Original Article

Comparative Study of Three Multiple Linear Regression Software Programs
การศึกษาเปรียบเทียบโปรแกรมการถดถอยแบบหลายตัวแปรเชิงเส้นสามโปรแกรม

Bodin Tuesuwan[1], Nisa Phutong[2], Mitr Pathipvanich[3]*
บดินทร์ ติวสุวรรณ[1], นิสา ภู่ทอง[2], มิตร ปทีปวณิช[3]*

Received: 7 April 2009 **Accepted**: 30 June 2009

Abstract

Two commercial and one free online multiple linear analysis programs were tested with simulated potentiometric titration data for accuracy in determination of equivalent point, dissociation constant and partition coefficient. In many cases, analysis of simulated data yielded inaccurate results, possibly due to programming algorithms. This study highlights a need for regression programs to be tested with simulated data to assess their limitations.

Keywords: multiple linear regression, dissociation constant, partition coefficient, equivalent point
IJPS 2009; 5(2): 131-139

บทคัดย่อ

โปรแกรมวิเคราะห์ความถดถอยแบบหลายตัวแปรเชิงเส้นที่มีการจำหน่ายเชิงพาณิชย์สองโปรแกรมและหนึ่งโปรแกรมทางอินเตอร์เนตที่ไม่มีค่าใช้จ่ายถูกนำมาใช้ทดสอบความถูกต้องด้วยข้อมูลจำลองโพเทนชิโอเมทริก-ไทเทรชันในการหาค่าจุดสมมูล ค่าคงที่การแตกตัว และค่าสัมประสิทธิ์การแบ่งส่วน ในหลายกรณีการวิเคราะห์ข้อมูลจำลองให้ผลไม่ถูกต้อง ซึ่งอาจจะเป็นผลมาจากการกำหนดลำดับขั้นตอนของคำสั่งและ/หรือข้อจำกัดของโปรแกรมที่ต่างกัน ในการศึกษานี้แสดงให้เห็นถึงความจำเป็นของการทดสอบโปรแกรมด้วยข้อมูลจำลองเพื่อทดสอบข้อจำกัด

คำสำคัญ: วิเคราะห์ความถดถอยแบบหลายตัวแปรเชิงเส้น ค่าคงที่การแตกตัว ค่าสัมประสิทธิ์การแบ่งส่วน ค่าจุดสมมูล
วารสารเภสัชศาสตร์อีสาน 2552; 5(2): 131-139

[1] Ph.D., Lecturer, [2] M.Pharm., [3] Ph.D., Assistant Professor, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok

[1] Ph.D. อาจารย์ [2] ภ.ม. [3] Ph.D. ผู้ช่วยศาสตราจารย์ คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร

* Corresponding author: Mitr Pathipvanich, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330 Tel. +66 2218 8310 E-mail: m.pathip@gmail.com

Introduction

In the course of scientific investigation, we often must make use of commercial linear regression programs to help us find the relationship between a criterion variable and predictor variables (Billo, 2001; Keeling and Pavur, 2007; Neter, 1983; Ryan, 1997). Most of the commercial linear regression programs such as those available with Excel® and Minitab® have been widely employed with satisfactory degree of accuracy in field of social sciences and business application. However, in the field of pharmaceutical sciences, we routinely encounter much more challenging data sets than that of the business and social sciences. In pharmaceutical quantitative analysis, concentrations data between 10^{-2} to 10^{-12} are the norm. We often take it for granted that all commercial programs will give us the correct calculated results, forgetting the fact that many of these programs may not have been written for application with numerical values as small as what we normally observed in pharmaceutical analysis.

With numerous commercial regression programs available in the market, it would be impossible for us to compare them all. Therefore, we will focus on only two widely used commercial programs: Excel® 2007, Minitab® 15 and a free software program, Wessa® 2009. It is a common acceptance that the calculated results of simple linear regressions with these three programs are virtually indifferent and thus we will limit the scope of this paper only in area of multiple linear regression. In this study, we will use simulated potentiometric titration data to test the multiple linear regression programs of these softwares.

Methods

Derivation of Multiple Linear Equation

Application of simple linear regression has been employed by various researchers to find aqueous dissociation constant and equivalent point of titration (Gran, 1952; Pathipvanich, 1990; Rossatti, 1965). Theoretically, it is also possible to simultaneously determine partition coefficient, aqueous dissociation constant and equivalent point of titration from a single titration experiment with the aid of multiple linear regression program.

To test the accuracy of Excel® 2007, Minitab® 15 and Wessa® 2009 multiple linear programs, we must simulate titration data to be analyzed by all three programs. We will begin with theoretical model of titrating neutral weak monoprotic acid in octanol/aqueous solvent system. In the dual-phase solvent system, at equilibrium, the distribution of the weak acid is

$$HA_o \rightleftharpoons HA \rightleftharpoons H^+ + A^-$$

Oil phase Aqueous phase

Assuming only the unionized weak acid can partition into octanol phase, the aqueous dissociation constant and partition coefficient are following:

$$K_a = \frac{[A^-][H_3O^+]}{[HA]_a} \quad (1)$$

$$P = \frac{[HA]_o}{[HA]_a} \quad (2)$$

If we start out by making stock solution of the weak acid in aqueous solvent and then mix volume V_a of this stock solution with volume V_o of octanol, the mass balance for the neutral weak acid is

$$[HA]_i V_a = V_e N = [HA]_o V_o + [HA]_a (V_a + V) + [A^-](V_a + V) \quad (3)$$

$[HA]_i$ = initial concentration of HA (stock solution), $[HA]_o$ = concentration of HA in octanol, $[HA]_a$ = concentration of HA remained in aqueous phase, V_o = volume of octanol, V_a = initial volume of aqueous, V_e = equivalent titrant volume, V is the cumulative volume of titrant added and N is the normality of the sodium hydroxide titrant.

Substituting equation 2 into Equation 3 and allowing $R = V_o/V_a$, follow by rearranging, we have

$$[HA]_a = \frac{V_eN - [A^-](V_a+V)}{\{PRV_a + (V_a+V)\}} \quad (4)$$

For the titration of weak acid with sodium hydroxide, the charge balance of the system is following:

$$[Na^+] + [H_3O^+] = [A^-] + [OH^-] \quad (5)$$

Substituting $[Na^+] = VN/(V_a + V)$ into equation 5, follow by rearranging, we have

$$[A^-] = \frac{VN}{(V + V_a)} + [H_3O^+] - [OH^-] \quad (6)$$

From equation 1, 4 and 6, we get

$$Ka = \frac{\left\{\frac{VN + \{[H_3O^+] - [OH]\}(V_a+V)}{(V + V_a)}\right\}[H_3O^+]}{V_eN - \left\{\frac{VN + \{[H_3O^+] - [OH^-]\}(V_a+V)}{PRV_a + (V_a+V)}\right\}} \quad (7)$$

Letting $G = VN + ([H_3O+] - [OH^-])(V_a+V)$, equation 7 can be rearranged into the form of multiple linear equation $(Y = a_o + a_1X_1 + a_2X_2)$.

$$\frac{GV[H_3O^+]}{(V_a+V)} = K_aV_eN - K_aG - \frac{(PR+1)\,GV_a[H_3O^+]}{(V_a+V)} \quad (8)$$

With $Y = GV[H_3O^+]/(V_a+V)$, $X_1 = G$, $X_2 = GV_a[H_3O^+]/(V_a+V)$, $a_o = K_aV_eN$, $a_1 = -K_a$ and $a_2 = -(PR+1)$. The equivalent point, dissociation constant and partition coefficient can be determined from a_o, a_1 and a_2.

Simulation of potentiometric titration data

By rearranging Equation 7, we can obtain

$$\{V_a(1 + PR) + V\}[H_3O^+]^3 + \left\{\frac{VN\{V_a(1+PR) + V\}}{V_a + V} + K_a(V_a+V)\right\}[H_3O^+]^2 + \left\{K_aVN - K_aV_eN - K_w\{V_a(1+PR) + V\}\right\}[H_3O^+] - K_wK_a(V_a + V) = 0 \quad (9)$$

Equation 9 is used to calculate $[H_3O^+]$ as a function of volume of titrant V for any given set of values for N, K_a, V_a, V_o, P and [HA]i. For our simulation, we fix N = 0.10000 normal, $pK_w = 14$, $V_a = 50$ mL, $[HA]_i = 0.010000$ molar and the volume of titrant was added in 0.100 mL increment, starting from V = 0 to V = 4.900 mL. The resulting data of $[H_3O^+]$ and volume of titrant V were then used to calculate values of Y, X_1 and X_2 in Equation 8 for each volume of titrant added, follow by analysis with Excel® 2007, Minitab® 15 and Wessa® 2009 multiple linear regression program to determine Ve, P and Ka.

Results and Discussion

In the cases of partition coefficients of 1 or less, Excel® 2007 multiple linear regression works well only for the cases with pK_a of 4 or lower. For these cases, Excel® 2007's computations yield the theoretical values for K_a, V_e and P. As P increases to 10, only $pK_a = 2$ and $pK_a = 3$ yield the theoretical values. With P = 100, only the case of $V_o = 5$ mL and $pK_a = 2$ gives the correct values. In all the cases which Excel® 2007 fails to compute the correct values for K_a, V_e and P, the X_2 data was rejected. Consequently Excel® 2007 determines the value of $a_2 = 0$ and apparently proceeds to fit the X_1 data to simple linear regression resulting in negative calculated values for partition coefficients as well as negative values for K_a and V_e (Table 1).

We thought of three possible explanations as to why Excel® 2007 would reject X_2 and proceed to employ simple linear regression between Y and X_1: (1) very low numerical values of X_2 as compared to X_1 such that if X_2/X_1 falls below certain value, X_2 would be rejected; (2) X_2 would be rejected if the numerical value of X_2 falls below certain threshold (simulated values of X_2 range from 10^{-3} to 10^{-14}); (3) X_2 was rejected because the correlation between Y and X_2 is lower than Y and X_1 or if variance inflation factor (VIF) between X_1 and X_2 exceed certain value.

Table 1 Summary of multiple linear regression results

V_o*	Theoretical		Excel® 2007			Wessa® 2009			MINITAB® 15		
	P	pK_a	V_e*	K_a	P	V_e*	K_a	P	V_e*	K_a	P
5	0.001	2	5.000	$1.000x10^{-2}$	0.001	5.000	$1.000x10^{-2}$	0.001	5.000	$1.000x10^{-2}$	0.001
		3	5.000	$1.000x10^{-3}$	0.001	5.000	$1.000x10^{-3}$	0.001	5.000	$1.000x10^{-3}$	0.001
		4	5.000	$1.000x10^{-4}$	0.001	5.000	$1.000x10^{-4}$	0.001	5.000	$1.000x10^{-4}$	0.001
		5	-95.1	$-8.109x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.001	5.000	$1.000x10^{-5}$	0.001
		6	-81.0	$-9.509x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.001	5.000	$1.000x10^{-6}$	0.001
		7	-76.6	$-1.004x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.001	5.000	$1.000x10^{-7}$	0.001
		8	-66.0	$-1.161x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.001	5.000	$1.000x10^{-8}$	0.001
		10	-8.63	$-7.851x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.001	5.000	$1.000x10^{-10}$	0.001
10	0.001	2	5.000	$1.000x10^{-2}$	0.001	5.000	$1.000x10^{-2}$	0.001	5.000	$1.000x10^{-2}$	0.001
		3	5.000	$1.000x10^{-3}$	0.001	5.000	$1.000x10^{-3}$	0.001	5.000	$1.000x10^{-3}$	0.001
		4	5.000	$1.000x10^{-4}$	0.001	5.000	$1.000x10^{-4}$	0.001	5.000	$1.000x10^{-4}$	0.001
		5	-95.1	$-8.109x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.001	5.000	$1.000x10^{-5}$	0.001
		6	-81.0	$-9.509x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.001	5.000	$1.000x10^{-6}$	0.001
		7	-76.6	$-1.004x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.001	5.000	$1.000x10^{-7}$	0.001
		8	-66.0	$-1.161x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.001	5.000	$1.000x10^{-8}$	0.001
		10	-8.63	$-7.851x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.001	5.000	$1.000x10^{-10}$	0.001
5	0.010	2	5.000	$1.000x10^{-2}$	0.010	5.000	$1.000x10^{-2}$	0.010	5.000	$1.000x10^{-2}$	0.010
		3	5.000	$1.000x10^{-3}$	0.010	5.000	$1.000x10^{-3}$	0.010	5.000	$1.000x10^{-3}$	0.010
		4	5.000	$1.000x10^{-4}$	0.010	5.000	$1.000x10^{-4}$	0.010	5.000	$1.000x10^{-4}$	0.010
		5	-95.0	$-8.111x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.010	5.000	$1.000x10^{-5}$	0.010
		6	-80.9	$-9.509x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.010	5.000	$1.000x10^{-6}$	0.010
		7	-76.5	$-1.004x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.010	5.000	$1.000x10^{-7}$	0.010
		8	-66.0	$-1.161x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.010	5.000	$1.000x10^{-8}$	0.010
		10	-8.63	$-7.849x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.010	5.000	$1.000x10^{-10}$	0.010
10	0.010	2	5.000	$1.000x10^{-2}$	0.010	5.000	$1.000x10^{-2}$	0.010	5.000	$1.000x10^{-2}$	0.010
		3	5.000	$1.000x10^{-3}$	0.010	5.000	$1.000x10^{-3}$	0.010	5.000	$1.000x10^{-3}$	0.010
		4	5.000	$1.000x10^{-4}$	0.010	5.000	$1.000x10^{-4}$	0.010	5.000	$1.000x10^{-4}$	0.010
		5	-94.9	$-8.113x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.010	5.000	$1.000x10^{-5}$	0.010
		6	-80.8	$-9.509x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.010	5.000	$1.000x10^{-6}$	0.010
		7	-76.5	$-1.004x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.010	5.000	$1.000x10^{-7}$	0.010
		8	-65.9	$-1.160x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.010	5.000	$1.000x10^{-8}$	0.010
		10	-8.62	$-7.847x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.010	5.000	$1.000x10^{-10}$	0.010
5	0.100	2	5.000	$1.000x10^{-2}$	0.100	5.000	$1.000x10^{-2}$	0.100	5.000	$1.000x10^{-2}$	0.100
		3	5.000	$1.000x10^{-3}$	0.100	5.000	$1.000x10^{-3}$	0.100	5.000	$1.000x10^{-3}$	0.100
		4	5.000	$1.000x10^{-4}$	0.100	5.000	$1.000x10^{-4}$	0.100	5.000	$1.000x10^{-4}$	0.100
		5	-94.0	$-8.128x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.100	5.000	$1.000x10^{-5}$	0.100
		6	-80.2	$-9.508x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.100	5.000	$1.000x10^{-6}$	0.100
		7	-75.9	$-1.003x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.100	5.000	$1.000x10^{-7}$	0.100
		8	-65.5	$-1.160x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.100	5.000	$1.000x10^{-8}$	0.100
		10	-8.56	$-7.831x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.100	5.000	$1.000x10^{-10}$	0.100

* all volumes are in mL with theoretical V_e = 5.000 mL,** negative values for P due to a_2 = 0, *** Wessa 2009 would not compute.

Table 1 (continued)

V_o^*	Theoretical		Excel® 2007			Wessa® 2009			MINITAB® 15		
	P	pK_a	V_e^*	K_a	P	V_e^*	K_a	P	V_e^*	K_a	P
10	0.100	2	5.000	$1.000x10^{-2}$	0.100	5.000	$1.000x10^{-2}$	0.100	5.000	$1.000x10^{-2}$	0.100
		3	5.000	$1.000x10^{-3}$	0.100	5.000	$1.000x10^{-3}$	0.100	5.000	$1.000x10^{-3}$	0.100
		4	5.000	$1.000x10^{-4}$	0.100	5.000	$1.000x10^{-4}$	0.100	5.000	$1.000x10^{-4}$	0.100
		5	-92.9	$-8.145x10^{-9}$	**	5.000	$1.000x10^{-5}$	0.100	5.000	$1.000x10^{-5}$	0.100
		6	-79.5	$-9.505x10^{-10}$	**	5.000	$1.000x10^{-6}$	0.100	5.000	$1.000x10^{-6}$	0.100
		7	-75.2	$-1.003x10^{-10}$	**	5.000	$1.000x10^{-7}$	0.100	5.000	$1.000x10^{-7}$	0.100
		8	-64.9	$-1.159x10^{-11}$	**	5.000	$1.000x10^{-8}$	0.100	5.000	$1.000x10^{-8}$	0.100
		10	-8.50	$-7.812x10^{-13}$	**	5.000	$1.000x10^{-10}$	0.100	5.000	$1.000x10^{-10}$	0.100
5	1.00	2	5.000	$1.000x10^{-2}$	1.00	5.000	$1.000x10^{-2}$	1.00	5.000	$1.000x10^{-2}$	1.00
		3	5.000	$1.000x10^{-3}$	1.00	5.000	$1.000x10^{-3}$	1.00	5.000	$1.000x10^{-3}$	1.00
		4	5.000	$1.000x10^{-4}$	1.00	5.000	$1.000x10^{-4}$	1.00	5.000	$1.000x10^{-4}$	1.00
		5	-85.5	$-8.218x10^{-9}$	**	5.000	$1.000x10^{-5}$	1.00	5.000	$1.000x10^{-5}$	1.00
		6	-74.4	$-9.432x10^{-10}$	**	5.000	$1.000x10^{-6}$	1.00	5.000	$1.000x10^{-6}$	1.00
		7	-70.7	$-9.911x10^{-11}$	**	5.000	$1.000x10^{-7}$	1.00	5.000	$1.000x10^{-7}$	1.00
		8	-60.9	$-1.145x10^{-11}$	**	5.000	$1.000x10^{-8}$	1.00	5.000	$1.000x10^{-8}$	1.00
		10	-7.99	$-7.659x10^{-13}$	**	5.000	$1.000x10^{-10}$	1.00	5.000	$1.000x10^{-10}$	1.00
10	1.00	2	5.000	$1.000x10^{-2}$	1.00	5.000	$1.000x10^{-2}$	1.00	5.000	$1.000x10^{-2}$	1.00
		3	5.000	$1.000x10^{-3}$	1.00	5.000	$1.000x10^{-3}$	1.00	5.000	$1.000x10^{-3}$	1.00
		4	5.000	$1.000x10^{-4}$	1.00	5.000	$1.000x10^{-4}$	1.00	5.000	$1.000x10^{-4}$	1.00
		5	-78.7	$-8.192x10^{-9}$	**	5.000	$1.000x10^{-5}$	1.00	5.000	$1.000x10^{-5}$	1.00
		6	-69.5	$-9.257x10^{-10}$	**	5.000	$1.000x10^{-6}$	1.00	5.000	$1.000x10^{-6}$	1.00
		7	-66.3	$-9.693x10^{-11}$	**	5.000	$1.000x10^{-7}$	1.00	5.000	$1.000x10^{-7}$	1.00
		8	-57.1	$-1.121x10^{-11}$	**	5.000	$1.000x10^{-8}$	1.00	5.000	$1.000x10^{-8}$	1.00
		10	-7.46	$-7.475x10^{-13}$	**	5.000	$1.000x10^{-10}$	1.00	5.000	$1.000x10^{-10}$	1.00
5	10.0	2	5.000	$1.000x10^{-2}$	10.0	5.000	$1.000x10^{-2}$	10.0	5.000	$1.000x10^{-2}$	10.0
		3	5.000	$1.000x10^{-3}$	10.0	5.000	$1.000x10^{-3}$	10.0	5.000	$1.000x10^{-3}$	10.0
		4	-69.8	$-5.543x10^{-8}$	**	5.000	$1.000x10^{-4}$	10.0	5.000	$1.000x10^{-4}$	10.0
		5	-57.7	$-6.744x10^{-9}$	**	5.000	$1.000x10^{-5}$	10.0	5.000	$1.000x10^{-5}$	10.0
		6	-53.7	$-7.236x10^{-10}$	**	5.000	$1.000x10^{-6}$	10.0	5.000	$1.000x10^{-6}$	10.0
		7	-51.6	$-7.515x10^{-11}$	**	5.000	$1.000x10^{-7}$	10.0	5.000	$1.000x10^{-7}$	10.0
		8	-43.2	$-8.927x10^{-12}$	**	5.000	$1.000x10^{-8}$	10.0	5.000	$1.000x10^{-8}$	10.0
		10	-5.03	$-6.339x10^{-13}$	**	5.000	$1.000x10^{-10}$	10.0	5.000	$1.000x10^{-10}$	10.0
10	10.0	2	5.000	$1.000x10^{-2}$	10.0	5.000	$1.000x10^{-2}$	10.0	5.000	$1.000x10^{-2}$	10.0
		3	5.000	$1.000x10^{-3}$	10.0	5.000	$1.000x10^{-3}$	10.0	5.000	$1.000x10^{-3}$	10.0
		4	-58.6	$-4.428x10^{-8}$	**	5.000	$1.000x10^{-4}$	10.0	5.000	$1.000x10^{-4}$	10.0
		5	-50.5	$-5.150x10^{-9}$	**	5.000	$1.000x10^{-5}$	10.0	5.000	$1.000x10^{-5}$	10.0
		6	-48.0	$-5.417x10^{-10}$	**	5.000	$1.000x10^{-6}$	10.0	5.000	$1.000x10^{-6}$	10.0
		7	-46.0	$-5.638x10^{-11}$	**	5.000	$1.000x10^{-7}$	10.0	5.000	$1.000x10^{-7}$	10.0
		8	-36.9	$-6.968x10^{-12}$	**	5.000	$1.000x10^{-8}$	10.0	5.000	$1.000x10^{-8}$	10.0
		10	-3.68	$-5.476x10^{-13}$	**	5.000	$1.000x10^{-10}$	10.0	5.000	$1.000x10^{-10}$	10.0

* all volumes are in mL with theoretical V_e = 5.000 mL,** negative values for P due to a_2 = 0, *** Wessa 2009 would not compute.

Table 1 (continued)

V_o*	Theoretical		Excel® 2007			Wessa® 2009			MINITAB® 15		
	P	pK_a	V_e*	K_a	P	V_e*	K_a	P	V_e*	K_a	P
5	100	2	5.000	$1.000x10^{-2}$	100	5.000	$1.000x10^{-2}$	100	5.000	$1.000x10^{-2}$	100
		3	-51.8	$-1.356x10^{-7}$	**	5.000	$1.000x10^{-3}$	100	5.000	$1.000x10^{-3}$	100
		4	-45.4	$-1.570x10^{-8}$	**	5.000	$1.000x10^{-4}$	100	5.000	$1.000x10^{-4}$	100
		5	-42.2	$-1.688x10^{-9}$	**	5.000	$1.000x10^{-5}$	100	5.000	$1.000x10^{-5}$	100
		6	-41.0	$-1.735x10^{-10}$	**	5.000	$1.000x10^{-6}$	100	5.000	$1.000x10^{-6}$	100
		7	-37.5	$-1.890x10^{-11}$	**	5.000	$1.000x10^{-7}$	100	5.000	$1.000x10^{-7}$	100
		8	-23.6	$-2.942x10^{-12}$	**	5.000	$1.000x10^{-8}$	100	5.000	$1.000x10^{-8}$	100
		10	-1.23	$-3.354x10^{-13}$	**	5.000	$1.000x10^{-10}$	100	5.000	$1.000x10^{-10}$	100
10	100	2	-41.3	$-8.132x10^{-7}$	**	5.000	$1.000x10^{-2}$	100	5.000	$1.000x10^{-2}$	100
		3	-48.9	$-7.601x10^{-8}$	**	5.000	$1.000x10^{-3}$	100	5.000	$1.000x10^{-3}$	100
		4	-43.0	$-8.694x10^{-9}$	**	5.000	$1.000x10^{-4}$	100	5.000	$1.000x10^{-4}$	100
		5	-40.8	$-9.144x10^{-10}$	**	5.000	$1.000x10^{-5}$	100	5.000	$1.000x10^{-5}$	100
		6	-39.6	$-9.414x10^{-11}$	**	5.000	$1.000x10^{-6}$	100	5.000	$1.000x10^{-6}$	100
		7	-34.2	$-1.082x10^{-11}$	**	5.000	$1.000x10^{-7}$	100	5.000	$1.000x10^{-7}$	100
		8	-18.0	$-1.989x10^{-12}$	**	5.000	$1.000x10^{-8}$	100	5.000	$1.000x10^{-8}$	100
		10	-0.64	$-2.646x10^{-13}$	**	5.000	$1.000x10^{-10}$	100	5.000	$1.000x10^{-10}$	100
5	1000	2	-48.9	$-1.562x10^{-7}$	**	5.000	$1.000x10^{-2}$	1000	5.000	$1.000x10^{-2}$	1000
		3	-43.4	$-1.790x10^{-8}$	**	5.000	$1.000x10^{-3}$	1000	5.000	$1.000x10^{-3}$	1000
		4	-40.4	$-1.923x10^{-9}$	**	5.000	$1.000x10^{-4}$	1000	5.000	$1.000x10^{-4}$	1000
		5	-39.3	$-1.975x10^{-10}$	**	5.000	$1.000x10^{-5}$	1000	5.000	$1.000x10^{-5}$	1000
		6	-36.3	$-2.132x10^{-11}$	**	5.000	$1.000x10^{-6}$	1000	5.000	$1.000x10^{-6}$	1000
		7	-23.6	$-3.205x10^{-12}$	**	5.000	$1.000x10^{-7}$	1000	5.000	$1.000x10^{-7}$	1000
		8	-7.58	$-9.000x10^{-13}$	**	5.000	$1.000x10^{-8}$	1000	5.000	$1.000x10^{-8}$	1000
		10	-0.08	$-1.606x10^{-13}$	**	5.000	$1.000x10^{-10}$	1000	5.000	$1.000x10^{-10}$	1000
10	1000	2	-47.6	$-8.156x10^{-8}$	**	5.000	$1.000x10^{-2}$	1000	5.000	$1.000x10^{-2}$	1000
		3	-42.0	$-9.301x10^{-9}$	**	5.000	$1.000x10^{-3}$	1000	5.000	$1.000x10^{-3}$	1000
		4	-39.9	$-9.781x10^{-10}$	**	5.000	$1.000x10^{-4}$	1000	5.000	$1.000x10^{-4}$	1000
		5	-38.8	$-1.006x10^{-10}$	**	5.000	$1.000x10^{-5}$	1000	5.000	$1.000x10^{-5}$	1000
		6	-33.8	$-1.148x10^{-11}$	**	5.000	$1.000x10^{-6}$	1000	5.000	$1.000x10^{-6}$	1000
		7	-18.1	$-2.065x10^{-12}$	**	5.000	$1.000x10^{-7}$	1000	5.000	$1.000x10^{-7}$	1000
		8	-4.79	$-6.671x10^{-13}$	**	5.000	$1.000x10^{-8}$	1000	5.000	$1.000x10^{-8}$	1000
		10	-0.02	$-1.363x10^{-13}$	**	5.000	$1.000x10^{-10}$	1000	5.000	$1.000x10^{-10}$	1000
5	10000	2	-43.2	$-1.815x10^{-8}$	**	5.000	$1.000x10^{-2}$	10000	5.000	$1.000x10^{-2}$	10000
		3	-40.2	$-1.950x10^{-9}$	**	5.000	$1.000x10^{-3}$	10000	5.000	$1.000x10^{-3}$	10000
		4	-39.1	$-2.002x10^{10}$	**	5.000	$1.000x10^{-4}$	10000	5.000	$1.000x10^{-4}$	10000
		5	-36.1	$-2.160x10^{-11}$	**	5.000	$1.000x10^{-5}$	10000	5.000	$1.000x10^{-5}$	10000
		6	-23.6	$-3.235x10^{-12}$	**	5.000	$1.000x10^{-6}$	10000	5.000	$1.000x10^{-6}$	10000
		7	-7.62	$-9.048x10^{-13}$	**	5.000	$1.000x10^{-7}$	10000	5.000	$1.000x10^{-7}$	10000
		8	-1.33	$-3.496x10^{-13}$	**	5.000	$1.000x10^{-8}$	10000	5.000	$1.000x10^{-8}$	10000
		10	0.000	$-1.092x10^{-13}$	**	***	***	***	5.000	$1.000x10^{-10}$	10000

* all volumes are in mL with theoretical V_e = 5.000 mL, ** negative values for P due to a_2 = 0, *** Wessa 2009 would not compute.

Table 1 (continued)

V_o*	Theoretical		Excel® 2007			Wessa® 2009			MINITAB® 15		
	P	pK_a	V_e*	K_a	P	V_e*	K_a	P	V_e*	K_a	P
10	10000	2	-41.9	-9.366×10^{-9}	**	5.000	1.000×10^{-2}	10000	5.000	1.000×10^{-2}	10000
		3	-39.8	-9.849×10^{-10}	**	5.000	1.000×10^{-3}	10000	5.000	1.000×10^{-3}	10000
		4	-38.7	-1.013×10^{-10}	**	5.000	1.000×10^{-4}	10000	5.000	1.000×10^{-4}	10000
		5	-33.7	-1.155×10^{-11}	**	5.000	1.000×10^{-5}	10000	5.000	1.000×10^{-5}	10000
		6	-18.1	-2.073×10^{-12}	**	5.000	1.000×10^{-6}	10000	5.000	1.000×10^{-6}	10000
		7	-4.80	-6.687×10^{-13}	**	5.000	1.000×10^{-7}	10000	5.000	1.000×10^{-7}	10000
		8	-0.67	-2.699×10^{-13}	**	5.000	1.000×10^{-8}	10000	5.000	1.000×10^{-8}	10000
		10	0.000	-1.048×10^{-13}	**	***	***	***	5.000	1.000×10^{-10}	10000

* all volumes are in mL with theoretical V_e = 5.000 mL, ** negative values for P due to a_2 = 0, *** Wessa 2009 would not compute.

All three possible explanations were proven fault. Excel® 2007 was able to calculate the correct theoretical values for data set of V_o = 10 mL, P = 10 and pK_a = 3 which showed average X_2/X_1 ratio of 0.000468. Whereas in the case of V_o = 10 mL, P = 100 and pK_a = 2 with higher average X_2/X_1 of 0.000598, Excel® 2007 was unable to return the correct theoretical values. Furthermore, for the same volume of titrant V, calculated X_2 values for the case of V_o = 10 mL, P = 100 and pK_a = 2 were higher than corresponding X_2 values of V_o = 10 mL, P = 10 and pK_a = 3. As for the correlation coefficient as criteria for rejection, we found that in all cases which Excel® 2007 rejected X_2, the coefficient of determination (r^2) between Y and X_2 are at least equal to or higher than that of Y and X_1. In the case of P = 100, V_o = 5 mL, pK_a = 2, Excel® 2007 accepted the multiple linear model and correctly calculate V_e, P and Ka even though VIF between X_1 and X_2 is very high at 140,639. Whereas in the case of P = 10, V_o = 10 mL, pK_a = 4 which shows VIF = 13,150, X_2 was rejected. Therefore it is unlikely that rejection of variable X_2 by Excel® 2007 was based on correlation coefficient criteria. Further investigation is needed to identify the exact criteria for rejection of multiple linear regression model in favor of apparently simple linear regression by Excel® 2007.

There are many free online statistics programs. Wessa® 2009 was chosen because it is widely used by many researchers. We found Wessa® 2009 multiple linear regression software program to be quite user friendly. One need only to copy data from the spreadsheet such as Excel® and just paste the data onto the data entry space on the website and click "compute". The output can be shown on the website or sent to Microsoft Excel® or Word® which is an extremely nice feature of the program. Wessa® 2009 gave accurate results in all but two of our simulated cases. Only in the two extreme cases where P = 10,000 and pK_a =10 (for both 5 and 10 mL octanol) that Wessa® 2009 could not yield the theoretical values of V_e, K_a and P. In these two cases, there was nothing shown in the output box.

Another main limitation of Wessa® 2009 lies in the online input of data. There are only spaces for 71 characters available for data inputs in each row. Since each variable must be delimited by a space or a tab, the higher the number of variables we have, the fewer significant figures we can have for each variable. This may not present any problem unless we have data set which contains relatively high number of variables with relatively high significant figures. For a free program, Wessa® 2009's performance exceeds our expectation.

Minitab® 15 has limit of only 8 decimal places for displaying the calculation results. Since we were unable to change the output display format to scientific notation, this means that if calculated value of a_o falls below 0.000000005, it would be displayed as zero which would lead to erroneous result of V_e = 0 mL. This problem was easily overcome by multiplying Y by an appropriate constant to obtain the desired number of significant figures for our outputs. In our calculations, we chose to multiply Y by 1.0 x 10^6. Therefore, to obtain the correct values of K_a, V_e and P, the coefficients a_o, a_1 and a_2 must also be divided by 1.0 x 10^6. This simple manipulation resulted in Minitab® 15 being the only multiple linear program which gives correct theoretical K_a, V_e and P for simulated cases with P = 0.001 to 10,000. Surprisingly, even the very challenging data set (pK_a = 10, V_o = 10 mL and P = 10,000) with very low numerical values for X_2 ($X_2 < 3x10^{-14}$) and highly significant disparity in the magnitude of X_1 values vs. Y and X_2 (X_1 are approximately 10^7 times larger than Y and approximately 10^{11} times larger than X_2) presented no problem in yielding the theoretical values (Table 1).

Minitab®15's website, www.minitab.com/support/answers/answer.aspx?id=721, has this message concerning multicollinearity: "If the correlation is moderately high, Minitab warns you in a message and continues with computations". However, there was no warning of multicollinearity from Minitab® 15 even with the simulated data pK_a = 10, V_o = 10 mL and P = 5.000 x 10^4 which yielded variance inflation factors (VIF) = 4.5 x 10^{15} resulting from correlation determination (r^2) between X_1 and X_2 = 1.00000000000000. In our opinion, this magnitude of correlation between X_1 and X_2 definitely far exceeds "moderately high". Even at this very high multicollinearity (and still no warning regarding multicollinearity), Minitab® 15 can still yield the correct theoretical values (round off to four significant figures). It is quite obvious that all of our simulated data exhibit high degree of multicollinearity between X_1 and X_2, yet, there was no warning regarding multicollinearity from Minitab® 15 until we tested simulated data for pK_a = 10, V_o = 10 mL and P = 1.000 x 10^5. In this case, we did get 2 warnings: "X_1 is highly correlated with other predictors" and "X_2 is highly correlated with other predictors". Despite the warnings, Minitab® 15 proceeds with the calculations with only slight errors: V_e = 5.000 mL, K_a = 1.001 x 10^{-10} and P = 1.001 x 10^5. The slight errors in calculations possibly are results of rounding off errors in simulation of titration data with Excel® 2007 or the regression operations by Minitab® 15 or both.

As Minitab® 15 has shown, even though very high degree of multicollinearity exists between the predictor variables X_1 and X_2, accurate values of a_o, a_1 and a_2 can still be computed. Since our main objective is primary mathematics perspective of the multiple linear regression, not multicollinearity, Minitab® 15's performance far exceeds our expectation in its ability to calculate coefficients of the predictor variables. However, it certainly would be a very useful feature if Minitab® would show variance inflation factor in the output to give clear indication of multi-collinearity of the data.

In actual experimental situations, it is possible that the ability of Minitab® 15 to accurately determine K_a, V_e and P may not differ significantly from Wessa® 2009 and possibly even Excel® 2007 due to random errors associated with measurement of pH and volume of titrant. In actual laboratory experiment situations, the limiting factor could be experimental errors not the accuracy of the multiple linear regression programs.

Conclusion

For those who are primary concerned with only pure mathematics application in calculating the coefficients of the multiple linear regressions, then

Minitab® 15 is definitely convenient to use and can yield accurate results even with very high degree of multicollinearity. In term of best value for money, Wessa® 2009 multiple linear regression is truly the winner since it performs very well in all but the most demanding cases and it is free. With challenging data such as those encountered with potentiometric titration where the numerical values of predictor variables can be very small (for our simulated data, the numerical values of X_2 ranges from 10^{-3} to 10^{-14}) or the disparity in magnitudes of the two predictor variables is very large, there is a definite need to test regression programs with simulated data (where the only error present is the rounding off error) if possible, to make certain that the programs are indeed capable of delivering expected results.

References

Billo EJ. *Excel® for chemists a comprehensive uide.* 2nd ed. New York (NY): J. Wiley & ons; 2001.

Gran G. Determination of the equivalence point in potentiometric titration. Part II. *Analyst* 1952; 77: 661-670.

Keeling KB, Pavur RJ. A comparative study of the reliability of nine statistical software packages. *Compt Statist Data Anal* 2007; 51: 381-431.

Microsoft Excel® 2007 [multiple linear regression], Microsoft Corporation, Redmond, VA.

Minitab Statistical Software Release 15 (2005-2008). [Multiple Linear Regression, ANOVA], State College, PA: Minitab, Inc.

Pathipvanich M, Rujirawat S, Sukbuntherng J. Quantitative determination of weak acidic drugs in mixed solvent using Gran's method. *Th J Pharm Sci* 1990; 15: 33-42.

Rossatti FJC, Rossatt H. Potentiometic titrations using Gran's plots. *J Chem Ed* 1965; 42: 75-378.

Ryan TP. *Modern regression methods.* New York (NY): J. Wiley & Sons; 1997.