Demonstrating Linear Regression and Error Using an Experiment with a Microwave Oven

Chistiane Mendes Feitosa, José Machado Moita Neto,* and Graziella Ciaramella Moita

Departamento de Química, da Universidade Federal do Piauí, 64.049-550 Teresina, Piauí, Brasil, jmoita@uol.com.br

Abstract: To provide an application for the method of linear least squares to data collected in a laboratory, a beaker with water is heated in a microwave oven, and the water temperature is measured as a function of heating variables (time and oven setting). This procedure enables a student to obtain a regression line for each oven setting, and to evaluate the intercept and slope of this line and compare them with the initial temperature of the water and the heating versus oven setting relationship described in the microwave's manufacturer's manual. They also are asked to identify any sources of errors observed in this experiment.

Introduction

Although interest has grown about fitting nonlinear data [1, 2], linear models ($y = a + bx$) are still extensively used in chemistry to treat data obtained in a laboratory [3–5]. In these cases, a straight line that is best fitted to a set of experimental data can be obtained graphically (by visual inspection) or by using the method of linear least squares. A graphical method is less accurate than a numerical approach because visual location of the best-fitting straight line involves subjective evaluation. Ready access to pocket calculators now allows the use of a numerical method instead of a graphical one. The method of linear least squares uses a mathematical approach to produce the parameters a and b of the straight line. This allows evaluation of the standard errors (s_a and s_b) associated with the parameters, the standard deviation of the straight line ($s_{y/x}$), and, also, a rough evaluation of the proposed linear model itself through the coefficient of linear correlation (r). For evaluation with more criteria, one must use variance analysis [3]. Also, the method of linear least squares is strictly valid only when the independent variable, x, is devoid measurement errors. Other methods of linear regression exist for use when both variables contain errors.

To provide a direct application of the method of linear least squares to data collected in a laboratory, we introduced a simple experiment into the course *Statistics Applied to Chemistry* (high school level) where data are acquired for subsequent mathematical treatment. For this purpose, we used a domestic microwave oven, which offers the possibility to explore important aspects of collection and interpretation of data obtained through an experiment in which a beaker with water is heated with microwaves at several oven settings. In this system the temperature is expected to increase linearly with the amount of heat absorbed through incident microwave radiation, as the heat capacity of water is almost constant within the temperature range of interest.

The main objectives of this work are:

- to obtain the regression lines using the method of linear least squares for each oven setting used where the independent variable, x, is the period of heating and the dependent variable, y, is the temperature of the water;
- to evaluate the constant, a, and linear coefficient, b, and to compare them with the initial temperature of the water and the heating "intensities"[1] described in the user manual provided with the microwave oven;
- to identify possible sources of error in this experiment and examine their effects on the standard errors in the constant and linear coefficients; and
- to prepare the students to construct and use a calibration curve.

Pre-Experimental Activity

Before the execution of this experiment, the students of *Statistics Applied to Chemistry* are provided with information on the importance of applying the method of linear least squares to experimental data, particularly for the generation of a calibration curve. Through numerical data provided by the instructor, the students learn how to calculate the parameters a and b and thereby the equation of the best straight line, as well as the correlation coefficient, r, and the associated standard errors (equations 1–6; x_i represents the ith individual value of variable x, $\bar{x}$ is the mean value of variable x, y_i represents the ith individual value of variable y, $\bar{y}$ is the mean value of variable y, $\hat{y}_i$ is the individual value of variable y estimated from regression and x_i, and N is the number of points).

$$b = \frac{\sum_i \left\{ \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \right\}}{\sum_i \left(x_i - \bar{x}\right)^2} \qquad (1)$$

$$a = \bar{y} - b\bar{x} \qquad (2)$$

---

[1] We have used the term heating "intensity" to conform to the wording used in the manufacture's manual.

S1430-4171(99)01272-2, 10.1007/s00897990272a, 410016jn.pdf

Table 1. Temperature (in ^{o}C) of Water Obtained at Several Oven Settings

Time, s	Observed Water Temperature at Oven Setting, °C					
	high	medium	medium-low	defrost	low	warming
0	31.0	31.0	31.0	31.0	31.0	31.0
10	39.0	38.0	38.0	32.0	36.0	33.0
20	46.0	43.0	40.0	34.0	36.0	33.0
30	55.0	48.0	45.0	36.0	41.0	35.0
40	61.0	54.0	48.0	42.0	41.0	35.0
50	69.0	58.0	52.5	45.0	46.0	37.0
60	74.0	65.0	57.0	48.0	46.0	37.0

Table 2. Statistical Parameters Associated with the Regression Straight Lines for Each Oven Setting

Parameter	Oven Setting					
	high	medium	medium-low	defrost	low	warming
A, °C	31.7	31.8	32.1	29.1	32.1	31.5
Sa, °C	0.7	0.5	0.7	1.0	1.0	0.4
B, °C s^{-1}	0.7	0.5	0.4	0.3	0.3	0.1
Sb, °C s^{-1}	0.0	0.0	0.0	0.0	0.0	0.0
R	0.998	0.998	0.995	0.979	0.971	0.969
Sy/x	1.0	0.8	1.0	1.5	1.5	0.6

$$r = \frac{\sum_i \left\{ \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \right\}}{\left\{ \left[\sum_i \left(x_i - \bar{x}\right)^2 \right] \left[\sum_i \left(y_i - \bar{y}\right)^2 \right] \right\}^{1/2}} \qquad (3)$$

$$s_{y/x} = \left\{ \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{N - 2} \right\}^{1/2} \qquad (4)$$

$$s_b = \frac{s_{y/x}}{\left\{ \sum_i \left(x_i - \bar{x}\right)^2 \right\}^{1/2}} \qquad (5)$$

$$s_a = s_{y/x} \left\{ \frac{\sum_i x_i^2}{N \sum_i \left(x_i - \bar{x}\right)^2} \right\}^{1/2} \qquad (6)$$

The students work with this data in class using pocket calculators and later outside of class using commercial software such as Origin or SigmaPlot.

After this introduction, the student is expected to undertake the exercise described above for just one oven setting in order to better assimilate the calculation of the regression parameters, to associate the results obtained with their physical meaning, and to identify the experimental errors responsible for the observed discrepancies. These data are acquired experimentally.

The results are presented in the form of a report, and subsequently the detected errors are discussed during the classroom time.

Experimental

A beaker (250 mL) containing water (200 mL) at room temperature (31.0 °C) was heated for 10 s in a microwave oven at the high setting. The heating period was measured using the timer on the microwave oven. After this, the temperature of the water was immediately measured (precision ±0.5 °C) with the beaker outside of the microwave oven. This procedure was repeated for 20, 30, 40, 50, and 60 s, beginning with the water and beaker at room temperature. Data were gathered at medium, medium low, defrost, low, and warming settings.

Results and Discussion

Table 1 presents the measured temperature of water at several heating settings on the microwave oven over specific time intervals. Table 2 presents the main statistical parameters obtained from linear regression analysis using the data presented in Table 1 and equations 1–6.

We observe that there is a better correlation (r near 1) between temperature and period of heating for the high, medium, and medium-low oven settings. The constant, a, of a straight line is the value of y for which $x = 0$. In this case, it is the temperature of the water for which the duration of heating is zero, that is, the initial temperature of the water. In all experiments the initial temperature was 31 °C. The values obtained from the method of linear least squares are all greater than 31 °C, except for that obtained at the defrost setting of intensity. To evaluate whether errors responsible for this disagreement are merely random, we can determine the limit of confidence [3, 4, 6] using equation 7 where n is the number of degrees of freedom and $T_{ambient}$ is the ambient temperature.

$$T_{ambient} = a \pm t s_a; \; n = N - 2 \qquad (7)$$

For all heating settings, the initial temperature (31 °C) is contained within the confidence interval. Therefore, it can be affirmed with 95% confidence that the difference between the initial temperature and the value of a is due to random error.

The linear coefficient, b, reflects the various “intensities” of microwave radiation. This value can be compared with that

Table 3. Comparison of Experimental Results with Values Given by the Manufactuer

	high	medium	medium-low	defrost	low	warming
Manual	100	70	55	35	30	10
b × 100/bhigh	100	75	56	42	34	13

indicated in the manual of the oven [7], for which the relationship among the “intensities” is supplied, with the high setting taken as 100% (Table 3).

In all cases, the “intensity” obtained through the method of linear least squares starting from experimental results exceeds that indicated in the manual of the oven, and relative errors are larger for the lower oven settings. In this case, the confidence limit of *b* can also be calculated (95% confidence) using equation 8 in which *t* is from Student’s *t* test.

$$I = b \pm ts_b \quad n = N - 2 \tag{8}$$

The value of the relative “intensity” obtained from the manual (I = manual × b_{high}/100) is between the low and high limits; we thus conclude with 95% confidence that disagreement between values of *b* and *I* is due merely to random error.

Although we reached the conclusion that disagreement between the values of the initial temperature and the “intensity” of radiation indicated in the manual and the statistical results (*a* and *b*) are due only to random errors (with 95% confidence), there is a possibility of up to 5% that the true values are outside this confidence interval. Moreover, one must identify all sources of errors to obtain the best conditions to conduct the experiment and to decrease the confidence limit.

The detected errors were the following:

- lack of reproducibility of the water volume,
- location of the beaker on the plate of the microwave oven,
- nonuniformity of the beaker walls,
- precision of the thermometer,
- temperature gradient in the vertical and horizontal directions,
- inertia of the oven to production of radiation, and
- precision of the heating period.

Only the first five items can be remedied; the others depend not only on the equipment itself but also on a variation of electric voltage that occurs frequently and to a variable extent. The use of a voltage stabilizer decreases these problems but does not eliminate them.

Temperature Gradient. A primary source of error is from the location of the thermometer in the beaker, if the water was not stirred to ensure a homogeneous heat distribution.

To evaluate the error caused by the position of the thermometer within the volume of water, we performed an experiment in which the temperature was measured in three different vertical positions for three heating periods at the high setting. A difference of 6 to 8 °C was observed between the temperature measured in the lower part of the volume of water and that at its surface. The reason for this observation is that the top of the beaker is more exposed to the radiation than the bottom. Although the correlation coefficients obtained for the straight line obtained from measurements in these two positions are near unity, the coefficient obtained for the straight line related to measurements in the middle of the volume was more deviant from unity, because it is more difficult to reproduce this position of the thermometer.

In the same way, if the temperature were measured in several horizontal positions, the part of the volume nearest the border of the plate in motion would be expected to have a temperature greater than that closest to the center. This difference can be minimized by positioning the beaker at the center of the oven plate.

The temperature of water in the beaker is meaningful only if the temperature is made homogeneous, that is, if all the water attains the same temperature as result of stirring before measurement.

Conclusion

A simple experiment involving the heating of water contained in a beaker by a microwave oven allows not only an introduction to the calculation of parameters characteristic to the method of linear least squares, but also the use and interpretation of a confidence limit and the identification of sources of experimental error.

References and Notes

1. Zielinski, T. J.; Allendoerfer, R. D. *J. Chem. Educ*. **1997**, *74*, 1001.
2. Lieb, S. G. *J. Chem. Educ.* **1997**, *74*, 1008.
3. Miller, J. C.; Miller, J. N.; *Estadística para Química Analítica*; Addison-Wesley Iberoamericana: Wilmington, DE, 1993.
4. Christian, G. D. *Analytical Chemistry,* 5th ed.; Wiley: New York, 1994; Chapter 2.
5. Miller, J. C.; Miller, J. N.; *Analyst* **1988**, *113*, 1351.
6. Miller, J. C.; Miller, J. N.; *Analyst* **1991**, *116*, 3.
7. Instruction manual accompanying the Panasonic NN-7209B microwave oven.