

## **An RM4Es Software System to Improve Data Analysis and Its Learning**

Alex Liu on March 5, 2009

Data analysis has been defined as a repetitive, cyclical search for understanding by many scholars (Lubinsky & Pregibon 1988, Young & Smith 1991). In other words, data analysis is a process that involves a sequence of repetitive actions. The order of these actions, as defined as statistical strategy by some scholars (Harrell 2001), has huge impacts on the results of any data analysis, as demonstrated by a few experiments such as that of professor Faraway (J. J. Faraway 2004).

We name the above mentioned sequence of data analysis actions as research flow as it is a term familiar to business analysts. We have created an RM4Es framework to represent all kinds of research flows. Based on the above mentioned RM4Es representation framework, we have developed an RM4Es software system that allows users to create research flows, to open saved research flows and modify them, then to run any research flows in the open source statistical software R. Research flows can fully represent all statistical strategies (Faraway 2004, Harrell 2001). Therefore, the RM4Es system allows analysts and researchers to manage their statistical strategies, so that they can re-use them and improve them in the future.

At the same time, our RM4Es software provides features facilitating research flow mapping, and intelligent guidance for research flow formation. As claimed by professor Young and professor Lubinsky (Young & Lubinsky 1995), in an environment where research flows can be structured and related intelligent guidance can be provided, data analysis productivity, accuracy, accessibility and satisfaction will all be improved. The users of our RM4Es software system are confirming this claim with their experience. Students of predictive modeling also found that they learned fast and learned more than before, with an RM4Es software system in place.

Statistical software packages are widely available, but all of them only handle statistical tactical computation. The RM4Es system as discussed in this note performs computing for statistical strategies that represents a new direction for statistical computing and has been considered as an important breakthrough.

### **Reference**

Faraway J. J. 2004 Linear Models with R, CRC Press

Harrell, F. E. 2001 Regression Modeling Strategies, New York: Springer-Verlag

Lubinsky, D. L. and Pregibon, D. 1988 Data analysis as search, Journal of Econometrics, 38: 247-268

Young, F. W. and Smith, J. B. 1991 Toward a structured data analysis environment: a cognition-based design, in Bujia, A. & Tukey, P. A. (eds.) Computing and Graphics in Statistics, 36, 253-279, New York: Springer-Verlag

Young, F. W. and Lubinsky, D. J. 1995 Guiding data analysis with visual statistical strategies, Journal of Computational and Graphical Statistics. 4: pp. 229-250