Matemático de Datos Físico-Químicos* (Moscow: Mi, 1973), 99–101, 111–112.
115. R. Caulcutt and R. Boddy, *Statistics for Analytical Chemistry*, (London: Chapman and Hall, 1983), 70–71.
116. C. I. Bliss, W. G. Cochran, and J. W. Tukey, A rejection criterion based upon the range. *Biometrika* 43 (1956):418–422.
117. V. L. Anderson and R. A. McLean, *Design of Experiments: A Realistic Approach* (New York: Marcel Dekker, 1974).
118. J. Neter and W. Wasserman, *Applied Linear Models* (Homewood, Ill.: R. D. Irving, 1974).
119. F. J. Rohlf and R. R. Soakal, *Statistical Tables*, 3rd ed. (New York: W.H. Freeman, 1995).
120. H. Levene, Robust tests for equality of variances, in *Contribution to Probability and Statistics*, *Essays in Honor of Harold Hotelling*; ed. I. Olkin et al. (Stanford: Stanford University Press, 1960), 278–292.
121. B. Schultz, On Levene's test and other statisticas of variation. *Evolutionary Theory* 6 (1983):197–203.
122. J. R. Green and D. Margerison, *Statistical Treatment of Experimental Data* (Amsterdam: Elsevier, 1977).
123. M. B. Brown and A. B. Forsythe, The small sample behaviour of some statistics which test the equality of several means. *Technometrics* 16 (1974):129–132.
124. R. G. O'Brien, Robust techniques for testing heterogeneity of variance effects in factorial designs. *Psychometrika* 43 (1978):327–342.
125. W. J. Conover, M. E. Johnson, and M. M. Johnson, A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data. *Tecnometrics* 23 (1981):351–361.
126. R. J. Carroll and H. Schneider, A note on Levene's tests for equality of variances. *Statistics and Probability Letters* 3 (1985):191–194.
127. T.-S. Lim and W. Y. Loh, A comparison of tests of equality of variances. *Comp. Stat. Data Anal.* 22 (1996):287–301.
128. H. O. Hartley, The use of range in analysis of variance. *Biometrika.* 37 (1950):271–280.
129. E. S. Pearson and H. E. Hartley, *Biometrika Tables for Statiscians*, 8th ed. (Cambridge: Cambridge University Press, 1966), Table 316.
130. 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*, L. A. Currie, ed. ACS Symposium Series 361 (Washington, D.C.: ACS, 1988). chap. 10, 194–207,
131. R. T. Leslie and B. M. Brown, Use of range in testing heteroscedasticity of variance. *Biometrika* 53 (1966):221–227.