SERVIÇO DE QUÍMICA E TOXICOLOGIA FORENSES

SELECTION CRITERIA FOR WEIGHTING FACTOR IN LINEAR REGRESSION

Pedro Costa[1], Paula Melo[1], Susana Simões[1], António Castañera[1], João Miguel Franco[1] , Helena M. Teixeira[1,2]

[1]Serviço de Química e Toxicologia Forenses do National Institute of Legal Medicine and Forensic Sciences, Portugal

[2]Faculty of Medicine, University of Coimbra, Coimbra, Portugal

INSTITUTO NACIONAL DE MEDICINA LEGAL E CIÊNCIAS FORENSES, I.P.

INTRODUCTION

Prior to fit mathematical functions to experimental data tests, it's required the proper establishment of a working range, regarding the verification of several assumptions before using calibration curves: visual inspection of the graphs, inspection of the influence of aberrant values and leverage points, the homocedasticity analysis, residual analysis and adjustment to the linear model.

With the study of the working range and linearity we seek to investigate if, for a certain range of concentrations, analytical signals are directly proportional to the concentrations used.

Therefore the use of weighting factors in linear regression should be studied in order to decrease the effects of non-linearity.

Linearity

- In the approach to the study of linearity the International Organization for Standardization, takes into account the following assumptions: visual inspection of graphs, inspection outliers and leverage points, the homocedasticity analysis, residues independence check and adjust the linear model. When the sample size is large compared with the number of parameters in the regression model the effect of dependence between residues is relatively insignificant and can be ignored.

Homocedasticity

- Due to the implicit need for assays in the Forensic Toxicology area, sometimes is not possible to confirm linearity in the desired **analytical range** that usually involves changes in order of magnitude (Peters, 2006), so the **heteroscedasticity** appears as an inescapable fact.

Correlation coefficient

- The evaluation of the correlation coefficient should not be made with the objective of evaluating the linearity, since values of r^2 very close to the unit do not presuppose a linear model (*Royal Society of Chemistry)*. However, r^2 indicates the proportion of the variance of the dependent variable, Y, which is explained in linear terms for the independent variable X. Thus, an **r^2 value close to unity indicate an adequate regression model.**

Weighting factors in linear regression

- When it is not possible to opt for a smaller range of work it appears as an option to apply **weighting factors**. The study of weighted linear regression should be made applying various weighting factors to the experimental data of the calibration curves.

MATERIAL AND METHODS

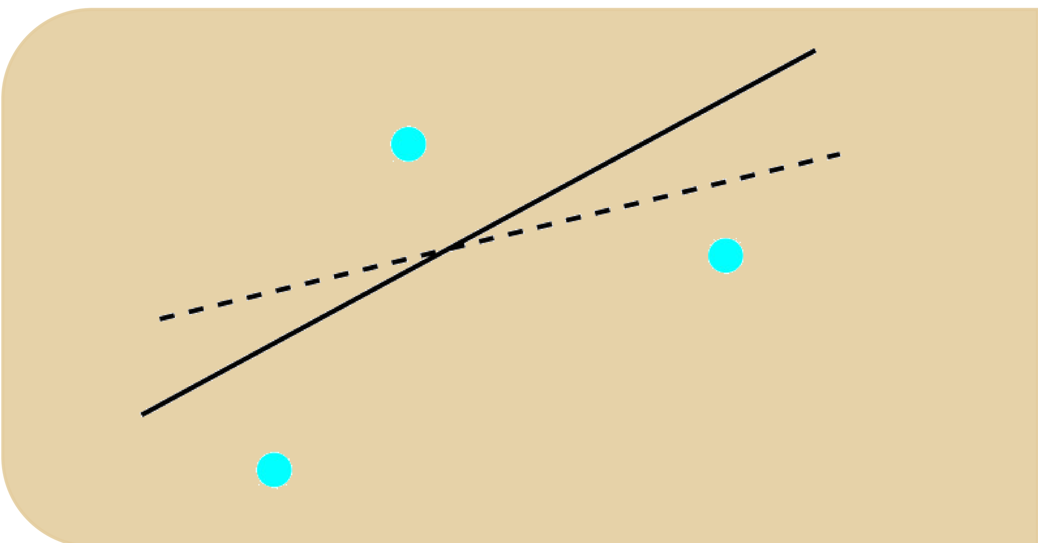

Fig 1- Assessing the Fit of Regression Model.

During the validation of the assay for quantification of benzodiazepines by LC-MS/MS (UPLC-TQD) were analyzed 9 calibration curves to 20 analytes with six weighting factors($w=1/x$; $w=1/x^2$; $w=1/x^{1/2}$; $w=1/y$; $w=1/y^2$; $w=1/y^{1/2}$).

In this paper we propose selection criteria for the most appropriate weighting factor. This selection is based in the sum of residuals and correlation coefficient in relation to the average values.

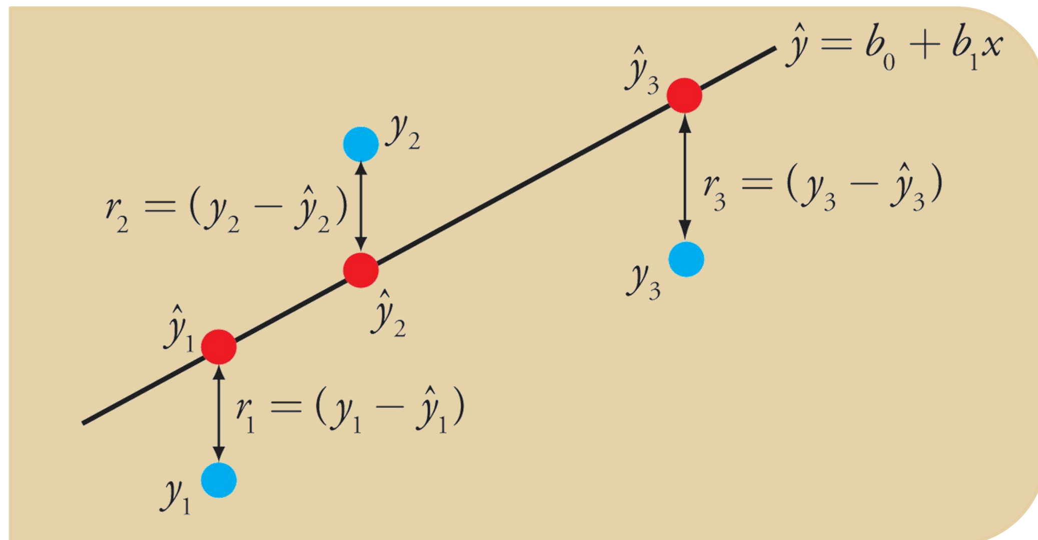

Fig 2- Ordinary Least Squares.

RESULTS

For 20 benzodiazepines it was selected a work range between 1 ng/mL and 500 ng/mL. To assess the homoscedasticity of variance were prepared 5 aliquots spiked with compounds in target concentrations corresponding to the extremes of the work range. The results were evaluated using an F test with a confidence level of 95% ($p = 0.05$). It was found heteroscedasticity for all benzodiazepines studied. Subsequently we studied the application of weighting factors to the performance of the method.

The selection of the weighting factor was intrinsically linked to the lower amount of residuals sum and the highest coefficient of determination. However, as expected, these items proved to be independent so it was necessary to normalize the values of each parameter to allow a proper evaluation.

The sum of the lowest normalized value of the residuals sum with the lowest normalized value of the equation ($1- r^2$) allows obtaining a ranking where the smallest value corresponds to the best weighting factor for each evaluation. The method developed demonstrated to be suitable for an objective selection of the best weighting factor relative to the sum of residues and coefficient of determination.

The large number of samples has made possible to treat the sum of relative errors as residue sum, Σ res. The weighting factor selection was linked to the lower value of residues sum.

Σ res

The coefficient of determination indicates adequacy of the regression model. Thus, the weighting factor selection was implicitly tied to the lower value of the transformed coefficient of determination, ($1- r^2$).

$1- r^2$

Diazepam

Calibration Curve		20130719_1	20130621_7	20130612_1	20130516_3	20130503_1	20130314_10	20130205_2	20130124_1	20121218_1	Media Σ res
Σ res	w = 1/x	58.99%	51.93%	30.23%	95.19%	83.48%	52.91%	59.37%	**66.82%**	75.99%	63.88%
	w = $1/x^2$	**44.69%**	**34.02%**	**24.22%**	**52.50%**	**36.46%**	**49.51%**	54.43%	66.97%	**40.64%**	**44.83%**
	w = $1/x^{1/2}$	107.97%	143.21%	83.81%	277.68%	185.70%	139.67%	125.52%	123.47%	224.81%	156.87%
	w = 1/y	59.46%	50.86%	31.06%	97.49%	85.46%	51.13%	59.63%	70.20%	79.49%	64.98%
	w = $1/y^2$	45.83%	34.45%	24.30%	59.32%	37.12%	50.84%	**53.64%**	68.85%	42.04%	46.27%
	w = $1/y^{1/2}$	104.76%	139.73%	87.13%	279.61%	189.38%	135.28%	117.16%	109.78%	227.39%	154.47%

Calibration Curve		20130719_1	20130621_7	20130612_1	20130516_3	20130503_1	20130314_10	20130205_2	20130124_1	20121218_1	Media r^2
r^2	w = 1/x	0.99760	0.99809	0.99862	0.99368	0.99789	0.99738	0.99572	0.99648	0.99475	0.99669
	w = $1/x^2$	0.99484	0.99748	0.99878	0.99173	0.99653	0.99360	0.99322	0.98759	0.99560	0.99438
	w = $1/x^{1/2}$	**0.99850**	0.99853	0.99875	0.99492	0.99861	0.99842	0.99620	0.99803	0.99616	0.99757
	w = 1/y	0.99763	0.99814	0.99866	0.99383	0.99797	0.99748	0.99585	0.99658	0.99512	0.99681
	w = $1/y^2$	0.99479	0.99731	**0.99880**	0.99041	0.99642	0.99331	0.99331	0.98636	0.99543	0.99402
	w = $1/y^{1/2}$	0.99850	**0.99856**	0.99878	**0.99508**	**0.99864**	**0.99846**	**0.99633**	**0.99810**	**0.99636**	**0.99765**

Normalization		Σ res	($1- r^2$)	Sum
Fator	w = 1/x	-0.47	-0.32	-0.7863
	w = $1/x^2$	-0.83	1.14	0.3061
	w = $1/x^{1/2}$	1.30	-0.87	0.4265
	w = 1/y	-0.45	-0.39	**-0.8384**
	w = $1/y^2$	-0.80	1.36	0.5592
	w = $1/y^{1/2}$	1.25	-0.92	0.3328

Tab 1-Application of criteria for weighting factor selection to Diazepam.

CONCLUSIONS

The Internal Quality Control applied in this assay requires the use of the coefficient of determination higher than 0.99 as acceptance criteria. Thus, proves to be coherent the use of this parameter in conjunction with the sum of evaluation of residues for the purpose of weighting factor selection.

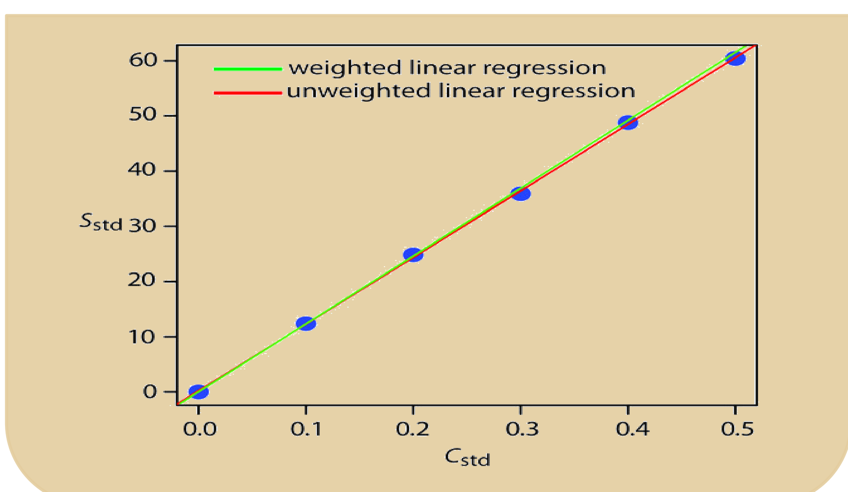

Fig 3- Weighting factor effects

The developed method demonstrated to be suitable for an objective selection of the best weighting factor based on the sum residues and the coefficient of determination. The regression models showed a better performance with particular evidence for the lower working ranges values.