P-4

Maximum Likelihood Estimation in Determination of Power of the Error in Bivariate Linear Models involving Generalized Gauss-Laplace Distributed Variables

L. Jäntschi[1] and S. D. Bolboacă[2]*

[1]*Department of Physics and Chemistry, Technical University of Cluj-Napoca, Romania*
[2]*Department of Medical Informatics and Biostatistics, Iuliu Haţieganu University of Medicine and Pharmacy, Romania*

*Author for correspondence e-mail: sbolboaca@umfcluj.ro

Normal distribution of errors is one assumption for linear regression analysis [1]. Due to its relative easy interpretation [2], linear regression analysis is used in structure-activity/property analyses to quantitatively relate chemical features to biological activity or property [3]. A new approach that maximizes the probability of observing the event according with the random error has been recently introduced to solve simple linear regression [4]. The reported results obtained on ten classes of compounds proved that in 93% of cases the power of the error was significantly different by the expected value of two [4]. An implementation for linear regression analysis with two independent variables is proposed in this research. The input values are the results obtained by classical regression analysis, named coefficients of the independent variables, population mean and population standard deviation obtained for the power of the error equal with two. The outputs are the optimal solution for above-coefficients identified by the implemented algorithm to maximize the likelihood of observation according to the random error. The proposed algorithm was tested on two sets of compounds, one with estrogen binding affinity (logRBA, n = 132) [5], and the other one with toxicity on *Tetrahymena pyriformis* (log1/IGC_{50}, n = 250) [6]. The optimization results are presented in the Table 1.

Table 1

Set	MLR					MLE				
	q	**σ**	**μ**	**a_1**	**a_2**	**q**	**σ**	**μ**	**a_1**	**a_2**
1	2	1.3938	-4.2841	$-2.6302\cdot10^{-2}$	$3.6793\cdot10^{-2}$	1.6604	1.3799	-4.5305	$-2.3326\cdot10^{-2}$	$3.7609\cdot10^{-2}$
2	2	0.5490	-0.4790	0.4950	$-2.4180\cdot10^{-1}$	0.9684	0.5554	-0.7980	0.5797	-0.3251

q = power of the error; σ = population standard deviation; μ = population mean (intercept); $a_{1,2}$ = coefficients of independent variables

On both sets of compounds, the power of the error proved significantly different by the expected value of two ($p < 0.001$). The proposed approach proved feasible for estimating the parameters of the bivariate linear regression under the assumption that the errors are Gauss-Laplace general distributed.

References

1. S. D. Bolboacă and L. Jäntschi, Biomath, 2, 2013, Article ID 1309089.
2. P. Liu and W. Long, International Journal of Molecular Sciences, 10, 2009, 1978-1998.
3. M. Goodarzi, B. Dejaegher, and Y. V. Heyden, Journal of AOAC International, 95, 2012, 636-651.
4. L. Jäntschi, L. L. Pruteanu, A. C. Cozma and S. D. Bolboacă, Computational and Mathematical Methods in Medicine, 2015, 2015, Article ID 360752.
5. J. Li and P. Gramatica, Molecular Diversity, 14, 2010, 687-696.
6. M. T. D. Cronin, A. O. Aptula, J. C. Duffy, T. I. Netzeva, P. H. Rowe, I. V. Valkova and T. W. Schultz, Chemosphere, 49, 2002, 1201-1221.