

■Poster presentation

Development of a state-of-the-art multiple regression KY-method corresponding to the big data era and its application to fish toxicity

Kohtaro Yuta

In Silico Data, Ltd., Narasino, Chiba, 275-0025, Japan

Key word; Computational Toxicology, Regression KY-method, Big data, Fish toxicity

Introduction: Even when the number of samples to be analyzed is extremely large, we have developed a multiple regression method which has high analytical reliability. In the case of evaluating the toxicity of a compound using a computer, a multivariate analysis / pattern recognition method is applied. However the conventional multivariate analysis / pattern recognition method is not designed to handle a large amount of sample data such as big data. We developed the "multiple regression KY-method" as a data analysis method corresponding to the big data era and analyzed fish toxicity data.

Method: In the KY-method, the sample group to be analyzed is divided into two groups. One is an in-liear group and the other is an out-liear sample group. The correlation coefficient (R) value of the in-liear sample group is greatly improved compared with the normal method.

Experiment and results: Fish toxicity data analysis was performed by two methods. One was normal linear multiple regression and the other was the regression KY-method.

1. Result by the normal linear multiple regression method using all samples.

Number of total samples: 779, Number of parameters: 28, Reliability index: 27.8,

R: 85.3, R2: 72.8, F value: 71.7, CV: 69.6

2. Results by the "multiple regression KY-method".

a) Number of in-liear samples: 398, Number of parameters: 22, Reliability index: 18.1,

R: 98.1, R2: 96.2, F value: 428, CV: 94.4

b) Number of out-liear samples: 393, Number of parameters: 29, Reliability index: 13.6,

R: 80.4, R2: 64.7, F value: 22.9, CV: 57.5

Summary: Currently, we have entered the era of big data where the number of samples extremely large. However the data analysis method currently deployed can not correspond to the big data era. Therefore, it is necessary to develop new and state-of-the-art data analysis methods. The multiple regression KY-method discussed in this poster is developed as a data analysis method corresponding to the big data era. Details will be discussed in the poster.