XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6–11, 2009, Lisbon, Portugal

LINEAR FITTING PROCEDURES APPLIED TO REFRACTOMETRY OF AQUEOUS SOLUTIONS

Olivier Pellegrino [1], *Andreia Furtado* [2], *Eduarda Filipe* [3]

[1] Instituto Português da Qualidade, Caparica, Portugal, opellegrino@mail.ipq.pt
[2] Instituto Português da Qualidade, Caparica, Portugal, afurtado@mail.ipq.pt
[3] Instituto Português da Qualidade, Caparica, Portugal, efilipe@mail.ipq.pt

Abstract – Two linear fitting procedures are applied to internationally reference data in refractometry of solutions. Least square linear regression is compared to linear interpolation in the intervals of the two successive referenced data. The formulas and the validity of the procedures are shortly presented. The results are comparable, but for faster results, the linear regression method is preferred to the interpolation by intervals, albeit their bigger uncertainties values.

Keywords: linear calibration, refractometry, uncertainties

1. INTRODUCTION

In calibration laboratories, it is frequent to interpolate data from a set of experimental pairs (x_i, y_i). Among the different linear interpolation procedures, one of the most established and used is the least square linear regression. From the resulting equation:

$$Y = a_0 + a_1 X \quad (1)$$

and a new experimental datum, X_j, it is possible to deduce the interpolated Y_j value. This is called the direct calibration. The indirect calibration consists on deducing X_j from the experimental datum Y_j [1]. However, there is another common practice that consists on deducing Y_j from the interval of values where X_j belongs, by the linear interpolation between the limits of the interval. This communication shortly investigates the difference of the procedures for the simple case of experimental pairs (x_i, y_i) without uncertainty. For the linear regression procedure, it is called the ordinary least square (OLS) [1] – [3]. In particular, uncertainties associated to the deduced values for the two procedures are presented with an application in refractometry of aqueous solutions.

Indeed, thanks to the values of refractive indexes and compound concentrations of aqueous solutions published in internationally accepted Tables [4], refractometry is an easy-to-use experimental technique to analyse diluted aqueous solutions of such compounds. As the data of these Tables present a linear behaviour, with the help of statistical functions included in commercial software, it is easy to deduce the characteristics of the OLS determined linear functions.

In this communication, Refractometry measurements results of potential alcoholic content, which a must can produce after fermentation, are presented, as obtained by the two interpolation procedures. Indeed, the Refractometry Laboratory of the Portuguese Institute for Quality is responsible for the Metrological Control of refractometers that measure the refractive index of grape must before the alcoholic fermentation. In particular, the Laboratory produces aqueous glucose solutions, which are reference materials used for the subsequent periodical verifications of the refractometers measuring the sugar content of grape must. The international Table of correspondence between refractive index and potential alcoholic content are used to deduce the respective quantities of the prepared reference materials. One of the objectives of this communication is also to compare the results of the two procedures, including the respective associated uncertainty, in order to suggest the most adequate method to a given task.

2. MEASUREMENT METHODOLOGY

2.1. The OLS procedure

From the experimental pairs (x_i, y_i), this procedure provides the parameters, the uncertainties associated to the parameters and a statistical measure of the goodness of the fit. Indeed:

$$a_1 = \frac{\sum_{k=1,N} (x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1,N} (x_k - \bar{x})^2} = \frac{Q_{xy}}{Q_{xx}} \quad (2)$$

$$a_0 = \bar{y} - a_1 \bar{x} \quad (3)$$

where the mean values of x_k and y_k are given by:

$$\bar{x} = \frac{1}{N} \sum_{k=1,N} x_k \text{ and } \bar{y} = \frac{1}{N} \sum_{k=1,N} y_k \quad (4)$$

with the respective uncertainties:

$$u^2(a_1) = \frac{S_R{}^2}{Q_{xx}} \tag{5}$$

$$u^2(a_0) = (\frac{1}{N} + \frac{\bar{x}^2}{Q_{xx}}) S^2{}_R \tag{6}$$

and:

$$u(a_0, a_1) = -\bar{x}\,\frac{S_R{}^2}{Q_{xx}} = -\bar{x}\,u^2(a_1) \tag{7}$$

where:

$$S_R{}^2 = \frac{\sum_{k=1,N} (y_k - a_0 - a_1\, x_k)^2}{N-2} \tag{8}$$

and the goodness of the linear fit being estimated by:

$$t = \frac{a_1}{u(a_1)} \tag{9}$$

In other words, the null hypothesis $a_1 = 0$ is rejected if, when comparing with the Student's t distribution:

$$t > t_{\frac{\alpha}{2}, N-2} \tag{10}$$

for N-2 degrees of freedom and the significance level α.

A commercial software, like Excel, provides all the functions allowing to estimate the parameters displayed above. Therefore, from the experimental pairs (x_i, y_i), all the expressions above are then easily evaluated.

From the experimental value X_j, with the associated uncertainty $u(X_j)$, it is possible to deduce the interpolated Y_j value using the OLS procedure. The associated uncertainty is then deduced by using the law of propagation of uncertainties [5] to $Y_j(a_0, a_1, X_j)$ in the equation (1):

$$u^2(Y_j) = \sum_k \left(\frac{\partial Y_j}{\partial Z_k}\right)^2 u^2(Z_k) + 2\sum_{k \neq l} \frac{\partial Y_j}{\partial Z_k}\frac{\partial Y_j}{\partial Z_l} u(Z_k, Z_l) \tag{11}$$

leading, with the help of equations (2) to (6), to:

$$u(Y_j) = \sqrt{a_1{}^2\; u^2(X_j) + \{\frac{(X_j - \bar{x})^2}{Q_{xx}} + \frac{1}{N}\} S_R{}^2} \tag{12}$$

Since the experimental values Y_i were obtained in different conditions that the pairs (x_i, y_i), the equation (12), including the term $(a_1\; u(Y_j))^2$, is preferred to the one published in the Standard [1] and generally used [6].

2.2. Interpolation by intervals procedure

From the experimental value X_j, lying in the interval of successive reference data [x_i, x_{i+1}], it is current to deduce Y_j belonging to the interval [y_i, y_{i+1}], according to:

$$Y_j = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}\,(X_j - x_i) + y_i \tag{13}$$

As, in this communication, we consider the experimental pairs (x_i, y_i) without any associated uncertainty, the uncertainty associated to Y_j, deduced from equation (13) is:

$$u(Y_j) = \left|\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right|\; u(X_j) \tag{14}$$

The model of this procedure is simpler than the previous one. However, it needs to be updated for all the different intervals, which is more time consuming.

2.3. Experimental methodology

The preparation of aqueous glucose solutions was performed gravimetrically with balances calibrated at the Mass Primary Laboratory of the IPQ, following the Recommendation of the Organization of the International Legal Metrology 124 [7]. Our standard refractometer is a Mettler Toledo RE 50, with a 0,00001 resolution and which is calibrated through LGC certified solutions. This reference measuring instrument also has 0,01 ºC resolution thermostat that maintains the temperature at the reference value of 20,00 ºC for the measurements.

For three different refractive index nominal values, corresponding to potential alcoholic volumetric fractions of 6 %, 11 % and 16 %, linear regressions are determined from 30 pairs (x_i, y_i) published in the international Tables [4]. For each refractive index nominal value, the refractive indexes of five different glucose solutions are measured three times in the 0,4 mL cell of our reference refractometer, from three different takes of the same sample.

Following the model of equation (12) or of equation (14), depending on the adopted procedure, the estimation of the uncertainty associated to the linear model always includes the uncertainty $u(X_j)$. In addition to its repeatability uncertainty component, the latter has a component due to the refractometer resolution, a component due to the temperature effect and another due to the refractomer calibration. The resulting quadratic sum of all these components is then equal to the square of $u(X_j)$.

Finally, the Welch-Satterthwaite's relationship is used to deduce the effective degree of freedom of the standard uncertainty $u(Y_j)$, leading to the covering factor, k, through the inverse t Student's distribution. The multiplication of $u(Y_j)$ by k, almost equal to 2,00 for all the systems considered here, gives the expanded uncertainty $U(Y_j)$. The final result can then be written as: $Y_j \pm U(Y_j)$.

3. RESULTS

Using the two procedures for the estimation of the potential alcohol volumetric fractions from the international Tables and refractive index measurements, the results allow some comparisons. In the following, the refractive index, *n*, is considered as the X_j variable. The potential alcohol volumetric fraction, X_v, is then the Y_j variable, following the notation adopted previously in the communication.

First of all, for the three refractive index nominal values considered, the inequality (10) is always verified, justifying the use of the linear regression for the considered intervals of the pairs (x_i, y_i), i.e. the tabulated (n, X_v).

Then, all the measurements display a repeatability uncertainty equal to zero, as the standard deviation is zero. So, the comparisons between the uncertainties of the two procedures are not screened by the effect of this uncertainty component.

On Table 1, the results of the determination of the potential alcohol volumetric fractions by the two procedures presented in this communication are displayed.

Table 1. Refractive index (n), potential alcohol volumetric fraction determined by linear regression ($X_{v,\text{ reg.}}$ (%)), potential alcohol volumetric fraction determined by linear interpolation in interval ($X_{v,\text{ int.}}$ (%)) and the corresponding expanded uncertainty values.

n	$X_{v,\text{ reg.}}$ (%)	$X_{v,\text{ interv.}}$ (%)	$U(X_{v,reg.}(\%))$	$U(X_{v,int.}(\%))$
1,37586	16,243	16,250	0,004	0,002
1,37594	16,275	16,277	0,004	0,002
1,37599	16,295	16,294	0,004	0,002
1,37582	16,227	16,237	0,004	0,002
1,37600	16,299	16,297	0,004	0,002
1,36389	11,425	11,425	0,003	0,002
1,36382	11,397	11,396	0,003	0,002
1,36384	11,405	11,404	0,003	0,002
1,36387	11,417	11,416	0,003	0,002
1,36385	11,409	11,408	0,003	0,002
1,35104	6,198	6,203	0,002	0,001
1,35118	6,255	6,258	0,002	0,001
1,35119	6,259	6,261	0,002	0,001
1,35122	6,271	6,273	0,002	0,001
1,35116	6,247	6,250	0,002	0,001

Table 1 evidences that the two linear procedures are comparable as the corresponding measurement results are alike. Since the resolution of the refractometers for sugar contents of grape musts correspond at the maximum to only two figures after the decimal comma, this result is satisfactory. However, for measurements involving more accurate results, it is recommended to use the linear interpolation method, as the uncertainty value is smaller than to the one using the linear regression method.

Contrarily to the interpolation by intervals procedure, from the results of Table 1, it seems that the linear regression method depends more on the characteristic nominal value of the intervals [n_j, n_{j+1}]. This proves less robustness than the other method.

4. CONCLUSIONS

Two frequently used linear fitting procedures, the least square linear regression and the linear interpolation in intervals were used in the field of internationally accepted data of aqueous solutions refractometry. The respective formulas and criterions of the procedures were displayed as being easy to use with commercial softwares. This communication evidenced the likeness of the methods for low precision measurements. As the linear regression method has the advantage of being faster than the other method, for low precision and fast tasks, this method is preferred. However, in case of higher precision results, it is recommended to use the interpolation by intervals. An alternative to the linear regression model may be the non linear regression model that would give faster and more precise results.

REFERENCES

[1] ISO 8466-1:1990 (E) *Water quality – Calibration and evaluation of analytical methods and estimation of performance characteristics. Part 1: Statistical evaluation of the linear calibration function*.

[2] ISO 11095:1996 (E) *Linear calibration using reference materials*.

[3] O. Pellegrino, A. Furtado, L. Cortez, E. Filipe, "Cálculo da incerteza num função de medição linear; aplicação em refractometria das soluções líquidas", *3º Encontro Nacional da Sociedade Portuguesa de Metrologia*, Porto, 2008

[4] Jornal Oficial das Comunidades Europeias, L 272 de 3.10.1990, 1.

[5] BIPM, IEC, IFCC, IUPAC, IUPAP and OIML, *Guide to the expression of Uncertainty in Measurement (GUM)*, ISO, 1993 amended 1995.

[6] D.B. Hibbert, "The uncertainty of a result from a linear calibration", *Analyst*, vol. 131, nº. 12, pp. 1273-1278, 2006.

[7] OIML R 124 *Refractometers for the measurement of the sugar content of grape must*, 1997.