E-LEARNING/E-TEACHING OF STATISTICS: A NEW CHALLENGE

Gökhan Aydınlı*, Wolfgang Härdle*, Bernd Rönz*
* CASE – Center of Applied Statistics and Economics, School of Business and Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin,Germany
{aydinli/haerdle/roenz}@wiwi.hu-berlin.de

Abstract
Travel arrangements and flight ticket booking via internet is widely used nowadays and follow already certain standards. Although increasing activity for multimedia/web education components can be observed, we are far away from standards in this important area. Statistics can possibly profit the most from e-learning since it requires a variety of skills including handling of quantitative data, graphical insights as well as mathematical ability. In this paper we take two positions – the student's view and the teacher's view – and discuss their relative coherence in order to propose standards for e-learning of statistics. The proposed standards are flexible with regard to content, multi-functionality, interactivity type, integration technology and design. Therefore the main focus may be directed on quality of e-learning tools in order to meet both teacher's and student's requirements. This is especially true for statistics which is taught in various disciplines.
*We present our thoughts and exemplify them via the e-learning/e-teaching tools MM*Stat, e-stat, MD*ReX and RExcel. The structure and the main characteristics of these multimedia tools will be explained. Then it will be described how such standards may be transferred to other cultures, languages or disciplines via the platform MD*Book.*

1 INTRODUCTION

Statistics as the science of extracting information from highly complex structured data is often difficult for students, since it requires a variety of skills incluing handling of quantitative data, graphical insights as well as mathematical ability. For an educated society, an effective education in statistics is a necessity for students of all subjects. For that, e-learning/e-teaching tools are an attractive and potentially powerful new way. Statistics can possibly profit the most from e-learning/e-teaching. Although increasing activity for multimedia/web education components can be observed, we are far away from standards in this important area. In section 2 several standards from both the student's view and the teacher's view will be proposed and their relative coherence are shown. These ideas are exemplified by four e-learning/e-teaching tools of statistics: MM*Stat (section 3), e-stat (section 4), MD*ReX amd RExcel (section 5). Finally it will be described in section 6 that MD*Book is an easy technical tool for generating e-learning documents.

2 TEACHERS' AND STUDENTS' VIEWS

An internet-based respectively a web enabled multimedia environment can only be an effective tool if both the teachers' and the users' requirements are fulfilled. Although teachers and students are positioned on opposite sides in the educational process, their standards are relative conformable.

Though most teachers and students are (in general) familiar with using browsers on the internet or using standard office applications like spreadsheets, the new medium of a (inter)networked computer and a web based learning/teaching tool will only be accepted and therefore widely used if its application is immediately understandable and easy for everybody. This requires a simple and not too profusely designed graphical user interface (GUI). On the other hand, the user interface must show all main features of the tool at a first sight. Furthermore, the navigation from one part to an other part of the tool must be fast and straight forward. If a lot of clicks are needed to reach an other desired statistical topic it would restrain the users from further usage of the tool: It would be too time consuming for the teacher to use it in the classroom; the student will get frustrated after a short time.

A clear and concise *structured format* in presenting the statistical content must be maintained throughout the whole system so that all users will recognize a consistent pattern of available materials. A well-designed e-learning tool can particularly make the learning process easier by allowing the students to develop their insights without getting bogged down in the mathematics. Therefore, a separation of the fundamentals of a statistical method and the more complicated mathematical proofs should exist. By that means the teachers can attract students with different mathematical preliminary knowledge and ability, and the students can gradually study the statistical problems in their learning process.

Examples are unalterable in an e-learning/e-teaching tool. They give the teacher the opportunity to explain in detail both the theoretical background and the practical application of statistical methods under varying conditions and to demonstrate the complex interconnection between the methods. Examples are unrenounceable for students. They must have the opportunity to actively practise the acquired knowledge and to deepen the problem-oriented understanding of statistical methods and models as well as the critical reflection of empirical data.

More elaborated *empirical applications* based on real-world data sets must be included in each e-learning/e-teaching tool to make clear the relevance and structural complexity of statistics in the daily life and to demonstrate the interplay of various statistical methods.

Interactive capability must enable the practice of statistical methods under varying conditions, graphical illustrations wherever possible and repeated examples with several variables or data sets. The teacher can thus demonstrate and explain in detail the assumptions connected with the various statistical methods and models. Additionally, the teacher can provide more elaborate statistical data analysis with larger data sets instead of simple classroom-examples. On the other hand, interactivity assures that the students actively participate in the learning process, rather than passively reading or hearing about something outside of their control. Interactivity capability would be superb if the students could use and analyse their own collected data.

Flexibility is required for teachers to tailor individual courses with varying content, profundity and applications according to the audiences to be addressed. Flexibility is also required for students to access the system at any point, to make their own way through the multimedia tool and to review (perhaps repeatedly) the content at their own pace.

Each e-learning/e-teaching tool must offer *self-assessment* components with automatic evaluation of the answers to give the students the opportunity of checking their acquired knowledge. By the correctness of the question responses the teacher receives a feedback on the students skills.

Furthermore, *hypertext functionality* should enable both teachers and students to jump between relevant context parts reinforcing previous ideas within the process of learning new ones. That also applies to the search for keywords.

Beyond all those required standards, there remains the question of how students and teachers conduct their analyses. What is the best medium to accomplish a seemingly impossible mission: combine modern and accurate statistical methodolgy with an intuitive and easy to use though open and flexible system. The requirements are obvious: a fully-fledged statistical computing language is required to stay up-to-date with modern methods. Furthermore *standardized inferfaces* are needed. Why inferfaces? First of all the requirements mentioned above should be kept in mind, hence the need for a well-designed user interface allowing for interactivity and structure. However we also need another interface to connect that GUI with a statistical language to fulfill further requirements of statistical teaching: reproducability and transparency. This is the point where another standard seems desirable: *integration* of statistical wisdom with various different systems (e.g. browsers or standard spreadsheet applications).

3 E-LEARNING/E-TEACHING: MM*STAT

MM*Stat (Härdle,W., Rönz,B., 2001) is an HTML-based multimedia tool for supporting teaching and learning statistics via the internet or from a CD. It contains all parts of descriptive and inferential statistics, usually taught in introductory courses at universities, and can therefore flexibly be used by teachers in the classroom and by students for reviewing the lecture subjects on their own. MM*Stat is now available in eight languages (German, English, French, Spanish, Italian, Czech, Polish, Indonesian; see http://www.md-stat.com) and soon in Japanese and Chinese too.

The user interface of MM*Stat (see Figure 1) is a composition of filing cards. It is very clearly arranged, shows the various available components and allows fast access from one component to an other. In this regard the teachers' and students' requirements are met by standardization to an HTML based filing card environment.

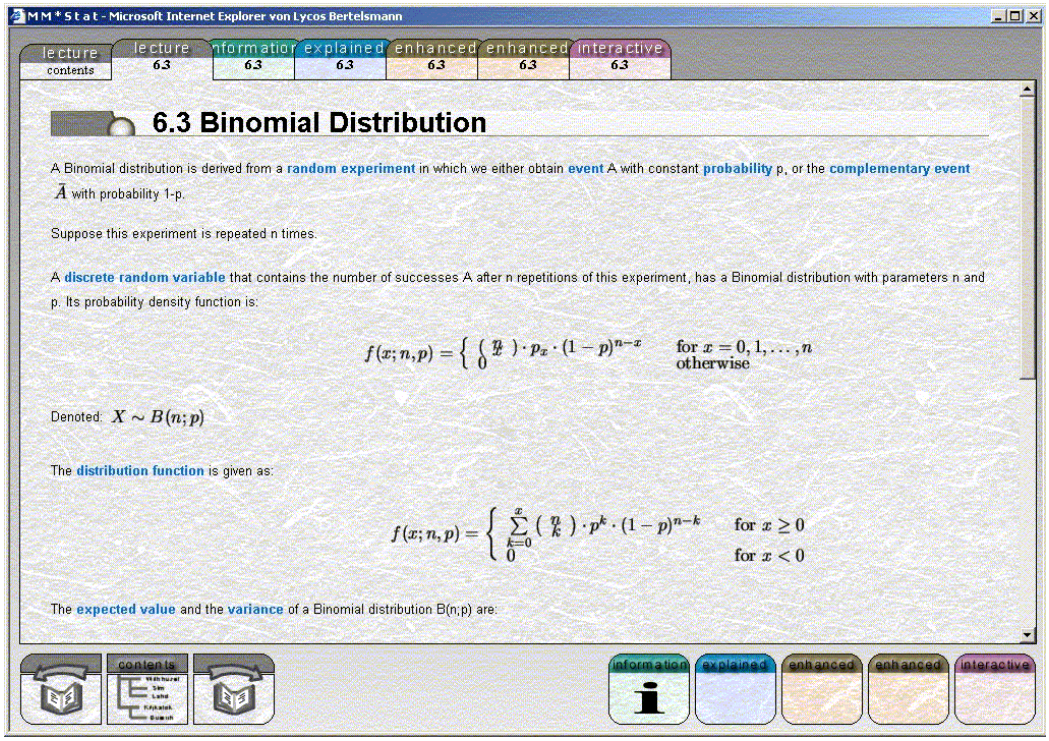

Figure 1: Graphical user interface of MM*Stat with lecture unit "Binomial Distribution"

A structured format with regard to the statistical content is realized through the components: *lecture units*, *additional information* and *various examples*. Each lecture unit provides an explanation of statistical methods and models considered which includes the basic concept, definitions, formulas and fundamental statistical requirements of its application (without any specific target area of application). The sequence of lecture units is canonical in statistical theory. The user can enter the course via the list of contents at any unit and may jump back to a previous one or forward to any following one.

The ramification in MM*Stat (Figure 1 and 2) arises from the inclusion of

- additional information, presenting additional explanations concerning the content of the lecture unit, which are not necessary for elementary use, but for a broader and deeper understanding of the statistical methods,
- three types of examples: *fully explained examples*, *enhanced examples* and *interactive examples*.

The user can easily switch between lecture unit, additional information and examples at any time. Especially the interactive examples allow teachers/students to present/study statistical methods under varying application scenarios, to use/analyse different data sets and to discuss/review the results at their own pace. The implementation of the interactive examples into MM*Stat is based on the XploRe Quantlet technology. For a detailed explanation of this technology see Härdle, Kleinow, Tschernig (2001) and Kleinow, Thomas (2000).

Multiple choice questions are provided at the end of each topic giving the students the opportunity for self-assessment. An automatic evaluation showing the correctness of the a nswers can be requested.

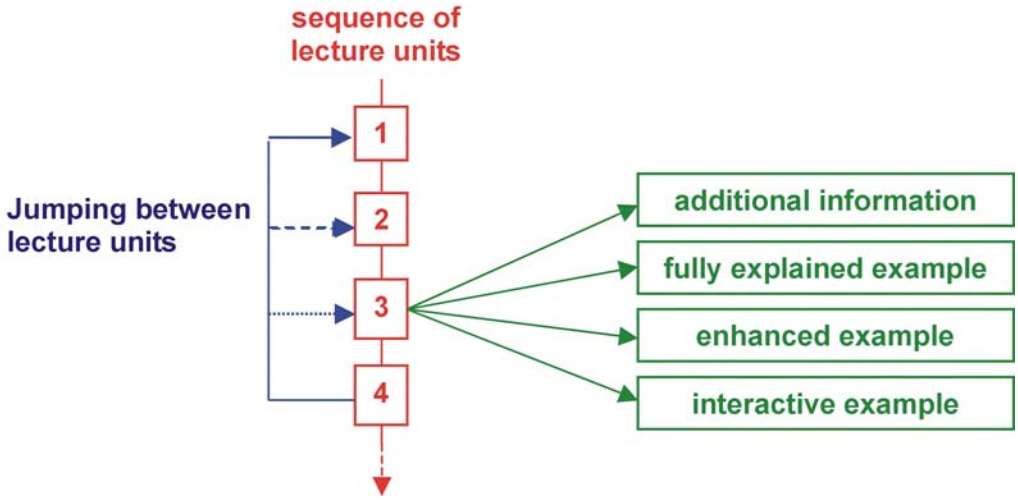

Figure 2: Basic structure of MM*Stat

In summarising, substantial advantages of MM*Stat both for teachers and students are:

- it provides an introduction into the fundamentals of statistics in a compact course;
- it can be used for teaching and learning statistics in various courses of studies;
- MM*Stat enables interactive examples.

On the other hand, we are aware of limitations of the MM*Stat technology:

- The specific course orientation might be considered as a disadvantage, because there is little flexibility for the teachers to include other statistical topics.
- The examples have mostly an economics flavor.
- A specific statistical engine is addressed for the interactive examples.

MM*Stat has been used now by two generations of students and has been highly evaluated by students in the Economics Department of the Humboldt-Universität zu Berlin. The filing card format and the information structure (as given in Figure 2) is called an MD*booklet. Many more of these MD*booklets have been created with the tool MD*book. On http://www.md-stat.com one finds for example a finance introductory course (FIC) as well a XploRe introductory course (XIC).

4 E-LEARNING/E-TEACHING: E-STAT

e-stat (http://www.e-stat.de), which is currently under development by several teams at seven German universities, is a step towards a more complex and deeper structured statistics e-learning/e-teaching system. e-stat is not linked to a special statistics course (such as MM*Stat) but it is being designed as an open source system with open architecture. To this end the content of statistical methods, models, examples and applications is broken down into small modules. Such a module may contain either a definition, a property, a derivation, a mathematical proof, a graphical illustration, a textual introduction, motivation or an explanation. This requires clear-cut separations and concise arrangements of the statistical topics. As a side effect, this modularisation avoids endless scrolling in an internet environment. Let us take the regression analysis as a simple example for this modularisation. All modules for this statistical method have the same name "regression analysis" and are consecutively numbered. Module 1 gives a textual motivation and module 2 an explanation of the general purpose and aim. Module 3 may contain the specification of the regression model and the definition of the variables involved. Module 4 comprises important properties and assumptions of the regression model. The estimation of the unknown regression parameters is explained in module 5, and so on.

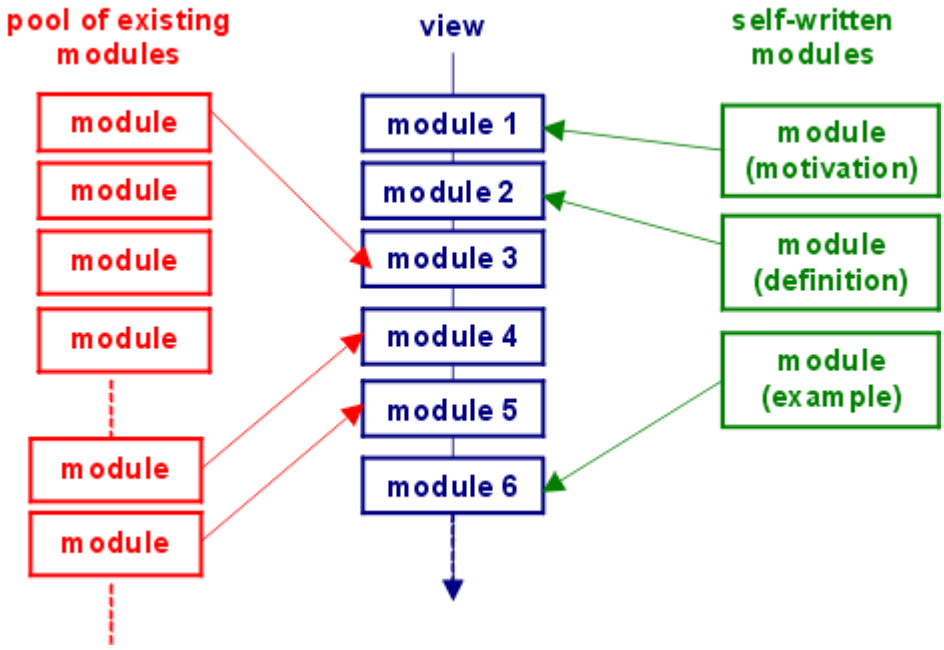

Figure 3: Modules and views in e-stat

This concept enables each teacher to compile his own statistics course of instruction (which is called view in e-stat) to match the different needs, interests and backgrounds of the audience. A view comprises a course which can be either more general in statistical theory (without a specific target of application) or for statistics within a specified subject area (e.g. business administration, economics, social sciences, psychology, biology, agriculture, medicine, computer science). With this technology it is possible to address students of many different fields. To compile a view, the teacher selects the appropriate modules from the pool of available modules and links them to the required sequence.

If a module with a special statistical concept, which is indispensable for the course, does not exist or if an other motivation is required, then the teacher can write new modules and include them in the view (see Figure 3). In the same way additional examples and/ or applications can be supplemented. Moreover these new modules are added to the pool so that they are available for all other users. Indeed, this concept provides a great flexibility for the teachers.

The consequence is however that the teacher who develops a view determines self-reliant the structural format of the course content. e-stat does not offer any specifications in this regard. The teachers can save their views in e-stat so that a pool of views will be produced in a short time. This pool of views provides flexibility for the students because they can select that view which is related to their main subject of study.

In addition, empirical applications (called scenarios in e-stat) with large real-world data sets are already in progress, e.g. production process in chemical industry, virtual enterprise, insurance and stock market game. Of course, statistical analyses are emphasized in these scenarios.

Another main feature, which guarantees great flexibility in instruction and encourages the motivation of self-learning statistics, is the presentation of nearly all statistical topics at three levels of difficulty (Cramer, Cramer, Kamps, 2002; Genschel, Gather, Busch, 2002). By this construction, the preliminary knowledge of the users should be taken into account. Level A is the lowest one and provides an elementary description of statistical methods in a popular scientific manner avoiding complicated mathematical formulas and derivations. Level A is thought to be for those users who have first contact with statistics and a low previous knowledge (e.g. pupils). The medium level B is equivalent to an introductory course for students who do not take statistics as principal subject, i.e. students of economics, social sciences and so on. This level provides a comprehensive knowledge of the statistical methodology, associated with elementary derivations of formulas and methods, and incorporates the practical use of statistical procedures together with a variety of examples. Level C is intended for advanced studies in statistics (i.e. students majoring

in statistics, scientific members of research institutions) and comprises the full statistical theory. It should be mentioned that it is very easy to switch from one level to another.

Furthermore, e-stat enables self-assessment by implementation of online exercises into the e-stat environment to provide immediate feedback for new gained knowledge (Bartels, K., 2002). There exist a diverse range of possible kinds of exercises:

- discrete choice or multiple choice questions,
- short entry exercises with fill-in-the-blank text, with demanded statement of relevant keywords or with a table to be completed,
- complex computational exercises.

Even free text answers should be enabled and automatically evaluated.

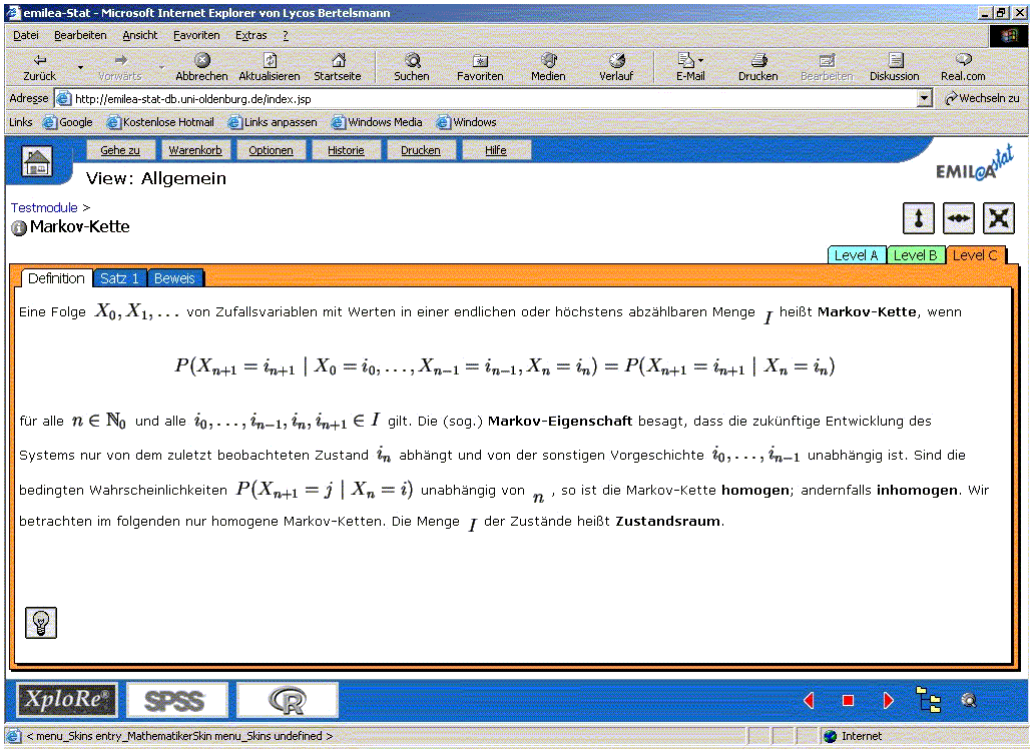

Figure 4: Graphical user interface of e-stat

Figure 4 shows the user interface of e-stat. The design of the user interface is simple, but it shows all the features mentioned above (view, module typ and content, level) and contains navigation buttons.

Throughout the whole multimedia tool interactive data analysis and interactive graphical visualization, using provided or user-collected data sets, is enabled either by interactive Java applets or by linking to different web machines for statistical computing, among others XploRe (http://www.i-xplore.de) and R-web.

5 E-LEARNING/E-TEACHING: MD*REX & REXCEL

As pointed out above, integration of statistical environments with other systems seems desirable. The tools MD*ReX and RExcel exactly do this. Both intergrate statistical languages with a completely different system, a standard spreadsheet application via well defined interfaces: MD*Crypt in case of the former and (D)COM in the latter (technical details are described in Aydınlı et al (2002), Aydınlı et al. (2003), Neuwirth (2001)). The rationale behind these efforts is straight-forward. It is useful in our opinion to combine matrix oriented procedural statistical languages (XploRe and R) with a spreadsheet application (Microsoft Excel). Literature suggests in no way to use Excel for statistical analyses, due to its methodological and numerical deficiencies. However the traits of Excel, which made this piece of software so exceedingly successful, are also desirable for our purposes: a intuitive GUI, interactivity with regard to graphics and data, a great deal of flexibility and customizability and the support for defined interfaces.

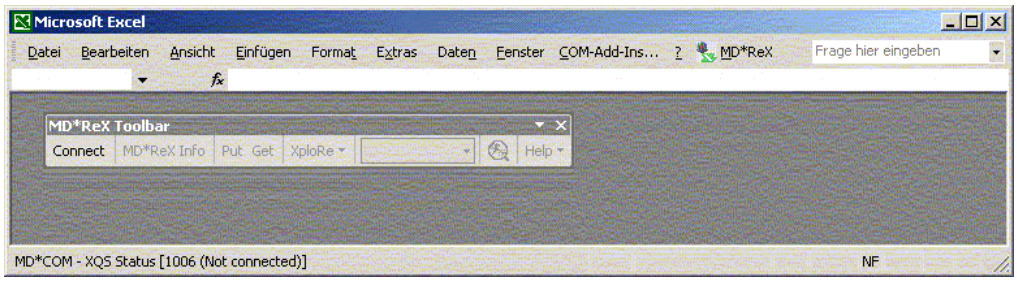

Figure 5: A standardized interface to MS Excel, the MD*ReX Toolbar

Hence the combination of both a standard office application and a statistical language creates a powerful, flexible and furthermore open tool for e-learning/e-teaching of statistics.

It is suitable for classroom use, where the teacher has access to 1) data, 2) methods and 3) interactive graphics. As well as the self-paced learning student who can 1) reproduce the classroom study, 2) interactively manipulate those examples and 3) can enhance them with self developed methods. As we eliminate the numerical drawbacks of Excel there is also little objection from this side.

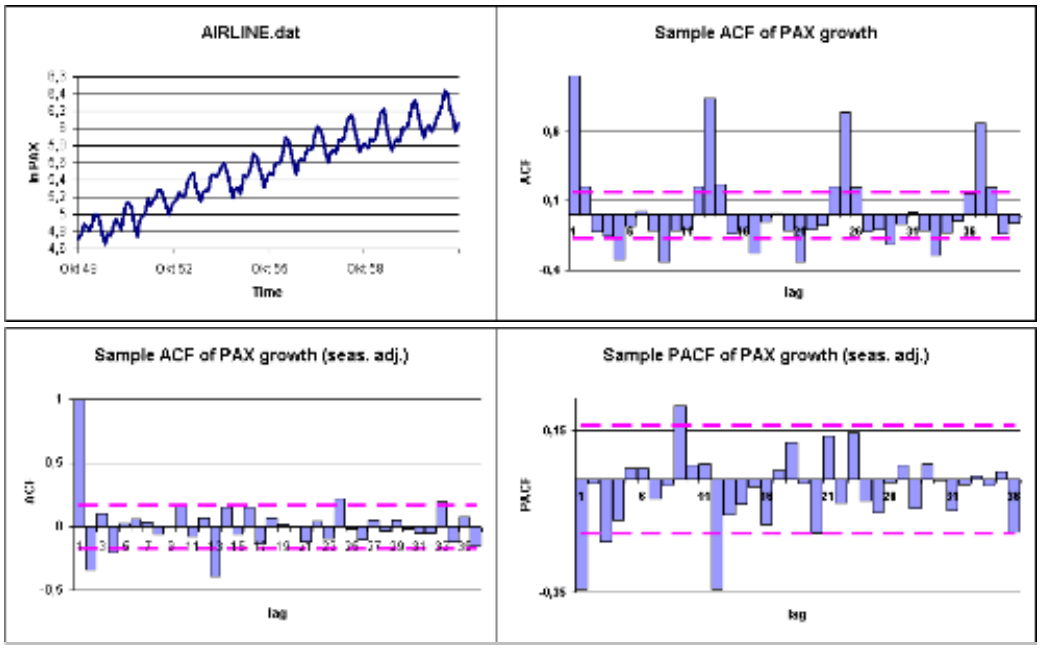

Figure 6: Time Series Analysis with MD*ReX

The use of abstract and well defined interfaces make this approach even further appealing, since the application is not restricted to a spreadsheet. A variety of applications like presentation programs and databases support these interface. Thus with only little efforts on the software development side, a whole bunch of applications can exploit the benefits of optimized statistical languages.

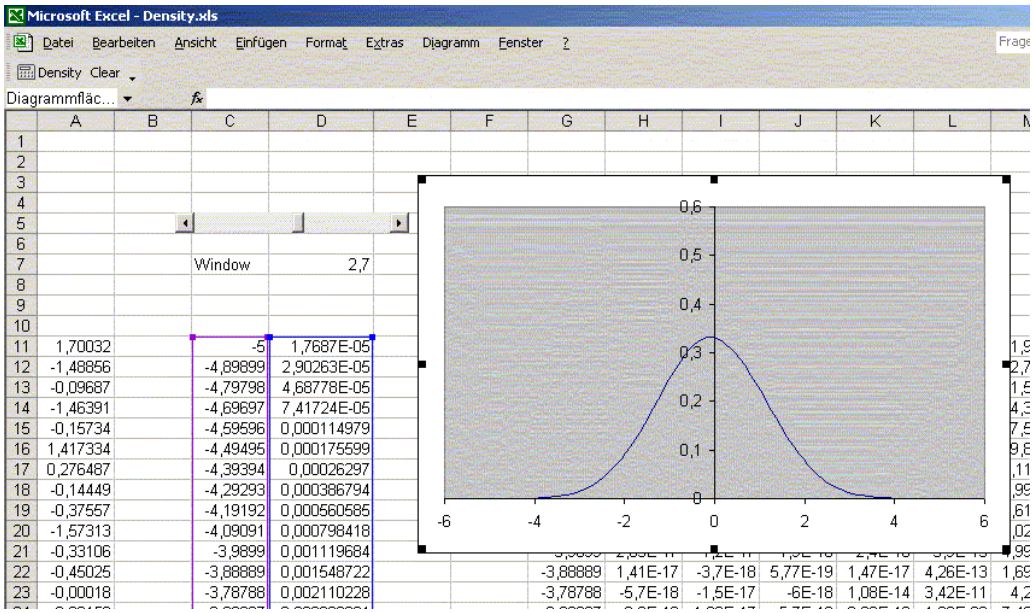

Figure 7: Interactive density estimation with Slider in RExcel

The most striking point however is the very little learning curve effect these clients exhibit. Evidently this is due to the large market impact of MS Excel. Almost every student and teacher is able to work with this spreadsheet and do basic steps like import / export data, simple calculations like summing and averaging, formatting, etc. Furthermore we are aware of different modes of usage. Hence it seems desirable to account for various user profiles: the methods developer (Teacher) who needs direct access to the statistical engine (e.g. through the command line utility in MD*ReX), the sophisticated methods user (Graduate Student) seeking for a macro editor (e.g. the XploRe Direct utility in MD*ReX) and of course the naïve user (Undergraduate Student) who is accustomed to a menu driven interface with dialogues and menu options (e.g. the MD*ReX Toolbar). RExel offers the same functionalities. Since the paradigm of a dialogue based window application is obeyed by both clients, the user can easily create interactive graphics or sophisticated statistical analyses.

GENERATING E-LEARNING/E-TEACHING DOCUMENTS

To ensure the above described flexibility without loss of interactivity, an easy technical generation of e-learning/e-teaching documents is essential. Text documents to be incorporated into MM*Stat must have HTML format, and formulas must be provided as gif's. A new module which should be included in e-stat has to be submitted in XML format and the formulas in MathML representation. If teachers are not familiar with these formats it will be impossible or very time consuming for them to input new ideas into these e-learning/e-teaching tool.

MD*Book (http://www.md-book.com, in co-operation with Springer-Verlag) fills this gap. MD*Book is a flexible user interface to create interactive e-learning documents and e-books. The only requirement is that the document, including the formulas and graphical presentations, must be available in LaTeX and that the module writer have to insert some structural information in the LaTeX source file. MD*Book converts the Latex source file to the desired format: PDF, HTML, Postscript, MM*Stat format or e-stat format. For more details of usage of MD*Book see Witzel, R., Klinke, S. (2002).

REFERENCES

[1] Härdle, W., Rönz, B. (2001), MM*Stat – Eine interaktive Einführung in die Welt der Statistik, Springer-Verlag, Berlin, Heidelberg, New York

[2] Härdle, W., Kleinow, T., Tschernig, R. (2001), Web Quantlets for Time Series Analysis. *Annals of the Institute of Statistical Mathematics*, 53, 1, 179 –188

[3] Kleinow, T., Thomas, M. (2000), Computational Resources for Extremes. *Franke, J., Härdle, W., Stahl, G. (Eds.), Measuring Risk on Complex Stochastic Systems, Lecture Notes in Statistics 147*, Springer-Verlag, 203 - 213

[4] Cramer, E., Cramer, K., Kamps, U. (2002), e-stat: A Web-based Learning Environment in Applied Statistics. *Härdle, W., Rönz, B., Compstat 2002, Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, New York , 309 - 314

[5] Genschel, U., Gather, U., Busch, A. (2002), EMILeA-stat: Structural and Didactic Aspects of Teaching Statistics through an Internet-based, Multi-medial Environment. *Härdle, W., Rönz, B., Compstat 2002, Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, New York, 340 – 342

[6] Bartels, Knut (2002), e-stat: Automatic Evaluation of Online Exercises. *Härdle, W., Rönz, B., Compstat 2002, Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, New York, 315 – 320

[7] Witzel, R. Klinke, S. (2002), MD*Book online & e-stat: Generating e-stat Modules from Latex. *Härdle, W., Rönz, B., Compstat 2002, Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, New York, 449 – 454

[8] Aydınlı, G. et al. (2002), MD*ReX, Linking modern statistical tool with standard office applications, *Computational Statistics*, Physica-Verlag, Heidelberg, New York

[9] Aydınlı, G. et al. (2003), Efficient and Secure Statistics with office Applications, *Interface 2003, Proceedings*

[10] Neuwirth, E. Baier, T. (2001), R in Office Applications. *DSC 2001, Proceedings in Computational Statistics*