UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



3. S.J. Haswell (ed.), *Practical Guide to Chemometrics*, Marcel Dekker Inc., New York (1992)
4. M. Meloun, J. Militký, M. Forina, *Chemometrics for Analytical Chemistry. Volume 2. PC-aided Regression and Related Methods*, Ellis Horwood, London (1994)
5. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam (1997)
6. S. Weisberg, *Applied Linear Regression*, 2nd Ed., John Wiley & Sons, New York (1985)
7. H. Working, H. Hotelling, *Journal of American Statistical Association*, **24** (1929) 73
8. R.G. Miller, *Simultaneous Statistical Inference*, McGraw-Hill, New York (1966)
9. G.W. Snedecor, W.G. Cochran, *Statistical Methods*, 8th ed., Iowa State University Press, Ames (1989)
10. P.D. Lark, B.R. Crowen, R.L.L. Bosworth, *The Handling of Chemical Data*, Pergamon Press, Oxford (1968)
11. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, *Analytical Chemistry*, **66** (1994) 909
12. J.O. Rawlings, *Applied Regression Analysis: A Research Tool*, Wadsworth & Brooks/Cole Advanced Books & Software, Belmont (1988)
13. J.C. Miller, J.N. Miller, *Statistics for Analytical Chemists*, Ellis Horwood, Chichester (1984)
14. J.N. Miller, *Analyst*, **116** (1991) 3
15. A.G. Asuero, A.G. González, *Microchemical Journal*, **40** (1989) 216
16. P.L. Bonate, *Analytical Chemistry*, **65** (1993) 1367
17. I.E. Frank, R. Todeschini, *The Data Analysis Handbook*, Elsevier, Amsterdam (1994)
18. P. Hall, *The Annals of Statistics*, **14** (1986) 1431

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


19. C.H. Spiegelman, R.L. Watters, L. Hungwu, *Chemometrics and Intelligent Laboratory Systems*, **11** (1991) 121
20. T. Lwin, C.H. Spiegelman, *Journal of the Royal Statistical Society Series C*, **35** (1986) 256
21. J. Mandel, *Journal of Quality Technology*, **16** (1984) 1
22. P.C. Meier, R.E. Zünd, *Statistical Methods in Analytical Chemistry*, John Wiley & Sons, New York (1993)

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



# Confidence intervals in linear regression taking into account uncertainties in both axes

F. Javier del Río, Jordi Riu* and F. Xavier Rius

*Departament de Química Analítica i Química Orgànica.*

*Universitat Rovira i Virgili.*

*Pl. Imperial Tàrraco, 1. 43005-Tarragona. CATALONIA, SPAIN.*

## ABSTRACT

This study reports the expressions for the variances associated to the response and predictor variables which are calculated with the bivariate least squares (BLS) regression technique, which takes into account the errors in both axes. The calculated results are compared to those obtained from a simulation process applied to six different real data sets. The mean error of the results found with the new expressions is between 4 and 5% whereas mean errors can be as high as 85%, 277%, 637% and 1697% when weighted least squares, ordinary least squares, constant variance ratio approach and orthogonal regression are used respectively. An important property of the confidence intervals calculated using the BLS regression technique is the invariance of the results when axes are switched.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


## INTRODUCTION

The remarkable mathematical properties of ordinary least squares, OLS, together with its practical performance characteristics are the main reasons why it is the regression technique which is most commonly used by the analytical chemistry community. However, this technique is based on a set of mathematical hypotheses such as the homoscedasticity on the *y* axis or the absence of errors on the *x* axis, that are sometimes not fulfilled. This leads to biased regression coefficients of the straight line and, consequently, to erroneous predicted results.[1,2] Method comparison studies, where the errors associated to both methods are usually of the same order of magnitude, or calibration lines, where the errors in the instrumental responses are comparable to the errors associated to the concentration values,[3] are situations in which the application of OLS often provides biased results.

An improvement on the OLS technique is the weighted least squares (WLS) technique[1,4] that takes into account heteroscedasticity in the *y*-axis. However, WLS still considers the *x* axis as being error free.

The errors-in-variables regression[5], also called constant variance ratio (CVR) approach,[6-8] considers the errors in both axes. It does not take into account the individual uncertainties of each experimental point but considers the ratio of the variances of the response to predictor variables to be constant for every experimental point ($\lambda=s_y^2/s_x^2$). A particular case of the CVR approach is the orthogonal regression (OR)[9], in which the errors are of the same order of magnitude in the response and predictor variable (i.e. $\lambda=1$). In the literature, this latter case is also called orthogonal distance regression (ODR)[2] or total least squares regression (TLS).[10]

The bivariate least squares (BLS) method[11,12] is a linear regression technique capable of overcoming the limitations of the previous methods i.e. the fact that the

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



individual uncertainties in both variables are not considered. This technique calculates the straight line regression coefficients by taking into account the heteroscedastic uncertainties in both axes. BLS has been applied in method validation studies to detect bias in newly developed analytical methodologies.[13]

The calculation of predicted values in regression analysis considering individual heteroscedastic errors in both axes is an important issue in practical instances that has merited little attention up to date. The calculation of the measurement results and the uncertainty of a newly developed method from the historical values recorded by using a previously established methodology, or the establishment of relationships between two dating methodologies, both incorporating uncertainty, so as to assign the chronological origin of archaeological samples, are two examples where confidence intervals from linear regression taking into account uncertainties in both axes should be considered.

This paper develops and validates new expressions for calculating the confidence intervals for predicted values of the response variable given a value of the predictor variable, and vice versa, using the BLS regression technique, i.e. by considering the individual uncertainties of every experimental point. The expressions for the predictor intervals considering errors in both axes have been derived from a generalisation of the existing OLS and WLS expressions. The same results have also been found using the error propagation theory.[14] To validate the appropriateness of the new confidence intervals, six real data sets were used. Random errors based on the individual uncertainties of each real point were added to the data sets using the Monte Carlo method. The values obtained with the new expressions based on BLS do agree with the theoretical values more than the results obtained using the expressions based on OLS, WLS, OR or CVR. One of the most important properties of the BLS confidence intervals is their invariance when axes are switched.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


## BACKGROUND AND THEORY

**Bivariate least squares technique.** Of the various regression techniques that consider errors in both axes, bivariate least squares (BLS) was chosen because it can readily provide the regression coefficients and their associated variances and covariances, and because of the simplicity of programming its algorithm. The prediction step using the straight line model is expressed in eq. 1:

$$\hat{y}_i = \hat{a} + \hat{b}x_i \tag{1}$$

where $\hat{a}$ represents the intercept, $\hat{b}$ the slope, and $\hat{y}_i$ is the predicted value for the observed value $x_i$. The method consists of minimising the sum of the weighted residuals of the regression straight line:

$$S = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{w_i} \tag{2}$$

where $n$ is the number of experimental data points and $w_i$ is the weighting factor that corresponds to the variance of the $i$th-residual:

$$w_i = s_{y_i}^2 + \hat{b}^2 s_{x_i}^2 - 2\hat{b}\,\mathrm{cov}(x_i y_i) \tag{3}$$

where $s_{x_i}^2$ and $s_{y_i}^2$ are, respectively, the variances of the $i$th-point for the predictor and response variables of the straight line expressed in the eq. 1, and $\mathrm{cov}(x_i y_i)$ is the covariance between the predictor and the response variable, which is normally set to zero.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



It is interesting to note that whenever the variances of the predictor variable values are zero and all the variances on the response variable are the same (i.e., all errors are constant and only due to the experimental measurement in the *y*-axis), the results obtained are identical to those obtained with the OLS method. Since in the BLS regression model the unobserved $\hat{x}$ and $\hat{y}$ values are affected by a random error, and it is only possible to observe the *x* and *y* values, BLS can be considered a structural regression model, in contrast to the functional models in which the *x*-variable is fixed and known to be without error.[5]

**Variance of the response variable.** In the OLS method, the well known expression for the variance of the predicted observation of the response variable $y_0$ obtained as the mean of $q$ observations performed at $x_0$ is given by eq. 4:

$$s_{y_0}^2 = \left[ \frac{1}{q} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i^2 - \bar{x})^2} \right] \cdot \hat{s}^2 \qquad (4)$$

where $\bar{x}$ is the mean value of the predictor variable and $\hat{s}^2$ is the estimation of the experimental error ($s^2$):

$$\hat{s}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2} \qquad (5)$$

Eq. 4 can also be expressed in matrix form:

$$s_{y_0}^2 = \left( \frac{1}{q} + \mathbf{X}_0' \cdot (\mathbf{X}' \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_0 \right) \cdot \hat{s}^2 \qquad (6)$$

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


where $\mathbf{X}_0$ is a two-element column vector formed by a 1 in the first row and the predictor variable ($x_0$) in the second row, and $\mathbf{X}$ is an $n$x2 matrix in which the first column is a column of ones and the second is formed by the $n$ values of the predictor variable corresponding to the experimental points.

For the WLS technique, which takes into account heteroscedastic errors in the response variable, the variance for the predicted observation $y_0$, calculated as the mean of $q$ observations performed at a selected value of $x_0$ is given by eq. 7:

$$s_{y_0}^2 = \left[\frac{1}{q} + \mathbf{X}_0' \cdot (\mathbf{X}' \cdot \mathbf{V}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_0\right] \cdot \hat{s}^2 \qquad (7)$$

where $\mathbf{V}$ is an $n$x$n$ diagonal matrix the $i$th element of which corresponds to the variance of $y_i$ ($s_{y_i}^2$), and where $\hat{s}^2$, the estimation of the experimental error, now takes into account the variances of the response variable as the weighting factor:

$$\hat{s}^2 = \frac{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{s_{y_i}^2}}{n-2} \qquad (8)$$

In the CVR approach, the expression for the variance of the prediction given a value of the predictor variable, is given by eq. 9:

$$s_{y_0}^2 = \hat{b}^2 s_\delta^2 + \left[\frac{1}{n} + \frac{(1+k\hat{b})^2 \cdot (x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 + 2k\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + k^2 (y_i - \bar{y})^2}\right] \cdot s_e^2 \qquad (9)$$

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


where $k$, $s_\delta^2$ and $s_e^2$ are defined in the process of finding the regression coefficients ($s_\delta^2$ refers to the error associated to the predictor variable and $s_e^2$ is associated to the estimate of the experimental error).[8] If λ, which appears in the coefficients $k$ and $\hat{b}$ in eq. 9 is chosen to be unity, then results for the OR method are obtained.

The expression for the variance of the prediction given a value of the predictor variable for the BLS regression technique, which takes into account heteroscedastic errors in both axes, is given by eq. 10:

$$s_{y_0}^2 = \mathbf{X}_0' \cdot (\mathbf{X}' \cdot \mathbf{W}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_0 \cdot \hat{s}^2 \tag{10}$$

where the matrix **W** is a *nxn* diagonal matrix the *i*th-diagonal element of which is the weighting factor $w_i$ defined in eq. 3. This weighting factor takes into account the errors in both axes. The estimation of the experimental error is now:

$$\hat{s}^2 = \frac{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{w_i}}{n-2} \tag{11}$$

However, in this expression there is still a term to consider when the errors of the predictor variable ($x_0$) are also taken into account. In order to correct the difference in ranges between the two axes, the factor corresponding to the square of the slope has to be introduced. The expression for the variance of the response true mean value at a given observation $x_0$ is shown in eq. 12:

$$s_{y_0}^2 = \left[\mathbf{X}_0' \cdot (\mathbf{X}' \cdot \mathbf{W}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_0 + s_{x_0}^2 \cdot \hat{b}^2\right] \cdot \hat{s}^2 \tag{12}$$

On the other hand, an independent expression for the variance associated to the predicted response variable can be found by applying the error propagation theory[14]

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


to eq. 1. This expression is presented in eq. 13. The covariances between the regression coefficients and $x_0$ are assumed to be negligible.

$$s_{y_0}^2 = \hat{s}_a^2 + x_0^2 \hat{s}_b^2 + \hat{b}^2 s_{x_0}^2 + 2x_0 \operatorname{cov}(\hat{a},\hat{b}) \qquad (13)$$

where $\hat{s}_a^2$ and $\hat{s}_b^2$ are the estimates of the variances of the intercept and the slope respectively, and $\operatorname{cov}(\hat{a},\hat{b})$ is the covariance between the two regression coefficients. The variances and covariances of the regression coefficients are easily obtained during the iterative process to find the regression coefficients provided by BLS.[12] The coincidence of the results when expressions 12 and 13 are used to calculate the variance of the response variable given a value of the predictor variable is an internal validation of the derived expressions.

The uncertainty of the predicted observation of the response variable using the BLS technique must take into account the variance of the regression line (eqs. 12 or 13) and the variance of the new observation. Eq. 14 gives the final matrix expression for the calculation of the variance of the response variable $y_0$ obtained as the mean of $q$ observations performed at $x_0$.

$$s_{y_0}^2 = \left[ \frac{1}{q} + \mathbf{X}_0' \cdot (\mathbf{X}' \cdot \mathbf{W}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_0 + s_{x_0}^2 \cdot \hat{b}^2 \right] \cdot \hat{s}^2 \qquad (14)$$

**Variance of the predictor variable.** The study of the variance associated to the predicted predictor variable given a value of the response variable is similar to the study of the prediction of the response variable. Only the new expressions developed for the BLS method are presented here. Taking into account the errors in both axes, the resulting expression is eq. 15:

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


$$s_{x_0}^2 = \left[ \mathbf{Y}_0' \cdot (\mathbf{Y}' \cdot \mathbf{W}^{-1} \cdot \mathbf{Y})^{-1} \cdot \mathbf{Y}_0 + s_{y_0}^2 \cdot \frac{1}{\hat{b}^2} \right] \cdot \frac{\hat{s}'^2}{\hat{b}^2} \qquad (15)$$

where $\mathbf{Y_0}$ is a two-element column vector with a 1 in the first row and the response variable $y_0$ in the second row, $\mathbf{Y}$ is an $n$x2 matrix in which the first column is a column of ones and the second is made up of the $n$ values corresponding to the response variables of the experimental points, $\mathbf{W}$ is an $n$x$n$ diagonal matrix whose $i$th-diagonal element is the weighting factor $w'_i$, and $\hat{s}'^2$ is the experimental error associated to the predictions on the $x$ axis, which corresponds to:

$$\hat{s}'^2 = \frac{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{w_i'}}{n-2} \qquad (16)$$

The weighting factor $w_i'$ is now defined in eq. 17.

$$w_i' = s_{x_i}^2 + \frac{1}{\hat{b}^2} s_{y_i}^2 - 2\frac{1}{\hat{b}} \mathrm{cov}(x_i y_i) \qquad (17)$$

where normally the covariance between the predictor and response variables is neglected. The variance of the predictor variable $x_0$ is the mean of $q$ observations at $y_0$ and is found according to eq. 18:

$$s_{x_0}^2 = \left[ \frac{1}{q} + \mathbf{Y}_0' \cdot (\mathbf{Y}' \cdot \mathbf{W}^{-1} \cdot \mathbf{Y})^{-1} \cdot \mathbf{Y}_0 + s_{y_0}^2 \cdot \frac{1}{\hat{b}^2} \right] \cdot \frac{\hat{s}'^2}{\hat{b}^2} \qquad (18)$$

**Predictor intervals.** The hypothesis of normality can be assumed in the distributions of both the intercept and the slope,[15] and the results are not appreciably biased when linear regression with errors in both axes is used. Furthermore, three methods for testing the normality (Kolmogorov test,[16] normal

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


probability plots[14] and Cetama method[17]) were applied to the Monte Carlo simulation data of the response and predictor variables and the results (not shown) indicate that the response and predictor variables, despite being non-normally distributed in most cases, are very close to normality. The hypothesis that their distribution is normal is, therefore, acceptable.

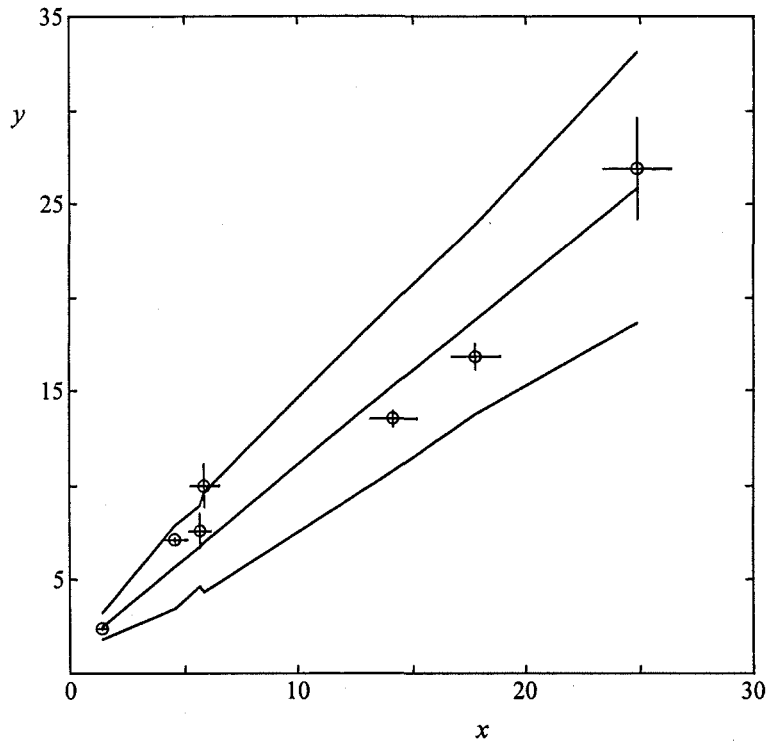



Figure 1* Experimental points for data set 3, calculated BLS regression line and confidence intervals associated to the response variable. A significance level of $\alpha = 0.1$ was selected.

The expressions of the confidence intervals for the response and predictor variables are then defined by the following equations:

$$y_0 \pm t_{\alpha,n-2} s_{y_0} \tag{19}$$

$$x_0 \pm t_{\alpha,n-2} s_{x_0} \tag{20}$$

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


where $t_{\alpha,n-2}$ is the $t$-value for a given level of significance $\alpha$ and $n$-2 degrees of freedom. As an example, the confidence interval associated to the prediction of the response variable which takes the uncertainties in both axes into account for data set 3 in the Experimental Section is shown in Figure 1. In linear regression taking into account errors in both axes, the patterns of the confidence intervals are very irregular. This is due to the variance of the predictor variable (i.e. the last term within brackets in eqs. 12 and 14, or in eqs. 15 and 18). If these terms were constant throughout the regression interval, then the confidence interval would have the shape of the classical hyperbola that is found in OLS. However, since heteroscedasticity is usually present, these terms are not constant and the pattern for the confidence intervals which take into account errors in both axes can be calculated at a given point provided that the individual uncertainty at this point is known. The continuous line for the confidence interval along the regression line is drawn by interpolating between contiguous points, since only the confidence interval at the points used for predicting can be strictly calculated.

## EXPERIMENTAL SECTION

**Data sets and software.** Six real data sets were used to validate the expressions for calculating the variance of the response variable given a value of the predictor variable and vice versa. In the data sets studied, mainly about method comparison studies, the established method is normally placed on the $x$ axis and the new method on the $y$ axis. Data sets 3 and 6, which are not related to method comparison studies, were introduced to show the usefulness of the new expressions in other fields. These six data sets are plotted in Figure 2. For the sake of clarity, only the BLS, OLS and WLS regression lines have been drawn in Figure 2.

*Data Set 1*: Concentrations of polycyclic aromatic hydrocarbons (PAHs) recovered from railroad bed soil after supercritical fluid extraction (SFE) with $CO_2$ as the modifier on the $x$ axis, and $CO_2$/10% toluene as the modifier on the $y$ axis.[18] The

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


standard deviations are the averages of three determinations at each of the 7 experimental points. The data set ranges between 1.4 and 26.9 μg/g of soil. The standard deviations associated to all experimental point are similar in both methods.

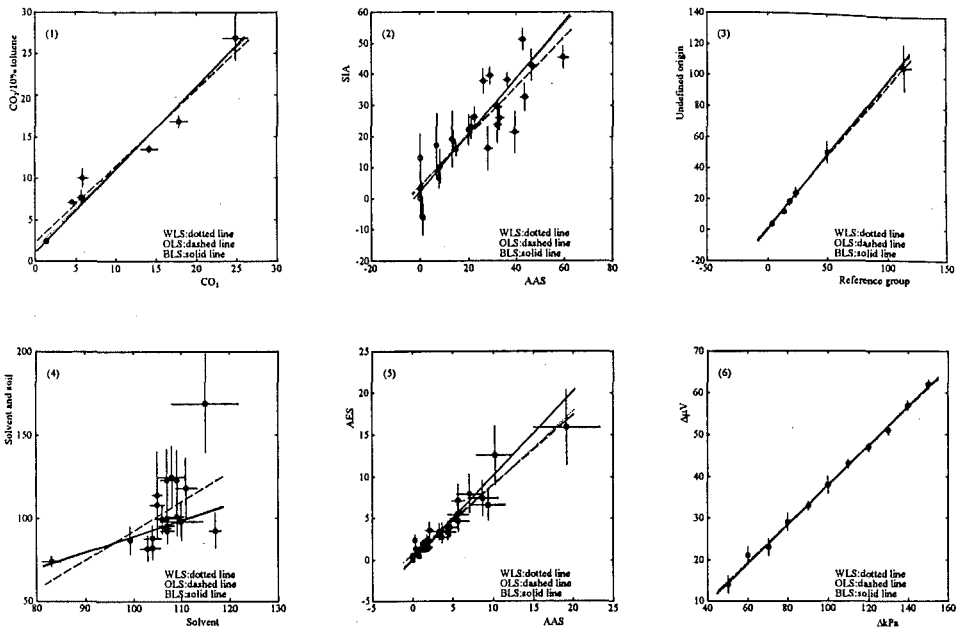



Figure 2 BLS (solid line), OLS (dashed line) and WLS (dotted line) regression lines for the six real data sets assayed. The experimental points are represented with their associated uncertainties.

*Data Set 2*: A method comparison study for analyzing $Mg^{2+}$ in natural waters with atomic absorption spectrometry (AAS) on the $x$ axis, and sequential injection analysis (SIA) on the $y$ axis.[19] The uncertainties in AAS are derived from the uncertainties associated to the univariate calibration line. The uncertainties in the SIA method are calculated from the multivariate regression model developed using the partial least squares (PLS) technique. The comparison consists of 26 data pairs within the range 0.4 and 46.3 mg/l. In all cases, the uncertainties associated to the SIA method are larger than the ones provided by AAS.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



*Data Set 3*: The composition of a set of archaeological samples of unknown origin (on the *y*-axis) is compared to a reference set of known origin (on the *x*-axis) with neutron activation analysis (NAA). Concentrations of six metal ions (Ce, Co, Cr, Fe, La and Sc) expressed in ppm, except for Fe which is in percent, are determined for a number of pottery jar handles found in Tell en-Nasbeh.[20]

*Data Set 4*: The percentage of recovery for several organochlorine pesticides after microwave-assisted extraction (MAE) with solvent (hexane/acetone 1:1) on the *x* axis, and solvent/soil suspensions spiked with the target compounds on the *y* axis.[21] The standard deviations are the average of three determinations at each point. The experiment consists of 20 points with recoveries ranging between 83 and 169%. The variances on both axes are quite large, and there is a possible outlier at high recovery values.

*Data Set 5*: A method comparison study for the determination of arsenic in natural water using continuous selective reduction and atomic absorption spectrometry (AAS) on the *x* axis, and reduction, cold trapping and atomic emission spectrometry (AES) on the *y* axis.[22] The study consists of 30 points ranging between 0 and 19.3 mg/l. The uncertainties are proportional to the concentration determined by both methods.

*Data Set 6*: Data from the measurement of the $CO_2$ Joule-Thompson coefficient.[23] The data correspond to thermocouple measured voltage differences ($\Delta$ $\mu$V) on the *y* axis, as a function of pressure increments ($\Delta$ kPa) on the *x* axis. There were 11 equally distributed data pairs with estimated unity *x* axis uncertainties. The *y* axis uncertainties were estimated to range between one and two units.

All calculations were performed using customized software using MATLAB.[24]

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


**Validation process.** The Monte Carlo simulation technique was applied[25] to validate the expressions derived to calculate the variances of the response and the predictor variables using regression considering errors in both axes (eqs. 12, 13 and 15). The Monte Carlo method generated 10,000 different data sets for each of the six initial real data sets using the individual uncertainties of each experimental point. For each of the 10,000 new data sets, the regression line was calculated, and used to predict a value of the response or the predictor variable. Finally, the variance of the 10,000 values for each original data set was calculated and compared to the predicted variance given by the theoretical expressions. This study was performed for two random values of each data set. The reversibility of the axes was also checked using the expressions for calculating both predictor and response variables. In the CVR approach, the $\lambda$ parameter was chosen to be the ratio between the average of the variances of the response variable and the average of the variances of the predictor variable for each data set.

## RESULTS AND DISCUSSION

**Variance associated to the response variable.** Table 1 shows the results of the variance of the response variable calculated using the BLS expressions at two randomly selected values for the six data sets described in the Experimental Section. It can be observed that all the results obtained from eq. 12 and eq. 13 coincide up to the eighth decimal place. Therefore, the two expressions, which were found independently, have to be considered equivalent.

**Reversibility of axes. Variance associated to the predictor and the response variables.** An interesting feature of the BLS regression technique is that it is invariant upon switching axes. OLS or WLS regression techniques do not have this feature, since only homoscedastic or heteroscedastic errors are taken into account on the *y* axis, and two different regression lines with different confidence intervals are obtained depending on which variable is placed on each axis. The CVR

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


approach, and its particular case of OR, are also invariant when the axes are switched.

| Data Set | $y_0$ Predicted | $s^2_{y_0}$ Equation 12 | $s^2_{y_0}$ Equation 13 | Differences (%) |
|---|---|---|---|---|
| 1 | 18.81 | 6.40655210 | 6.40655210 | 0.00 |
| 1 | 66.97 | 1.71674460 | 1.71674460 | 0.00 |
| 2 | 31.69 | 8.01060550 | 8.01060550 | 0.00 |
| 2 | 2.74 | 5.69862276 | 5.69862276 | 0.00 |
| 3 | 13.20 | 0.57148232 | 0.57148232 | 0.00 |
| 3 | 3.59 | 0.03850903 | 0.03850903 | 0.00 |
| 4 | 92.03 | 7.63379592 | 7.63379592 | 0.00 |
| 4 | 93.93 | 7.06714730 | 7.06714730 | 0.00 |
| 5 | 7.07 | 4.17691152 | 4.17691152 | 0.00 |
| 5 | 4.68 | 1.88624149 | 1.88624149 | 0.00 |
| 6 | 55.72 | 0.14971572 | 0.14971572 | 0.00 |
| 6 | 23.84 | 0.21554948 | 0.21554948 | 0.00 |

Table 1.- Comparison of expressions 12 and 13, for the calculation of the variance associated to the prediction of the response variable.

To check the reversibility of the axes, the variance corresponding to the response variable on the *y* axis (e.g., for a namely new method) was calculated for a fixed value of the predictor variable on the *x* axis (corresponding to a namely established method) using eqs. 12 or 13. Then, both methods were switched upon the axes and the variance of the variable on the *x* axis (the former so called new method) was calculated at the same value of the predictor variable (the established method on the *y* axis) using eq. 15. The process can be seen in Figure 3. The reversibility of the axes was tested for two random points in each of the six data sets studied. It can be observed in Table 2 that placing the methods on either of the two axes does not change the results for the variances of the predicted value. Table 2 also shows the agreement between the expressions for calculating the variance of the predictor and response variables, since the results are identical.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008

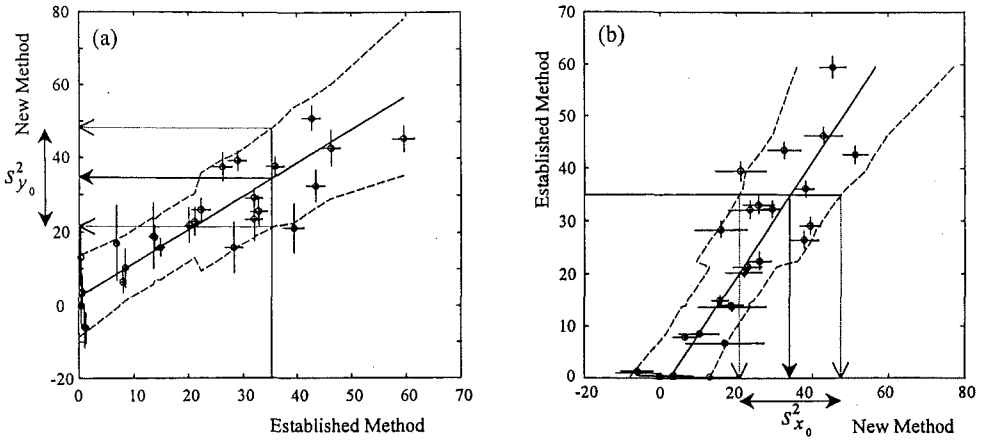




Figure 3 a) Variance associated to the predicted value of a new method (response variable) at a given value of an established method (predictor variable) b) Predicted values upon switching axes. In this case the calculated variance is associated to the predicted value of the new method (predictor variable), at a given value of the established method (response variable).

**Prediction of the variances taking into account errors in both axes.** The expressions which were derived to calculate the variance of the true mean associated to the predicted variables were validated by comparing the calculated variance values and those found by means of Monte Carlo simulations. The values obtained by means of the simulation process were also compared to the values obtained using the expressions for OLS, WLS, CVR and OR, and it was shown that errors may be significant if these techniques are used in situations in which there are heteroscedastic errors in both axes. The results obtained for the variance of the true mean of the response variable are shown in Table 3a, and the results for the variance of the true mean of the predictor variable are shown in Table 3b. In all the individual cases assayed (except two for the response variable and one for the predictor variable), the agreement between the simulated and predicted variances associated to the response and predictor variables obtained with BLS is significantly better than the agreement obtained with the other four methods. The variance for the response variable in data set 2 using the BLS expressions is

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



overestimated by up to 11% whereas the error obtained using the WLS expression was unusually low, 2.3%. The mean errors for the response and predictor variables found with BLS, WLS, OLS, CVR and OR are between 4-5%, 57-85%, 277-205%, 444-637% and 1697-462% respectively.

| Data Set | $x_0$ / $y_0$ Predicted | $s^2_{x_0}$ Equations 12 and 13 | $s^2_{y_0}$ Equation 15 | Differences (%) |
|---|---|---|---|---|
| 1 | 15.78 | 3.45781347 | 3.45781347 | 0.00 |
| 1 | 8.94 | 4.95883603 | 4.95883603 | 0.00 |
| 2 | 29.48 | 20.99017650 | 20.99017650 | 0.00 |
| 2 | 0.91 | 100.63858697 | 100.63858697 | 0.00 |
| 3 | 12.44 | 0.53514345 | 0.53514345 | 0.00 |
| 3 | 4.01 | 0.07533935 | 0.07533935 | 0.00 |
| 4 | 92.36 | 79.88020685 | 79.88020685 | 0.00 |
| 4 | 126.10 | 881.92789365 | 881.92789365 | 0.00 |
| 5 | 7.85 | 7.90043654 | 7.90043654 | 0.00 |
| 5 | 3.86 | 1.92173104 | 1.92173104 | 0.00 |
| 6 | 140.60 | 1.67904956 | 1.67904956 | 0.00 |
| 6 | 68.22 | 5.84275834 | 5.84275834 | 0.00 |

Table 2.- Comparison between the expressions for the variances of the predictor and response variables when these variables are switched upon the axes.

| | $y_0$ Pred. | $s^2_{y_0}$ Simul. | $s^2_{y_0}$ BLS | Error (%) | $s^2_{y_0}$ OLS | Error (%) | $s^2_{y_0}$ WLS | Error (%) | $s^2_{y_0}$ CVR | Error (%) | $s^2_{y_0}$ OR | Error (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18.81 | 6.9337 | 6.4066 | 7.60 | 0.8715 | 87.43 | 3.1740 | 54.22 | 1.9556 | 71.80 | 2.4708 | 64.37 |
| 1 | 6.97 | 1.7601 | 1.7167 | 2.47 | 0.6510 | 63.01 | 0.2547 | 85.53 | 1.7330 | 1.54 | 2.2459 | 27.60 |
| 2 | 31.69 | 8.0788 | 8.0106 | 0.84 | 2.9733 | 63.20 | 2.0940 | 74.08 | 5.0474 | 37.52 | 30.1071 | 272.67 |
| 2 | 2.74 | 6.4005 | 5.6986 | 10.97 | 5.5663 | 13.03 | 6.2511 | 2.33 | 7.6424 | 19.40 | 32.9836 | 415.33 |
| 3 | 13.20 | 0.5568 | 0.5715 | 2.64 | 0.7833 | 40.68 | 0.0366 | 93.43 | 1.1714 | 110.38 | 2.3364 | 319.61 |
| 3 | 3.59 | 0.0387 | 0.0385 | 0.52 | 1.0131 | 2517.83 | 0.0143 | 63.05 | 1.4007 | 3519.38 | 2.5659 | 6530.23 |
| 4 | 92.03 | 7.7837 | 7.6338 | 1.93 | 20.3506 | 161.45 | 5.0557 | 35.05 | 59.8349 | 668.72 | 976.4771 | 12445.15 |
| 4 | 93.93 | 7.3417 | 7.0671 | 3.74 | 17.0338 | 132.01 | 5.8768 | 19.95 | 56.3773 | 667.91 | 9.5010 | 29.41 |
| 5 | 7.07 | 4.4771 | 4.1769 | 6.71 | 0.0610 | 98.64 | 0.0289 | 99.35 | 0.4397 | 90.18 | 0.5412 | 87.91 |
| 5 | 4.68 | 2.0327 | 1.8862 | 7.21 | 0.0383 | 98.12 | 0.1176 | 94.21 | 0.4165 | 79.51 | 0.5178 | 74.53 |
| 6 | 55.72 | 0.1486 | 0.1497 | 0.74 | 0.1633 | 9.89 | 0.0832 | 44.01 | 0.2218 | 49.26 | 0.2873 | 93.34 |
| 6 | 23.84 | 0.2153 | 0.2186 | 1.53 | 0.1193 | 44.59 | 0.1622 | 24.66 | 0.1778 | 17.42 | 0.2433 | 13.01 |
| Mean error (%): | | | | 3.91 | | 277.49 | | 57.49 | | 444.42 | | 1697.76 |

Table 3a.- Comparison of the variance values of the new method (response variable) calculated using equations 12 or 13 with the experimental values from the simulation process on the six real data sets and the results obtained using OLS, WLS, CVR and OR.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


| | $x_0$ Pred. | $s^2_{x_0}$ Simul. | $s^2_{x_0}$ BLS | Error (%) | $s^2_{x_0}$ OLS | Error (%) | $s^2_{x_0}$ WLS | Error (%) | $s^2_{x_0}$ CVR | Error (%) | $s^2_{x_0}$ OR | Error (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 15.78 | 3.6885 | 3.4578 | 6.25 | 0.7942 | 78.47 | 2.4092 | 34.68 | 3.6207 | 1.84 | 2.7991 | 24.11 |
| 1 | 8.94 | 5.0595 | 4.9588 | 1.99 | 0.6060 | 88.02 | 0.6077 | 87.99 | 3.3658 | 33.48 | 2.6145 | 48.32 |
| 2 | 29.48 | 21.3897 | 20.9902 | 1.87 | 4.2472 | 80.14 | 2.3623 | 88.96 | 87.7860 | 310.41 | 43.8907 | 105.20 |
| 2 | 0.91 | 103.0027 | 100.6386 | 2.30 | 7.8508 | 92.38 | 7.1846 | 93.02 | 91.7295 | 10.94 | 47.1192 | 54.25 |
| 3 | 12.44 | 0.5340 | 0.5351 | 0.21 | 0.9533 | 78.52 | 0.0289 | 94.59 | 4.3668 | 717.75 | 3.0540 | 471.91 |
| 3 | 4.01 | 0.0740 | 0.0753 | 1.76 | 1.1956 | 1515.68 | 0.0148 | 80.00 | 4.6337 | 6161.76 | 3.2965 | 4354.73 |
| 4 | 92.36 | 91.1230 | 77.2692 | 15.20 | 11.0476 | 87.88 | 5.2274 | 94.26 | 69.8071 | 23.39 | 4.5703 | 94.98 |
| 4 | 126.10 | 994.3954 | 870.0950 | 12.50 | 6.7616 | 99.32 | 23.7078 | 97.62 | 69.0507 | 93.06 | 3.2380 | 99.67 |
| 5 | 7.85 | 8.4438 | 7.9004 | 6.44 | 0.1242 | 98.53 | 0.4586 | 94.57 | 1.0325 | 87.77 | 0.9325 | 88.96 |
| 5 | 3.86 | 2.0086 | 1.9217 | 4.33 | 0.0510 | 97.46 | 0.1108 | 94.48 | 1.0018 | 50.12 | 0.8615 | 57.11 |
| 6 | 140.60 | 1.7145 | 1.6790 | 2.07 | 0.7621 | 55.55 | 0.3849 | 77.55 | 3.6430 | 112.48 | 3.3568 | 95.79 |
| 6 | 68.22 | 5.8198 | 5.8428 | 0.40 | 0.5833 | 89.98 | 0.7755 | 86.67 | 3.4417 | 40.86 | 3.1780 | 45.39 |
| **Mean error (%):** | | | | **4.61** | | **205.16** | | **85.37** | | **636.99** | | **461.70** |

Table 3b.- Comparison of the variance values of the reference method (predictor variable) calculated using equation 15 with the experimental values from the simulation process on the six real data sets and the results obtained using OLS, WLS, CVR and OR.

The lowest errors using the BLS expressions are obtained with data sets 3 and 6 that seem to present the best goodness of fit of the regression line to the experimental points, which confirms the assumption that the closeness of the experimental points to the regression line is an important factor for predicting the correct variances of the response and predictor variables. On the other hand, the errors are highest for data sets 2 and 4, the maximum being around 15% for the predictor variable. Data set 4 enables the behaviour of the BLS technique to be examined in presence of data sets with a low correlation between the variables, and with two possible outliers which have a very different degree of uncertainty at the extremes of the regression range. Since the BLS technique negatively weights the influence of points with high uncertainties, the point at the furthest extreme of the range relatively affects the value of the calculated regression coefficients. This feature is partially present in WLS but absent in the other methods because they do not take into account the individual uncertainties. Therefore, the resulting regression coefficients and associated variances of the five techniques are quite different, and again, the variances corresponding to the variables predicted using BLS are closer to the simulated results than the ones calculated using the other methods. Data sets 1 and 5 give errors ranging from 1% to 7% for the response and

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


predictor variables, although data set 1 is made up of seven experimental points and data set 5 of 30, which suggests that the number of experimental points does not significantly influence the correct estimation of the variances of the response and predictor variables.

## CONCLUSIONS

The new expressions for calculating the variance of the predicted values in the $x$ and $y$ axes taking into account heteroscedastic individual errors in both axes have been developed and validated by means of simulation studies on six real data sets. The possibility of deriving two predicting expressions that give identical results for the calculation of the variances of the $y$ axis using BLS shows the reliability of the results obtained and the validity of the mathematical hypotheses which were assumed to find these expressions. The expressions developed are of a general nature and can be applied to predict values and associated uncertainties of any type, such as measurement results when using two different methods, analytical techniques, observers or laboratories.

BLS-based calculations can be made rapidly with an iterative process. The main limitation of this technique is that the uncertainties in both axes of each experimental point used in the regression analysis need to be known. However, this will probably not be unusual in the future since the international standards recommend that uncertainties be stated for every measurement result. Nevertheless, it is important to note that, in those cases where only the errors in one variable are considered, BLS gives results which are identical to those obtained using OLS or WLS regression techniques.

CVR and OR appear to produce acceptable results when the data structure meet their requirements, but as the individual errors are not taken into account, their results may be far from the real ones. It should be pointed out that despite the high

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


mean errors shown by CVR and OR methods in Tables 3a and 3b, these are mainly due to their application to data set 3. If this data set had not been taken into account, the results obtained using CVR and OR methods would have been more similar to those obtained using WLS and OLS expressions.

A feature of the BLS method is that it provides results that are invariant upon switching axes. This property is of practical significance since, in method comparison studies, for instance, it should be of no importance which of the axes is used to represent the method to be compared, as long as all the uncertainties are considered in both axes. Further studies are in progress based on the present results. Of particular interest may be the development of estimators for detection and quantitation limits.

## ACKNOWLEDGMENTS

The authors thank the Spanish Ministry of Education and Science (DGICyT project no. BP96-1008) for financial support.

## REFERENCES

(1) Draper, N.; Smith, H. *Applied regression analysis*, 2nd ed.; John Wiley: New York, 1981; pp 8-70, 108-17.

(2) Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C., de Jong, S., Lewis, P.J., Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part A*; Elsevier: Amsterdam, 1997; pp 75-8.

(3) Watters, R.L.; Carroll, R.J.; Spiegelman, C.H. *Anal. Chem.*, **1987**, *59*, 1639-43.

(4) Rawlings, J.O. *Applied Regression Analysis*; Wadsworth & Brooks/Cole: Belmont, 1988; pp 315-8.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



(5) Fuller W.A. *Measurement Error Models*; John Wiley & Sons: New York, 1987; pp 1-5, 30-6, 74-9.

(6) Anderson R.L. *Practical Statistics for Analytical Chemists*, Van Nostrand Reinhold: New York, 1987.

(7) Creasy M.A. *J. Roy. Stat. Soc. B*, **1956**, *18*, 65-9.

(8) Mandel J. *J. Qual. Tech.*, **1984**, *16*, 1-14.

(9) Hartmann C.; Smeyers-Verbeke J.; Penninckx W.; Massart D.L. *Anal. Chim. Acta*, **1997**, *338*, 19-40.

(10) Van Huffel S.; Vandewalle J. *The Total Least Squares Problems. Computational Aspects and Analysis*, Siam: Philadelphia, 1991; pp 1-18.

(11) Lisý, J.M.; Cholvadová, A.; Kutej, J. *Computers Chem.*, **1990**, *14*, 189-92.

(12) Riu, J.; Rius, F.X. *J. Chemom.*, **1995**, *9*, 343-62.

(13) Riu, J.; Rius, F.X. *Anal. Chem.*, **1996**, *68*, 1851-7.

(14) Meloun M.; Militký J.; Forina M. *Chemometrics for Analytical Chemistry. Volume 1: PC-aided statistical data analysis*, Ellis Horwood: Chichester, 1992; pp 61-8.

(15) Martínez, A.; del Río, F. J.; Riu ,J.; Rius, F.X. In preparation.

(16) Kateman G.; Pijpers F.W. *Quality Control in Analytical Chemistry*, John Wiley & Sons: New York, 1981; pp 135-41.

(17) Commission d'Établissement des Méthodes d'Analyses du Commissariat à l'Énergie Atomique (Cetama) *Statistique Appliquée a l'exploitation des Mesures*, Masson: Paris, 1986; pp 55-63.

(18) Langenfeld, J.J.; Hawthorne, S.B.; Miller, D.J.; Pawliszyn, J. *Anal. Chem.*, **1994**, *66*, 909-16.

(19) Ruisánchez, I.; Rius, A.; Larrechi, M.S.; Callao, M.P.; Rius, F.X. *Chemom. Intell. Lab. Syst.*, **1994**, *24*, 55-63.

(20) Yellin, J. *Trends Anal. Chem.*, **1995**, *14*, 37-44.

(21) López-Ávila, V.; Young, R.; Beckert, W.F. *Anal. Chem.*, **1994**, *66*, 1097-106.

(22) Ripley, B.D.; Thompson, M. *Analyst*, **1987**, *112*, 377-83.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


(23) Ogren, P.J.; Norton, J.R. *J. Chem. Edu.*,**1992**, *69*, A130-1.

(24) Mathworks Inc., Natick, Massachussets, USA.

(25) Meier, P.C.; Zünd, R.E. *Statistical Methods in Analytical Chemistry*; John Wiley & Sons: New York, 1993; pp 145-50.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008

# Capítol 7

# Conclusions

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


## 7.1 Conclusions generals

El principal objectiu de la comparació de mètodes analítics emprant regressió lineal és comprovar si la sèrie de mètodes que es comparen produeixen resultats que no difereixen estadísticament entre ells a diversos nivells de concentració de l'analit que es vol determinar, considerant sempre que sigui possible les probabilitats d'error $\alpha$ i $\beta$ associades. Altres objectius són trobar el valor de la concentració i la incertesa associats a una mostra qualsevol analitzada per un nou mètode analític a partir dels resultats obtinguts amb un altre mètode de rutina, o detectar la presència d'errors sistemàtics proporcionals o constants.

És important tenir present que en la comparació de mètodes, el model de regressió construït a partir dels resultats obtinguts amb els mètodes analítics en comparació (línia recta si es comparen dos mètodes, hiperplà si es comparen més de dos mètodes) ha de elaborar-se tenint en compte els errors associats a tots els mètodes, ja que normalment aquests seran del mateix ordre de magnitud. En calibració lineal, on normalment es vol relacionar la concentració de l'analit que es vol determinar (variable predictora) amb la resposta instrumental proporcionada per una determinada tècnica analítica (variable resposta), tal com s'ha exposat a l'apartat 1.4 hi ha una sèrie de casos on no es poden negligir els errors associats a les dues variables.

En el model de regressió de línia recta, el mètode tradicionalment més emprat és el mètode OLS degut a les seves bones propietats matemàtiques i a la rapidesa en la obtenció dels coeficients de regressió i altres paràmetres relacionats. Però sota certes condicions (per exemple l'existència d'errors associats al mètode situat a l'eix de les $x$ o d'errors no constants en el mètode situat a l'eix de les $y$ en processos de comparació de mètodes analítics) OLS condueix a estimacions incorrectes dels coeficients de regressió i paràmetres relacionats com poden ser les

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


variàncies dels coeficients. El mètode WLS constitueix una millora respecte al mètode OLS, doncs ja té en compte la possible heteroscedasticitat al mètode situat a l'eix de les $y$. No obstant, en cas d'existència d'errors associats a la variable predictora també pot donar lloc a estimacions incorrectes dels coeficients de regressió donat que continua considerant que aquesta variable no té error. No obstant això, OLS o WLS poden ser aplicats perfectament en processos de comparació de mètodes on es tingui la certesa de que els errors proporcionats per un dels dos mètodes en comparació són molt més petits que els errors proporcionats per l'altre mètode. En aquest cas el mètode que dóna menors errors hauria de situar-se a l'eix de les $x$ i l'altre a l'eix de les $y$. Si el mètode situat a l'eix de les $y$ proporciona errors homoscedàstics al llarg de tot l'interval de comparació, es podrà emprar el mètode OLS, mentre que si els errors són heteroscedàstics, s'haurà d'utilitzar el mètode WLS. En cas de que hi hagi errors presents en els dos mètodes en comparació, s'haurà d'utilitzar algun mètode que consideri els errors en els dos eixos. Si es disposa de la informació de que la relació entre els errors dels dos mètodes es constant al llarg de l'interval de comparació, es pot emprar l'aproximació CVR, on cal determinar el paràmetre $\lambda$ (apartat 2.2), corresponent a la relació entre els errors dels dos mètodes. Un cas particular de l'aproximació CVR el constitueix el mètode OR, el qual es pot emprar quan els errors dels dos mètodes analítics són iguals per a cada punt experimental ($\lambda$=1). Per últim, si hi ha present errors en els dos mètodes i no es pot assegurar que hi hagi una relació constant entre els errors dels dos mètodes al llarg de l'interval de comparació, s'hauria d'utilitzar un mètode que considerés els errors individuals en els dos mètodes analítics. Dins de tots els mètodes de regressió que consideren els errors individuals en els dos eixos, hem triat el mètode de Lisý i col·laboradors, anomenat també mètode BLS, ja que proporciona les estimacions correctes dels coeficients de la recta de regressió. Altres mètodes que consideren els errors en els dos eixos també arriben a les estimacions correctes dels coeficients de regressió (Taula 2, pàgina 86), però l'algorisme de càlcul emprat en el mètode BLS proporciona la

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


matriu variància-covariància dels coeficients de regressió, que és de gran utilitat pel posterior desenvolupament de tests estadístics associats.

Cal tenir present que en aquells casos on es cregui que l'estructura dels errors present a les dades és una quan en realitat n'és un altre (és a dir, que per exemple s'apliqui el mètode de regressió OR creient que la relació entre els errors dels dos mètodes és constant i igual a 1 en tots els punt, quan en realitat aquesta relació no és constant al llarg de l'interval de comparació, i per tant es tindria que haver aplicat per exemple el mètode BLS), la recta de regressió i les conclusions obtingudes amb l'aplicació d'algun test estadístic basat en algun dels seus coeficients poden estar lluny de les conclusions o resultats reals, tal com s'ha demostrat a bastament al llarg d'aquesta tesi doctoral.

En processos de calibració, la discussió seria completament anàleg, havent-se d'utilitzar el mètode BLS en aquells casos que presenten errors heteroscedàstics no constants tant a la variable predictora com a la variable resposta. Aquest procés es veu reflectit a l'esquema 7.1.

En la comparació dels resultats proporcionats per més de dos mètodes analítics, els coeficients de l'hiperplà de regressió s'han trobat mitjançant el mètode MLS, que és l'extensió del mètode BLS al camp multivariant, i que per tant considera els errors individuals associats a cada punt experimental de tots els mètodes analítics.

Una de les propietats importants dels mètodes BLS i MLS és que la seva aplicació a un conjunt de dades que no presenti errors en la variable o variables predictores i els errors en la variable resposta siguin homoscedàstics, condueix als mateixos resultats que mitjançant l'aplicació dels mètodes OLS i MLR respectivament.

Tots els tests estadístics desenvolupats en aquesta tesi doctoral es basen en

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008

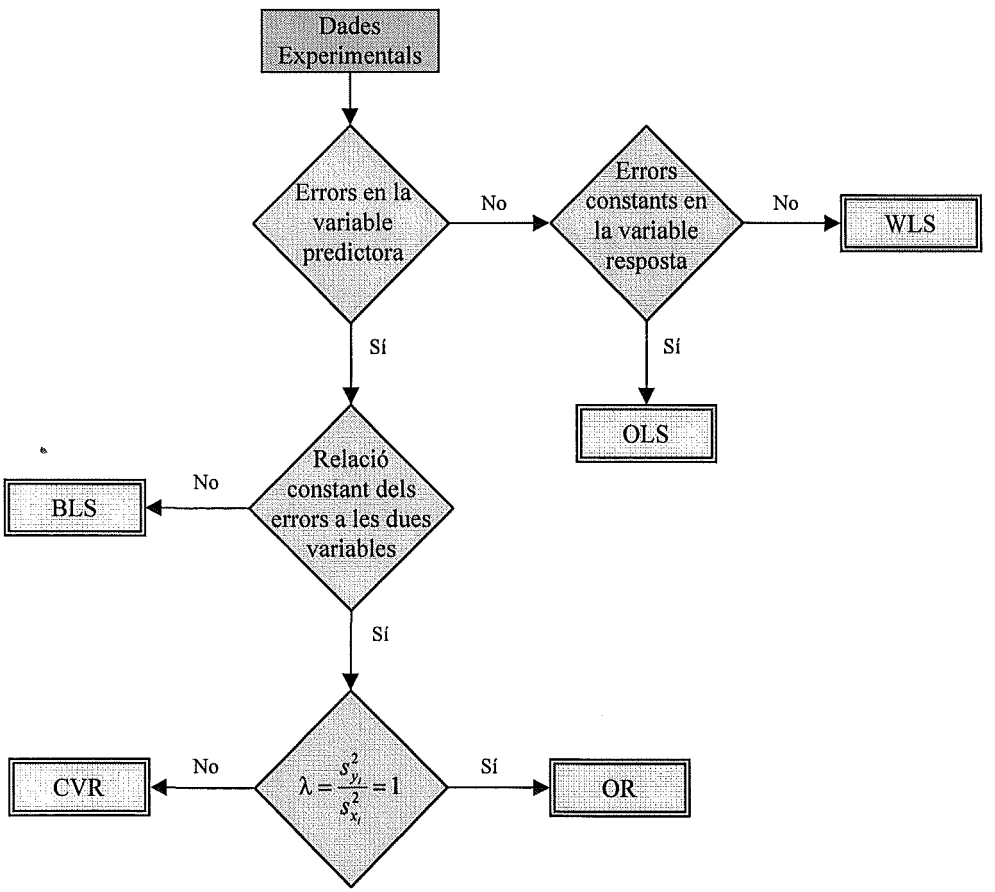

l'assumpció de que la distribució dels coeficients de la recta i l'hiperplà de regressió segueixen la distribució normal. En la recta de regressió trobada amb el mètode BLS, els coeficients de regressió no solen seguir la distribució normal (Taula 1, pàgina 132), però els errors comesos acceptant aquesta hipòtesi són petits i en tot cas menors que emprant els mètodes OLS o WLS quan es tenen errors en els dos eixos. No s'ha estudiat encara la normalitat dels coeficients de l'hiperplà de regressió trobats amb el mètode MLS, però en analogia amb el mètode BLS es pot dir que els coeficients de regressió probablement tampoc seguiran la distribució normal. Però malgrat aquest fet, probablement la seva desviació no serà significativa, ja que els resultats obtinguts en l'etapa de validació dels processos de comparació de múltiples mètodes analítics assumint la normalitat en els coeficients de l'hiperplà de regressió concorden bastant bé amb els valors teòrics (Taula 6, pàgina 237).

Malgrat la complexitat matemàtica que pot suposar passar dels mètodes OLS o WLS a BLS en regressió univariant, o de MLR a MLS en regressió multivariant, el temps de càlcul necessari per tal de trobar els coeficients de la recta o hiperplà de regressió considerant els errors individuals en tots els eixos no és gaire elevat. Per exemple, el procés de càlcul per tal d'obtenir la representació gràfica del test conjunt per a l'ordenada a l'origen i el pendent basat en el mètode BLS per a diverses sèries de dades sol tardar entre 1 i 2 segons en un ordinador HP Vectra VE Pentium 75. En regressió multivariant, l'algorisme MLS programat sobre el mateix ordinador sol trobar els coeficients de l'hiperplà de regressió de diverses sèries de dades aplicades a la comparació d'entre 4 i 8 mètodes analítics en un temps inferior als 3 segons per a la majoria de conjunts assajats.

Una de les principals dificultats a l'hora d'aplicar els tests estadístics basats en les mètodes de regressió BLS o MLS és la necessitat de conèixer el valor de la variància individual de cada variable en cada un dels punts experimentals, ja que

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


les variàncies normalment impliquen repeticions i per tant més temps d'anàlisi i un cost econòmic superior. Cal recordar, però, que la sèrie de normes EN 45000/UNE 66500 especifiquen que cada resultat analític ha d'anar acompanyat del valor de la seva incertesa. De totes maneres, els analistes són, a la llarga, els qui poden escollir si volen ser més o menys restrictius a l'hora d'estimar o calcular les incerteses associades a cada resultat analític, tot i tenir clar que el fet de considerar adequadament les incerteses o no, pot arribar a fer que s'acceptin com a bons mètodes analítics esbiaixats o que es rebutgin mètodes analítics no esbiaixats, tal com s'ha pogut comprovar al capítol 4.




Esquema 7.1. Procés de trobada del mètode de regressió lineal adequat segons les característiques del conjunt de dades experimental.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


## 7.2 Conclusions del capítol 3

En aquest capítol s'ha procedit a determinar la distribució dels coeficients de la recta de regressió calculada segons el mètode BLS. S'ha demostrat, per a una sèrie de conjunts de dades que abasten diverses situacions que es poden donar en processos de comparació de mètodes, que tot i obtenir-se distribucions no normals, es pot acceptar la hipòtesi de normalitat en els coeficients de la recta de regressió trobada segons el mètode BLS sense cometre un error significatiu. D'altra banda, s'han desenvolupat i validat les expressions per a la detecció d'un error sistemàtic constant o proporcional en processos de comparació de dos mètodes analítics (o per exemple per a la detecció d'efectes de matriu o correccions del blanc) considerant les probabilitats d'error $\alpha$ i $\beta$ associades i tenint en compte el biaix màxim que no es considera significatiu en el procés de comparació. També s'han desenvolupat les expressions (en forma de procediment iteratiu) pel càlcul *a priori* del número de punts que ha de tenir la recta de regressió per tal d'obtenir els coeficients de regressió amb probabilitats prefixades d'errors $\alpha$ i $\beta$.

Un aspecte posterior a considerar seria l'avaluació del biaix màxim que no es considera significatiu, ja que en la detecció d'errors sistemàtics o constants aquest biaix ve donat en termes d'ordenada a l'origen o de pendent, i pot arribar a ser difícil la seva translació a unitats de concentració.

## 7.3 Conclusions del capítol 4

En aquest capítol s'ha desenvolupat el test conjunt per a l'ordenada a l'origen i el pendent de la recta de regressió trobada segons el mètode BLS (equació 11, pàgina 169) per tal de comparar els resultats de dos mètodes analítics. S'ha comparat aquest test conjunt basat en el mètode BLS amb els tests conjunts basats en els

UNIVERSITAT ROVIRA I .VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


mètodes OLS i WLS, així com amb altres tests estadístics, i s'ha demostrat que el fet d'ignorar els errors associats als dos mètodes analítics en comparació a l'hora d'establir la recta de regressió pot donar lloc a conclusions incorrectes: es poden acceptar com a bons, mètodes incorrectes, i es poden rebutjar bons mètodes. La seva validació mitjançant el procés de simulació de Monte Carlo ha donat lloc a resultats que concorden molt aproximadament amb els teòrics.

Pel que fa al programa desenvolupat pel càlcul i representació del test conjunt per a l'ordenada a l'origen i el pendent en Matlab 4.0 per a Windows 3.1 o superior, permet el càlcul i la visualització del test conjunt en pocs segons (normalment entre 1 i 2) escollint diverses probabilitats d'error $\alpha$ i pels tests conjunts basats en els mètodes OLS, WLS i BLS.

## 7.4 Conclusions del capítol 5

En aquest capítol s'ha desenvolupat un test per a la comparació simultània dels resultats de múltiples mètodes analítics a diversos nivells de concentració: el test conjunt per a l'ordenada i la suma de pendents de l'hiperplà de regressió. Aquest test està basat en el mètode de regressió MLS, que és l'ampliació del mètode BLS al camp multivariant. Aquest test també ha estat validat emprant el mètode de simulació de Monte Carlo, i els resultats també coincideixen en un grau elevat amb els valors teòrics.

Els resultats del procés de comparació de múltiples mètodes analítics per a la majoria de la sèrie de conjunts de dades comprovats s'obtenen en un període curt de temps.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



## 7.5 Conclusions del capítol 6

En aquest capítol s'han desenvolupat els intervals de confiança associats a la predicció de la variable predictora donat un valor de la variable resposta o viceversa en regressió lineal considerant els errors individuals en dos eixos. Aquests intervals de confiança, apart de ser útils en processos de comparació de mètodes, poden aplicar-se a altres camps com per exemple la datació per radiocarboni o l'assignació d'orígens a peces arqueològiques desconegudes. Les variàncies de la variable predictora o resposta, peça clau en la construcció dels seus intervals de confiança, han estat trobats seguint dos camins independents que condueixen a resultats idèntics, el que constitueix una validació interna. A més, aquestes expressions han estat també validades mitjançant el mètode de simulació de Monte Carlo, assolint-se resultats que concorden aproximadament amb els teòrics, i molt millors que els intervals de confiança obtinguts amb l'aplicació dels mètodes de regressió OLS, WLS, OR o CVR quan hi ha presents errors heteroscedàstics individuals en dos eixos.

## 7.6 Perspectives de futur

Com a perspectiva de futur, i dins d'un objectiu global que consistiria en millorar la qualitat de la informació proporcionada pels mètodes que consideren errors en ambdós eixos, la recerca es pot enfocar cap al desenvolupament de tècniques de regressió robusta que consideressin els errors heteroscedàstics individuals en tots els eixos. Una altra alternativa consistiria en el desenvolupament de tècniques de detecció de punts discrepants per ambdós mètodes, així com de tests estadístics per tal de comprovar el bon ajust dels punts experimentals a la recta o hiperplà de regressió. Aquest punt és important perquè una recta o hiperplà de regressió que presenti algun punt discrepant o un mal ajust dels seus punts experimentals, donarà

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


com a resultat un increment en el valor de l'error experimental i en conseqüència de tots els intervals de confiança dels coeficients de regressió o paràmetres derivats, essent aquest increment només degut a la falta d'ajust.

Fins a la data, el test conjunt per a l'ordenada a l'origen i el pendent considerant errors en dos eixos s'ha desenvolupat i validat per a la comparació dels resultats de la determinació d'un sol analit per dos mètodes diferents. Però hi ha tot un seguit de mètodes analítics que poden proporcionar informació simultània de diversos analits en una sola mostra (per exemple les tècniques cromatogràfiques). Per tant, una extensió del test conjunt per a l'ordenada a l'origen i el pendent considerant errors en dos eixos seria l'estudi de la seva aplicació per tal de comparar els resultats proporcionats per dos mètodes de simultàniament més d'un analit.

De la mateixa manera que s'ha introduït el càlcul de les probabilitats d'error $\beta$ en els tests individuals per a l'ordenada a l'origen i el pendent, un següent pas consistiria en el càlcul de les probabilitats d'error $\beta$ associades al test conjunt per a l'ordenada a l'origen i el pendent basat en el mètode BLS i al test conjunt per a l'ordenada a l'origen i la suma de pendents basat en el mètode MLS, és a dir, a la quantificació de les probabilitats d'acceptar com a bons, mètodes esbiaixats en un estudi de comparació de mètodes analítics. Dins del mètode de regressió MLS també caldria estudiar els coeficients de l'hiperplà de regressió per tal de comprovar si la seva distribució és normal, o en cas que no ho sigui, si s'allunya significativament de la normalitat, així com també fer l'algorisme de trobada dels coeficients de l'hiperplà regressió menys sensible al punt inicial escollit per tal de poder executar amb més fiabilitat el procediment iteratiu que condueix a la seva obtenció.

Altres actuacions dins de la calibració univariant consistirien en el desenvolupament de límits de decisió, detecció i quantificació considerant els

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



errors en ambdós eixos, i en el desenvolupament de tècniques de regressió no lineal, útil en casos com la datació per radiocarboni de materials arqueològics mitjançant mesura per centelleig líquid, on la relació entre concentració i resposta es sol ajustar a un polinomi, normalment de tercer grau.

Per últim, i dins de la calibració multivariant, una futura actuació consistiria en intentar desenvolupar un nou mètode de calibració multivariant basat en la descomposició per components principals que tingués en compte els errors individuals en tots els eixos: primer dur a terme una descomposició de les dades inicials segons la tècnica MLPCA i llavors aplicar el mètode MLS a les dades descompostes.

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008

# Apèndix

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



Expressions emprades en la comprovació de la normalitat de les distribucions de l'ordenada a l'origen i el pendent mitjançant el mètode de Cetama

**Coeficients generals**

$$k_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = s_x^2 \qquad \text{(A.1)}$$

$$k_3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i - \bar{x})^3 \qquad \text{(A.2)}$$

$$k_4 = \frac{n(n+1)\sum_{i=1}^{n}(x_i - \bar{x})^4 - 3(n-1)\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2}{(n-1)(n-2)(n-3)} \qquad \text{(A.3)}$$

***Corba simètrica limitada pel domini*** $\bar{x} \pm d_1$

La distribució segueix una corba simètrica limitada pel domini $(\bar{x} - d_1, \bar{x} + d_1)$, amb el coeficient $d_1$ calculat segons:

$$d_1 = \sqrt{\frac{2b_2 s_x^2}{3 - b_2}} \qquad \text{(A.4)}$$

El paràmetre $m_1$ de l'equació 3.12 ve definit per:

$$m_1 = \frac{5b_2 - 9}{2(3 - b_2)} \qquad \text{(A.5)}$$

i $f_0$ de la mateixa equació 3.12 és una constant tal que:

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008


$$\int_{\bar{x}-d_1}^{\bar{x}+d_1} f(x)dx = 1 \qquad \text{(A.6)}$$

***Corba simètrica il·limitada pels dos costats*** $(-\infty,+\infty)$

Els paràmetres $m_2$ i $m_3$ de l'equació 3.13 venen definits segons:

$$m_2 = \sqrt{\frac{2b_2 s_x^2}{b_2 - 3}} \qquad \text{(A.7)}$$

$$m_3 = \frac{5b_2 - 9}{2(b_2 - 3)} \qquad \text{(A.8)}$$

i $f_0$ és una constant tal que:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \qquad \text{(A.9)}$$

***Corba no simètrica limitada als dos costats*** $(d_2...d_3)$

Els paràmetres $d_2$ i $d_3$ es calculen segons les següents expressions:

$$d_2 = \bar{x} - \frac{w \cdot q_1}{r_1} \qquad \text{(A10)}$$

$$d_3 = w + d_2 \qquad \text{(A.11)}$$

amb els paràmetres $r_1$, $w$ i $q_1$ que venen donats per:

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



$$r_1 = \frac{6(b_2 - b_1 - 1)}{3b_1 - 2b_2 + 6} \tag{A.12}$$

$$w = 2\sqrt{s_x^2(1-k)(1+r_1)} \tag{A.13}$$

$$q_1 = \frac{r_1}{2}\left(1 - \sqrt{\frac{-k}{1-k}}\right) \tag{A.14}$$

El paràmetre $q_2$ de l'equació 3.15 ve expressat segons:

$$q_2 = r_1 - q_1 \tag{A.15}$$

i $f_0$ de l'equació 3.15 és una constant tal que:

$$\int_{d_2}^{d_3} f(x)dx = 1 \tag{A.16}$$

El valor màxim (moda) de la distribució, correspon a:

$$m = \bar{x} + \frac{w \cdot (q_2 - q_1)}{r_1(r_1 - 2)} \tag{A.17}$$

***Corba no simètrica il·limitada als dos costats*** $(-\infty,+\infty)$

Els diferents paràmetres de l'equació 3.16 es troben definits segons:

$$r_2 = \frac{6(b_2 - b_1 - 1)}{2b_2 - 3b_1 - 6} \tag{A.18}$$

$$q = 1 + \frac{r_2}{2} \tag{A.19}$$

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



$$p = r_2\sqrt{\frac{k}{1-k}} \tag{A.20}$$

$$v = \sqrt{s_x^2(r_2 - 1)(1-k)} \tag{A.21}$$

$$c_1 = \bar{x} - \frac{p \cdot v}{r_2} \tag{A.22}$$

i $f_0$ és una constant tal que:

$$f_0 = \int_{-\infty}^{\infty} f(x)dx = 1 \tag{A.23}$$

La moda d'aquest tipus de distribucions correspon a:

$$m = \bar{x} + \frac{2p \cdot v}{r_2(r_2 + 2)} \tag{A.24}$$

***Corba no simètrica limitada a un costat*** $(d_4 \ldots +\infty)$

La corba es troba limitada per l'esquerra segons el següent paràmetre $d_4$:

$$d_4 = \bar{x} - c_2\frac{q_3}{r_2} \tag{A.25}$$

i la distribució definida a l'equació 3.17 presenta la següent moda:

$$m = \bar{x} - \frac{c_2(q_3 + q_4)}{r_2(r_2 + 2)} \tag{A.26}$$

Els nous coeficients de les equacions A.25–A.26 es defineixen segons:

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



$$q_3 = \frac{r_2}{2}\left[\sqrt{\frac{k}{k-1}} - 1\right] \tag{A.27}$$

$$q_4 = q_3 + r_2 \tag{A.28}$$

$$c_2 = 2\sqrt{s_x^2 (r_2 - 1)(k-1)} \tag{A.29}$$

i $f_0$ de l'equació 3.17 és una constant tal que:

$$\int_{d4}^{\infty} f(x)dx = 1 \tag{A.30}$$

***Corba no simètrica limitada a un costat (- ∞ ... $d_5$)***

La corba es troba limitada per la dreta segons el següent paràmetre $d_5$:

$$d_5 = \bar{x} + c_2 \frac{q_3}{r_2} \tag{A.31}$$

El paràmetre $f_0$ de l'equació 3.18 és una constant tal que:

$$\int_{-\infty}^{d_5} f(x)dx = 1 \tag{A32}$$

i la distribució presenta la següent moda:

$$m = \bar{x} + \frac{c_2(q_3 + q_4)}{r_2(r_2 + 2)} \tag{A33}$$

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



# GLOSSARI

| | |
|---|---|
| APM | Mètode paramètric aproximat<br>*Approximate parametric method* |
| BLS | Mínims quadrats bivariants<br>*Bivariate least squares* |
| CVR | Relació constant de variàncies<br>*Constant variance ratio* |
| EPM | Mètode paramètric exacte<br>*Exact parametric method* |
| GLS | Mínims quadrats generalitzats<br>*Generalized least squares* |
| IANOVA | Anàlisi d'informació de la variància<br>*Informational analysis of variance* |
| ILF | Funció lineal implícita<br>*Implicit linear function* |
| IRWLS | Mínims quadrats iterativament ponderats<br>*Iteratively reweighted least squares* |
| LSM | Mínims quadrats de la mediana<br>*Least median squares* |
| MLLRR | Regressió per arrel latent de màxima versemblança<br>*Maximum likelihood latent root regression* |
| MLPCA | Anàlisi per components principals de màxima versemblança<br>*Maximum likelihood principal component analysis* |
| MLPCR | Regressió per components principals de màxima versemblança<br>*Maximum likelihood principal components regression* |
| MLR | Regressió lineal múltiple<br>*Multiple linear regression* |
| MLS | Mínims quadrats multivariants<br>*Multivariate least squares* |
| ODR | Regressió de la distància ortogonal<br>*Orthogonal distance regression* |
| OLS | Mínims quadrats<br>*Ordinary least squares* |
| OR | Regressió ortogonal<br>*Orthogonal regression* |
| PCA | Anàlisi per components principals<br>*Principal components analysis* |
| PCR | Regressió per components principals<br>*Principal components regression* |
| PLS | Regressió per mínims quadrats parcials<br>*Partial least squares* |

UNIVERSITAT ROVIRA I VIRGILI
REGRESSIÓ LINEAL AMB ERRORS EN AMBDÓS EIXOS. APLICACIÓ A LA CALIBRACIÓ I A LA COMPARACIÓ DE MÈTODES ANALÍTICS
Jordi Riu Rusell
ISBN: 978-84-691-1897-9/D.L:T-353-2008



| | |
|---|---|
| TLS | Mínims quadrats totals<br>*Total least squares* |
| WLS | Mínims quadrats ponderats<br>*Weighted least squares* |