edited by
Steven D. Gammon
Western Washington University
Bellingham, WA 98225-9150

Multiple Linear Regression Using a Graphing Calculator

Applications in Biochemistry and Physical Chemistry

W

Seán P. Madden
Department of Mathematics, Frontier Academy Middle and High School, Greeley, CO 80639

Wayne Wilson*
Student Health Service Pharmacy, Humboldt State University, Arcata, CA 95521; **waynewilson@sprintmail.com*

Aichun Dong
Department of Chemistry and Biochemistry, University of Northern Colorado, Greeley, CO 80639

Lynn Geiger
Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215

Christopher J. Mecklin
Department of Mathematics and Applied Statistics, Murray State University, Murray, KY 42071

The use of computers, spreadsheets, and statistical software in the chemistry curriculum is well established. The ready availability of graphing calculators has been largely ignored. We intend to demonstrate that the graphing calculator can greatly supplement the computer and often eliminates the need for computers and software. In many instances these devices can perform the calculations and manipulation of data that many people assume can only be performed with a desktop computer. The graphing calculator is in fact a handheld computer, more powerful than the most sophisticated computers of 25 years ago. Programs may be written for these devices or downloaded off the Internet, and data collection probes and devices are available to use with these calculators.

A recent survey of high school and undergraduate math texts reveals that graphing calculators have become required equipment in the math classroom *(1–3).* This suggests that these handheld computers are readily available for use in the chemistry classroom and laboratory.

We suggest that the availability, portability, and low cost of these devices should motivate chemical educators to incorporate their use to enhance learning. Recent conferences on chemical education have included sections on the use of graphing calculators in the chemistry curriculum *(4–11).* Articles published in this *Journal* have demonstrated the usefulness of graphing calculators in solving chemical equilibria problems and calculating units, collecting and analyzing data, as well as comparing the capabilities of various calculators *(12–15).* We have used these calculators to simulate and model chemical processes, analyze complex data, and explore the effects of changing parameters on process outcomes. In this article we demonstrate the power of the graphing calculator in complex data analysis.

The idea of performing multiple linear regression (MLR) on the graphing calculator originated when one of the authors (Madden) was working with a group of undergraduate students at the University of Northern Colorado trying to analyze IR spectroscopy data of CO vibrations. The computer facilities were not always available to students, especially after 10 p.m. when most students did their homework. A statistics instructor (Mecklin) was consulted and the method of matrix algebra to perform MLR was explained. It was quickly noted that the same method described for the spreadsheet could be executed on the graphing calculator. The method was described to the undergraduate students, who felt empowered by use of their own handheld graphing calculator to do complex data analysis without having to wait for time in the computer laboratory.

MLR is one of the most powerful methods of data analysis. In this article we explore two uses of MLR using data from current research. These examples demonstrate just two of the many uses of MLR: determining the relative contribution of structural properties of amino acids to the formation of beta sheets in proteins and predicting the properties of a molecule using parameters derived from IR spectroscopy.

MLR is a mathematical technique that can be applied to a variety of problems in statistics, chemistry, business, and many other fields. For those of us who have competed with our peers for admission to college or graduate school, the admissions office may have used the following scenario: an admissions officer at a college is faced with selecting the most qualified candidates for the available positions. The data she has to work with include an applicant's grade point average, Graduate Record Examination (GRE) scores, and letters of recommendation. The task of selecting the correct candidates in an unbiased manner is a challenging one. What if a mathematical technique could be applied to the data to facilitate this decision? Such a technique exists in the form of MLR. If we knew what magnitude of weight to assign to each of the pieces of data before adding them together, the total scores would be a measure of the best candidates. The higher the score, the better the candidate. This is of course only one of many possible applications of MLR.

MLR and Matrices: Mathematical Concepts

Consider an equation of the following form *(16)*:

$$y = m_1x_1 + m_2x_2 + m_3x_3 + \ldots + m_nx_n + b \quad (1)$$

The dependent variable, *y*, is linearly related to each of the independent variables, *x*. Each independent variable is linearly related to the dependent variable because the power on each *x* variable is 1. In chemistry, one often collects data on the dependent variable and a *variety* of independent variables to establish the relationship between dependent and independent variables. Each of the independent variables will have its own slope, *m*, associated with it. The entire equation, however, will have only one intercept, *b*. The goal, from a statistical point of view, is to find the slope associated with each variable and the intercept such that the resultant line does the best job of describing the relationship between *x* and *y*.

The matrix approach is best to solve these kinds of problems. In matrix notation, we can represent eq 1 in the form of the general linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

where **y** is a vector (one-column matrix) containing the data corresponding to the dependent variable, **X** is a matrix containing the data corresponding to the independent variables, $\boldsymbol{\beta}$ is the vector containing the *parameters* (or the regression coefficients) corresponding to the intercept and the slopes, and $\boldsymbol{\varepsilon}$ is the vector containing the residuals for each of our independent variables. To solve for the parameter vector, we will utilize the least-squares principle to minimize the sum of the squared residuals. The resulting equation (in matrix form) to solve for **b**, the estimate of the parameter vector $\boldsymbol{\beta}$, is:

$$\mathbf{b} = \left[\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}}\right] \cdot \mathbf{y} \quad (3)$$

where X^T is the transpose of the X matrix *(16)*.

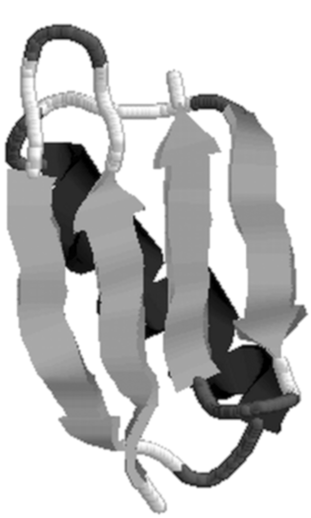


Figure 1. The Immunoglobulin binding domain of Protein G from *Streptomyces griseus*. Arrows represent beta-sheets. PDB code: 2GB1.

What Is MLR Used For?

Modern drug discovery relies heavily on the use of MLR. After a promising molecule is developed, it must be tested in vivo. Drug testing is the most expensive step of the development process and any method that can reduce the number of molecules tested is of great value. Determining which molecular structure will best reach the desired site of action, not accumulate in the body, avoid interacting with other molecules, be the most potent, cause the least toxicity, and achieve the desired result is a prolonged and daunting process. Prior to the 1970s, drug development was primarily a matter of trial and error. Beginning in the 1980s the method of QSAR (Quantitative Structure–Activity Relationships) was developed. With this method the physicochemical attributes of a molecule are quantified and an attempt is made to statistically correlate these attributes with desired biological activity *(17)*. For example, the lipophilicity of a drug determines how it will be distributed in the body. The stereochemistry may determine activity or toxicity. Molecular weight, electrical charge, density, and refractive index of a molecule may also be predictive of its potential. Using MLR, multiple properties can be analyzed simultaneously and guide the synthesis of the most promising molecules *(17, 18)*.

Example 1: Proteins Structure

Chemical Problem

Niwa and Ogino, using data from several studies, performed a multiple regression analysis of the physicochemical properties of 19 amino acids to determine which, if any, of the properties were important to beta sheet formation in proteins (Figure 1). The physicochemical properties producing

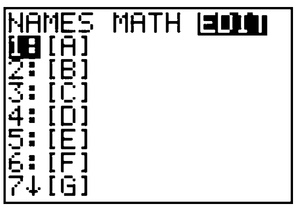
Table 1. Data Used in Matrix A

Amino Acid	$\Delta\Delta G$/ (kcal mol^{-1})	B1	B5
Ala	-0.35	1.52	2.04
Arg	-0.44	1.52	6.24
Asn	-0.38	1.52	4.37
Asp	-0.41	1.52	3.78
Cys	-0.47	1.52	3.41
Gln	-0.40	1.52	3.53
Glu	-0.41	1.52	3.31
Gly	0.00	1.00	1.00
His	-0.46	1.52	5.66
Ile	-0.56	1.90	3.49
Leu	-0.48	1.52	4.45
Lys	-0.41	1.52	4.87
Met	-0.46	1.52	4.80
Phe	-0.55	1.52	6.02
Ser	-0.39	1.52	2.70
Thr	-0.48	1.73	3.17
Trp	-0.48	1.52	5.90
Tyr	-0.50	1.52	6.72
Val	-0.53	1.90	3.17

a small value of ΔΔ*G* were considered important to beta sheet formation *(19)*. Two of Niwa and Ogino's properties, Sterimol parameters reflecting the size and bulk of amino acids (referred to as B1 and B5 in this article) were used in the graphing calculator (TI-83) MLR analysis described below.

MLR Analysis

Using B1 and B5 as the independent variables construct matrix **A** (shown as [A] on the screen shots) as a 19 × 3 matrix. The first column of the matrix should be all ones to make the matrix conform to the regression equation. First, enter the matrix option by pressing **2nd, MATRIX,** then choose "EDIT",



(i)

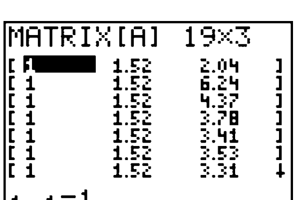
select matrix **A** by placing the cursor on "1:" and press **ENTER**. The screen should now look like:

(ii)

Matrix **A** will be the matrix containing the *x*-variable data (B1 and B5). Use the arrow keys and assign the dimensions of 19 × 3, then press **ENTER**. The screen should now look like:

(iii)

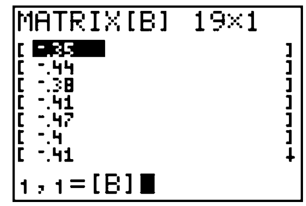
Press the **ENTER** key and enter the data into the matrix (the third and fourth columns of Table 1). Ones should also be added in the first column to ensure the matrix is conformable with the parameter vector. After entering the data the screen will appear as:



(iv)

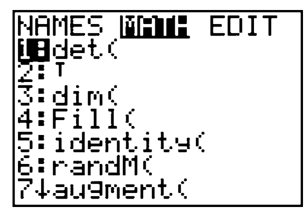
Note that not all of the matrix is visible on the screen. The cursor position is indicated in the lower left of the screen, along with the data value for that position.

Once the independent variables have been inputted into matrix **A**, press **2nd, QUIT** returning to a blank screen referred to as the home screen. Construct Matrix **B** as a 19 × 1 matrix using the dependent variable ΔΔ*G*.



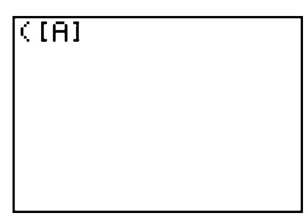
(v)

Next, press **2nd, QUIT**, returning to the home screen. Here the matrix algebra is performed. Press **2nd, MATRIX** and using the options under the MATH key menu



(vi)

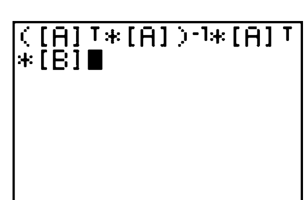
construct an equation of the same form as eq 3. (Note: do not press **ENTER**, or "det (" will be selected.) Press **2nd, QUIT** to return to the home screen and enter "(". Then press **2nd, MATRIX, ENTER** and the **A** matrix will be selected. The screen should now appear as:



(vii)

Please note that you do not type in each character, that is, [, A, and]: the matrix is selected and entered via the matrix screen *(20)*.

Continue entering the matrix equation. The transpose character, ^{T}, is accessed through the matrix math screen (screen vi), **2nd, MATRIX**, highlight "MATH", and press "2". The final equation should appear as:



(viii)

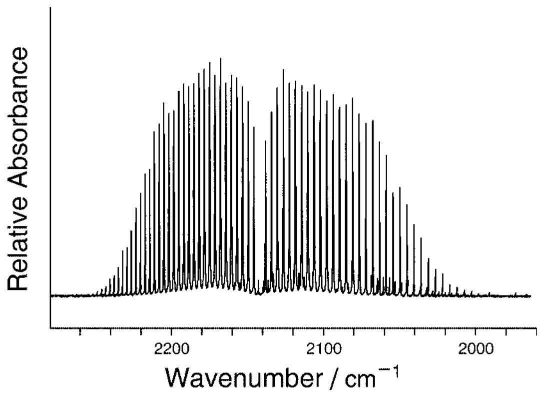
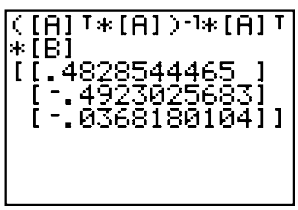


Figure 2. IR spectrum of carbon monoxide.

Table 2. IR Absorbance Data from Carbon Monoxide

J''	R Branch Absorbance/cm^{-1}
0	2147.084
1	2150.858
2	2154.599
3	2158.301
4	2161.971
5	2165.602
6	2169.200
7	2172.759

Now press **ENTER** and the result will be displayed as:



(ix)

The first result, 0.483, reflects the *y* intercept of the equation, the second result, −0.492, is the slope of the independent variable B1, and the third result, −0.0368, is the slope of the independent variable B5. These values agree exactly with the literature values *(19).* The equation for predicting $\Delta\Delta G$ from the B1 and B5 structural descriptors is

$$\Delta\Delta G = -0.492\,\text{B1} - 0.037\,\text{B5} + 0.483 \qquad (4)$$

Statistical analyses of the reliability of the regression equation are usually performed. For our purposes, we wish to simply present the method of MLR.

Example 2: Good Vibrations—Molecular Style

Chemical Problem

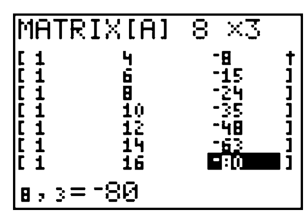
The following example from a college-level physical chemistry course also illustrates how to use the MLR technique described above *(21).* Transitions between various rotational and vibrational states in carbon monoxide occur at discrete frequencies as the sample absorbs infrared (IR) radiation. An IR spectrometer can be used to measure the frequencies at which these vibrational and rotational transitions take place (Figure 2). The resulting spectrum is a pattern of absorbance peaks *(22).* Each IR peak corresponds to the energy required for a specific transition. A typical set of data is shown in Table 2. The relationship between wavenumber of the absorbance, ν_R, measured in cm^{-1} and the band origin (ν_0), the rotational constant (B_e), and the rotation–vibration interaction constant α_e can be described as

$$\nu_R = \nu_0 + 2(J'' + 1)B_e - (J'' + 1)(J'' + 3)\alpha_e \qquad (5)$$

The values of the coefficients (the values in parentheses) are shown in Table 3. These data can be entered into the TI-83 graphing calculator as matrices and values for ν_0, B_e, and α_e can be obtained as described below.

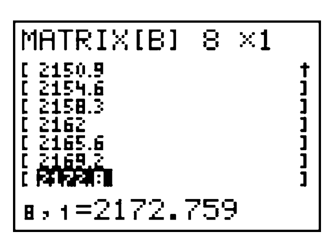
MLR Analysis

Create an 8 × 3 matrix as described in the previous example. This will be the *x*-variable data matrix. The first column will be all ones to conform to the parameter vector. The second and third columns of Table 3 are entered into the matrix as:



(x)

After entering the *x*-variable data, press the **2nd** key, move to matrix **B**, and establish its dimensions as 8 × 1. Press **ENTER** and enter the *y*-variable data, the R-branch absorbance data in Table 2.



(xi)

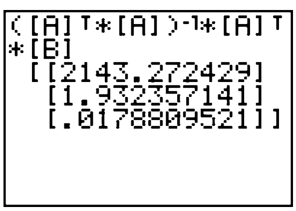
Note that the numbers are rounded off to one decimal place in the display. However, the entire number entered is retained in memory and used in calculations, as shown by the highlighted number: the full value displayed at the bottom of the screen.

Return to the home screen by pressing **2nd, QUIT**. Here we will perform the matrix algebra. The regression equation will take the same form as screen viii. If the equation is still

Table 3. Calculated Values for the Coefficients in Eq 5

J''	$2(J''+1)$	$(J''+1)(J''+3)$
0	2	3
1	4	8
2	6	15
3	8	24
4	10	35
5	12	48
6	14	63
7	16	80

on the home screen, it can be "recycled" with the new data. The final result should appear as:



(xii)

Press the **ENTER** key and the calculator will return the parameter vector, which contains, in order, the intercept, the first slope, and the second slope. In this example, these parameters correspond to ν_0, B_e, and α_e These values match perfectly with results obtained using a spreadsheet and agree with the literature values for these constants (*22,* p 835). Using the spectroscopic constants obtained from this procedure other interesting values, such as the interatomic distance, can be obtained *(21).*

Conclusion

With a little practice this method of finding the MLR for a data set can become quite routine. The general form of the matrix solution can be saved in memory for ready use with other problems. Thus, the graphing calculator provides a convenient alternative to the usual method of solving these problems using a computer spreadsheet program.

Acknowledgment

We would like to acknowledge the advice and assistance of Richard Schwenz in the prepartion of this article.

ᵂSupplemental Material

Another example to apply the MLR method is available in this issue of *JCE Online*. This example is also worked out using a Casio CFX 9850 calculator. Some information is included about using menus on the TI-89 calculator.

Literature Cited

1. Dugoploski, M. *Elementary Algebra,* 3rd ed.; McGraw-Hill Higher Education: New York, 2000; p 500.
2. Bittinger, M. L.; Ellenbogen, D. J.; Johnson, B. L. *Intermediate Algebra Graphs and Models,* 1st ed.; Addison-Wesley: Reading, MA, 1999; p 698.
3. Anton, H. *Calculus: A New Horizon,* 6th ed.; John Wiley & Sons: New York, 1998; p 1312.
4. Iriate-Gross, J.; Grissom, D. Chemistry Beyond Chemistry: Introducing New Technology-Based Labs for the Science Major. In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
5. Madden, S. P.; Wilson, W. I Love Lucy, Tadpoles to Pollywogs, and Good Vibrations: Brief Tour of Graphing Calculator Applications in Chemistry. In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
6. Sconzo, P. Investigating Ideal and Non-Ideal Gases. In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
7. Madden, S. P.; Willis, S.; Milne, R.; Jones, L.; Dong, A. The Use of Multiple Representations and Graphing Calculators in the Undergraduate Biochemistry Lab. In *15th Biennial Conference on Chemical Education;* University of Waterloo: Ontario, Canada, 1998.
8. Sinex, S. Technology as a Tool to Enhance Inquiry Science: General Chemistry Activities with Graphical Analysis. In *15th Biennial Conference on Chemical Education;* University of Waterloo: Ontario, Canada, 1998.
9. DeMauro, E. Is the Quadratic Going the Way of the Slide Rule? In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
10. Eaker, C. Rate Laws, Rate Constants, and Activation Energies with the CBL. In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
11. Gendreau, H. Math 0: Using Equation Solver to Calculate LeChatalier Shifts. In *17th Biennial Conference on Chemical Education;* Western Washington University: Bellingham, WA, 2002.
12. Cortes-Figueroa, J. E.; Moore, D. A. *J. Chem. Educ.* **1999**, *76,* 635.
13. Morgan, M. E. *J. Chem. Educ.* **1999**, *76,* 631.
14. Kim, M.-H.; Ly, S.-Y.; Hong, T.-K. *J. Chem. Educ.* **2000**, *77,* 1367.
15. Donato, H., Jr. *J. Chem. Educ.* **1999**, *76,* 632.
16. Myers, R. H.; Milton, J. S. A First Course in the Theory of Linear Statistical Models. In *The Duxbury Advanced Series in Statistics and Decision Sciences;* PWS-Kent Pub. Co.: Boston, MA, 1991; Vol. X, p 342.
17. Foye, W. O.; Lemke, T. L.; Williams, D. A. *Principles of Medicinal Chemistry,* 4th ed.; Williams & Wilkins: Baltimore, MD, 1995; Vol. XII, p 995.
18. *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 9th ed.; Goodman, L. S., Hardman, J. G., Limbird, L. E., Eds.: McGraw-Hill Health Professions Division: New York, 1996; Vol. XXI, p 1905
19. Niwa, T. S.; Ogino, A. *J. Mol. Struc. (Theochem)* **1997**, *419,* 155–160.
20. *TI-83 Plus Graphing Calculator Guidebook.* Texas Instruments Incorporated: Dallas, TX, 1999.
21. Schwenz, R. W.; Polik, W. F. *J. Chem. Educ.* **1999**, *76,* 1302.
22. Noggle, J. H. *Physical Chemistry,* 3rd ed.; HarperCollins College Publishers: New York, 1996; Vol. XII, p 1108.