RESEARCH PAPER

Accepted 19 June 1998
Received 16 February 1998
Indian J. Pharm. Sci., 1998, 60 (6) pp. 371-374

SMRAILS : A Lotus 1-2-3 Spreadsheet for Stepwise Multiple Parameter Regression Analysis Applied to QSAR Analysis

S. K. CHAKRAVARTI, S. AJMANI AND S.C. CHATURVEDI*
Department of Pharmacy, Shri Govindram Seksaria Institute of Technology and Science,
23-Park Road, Indore-452003 (M.P.)

An original LOTUS 1-2-3 spreadsheet has been developed to perform quantitative structure activity/structure property relationship analysis using stepwise multiple parameter regression analysis method. 1-2-3's advanced macro commands have been used to prepare the spreadsheet to perform this statistical treatment of the data. Linear as well as parabolic relationships can be searched using the spreadsheet. The results can be directed to a text file and graphs can be plotted using an automatic graphing macro. Correlation matrix can also be obtained using a separate macro.

THE study of quantitative structure activity/structure property relationships comprises an integral part of the drug designing process. The problem is that any one structural change alters many physicochemical properties. How does one decide which properties are relevant to biological potency? To do so each chemical structure is transformed into a list of physicochemical properties : the relationship between these properties and biological property is examined by various methods. Multiple parameter regression analysis is one of such popular methods[1]. If the number of calculated properties are large then the best combination of properties that are significant to biological properties can be found by stepwise multiparameter regression analysis.

LOTUS 1-2-3 is a popular spreadsheet package and extensively used in applications ranging from business to scientific applications. Its advanced command language macro is extremely helpful to develop such applications. Currier *et al* reported a LOTUS 1-2-3 spreadsheet for variance stabilized linear regression applied to the analysis of drugs in biological fluids[2]. The present paper describes a LOTUS 1-2-3 spreadsheet to perform systematic multiparameter regression analysis that can be used to find combinations of physicochemical properties that are significant to biological activity in a large data set.

*For Correspondence

EXPERIMENTAL

SMRAILS was written using advanced macro commands of LOTUS 123 (version 2.2)[3]. The spreadsheet basically contains four macros: a stepwise regression macro (STEPREG), an output to file macro (OUTFILE), a correlation matrix generating macro (CORMAT) and a graphing macro (AGRAPH). STEPREG can be executed in linear as well as parabolic mode after pressing ALT and E keys simultaneously. CORMAT produces a correlation matrix of all the variables when executed by pressing ALT-X. The graphing macro AGRAPH produces graph between any two selected variables. AGRAPH can be executed by pressing ALT-G and OUTFILE can be made run by pressing ALT-F.

Before executing any macro the data for dependent and independent variables has to be organized in the worksheet properly. The independent variables should be put in area ranging from column A to column AP. In the first row (row-1) of this area the names of the independent variables should be entered. The second row should be left empty. The dependent variable (e.g., biological activity) should be entered in column AS and like the independent variables its name should be entered in first row of column AS and the second row should be left empty.

The STEPREG is an interactive macro. When executed it asks for four inputs, the number of independent variables, the number of data points (the number of compounds), degree of polynomial (1 for linear and 2 for parabolic) and the relaxation factor (the maximum allowable squared cross correlation between any two selected independent variables). After entering the required information, the macro starts scanning through all the independent variables for significant contribution towards the dependent variable. After each step it adds one term to the equation. Before adding the term the program checks for its correlation with already selected terms in previous steps. The term that has maximum contribution to the dependent variable and having cross-correlation less than the relaxation factor is finally entered in the equation. This process continues until the number of selected terms is one fifth of the data points. In each step the macro produces one equation. It also calculates following statistical measures ; n, the number of samples in the regression; r, correlation coefficient; F-test; and std, standard deviation. The standard errors for the derived coefficients are also calculated. When STEPREG is executed in parabolic mode, the square term of the first parameter is also added to the equation.

The output of STEPREG and CORMAT is produced in various areas of the worksheet. The parameters selected in the regression analysis and various statistical measures are placed in the area ranging from BH1 to BT7. Regression coefficients and their standard errors are placed in the area started from the cell BM11 and the regression constants are placed in the area started from the cell BM41. The correlation matrix is placed from the cell BM46.

The macro OUTFILE when executed, asks for the number of selected parameters and the name of the desired output file. It produces a text file containing the equations and the statistical measures in a systematic manner. The text file can be viewed using any editor or by the DOS command "TYPE".

AGRAPH is a menu driven macro. It is helpful to create XY graphs, bar graphs and pie graphs.

The listing of macros is available from the authors on request.

RESULTS AND DISCUSSION

To validate SMRAILS, the QSAR work reported by Garcia *et al* was taken as an example[4]. In the above paper the log of cyclooxygenase inhibition, IC50 (μM) of twenty analgesic drugs and their connectivity parameters (Table 1) was subjected to QSAR analysis. The same data set was used to validate SMRAILS. After arranging the dependent and independent variables as previously mentioned, STEPREG was executed. the number of parameters and the number of compounds were set to 8 and 20 respectively, the degree of polynomial was set to 1 and the relaxation factor was set to 0.3. After the execution of OUTFILE the following equations were obtained exactly in the following format :

LOG (IC50)=E*(0.8590795578) [±0.4795732511] +(0.0921629036)

n=20 r=0.3889728683 F=3.2089050487 STD = 1.0076808626

LOG(IC50)=E*(0.9099159243) [±0.4456185007] + V4*(0.2450276789)

[±0.1238035955] +(1.2825030839)

n=20 r=0.5569864095 F=8.0958011208 STD =0.934778273

LOG(IC50)=E*(1.0354979672) [± 0.4857916679] +V4*(0.2338395819)

[±0.1266467731] +J1*(3.4317831853) [± 4.8441763751] + (0.0049269421]

n=20 r=0.5755102616 F=8.9143609742 STD = 0.948781831

LOG(IC50)=E*(1.0393032069) [± 0.5024776116] + V4*(0.2302188251)

[±0.1341552677]+J1*(3.476622189) [± 5.0144957679] +X4c*(-0.4440891707) [±3.6906225484]+(-0.024741698)

n=20 r=0.5760703092 F=8.9403406614 STD = 0.9794250695

When polynomial degree was raised to 2, the following equations were obtained :

LOG(IC50) = X 4p_SQ*(0.7473771772) [± 0.2558652474] +X4p*(-3.8462628848) [±1.3163210591]+(5.8814445742)

Table 1: Cyclooxgenase inhibition, IC_{50} (μM) and connectivity parameters from ref.(4) used to test "STEPREG"

Compd No.	IC50 (μM)	X3pv	X4p	X4c	X4pc	Glv	JI	V4	E
1	700	1.37	1.42	0.00	0.79	7.85	0.33	4	2.03
2	6	2.63	2.46	0.25	2.06	9.55	0.32	6	1.74
3	15	3.58	2.56	0.00	1.89	9.15	0.31	7	2.19
4	4	3.03	2.30	0.00	1.59	8.30	0.28	6	2.17
5	1000	1.13	1.05	0.00	0.71	4.55	0.28	3	2.80
6	18	1.74	1.44	0.00	0.54	9.45	0.33	3	0.74
7	10000	2.97	3.19	0.00	2.20	13.83	0.27	4	2.00
8	5	3.27	2.55	0.00	1.45	9.40	0.32	6	1.73
9	169	4.06	4.01	0.00	1.88	10.90	0.18	4	1.95
10	100	2.71	2.09	0.00	1.35	9.05	0.28	4	1.39
11	0.34	2.95	2.18	0.00	1.62	6.92	0.26	5	1.39
12	4	2.49	1.70	0.00	1.38	5.80	0.36	3	0.93
13	11	4.14	3.56	0.00	2.04	13.63	0.30	8	1.69
14	6	3.06	2.09	0.00	1.56	6.80	0.24	5	1.40
15	32	2.94	2.09	0.00	1.58	9.20	0.31	5	1.38
16	5	3.31	2.97	0.00	1.25	9.80	0.23	6	1.94
17	662	1.19	1.05	0.00	0.51	7.70	0.40	3	1.45
18	120	3.95	3.94	0.14	3.46	13.32	0.26	7	1.69
19	28	4.30	3.57	0.00	2.28	18.38	0.31	9	1.37
20	12	3.10	2.53	0.00	1.64	12.78	0.33	6	1.07

n=20 r=0.5797155861 F=4.3025575856 STD 0.9171043804

LOG(IC50)=X4p_SQ*(0.6743432472) [± 0.2512806474] + X4p*(3.5031677053)

[±1.2887817323]+E* (0.6517009561) [± 0.4298589125] + (4.4596353155)

n=20 r=0.6476630038 F=6.1417264255 STD = 0.8839657987

LOG(IC50)=X4p-SQ*(0.6719272982) [± 0.2584129673] + X4p*

(-3.4730010459) [±1.3274209542]+E*(0.6582359229) [±0.442269156]+X4c* (-1.2670857482) [±3.4044031299]+ (4.4162092486)

n=20 r=0.6517511416 F=6.276943379 STD = 0.9087693342

Where X4P_SQ is the square term of X4P.

The squared correlation matrix produced by CORMAT can be seen in Table 2.

The programs were run on a computer with a Pentium processor of 120 MHz clock speed, DOS version 6.22 and

Table 2 : Squared correlation matrix produced by "CORMAT" for the data used as example

	$LogIC_{50}$	X3pv	X4p	X4c	X4pc	G1v	J1	V4	E
$LogIC_{50}$	1.000								
X3pv	0.116	1.000							
X4p	0.003	0.843	1.000						
X4c	0.007	0.005	0.042	1.000					
X4pc	0.006	0.635	0.705	0.218	1.000				
G1v	0.020	0.459	0.591	0.012	0.403	1.000			
J1	0.001	0.249	0.324	0.001	0.143	0.011	1.000		
V4	0.141	0.610	0.457	0.048	0.448	0.469	0.019	1.000	
E	0.151	0.004	0.006	0.003	0.002	0.027	0.136	0.003	1.000

LOTUS 1-2-3 version 2.2. In each run STEPREG has taken less than 5 seconds to complete the analysis and CORMAT has taken less then 3 seconds the produce the squared correlation matrix. It can be seen that these macros are quite fast. Besides the facilities provided by these macros, every other facility present in LOTUS 1-2-3 is accessible to facilitate the analysis of the data, such as easy storage and retrieval of data, sorting of data etc. One advantage of these macros are that they can be very easily modified with little knowledge of macro programming.

These macros are interactive, easy to modify and consist of useful tools for QSAR analysis. Since LOTUS 1-2-3 is one of the most easily available and widely used spreadsheet software, therefore this macro package might prove to be valuable for QSAR/QSPR analysis and use of costly software can be avoided.

ACKNOWLEDGMENTS

Authors are grateful to the Director, S. G.S. I. T. S., Indore for providing computer and software facilities to carry out this work. Mr. S.K. chakravarti and Mr.S. Ajmani thank to the CSIR, New Delhi for Providing senior research fellowships to support this work.

REFERENCES

1. Draper, N.R. and Smith, H., In; Applied Regression Analysis, 2nd Ed, Wiley, New York, 1981, 10.
2. Currier, F. P., Cefali, E. A. and Mclean, A. M., **J. Pharm. Sci.**, 1991, 80, 1003.
3. Nielson, J. J., In; Using 1-2-3 Release 2.4, Shaw, S. M. Ed., Special Edition, Prentice-Hall of India Pvt. Ltd., New Delhi, 1995, 631.
4. Garcia-Domenech, R., Garcia-March, F. J., Soler, R. M., Galvez, J., Anton-Fos, G. M. and de Jolian-Ortiz, J.V., **Quant. Struct.-Act. Relat.**, 1996, 15, 201.