

Interactive analysis system on the web for data-oriented approaches

Yuichi Mori

Okayama University of Science, Department of Socio-Information
Okayama 700-0005, Japan, mori@soci.ous.ac.jp

Keisuke Honda

Okayama University of Science, Graduate School
Okayama 700-0005, Japan, honda@soci.finfo.ous.ac.jp

Hiroshi Yadohisa

Kagoshima University, Department of Mathematics and Computer Science
Kagoshima 890-0065, Japan, yado@sci.kagoshima-u.ac.jp

Yoshiro Yamamoto

Tokai University, Department of Mathematics
Hiratsuka 259-1292, Japan, yamamoto@sm.u-tokai.ac.jp

1. Data-oriented statistical system

The authors have developed a web-based statistics system (mainly for educational use) that archives data sets and the corresponding analysis stories and allows statistical analysis to be performed directly on these data sets on the web.

The analysis story, which is a kind of document or record that describes the actual process used in the original analysis, is especially thought necessary as well as an original data set for data-oriented education because this information takes on the role of the manual or chart typical in a similar situation to original one. Furthermore, it is desirable that anyone can use any data set anytime and anywhere and that the analysis of the data can be performed automatically to trace the process and interpret the results in the original analysis. The environment where the above can be realized has been established in our system, which has been named the “Data-oriented Statistical System” or called **DoSS@d** (Mori et al., 2003). Data sets and analysis stories are collected from various fields and are stored as XML documents in the database. The XML format allows the data to be categorized and characterized easily and also allows various XML technologies to be applied.

DoSS@d is located at <http://mol61.soci.ous.ac.jp/@d/index.html> and consists of three subsystems, **DoDStat@d** (data and story database), **DoAStat@d** (online analysis system) and **DoLStat@d** (learning courses). This paper presents the details of the online analysis system **DoAStat@d** which has been refined with the latest web-based and XML-based technologies.

2. Online analysis system “DoAStat@d”

DoAStat@d is a web-based application for the analysis of any data set stored in **DoDStat@d**, as well as data sets stored on the local computer. Currently this system executes data analysis using R and XploRe Quantlet Server (XQS) as statistical engines. Both communicate with XML files describing the data to perform the analysis of the data.

2.1 CGI version “DoA_R”

The R Server-based system **DoA_R** is controlled by CGI. When the user chooses a data set and an analysis method on the ordinary web browser, **DoA_R** selects appropriate variables for the selected method among all variables by interpreting the attributes described in the data XML file. For example, when choosing “regression

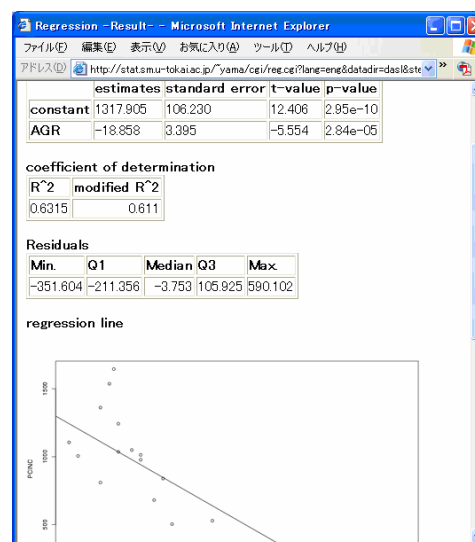


Figure 1. DoA_R (regression analysis)

analysis”, the CGI program in DoA_R selects only continuous variables and displays them on the browser which allows the user to specify parameters for regression analysis. The results including graphics based on the specified parameters are obtained from the R-server and reformed and displayed on the browser by the CGI program (Fig. 1).

2.2 Java version using XQS “DoA_X”

The XQS-based system DoA_X provides new GUIs constructed based on Java Swing as a standard component. At first the user selects a place where a data set that you wish to analyze is stored, [local computer] or [DoDStat@d]. The ordinary file selection window and a special window written in Java appear for local computer and DoDStat@d, respectively. The Java window lists all data sets stored in DoDStat@d and displays the description of a highlighted data set reading from the <description> part in the data XML file (Fig. 2). If you want to see more information on the data, click the [More information] button to open the page displaying all the information in the data XML file. The method that you want to apply to the data is specified on the same window. Clicking the [Start] button, the selection window disappears and the GUI is divided into four parts: data sub-window in which variables to be analyzed are selected; parameter sub-window in which parameters for the selected method are specified, output sub-window where the results are output, and history sub-window where all data sets and operations in the analysis are displayed as objects (Fig. 3). When this GUI is called from the analysis story page, variables and parameters have been pre-set according to the <method> part in the story XML file. Then clicking the [Execute] button, a network handling section @d_Server communicates with the database to get data body and other necessary information and the MD*Crypt communicates with XQS to analyze the data based on parameters receiving from the GUI. Then the analysis results are displayed in the output sub-window. Graphic window(s) such as a scatter plot appears in case of need.

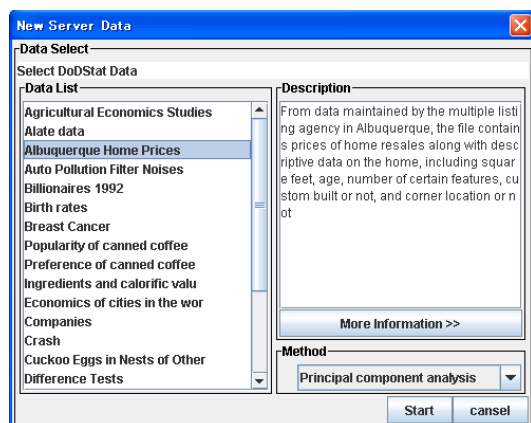


Figure 2. DoA_X (data and method selection window)

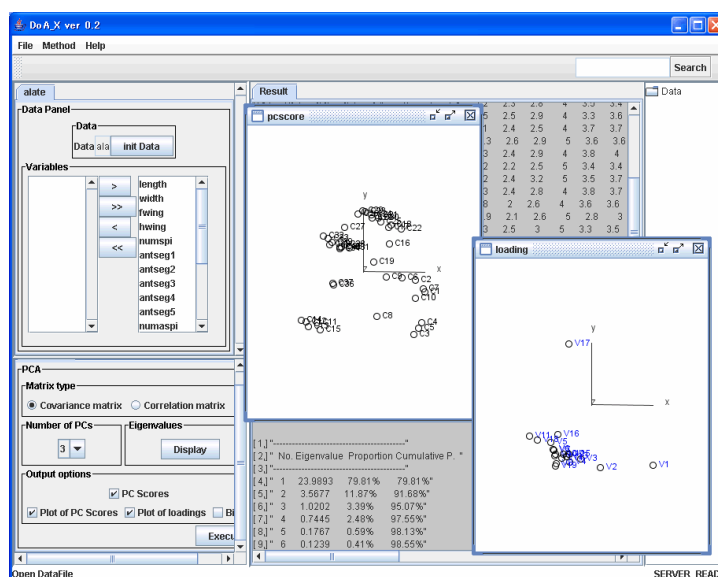


Figure 3. DoA_X (GUI for PCA)

3. Discussion

We implemented two different systems according to their distinctive functions. These systems provide more intuitive and interactive user interfaces to analyze data on the web. Using the systems users can learn the skills of real-world data analysis based on the data attributes and analysis story without needing to have knowledge of macros or script languages.

REFERENCES

Mori, Y., Yamamoto, Y. and Yadohisa, H. (2003). Data-oriented Learning System of Statistics based on Analysis Scenario/Story (DoLStat). *Bulletin of the International Statistical Institute, 54th Session Proceedings, Volume LX Two Books*, Book 2, 74-77.

Honda, K., Mori, Y., Yamamoto, Y. and Yadohisa, H. (2004). Web-Based Analysis System in Data-oriented Statistical System "DoSS@d". In: Antoch, J (ed), *COMPSTAT2004 Proceedings in Computational Statistics*, 1209-1216, Physica-Verlag.